Clemson University
TigerPrints

All Dissertations | Dissertations

12-2010

DEVELOPMENT OF GUIDANCE FOR STATES TRANSITIONING TO NEW SAFETY ANALYSIS TOOLS

Priyanka Alluri
Clemson University, palluri@clemson.edu

Follow this and additional works at: https://tigerprints.clemson.edu/all_dissertations

Part of the Transportation Commons

Recommended Citation
Alluri, Priyanka, "DEVELOPMENT OF GUIDANCE FOR STATES TRANSITIONING TO NEW SAFETY ANALYSIS TOOLS" (2010). *All Dissertations*. 663.
https://tigerprints.clemson.edu/all_dissertations/663

This Dissertation is brought to you for free and open access by the Dissertations at TigerPrints. It has been accepted for inclusion in All Dissertations by an authorized administrator of TigerPrints. For more information, please contact kokeefe@clemson.edu.

DEVELOPMENT OF GUIDANCE FOR STATES TRANSITIONING TO NEW SAFETY ANALYSIS TOOLS

A Dissertation

Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Civil Engineering

by
Priyanka Alluri
December 2010

Accepted by:
Dr. Jennifer Ogle, Committee Chair
Dr. Wayne Sarasua, Committee Member
Dr. William Davis, Committee Member
Dr. Karen Dixon, Committee Member

ABSTRACT

With about 125 people dying on US roads each day, the US Department of Transportation heightened the awareness of critical safety issues with the passage of SAFETEA – LU (Safe Accountable Flexible Efficient Transportation Equity Act - a Legacy for Users) legislation in 2005. The legislation required each of the states to develop a Strategic Highway Safety Plan (SHSP) and incorporate data-driven approaches to prioritize and evaluate program outcomes: Failure to do so resulted in funding sanctioning. In conjunction with the legislation, research efforts have also been progressing toward the development of new safety analysis tools such as IHSDM (Interactive Highway Safety Design Model), SafetyAnalyst, and HSM (Highway Safety Manual). These software and analysis tools are comparatively more advanced in statistical theory and level of accuracy, and have a tendency to be more data intensive.

A review of the 2009 five-percent reports and excerpts from the nationwide survey revealed astonishing facts about the continuing use of traditional methods including crash frequencies and rates for site selection and prioritization. The intense data requirements and statistical complexity of advanced safety tools are considered as a hindrance to their adoption. In this context, this research aims at identifying the data requirements and data availability for SafetyAnalyst and HSM by working with both the tools. This research sets the stage for working with the Empirical Bayes approach by highlighting some of the biases and issues associated with the traditional methods of selecting projects such as greater emphasis on traffic volume and regression-to-mean phenomena. Further, the not-so-obvious issue with shorter segment lengths, which effect the results independent of the methods used, is also discussed. The more reliable

and statistically acceptable Empirical Bayes methodology requires safety performance functions (SPFs), regression equations predicting the relation between crashes and exposure for a subset of roadway network. These SPFs, specific to a region and the analysis period are often unavailable. Calibration of already existing default national SPFs to the state's data could be a feasible solution, but, how well the state's data is represented is a legitimate question. With this background, SPFs were generated for various classifications of segments in Georgia and compared against the national default SPFs used in SafetyAnalyst calibrated to Georgia data.

Dwelling deeper into the development of SPFs, the influence of actual and estimated traffic data on the fit of the equations is also studied questioning the accuracy and reliability of traffic estimations.

In addition to SafetyAnalyst, HSM aims at performing quantitative safety analysis. Applying HSM methodology to two-way two-lane rural roads, the effect of using multiple CMFs (Crash Modification Factors) is studied. Lastly, data requirements, methodology, constraints, and results are compared between SafetyAnalyst and HSM.

DEDICATION

To my mother and my husband for believing in my goals and passion

ACKNOWLEDGMENTS

It is my pleasure to thank all the people who in one way or the other helped me in completing this journey. Primarily, it is my privilege to thank my advisor and mentor Dr. Jennifer Harper Ogle for her expert advice and continuous encouragement. During my four-and-a-half years of study at Clemson, Dr. Ogle has helped me become a strong, responsible, and a complete person. She has helped me realize what I want to do in life by uncovering my inner passion toward teaching, service, and research. I personally think that an opportunity for working with her is by far the best thing that has happened to me professionally. Lastly, I want to thank her for her continuous support and guidance.

I would like to acknowledge my committee member, Dr. Wayne Sarasua, for helping me with the GIS and for keeping me active in the professional community.

My sincere thanks also go to Dr. Karen Dixon for her expert advice and help with the research related to the Highway Safety Manual. Long conversations with her helped me decide on my course of action and made me rethink my approach toward addressing the research questions.

It is my pleasure to thank Dr. William Jeff Davis for helping me live through hard times throughout the course of the work since the time I have known him in 2007 and for his friendly advice.

My sincere thanks also go to GDOT (Georgia Department of Transportation) for funding this project and for giving me an opportunity to work with SafetyAnalyst through this project. I would like to express my deepest appreciation to Norm Cressman from GDOT

for his timely response to all my queries about crash data. A part of this dissertation involving survey to states wouldn't have been completed without the help of the 25 participating states. I would like to thank all the DOT officials for completing the survey in a timely manner.

This project became very interesting and challenging with the use of software SafetyAnalyst. I would also like to thank AASHTO for giving access to the SafetyAnalyst software for free for the research purpose. I would like to express my sincere thanks to SafetyAnalyst development team for continually making improvements to the software. I take this opportunity to thank Tom Robbins, Peter Holm, and Daniel Tomich from ITT for helping me understand SafetyAnalyst and for patiently answering all my questions. Their continuous and timely support made it possible to finish the data analysis in time. Also, I would like to thank Karen Richard from Midwest Research Institute who helped me in developing the safety performance functions and further helped in understanding the calculations within the software. Her suggestions are invaluable to this work. I would also like to thank Tom Gaines for helping me with hours of data retrieval.

To all my friends and well-wishers, my heartiest thanks for being the best source of encouragement.

TABLE OF CONTENTS

Table of contents Page number

Table of contents Page number

Table of contents Page number

Table of contents | Page number

Table of contents Page number

List of figures Page number

List of figures Page number

List of figures Page number

List of tables

Page number

List of tables Page number

List of tables Page number

NOMENCLATURE

AADT	Average Annual Daily Traffic
AASHTO	American Association of State Highway Transportation Officials
ADT	Average Daily Traffic
CARE	Critical Analysis Reporting Environment
CMF	Crash Modification Factor
CRF	Crash Reduction Factor
CSV	Comma Separated Value format
CV	Coefficient of Variation
DBMS	Database Management System
DD	Driveway Density
DOT	Department of Transportation
EB	Empirical Bayes
EMS	Emergency Management System
EPDO	Equivalent Property Damage Only
EXPO	Exposure
FARS	Fatality Analysis Reporting System
FHWA	Federal Highway Administration
FI	Fatal and Injury crashes
FMCSA	Federal Motor Carrier Safety Administration
GA	Georgia
GDOT	Georgia Department of Transportation
GHSA	Governors Highway Safety Association
GIS	Geographic Information System
GPS	Global Positioning System
HCL	High Crash Location
HPMS	Highway Performance Monitoring System
HS	Homogeneous Segments
HSIP	Highway Safety Improvement Program
HSM	Highway Safety Manual
IHSDM	Interactive Highway Safety Design Model
IT	Information Technology
LOSS	Level Of Service of Safety
LRS	Location Referencing System
MBB	Most Bang for the Buck
MIRE	Model Inventory of Roadway Elements
MMUCC	Model Minimum Uniform Crash Criteria
MVMT	Million Vehicle Miles of Travel
NB regression	Negative Binomial Regression

NCHRP	National Cooperative Highway Research Program
NEMSIS	National EMS Information System
NHTSA	National Highway Traffic Safety Administration
ODP	Overdispersion Parameter
PDO	Property Damage Only
PIL	Priority Investigation Location
PSI	Potential for Safety Improvement
R^2_{FT}	Freeman Tukey's R square
RC	Roadway Characteristics
RCLINK	Roadway Characteristics link id
RDBMS	Relational Database Management System
RHR	Roadside Hazard Rating
RTM	Regression To the Mean
SA	SafetyAnalyst
SAFETEA-LU	Safe Accountable Flexible Efficient Transportation Equity Act: A Legacy for Users
SAS	Statistical Analysis Software
SHSP	Strategic Highway Safety Plan
SPF	Safety Performance Function
SQL	Structured Query Language
SWiP	Sites With Promise
TOT	Total crashes
TRB	Transportation Research Board
TWLTL	Two-Way Left Turn Lane
VMT	Vehicle Miles of Travel

CHAPTER 1 **INTRODUCTION**

1.1 Introduction:

Dwight Eisenhower, the 34th President of the United States signed the Federal Aid Highway Act of 1956 initializing the Interstate system – a system with the potential to address to the previously identified five main causes of obsolete road network. The issues identified by the President as a young Army officer crossing the country in the 1919 Army Convoy were "*annual death and injury toll, the waste of billions of dollars in detours and traffic jams, the clogging of the nation's courts with highway-related suits, the inefficiency in the transportation of goods, and 'the appalling inadequacies to meet the demands of catastrophe or defense'*" (Weingroff,). Even after 9 decades, annual death and injury toll is still a main point of concern with 34,017 fatal crashes and approximately 1.63 Million injury crashes in the year 2008 (National Highway Traffic Safety Administration, 2010).

Today, traffic fatalities are found to be the leading cause of death between the ages of 3 and 33 in the United States (Kraft, 2009). With about 40,000 fatalities, 3 Million injuries, 6 Million crashes, and a total cost of $164.2 Billion annually (Clifford, 2008), the seriousness of the safety problem is paramount. Greater attention needs to be given to highway safety and the 4 E's of traffic safety (National Highway Traffic Safety Administration, NA)(National Highway Traffic Safety Administration, NA)(National Highway Traffic Safety Administration, NA): Engineering, Education, Enforcement, and Emergency Medical Services (Kraft, 2009). Of the 4E's, Engineering is considered to play a crucial and significant role in reducing the frequency and severity of crashes. In the year 2005, fatal traffic crashes had reached its highest 39,252 since 1994 (National

Highway Traffic Safety Administration, NA), thus, urging the government to take formal steps towards improving safety (see Figure 1).

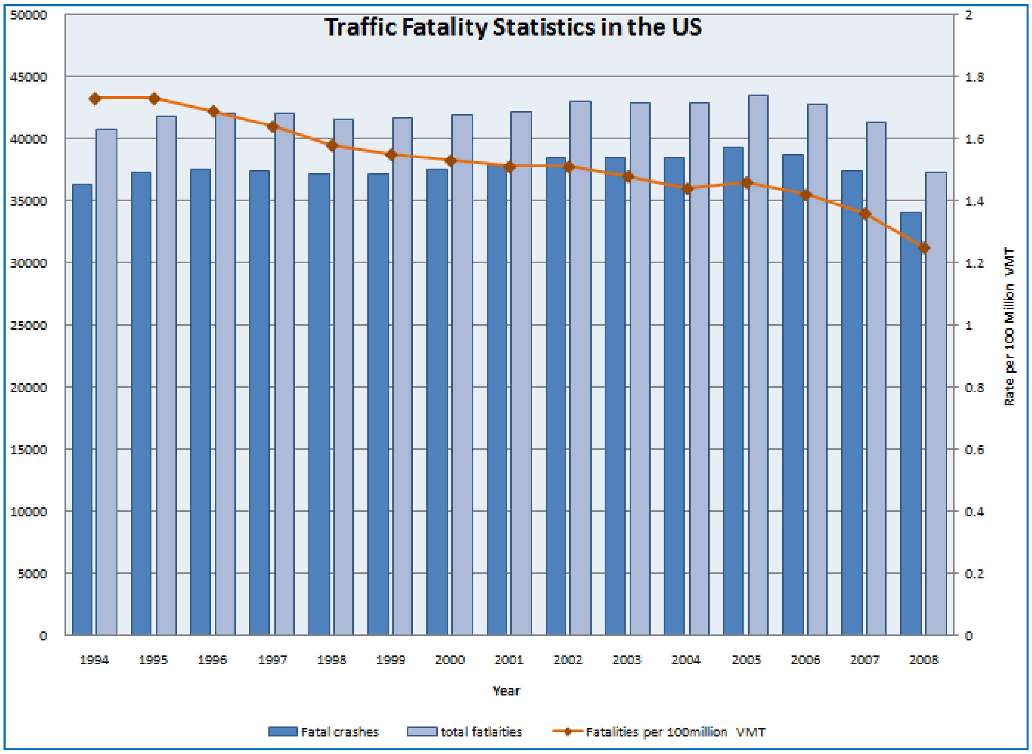

Figure 1: Traffic fatality statistics in the US from 1994-2008 ((National Highway Traffic Safety Administration, NA))

On August 10, 2005, SAFETEA–LU (*Safe Accountable Flexible Efficient Transportation Equity Act: A Legacy for Users*) was signed by President, George W. Bush governing spending of federal money on surface transportation (Federal Highway Administration, 2005). The bill's name reflects the focus on improving safety on all public roads. SAFETEA-LU established Highway Safety Improvement Program (HSIP) as a core federal program which required each of the states to develop an annual Strategic Highway Safety Plan (SHSP) to receive federal funding. The intent of the SHSP requirement was to involve all 4E stakeholders in a data driven process to identify safety problems, potential countermeasures, and develop measures by which performance will be evaluated (Federal Highway Administration, 2005).

Under SAFETEA-LU, states were also required to develop annual five-percent transparency reports which identify the top 5% of its roadway network currently exhibiting the most severe highway safety needs. Each state's report is to include potential remedies to the hazardous locations identified; estimated costs of the remedies; and impediments, if any, to the implementation of remedies. The methods used to identify these top 5% locations were also explained in the five-percent reports (Federal Highway Administration, 2010a).

In conjunction with new requirements for states, federal agencies also began the development of advanced safety analysis tools to overcome many of the biases uncovered during research associated with traditional methods. A review of the states' five-percent reports indicate that most states are still using traditional methods for safety analysis, and only few are moving toward more advanced methods. This is mainly due to the states' misconception about the data and expertise requirements of newer methods. The need of the hour is to help states begin to implement these newer methods with minimal problems/ roadblocks and also shorten the learning curve.

<u>Site Safety Improvement Process:</u> Mere identification of problematic sites by either traditional or advanced methods does not constitute a comprehensive roadway safety analysis procedure. While new safety analysis methods have been developed, the safety improvement process is still the same. According to the most bang for the buck theory, money needs to be spent where it achieves maximum benefit (Hauer, Kononov, Allery, & Griffith, 2002)(Hauer, Kononov et al., 2002). It is not advisable to spend money to

improve a site when the same amount would save more lives at another similar location. Instead, sites should be identified and prioritized based on their potential for safety improvement (PSI). Site selection is the first step in the highway safety improvement process, a fourfold approach involving site identification, detailed engineering survey, treatments selection, and prioritization as shown in Figure 2 (Hauer, Allery, Kononov, & Griffith, 2004).

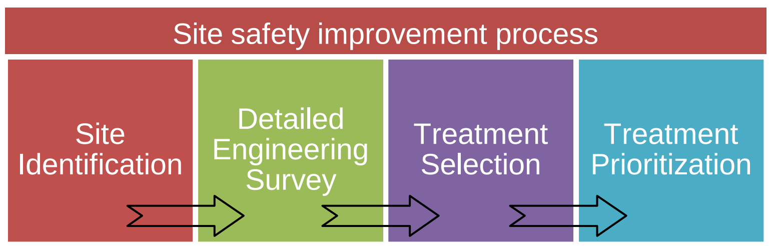

Figure 2: Site safety improvement process (Hauer et al., 2004)

Of all the aforementioned steps, identification of sites is the most fundamental building block for a successful safety improvement program, since the improper identification of high priority sites results in less cost-effective solutions (Hauer, Kononov et al., 2002).

Over the last 50 years, there have been many methods, tools and measures in practice to help in the process of identification and prioritization of sites. These methods are referred to as traditional methods. The traditional methods use accident counts or their proportions to identify unsafe sites. Today, superior methods are available for use employing advanced statistical methods (i.e. Empirical Bayes method and Full Bayesian approach). These methods have been developed over the last decade and have recently been made available through the Interactive Highway Safety Design Model (IHSDM) in the year 2003, and, SafetyAnalyst and Highway Safety Manual (HSM) in the year 2010.

While evaluating the pros and cons of traditional and advanced methods, it was found that the traditional methods require little data, but are fraught with problems and false assumptions including site selection bias, false assumption of a linear relationship between crash count and traffic volume, bias towards heavier volume roads and smaller segment lengths, etc (Alluri, 2008). Though superior safety analysis tools address the biases associated with traditional methods, they tend to require more complete and comprehensive data for crashes, roadway characteristics, and traffic to be fully utilized. However, these advanced methods have the flexibility of performing incremental analysis depending on the current data availability and technical expertise within the states. Thus, as states are ramping up data collection and analysis procedures, they can still make use of the new tools.

HSM, SafetyAnalyst and IHSDM are the three advanced safety analysis tools developed by NCHRP (National Cooperative Highway Research Program), TRB (Transportation Research Board) and FHWA (Federal Highway Administration) through research mechanism and state involvement. HSM was released in July 2010 while SafetyAnalyst and IHSDM were released in March 2010 and 2003 respectively.

The Highway Safety Manual provides analytical tools for quantifying effects of potential changes at individual sites (American Association of State Highway and Transportation Officials, 2010b). SafetyAnalyst software provides a suite of analytical tools to identify and manage system-wide safety improvements (American Association of State Highway and Transportation Officials, 2010c). Both the federal projects were developed to address two diverse aspects of road safety. SafetyAnalyst is considered to be companion software to the Highway Safety Manual, yet SafetyAnalyst is designed for more system-wide analysis, and HSM is better suited for site specific analysis – although HSM can be used for statewide analysis, but the data needs are significant. It is expected that HSM and SafetyAnalyst working together would constitute a more comprehensive set of safety improvement tools for an agency.

IHSDM is also a set of software tools aimed at improving safety on specific sections of roads by evaluating safety and operational effects of geometric design decisions on these sections (Chen, 2009). Table 1 gives a summary of the data requirements for the basic (crash frequency, crash rate, and rate quality control) and the three advanced safety analysis tools.

Table 1: Data requirements for various safety analysis tools

Methods	Crash data by type and location	Traffic Volume	Basic Roadway Characteristics by location	Full geometric roadway characteristics	Safety Performance Functions
Category A - Screening Based on Counts					
Frequency[1]	Yes		Yes		
Rate[1]/ Rate Quality Control[1]	Yes	Yes	Yes		
Category B - Screening Based on Potential for Safety Improvement					
IHSDM[2]		Yes	Yes	Yes	
SafetyAnalyst[3]	Yes	Yes	Yes		Yes
HSM[1]	Yes	Yes	Yes	Yes[4]	Yes

Source:
[1] (American Association of State Highway and Transportation Officials, 2010b)
[2] (Federal Highway Administration, 2010b)
[3] (American Association of State Highway and Transportation Officials, 2010c)
[4] Sample of roads required for calibration purposes

The various types of data that are required include: Crash data by type and location, traffic volume data, basic roadway characteristics, complete geometric roadway characteristics data, and safety performance functions (SPFs). SPFs represent the relation between crashes and exposure (usually traffic volume) for a group of reference sites.

The method of site selection by crash frequency requires minimal information on crashes and roadway characteristics. Crash rates and critical crash rates (used to perform rate quality control) are the most commonly used methods and require crash data along with traffic volume, roadway characteristics data, and segment length.

For their complete implementation, advanced tools require a wide range of data in comparison to basic methods. For example, SafetyAnalyst and HSM require SPFs which are rarely available at the state level. As such, both tools come with a set of default SPFs. The default SPFs for SafetyAnalyst were developed using multiple year data from California, Minnesota, Ohio, and Washington. The default SPFs for HSM came from various states and different analysis periods for each roadway type.

The individual SPFs included as defaults were chosen as most representative based on R^2_{FT} (Freeman Tukey's R square) and overdispersion values. Some researchers have questioned the reliability of these default SPFs in representing other states' safety patterns and for representing crash patterns for different analysis periods. On another note, IHSDM and HSM require complete geometric alignment information. For IHSDM, this requirement only includes geometric data for the sections under evaluation. HSM requires complete geometric and roadside information for a minimum of 30-40 roadway sections totaling 100crashes/year for SPF calibration purposes. Given the changes in data requirements from traditional to advanced methods, many states will be challenged. Shifting of analysis methods from traditional to advanced would be more gradual with states planning on a few years in the transition process.

1.2 Problem Statement:

For many decades, states have been using traditional methods like crash frequencies, crash rates, and safety indices for crash data analysis and site selection which have their own advantages and limitations. However, these have been shown to be subpar to their advanced counterparts that include the HSM, and SafetyAnalyst. Understanding the constraints and issues with traditional methods, states are looking to shift to newer and advanced tools which require safety performance functions and geometric alignment data in various steps of the site safety improvement process.

This dissertation could act as a guide to help states transitioning to newer safety analysis tools by providing a thorough discussion on the data requirements, the requirement of state specific SPFs, and the expertise required to shift to the advanced methods.

Various diverse problems are addressed in this research broadly dealing with data accuracy and availability, influence of segment length on site selection methods, issues with the deployment of network screening module of SafetyAnalyst, fit of national SPFs to the state's data, influence of variations in traffic counts on the fit of Georgia specific SPFs and crash predictions, detailed application of HSM procedure for identifying top ranked two-lane two-way rural sites along with the calculation of calibration factors, effect of the use of multiple crash modification factors on crash predictions, and the differences between SafetyAnalyst and the Highway Safety Manual.

1.3 Goal and Objectives:

The goal of this research is to document and methodically identify the issues associated with traditional rates and frequencies, understand the present stand of the states in relation to their acceptance of newer safety analysis tools, and to shorten the learning curve for the states implementing SafetyAnalyst and the Highway Safety Manual by developing guidance for states with questions such as: Do I need to develop my own SPFs? What issues am I likely to face in the initial stages of implementing SafetyAnalyst/ HSM? Are the processes of SafetyAnalyst and HSM interchangeable?

The hypothesis considered for this research is that the states are ready to shift to the newer safety analysis tools.

The objective of this research is to provide guidance to states to aid in transitioning to the new methods by:

- Developing guidance for overcoming common data issues – data completeness, data inaccuracy and interoperability, and data sensitivity
- Providing specific examples of the issues, constraints, and biases with traditional site selection methods for training for upper level management
- Providing toolbox with solutions for common problems and documenting the lessons learned while implementing the network screening module of SafetyAnalyst using Georgia data
- Conducting survey across states to understand the present safety analysis procedures, safety data availability and constraints within each state and thus identifying critical gaps in safety analyses

- Generating state specific safety performance functions for various site subtypes, and comparing the fit of state specific SPFs to the calibrated SPFs used within SafetyAnalyst
- Determining the influence of AADT estimations on the fit of SPFs
- Documenting the process of generating calibration factors for two-way two-lane rural roads to be used within HSM and to study the influence of multiple CMFs on crash predictions.
- Comparing the two SPFs (calibrated SafetyAnalyst SPF, and calibrated HSM SPF) for two-way two-lane rural roads and recommending a SPF with the best fit.

1.4 Organization of the Dissertation:

Following this introduction, the remaining dissertation describes the work completed to meet the research goals and objectives.

Chapter 2 provides a brief literature review related to the new safety analysis tools (HSM and SafetyAnalyst), their data requirements, and data issues while performing safety analysis. Extant literature about the safety performance functions, their functional form, dependant and independent variables, and the Empirical Bayes approach is also discussed.

Chapter 3 discusses the approach and methodology used for meeting the research objectives. This chapter deals with preparing data files for use by EB methodology, identifying data issues, and the minimum data requirements to perform advanced methods. The procedure used for generating SPFs, calculating the goodness-of-fit, identifying and prioritizing sites based on Empirical Bayes approach are also explained in detail. A complete descriptive record of the calibration process based on the procedure described in HSM is documented in this chapter.

Chapter 4 presents various problems and issues identified with the data, the survey results, Georgia specific SPFs for various site subtypes, various CMFs, and the top list of sites identified by Empirical Bayes approach.

Chapter 5 summarizes conclusions of this dissertation and provides recommendations for future research.

CHAPTER 2 **LITERATURE REVIEW**

A crash is defined as "a set of events that result in injury or property damage, due to the collision of at least one motorized vehicle and may involve collision with another motorized vehicle, a bicyclist, a pedestrian or an object" (American Association of State Highway and Transportation Officials, 2010b). In his book, Observational before and after studies, Hauer, E defined safety of a location as "the number of crashes or crash consequences, by kind and severity, expected to occur on the entity during a specified period of time" (Hauer, 1997). Depending on the persons/ agencies involved, safety could be considered as two types: nominal and substantive. Nominal safety "adheres to design practices, standards, and warrants etc" (iTRANS Consulting Ltd & Human Factors North INC, 2003) and could be measured by comparing the roadway with design standards. On the other hand, substantive safety refers to "the actual or expected performance as defined by the frequency and severity of crashes" (iTRANS Consulting Ltd & Human Factors North INC, 2003). Thus a roadway can have nominal safety without having substantive safety. Due to random and infrequent nature of crashes, substantive safety is the most difficult and complicated gauge for assessing the safety improvements and the performance of a roadway from safety perspective.

A roadway safety management process is "a quantitative, systematic, process for studying roadway safety on existing transportation systems, and identifying potential safety improvements" (American Association of State Highway and Transportation Officials, 2010b). Figure 2 explains the various steps involved in this process (American Association of State Highway and Transportation Officials, 2010b).

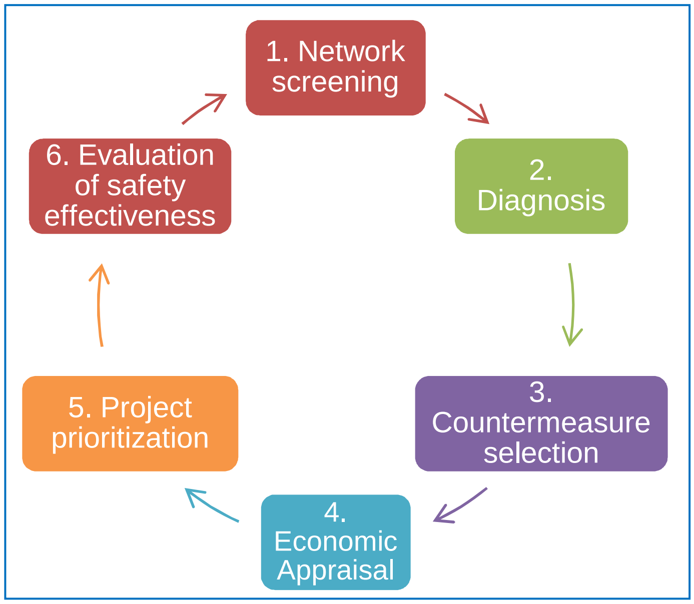

Figure 3: Roadway Safety Management Process

Network screening, the first step in roadway safety management process, is the process of identifying and prioritizing sites for further engineering study and potential countermeasure implementation which have a greater potential for safety improvement. Next to network screening, diagnosis involves the process of identifying the reasons and factors resulting in crashes. In this step, crash patterns and possible causes of the collisions are identified for further evaluation. The third step, countermeasure selection, deals with identifying contributing factors and suggesting the most effective countermeasures. Next to countermeasure selection step is economic appraisal step which deals with evaluating the countermeasures from an economic perspective dealing with project costs either by performing benefit cost analysis or cost-effectiveness analysis. Once the monetary factor is incorporated in the assessment of safety improvements at problematic sites, project prioritization is carried out which constitutes the fifth major step within the process. Following project prioritization, safety effectiveness evaluation is considered to be the important and final step. It deals with assessing the safety improvements in some period after the suggested countermeasures are implemented.

All the steps in the process are equally important and could be independently pursued depending on the requirements of the state agency. The present research deals with network screening – data requirements and advanced methods used to identify and prioritize sites. SafetyAnalyst and HSM procedures are also explored in detail. Therefore, the literature review is divided into four sections: data requirements, issues with traditional methods, generation of SPFs to be used by advanced methods, and the background behind SafetyAnalyst.

2.1 Data requirements and issues:

According to SAFETEA-LU signed in August 2005, in order to be able to use federal funds, states are required to develop a Strategic Highway Safety Plan (SHSP) and incorporate *data-driven* approaches to prioritize and evaluate program outcomes to obtain/ use federal funding (Federal Highway Administration, 2005). States also have to be able to perform various steps in roadway safety management process on all public roads.

The backbone of any safety management system is data collection and maintenance (Ludwig, 2007). Whether dealing with just site selection methods or various steps in the site safety improvement process, at a minimum, three databases, crash, roadway inventory and traffic operations are required (Ogle, 2007); (American Association of State Highway and Transportation Officials, 2010b). However, for a complete and comprehensive roadway safety analysis, in addition to the earlier mentioned databases, various other databases including but not limited to driver history information, citation records, FARS statistics, VMT numbers, census information, trauma registry, observational safety belts and child safety seat surveys, telephone and driver facility surveys are also required (Federal Highway Administration,)(Council & Harkey, 2006); (CH2MHill, 2009)(National Highway Traffic Safety Administration, 2003; National Highway Traffic Safety Administration, 2003).

Over the past ten years, the federal government has been spending considerable amount of time and resources in developing guidance for identifying the data requirements for various datasets with the main goal of making “accurate, reliable and credible highway safety decisions within a state, between states and at a national level” (WSDOT, 2010). Model Minimum Uniform Crash Criteria (MMUCC) for crash database, Model Inventory of Roadway Elements (MIRE) for roadway inventory database and National EMS Information System (NEMSIS) for EMS data are the three guidelines currently available for use (Council, Harkey, Carter, & White, 2007)(National Highway Traffic Safety Administration, 2008). Over a hundred data elements are recommended in MMUCC and about 200 are recommended in MIRE which make the overall process of data collection and maintenance more intensive.

In the context of the magnitude of this task, lack of funding and good data collection infrastructure are considered to be the barriers for collecting and maintaining data (Ogle, 2007)(WSDOT, 2010). Collaboration between agencies and organizations could potentially result in mutual support in data collection and analysis, and minimized duplication of efforts (WSDOT, 2010). The efficiency of data collection could be improved by minimizing the number of data variables collected at the crash site, integrating one or more databases to obtain some of the variables, and also by using technologies like Global Positioning System (GPS) (Council & Harkey, 2006). Understanding the importance of data collection, in their report, Traffic Safety Information Systems International Scan: Strategy Implementation White Paper, Council, F.M and Harkey,D.L suggested a few strategies to improve highway safety data (Council & Harkey, 2006):

- Increase support for both safety programs and safety information systems (the data) from top-level administrators in state and local transportation agencies
- Define good inventory data and institutionalize continual improvement toward established performance measures
- Make it easier to collect, store, and use
- Increase the use of critical safety analysis tools, which themselves require good data
- Improve and protect safety data by storage and linkage with critical non safety data

In the past three years, since the release of this white paper report, many of the aforementioned strategies have been followed, at least to an extent. Many states have identified the importance of data and consequently incorporated state-wide safety programs that are interoperable (INDOT, 2006); (FDOT, 2006); (Iowa DOT, 2007); (MDOT, 2008). FHWA identified and defined good inventory data by releasing MIRE and is working towards establishing performance measures (Federal Highway Administration,). The newer safety analyses tools recommended by FHWA are data intensive requiring the states to collect and maintain complete and accurate data (American Association of State Highway and Transportation Officials, 2010b).

Limitations within data collection and maintenance that hinder the process of addressing the safety issues identified by Pfefer et al include (Pfefer, Neuman, & Raub, 1999):

- Lack of precision measurement and reporting
- Lack of automated tools in data collection and management
- Inadequate coverage of traffic data
- Incomplete and missing data
- Lack of adequate documentation on the dynamic nature of the roadway inventory database
- Issues with data integration and interoperability

Delucia, B. and Scopatz, R also acknowledged the earlier identified issue related to the dynamic nature of the roadway inventory database by recognizing the inadequate maintenance and linkage of roadway characteristics associated with specific locations even with the increased use of GIS technology (Delucia & Scopatz, 2005).

Similar idea had been reinforced in NCHRP Synthesis report 350, where, the authors Delucia, B. and Scopatz, R have identified three broad areas that define the success of a crash record system. They are (1) data collection, (2) data processing and management, and (3) data linkages for reporting and analysis (Delucia & Scopatz, 2005).

Due to the continuous increase in the costs of crash data collection, over the past two decades, some states have reduced the quantity and quality (including the accuracy, precision, timeliness, and completeness of the data (O'Day, 1993)) of data being collected at crash locations instead of increasing them (Council & Harkey, 2006; Delucia & Scopatz, 2005); (Ogle, 2007). The lack of sufficient number of police officers led to an increase in the crash thresholds resulting in fewer number of reported PDO crashes (Property Damage Only crashes). However, these trends are being reversed in the recent years with the federal requirement of states to collect more data. In addition to the level of reporting, the quality of crash data is also influenced by its uniformity and accuracy. The quality of data collection could be improved mainly by using an automated field data collection tool that runs on laptops with barcode readers and GPS etc.

"Capability to accept data electronically" is considered to be another major hindrance to data management, the next major concern. Data interoperability is by far the most complicated issue. As mentioned earlier, it is observed that a number of agencies are responsible for various data files that are required for a comprehensive roadway safety management process. The process of linkage between the databases hasn't been given much attention in the past. Even within the same database, inconsistencies exist between the data items collected by local agencies, state officials and the federal requirements, mainly due to the flexibility within the agencies. It is thus agreed upon that a "comprehensive traffic records system is required with linkable components to support reporting and analyses of all types of data" (Delucia & Scopatz, 2005).

Addressing the above discussed concerns, NCHRP Synthesis report 350 identifies establishing a statewide traffic records coordinating committee, developing data-for-data partnerships, developing a knowledge base for traffic records system and simplifying crash data collection as approaches towards implementing a successful and comprehensive crash records system. (Delucia & Scopatz, 2005).

From the review of extant literature, the role of data quality and quantity in improving highway safety is evident. Having discussed about the data quality in a broad sense, the following paragraphs explain the issues with data in particular to each dataset.

<u>2.1.1 Crash data:</u>

"Crash data represent a sort of window on the world of the untoward things that happen in the traffic system" (O'Day, 1993), making its interpretation a basis for improving highway safety which is often measured as the frequency of expected crashes (iTRANS Consulting Ltd & Human Factors North INC, 2003).

Quality of crash data is a generic term constituting of various components - data completeness, consistency of coverage and interpretation, appropriate level of detail, missing data, the right data, correct entry procedures and freedom from response error (O'Day, 1993). Understanding the importance of data interoperability, high data quality and consistency within a state and between states, Governors Highway Safety Association (GHSA), the Federal Highway Administration (FHWA), the Federal Motor Carrier Safety Administration (FMCSA), and the National Highway Traffic Safety Administration (NHTSA) have collaboratively created MMUCC, Model Minimum Uniform Crash Criteria in 1998 and later updated in 2003 (National Highway Traffic Safety Administration, 2008). MMUCC provides a list of recommended data elements to be collected and maintained in a state's crash database.

Although the use of MMUCC is voluntary, it would be highly beneficial to the states in the long run from data quality, interoperability and consistency points of view. With pressure from the federal government, it is also observed that the crash data collection and maintenance have had a major transformation for better in the past decade. However, there still exist a few critical gaps in the area of processing and maintaining of data which would be addressed to in later sections of this dissertation.

2.1.2 Roadway characteristics data:

Next to crash data, roadway inventory information is the core area of safety data that is required for any type of highway safety analysis. Roadway inventory information includes all the "physical features within a road's right-of-way" (Ogle, 2007). Information on geometric data, cross-sectional elements, traffic control devices, pavement-related data etc on all public roads constitute an ideal roadway inventory file. MIRE (Model Inventory of Roadway Elements), a companion to MMUCC helps the states in defining the data elements and attributes to be collected and maintained. Similar to the issues relating to crash database, data quality and quantity are the major deterrents to accurate and comprehensive roadway safety analysis. In the NCHRP Synthesis 367, Ogle, J.H. had successfully captured the major issues and considerations with roadway inventory data which are briefly discussed below:

- Route milepost and node-to-node are the two primary linear referencing systems being used by 75% and 25% of the states respectively.
- Dynamic nature of roadway inventory database is seldom addressed when a route milepost system is used as it is difficult to readily access multiple years of roadway inventory data. This proves to be a major hurdle while analyzing safety

on roadways with major reconstruction. “Using the route-milepost system, there is no good way to manage change dates for specific pieces of information (e.g., the date that raised pavement markers were added to a section of roadway or the date when a traffic signal was added to an intersection)” (Ogle, 2007).

- Even though most states collect hundreds of data elements, there are still a number of required fields that are not generally collected on a regular basis (ex: cross slope). This issue has been addressed with the release of MIRE which constitutes the recommended list of data variables to be collected and maintained for comprehensive safety analyses. Similar to MMUCC, MIRE is a guideline and not a standard (Federal Highway Administration,).

2.1.3 Traffic operations data:

Theoretically, many variables including but not limited to AADT, speed, volume, density, axle load, and vehicle classification constitute traffic operations data (Ogle, 2007). However, Annual Average Daily Traffic (AADT) is one of the most important and required data elements without which network screening is not possible except for the use of crash frequency. Yearly AADT values, either measured or estimated from counts, are used in various steps of the roadway safety management process. Irrespective of the quality of roadway segment data, the inclusion or exclusion of segments depends a lot on the completeness and correctness of traffic data. The most common reason for maintaining traffic data is HPMS (Highway Performance Monitoring System)

Having discussed about the three core areas of safety data, it is evident that data quality and completeness play a vital role in defining the success of highway safety analyses and improvements. Regular data quality and consistency checks help identify the hidden issues which otherwise go unidentified.

2.2 Traditional methods and their issues:

Network or site screening identifies sites with potential for safety improvement and results in a number of sites that are priority ranked. Over the years, these sites have been referred to as Black Spots, High Crash Locations (HCLs), Hazardous Locations, Priority Investigation Locations (PILs), or Sites With Promise (SWiP) depending on the researcher (Hauer, Kononov et al., 2002; Hauer et al., 2004). “Sites With Promise (SWiP)”, the most recent term, identifies sites in which safety can be improved cost-effectively based on Empirical Bayes methods and using Safety Performance Functions (Hauer et al., 2004). In order to identify and prioritize problematic sites, there are numerous methods of safety analysis in existence today, but, the most commonly used methods, known as traditional methods, rely on accident counts. Newer and more advanced tools use safety performance functions and Empirical Bayes approach, and identify and prioritize sites based on their potential for safety improvement (PSI). Traditional methods, even though most widely used are fraught with problems and false assumptions most of which are addressed by EB approach.

As mentioned in the introduction chapter, highway safety had been a problem since early 1900s well before the construction of the Interstates. Due to comparatively lesser complexity of crashes and fewer numbers of incidents, safety was the responsibility of the local agencies that used colored pins to mark a traffic incident on a map. Safety improvements and stricter law enforcements were performed at the locations with greater “pin” density (O’Day, 1993). This concept was now termed as “network screening by crash frequency". As identified by many researchers till date, the crash frequencies will be comparatively higher for sites with heavier traffic such as urban roads and interstates resulting in a biased estimate.

Further, ranking based on accident rates has its own disadvantages. "Rate measures the risk road users face while driving on specific roads" (Hauer, 1996). Crash rate is defined as the number of crashes per unit exposure per unit of time (Hauer, 1997). Crash rates assume a linear relationship between crash frequency and exposure, while the actual relation is non-linear, thus, resulting in incorrect identification of "problematic sites" (iTRANS Consulting Ltd & Human Factors North INC, 2003). Due to this incorrect assumption, crash rates tend to identify sites that have lower exposure. When traffic volumes are considered as exposure, any crash on the segment with lower traffic will produce a large rate. In addition, crash rates are dependent on segment length, and very short segments have the same effect on rates as do small traffic volumes - thus leading to high rate. Hence, it is observed that crash rates and frequencies, the most frequently used site selection methods produce biased results making the safety improvements less cost-effective (Alluri, 2008; Hauer, 1997).

Irrespective of the type of the network screening method used, one of the major shortcomings is the use of few years of historical crash data, resulting in regression-to-mean effect (RTM). This is defined as "the phenomenon of repeated measures of data in the long run drifting towards a mean value" (iTRANS Consulting Ltd & Human Factors North INC, 2003).

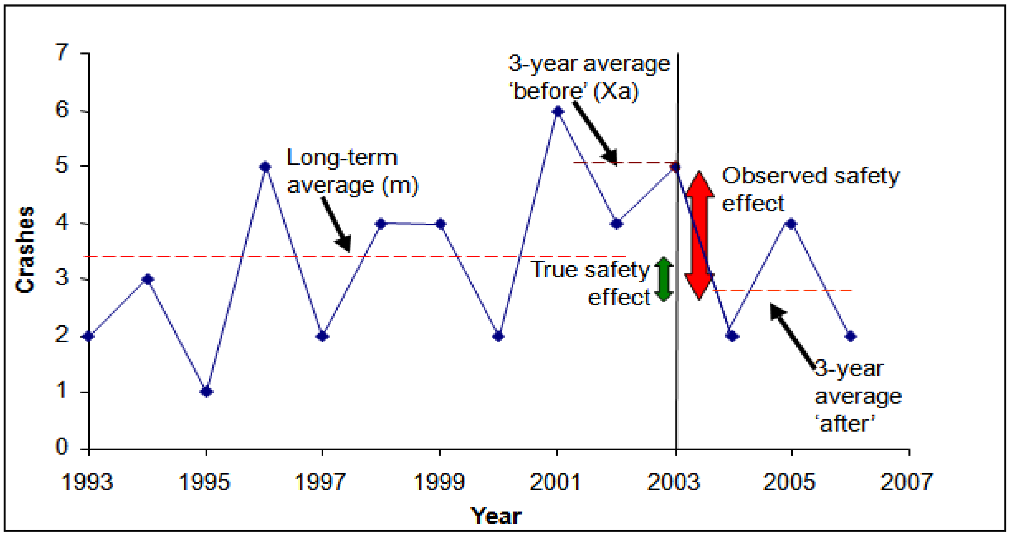

Figure 4: Regression-to-Mean effect (iTRANS Consulting Ltd & Human Factors North INC, 2003)

Due to the random nature of crashes, it is observed that the short term average crash frequency at a site is independent of its long term average, the true safety characteristic of the site, thus questioning the reliability of safety predictions made with few years of crash data. In practice, this issue, also known as "selection bias" (American Association of State Highway and Transportation Officials, 2010b) might not be addressed depending on the site selection method used by the state DOTs. Traditional or basic methods like crash frequency, crash rate, and safety indices do not address the aforementioned issue of RTM.

Another major limitation among the screening methods based on accident counts is their inability to predict the future expected performance of crashes. It is believed that In comparison to the past, the present and future safety of a roadway is of primary importance (Harwood, Council, Hauer, Hughes, & Vogt, 2000). In this context, it is observed that methods based on accident counts rank sites based on just the past performance of sites with no information on the future. On the contrary, advanced methods have the capability of calculating the expected and predicted crashes at sites based on crash history at similar sites and the current observed crash frequency at the intended sites (American Association of State Highway and Transportation Officials, 2010b). Even more importantly, other statistical variables including variance could be calculated using advanced methods that add reliability to the results. Table 2 gives a summary of various site selection methods and the considerations that they address.

Table 2: Various site selection methods and the considerations that they address

Considerations						
Methods	Categorize Sites	Descriptive Information	Accounts for RTM	Does not assume a linear crash-exposure relationship	Predicts Expected Performance	Need SPF
Category 1 - Screening Based on Counts						
Frequency[1]	Yes	Yes	No	Yes	No	No
Crash rate[1]/ Rate Quality Control[1]	Yes	Yes	No	No	No	No
Category 2 - Screening Based on Potential for Safety Improvement						
IHSDM[3]	Yes	Yes	Yes	Yes	Yes	Yes
SafetyAnalyst[2]	Yes	Yes	Yes	Yes	Yes	Yes
HSM[1]	Yes	Yes	Yes	Yes	Yes	Yes

Source: [1] (American Association of State Highway and Transportation Officials, 2010b)
[2] (American Association of State Highway and Transportation Officials, 2010c)
[3] (Federal Highway Administration, 2010b)

In summary, network screening, the first step in road safety management process, could be performed using a number of methods – both traditional and advanced. Traditional methods use less data, but, are fraught with biases, issues and limitations. Regression-to-mean effect, false assumption of a linear relation between crashes and exposure, lack of predictive power, and inability to calculate and rank sites based on expected and predicted crashes are some of the major concerns which are mostly addressed by the advanced site selection methods. The advanced methods use longer periods of crash data and safety performance functions developed from numerous peer sites for calculating substantive safety and prioritizing sites. The intricate statistical procedures used by the advanced methods rank sites based on the expected crashes and also provides a measure of variance that explains the reliability of the results. Even though the general consensus of many transportation officials at state DOTs and local offices is that the newer advanced site selection methods have too stringent data requirements, it is feasible for the safety officials to use advanced methods on an incremental basis depending on data availability (American Association of State Highway and Transportation Officials, 2010b).

HSM provides analytical tools for quantifying effects of potential changes at individual sites (American Association of State Highway and Transportation Officials, 2010b). SafetyAnalyst software provides a suite of analytical tools to identify and manage system-wide safety improvements (American Association of State Highway and Transportation Officials, 2010c). Both the federal projects were developed to address two diverse aspects of road safety. SafetyAnalyst is considered to be companion software to Highway Safety Manual, yet SafetyAnalyst is designed for more system-wide analysis, and HSM is better suited for site specific analysis – although HSM can be used for statewide analysis, but the data needs are great. HSM and SafetyAnalyst working together constitute a more comprehensive set of safety improvement tools for an agency.

Although HSM includes procedures for both traditional and advanced site selection methods, greater emphasis is given to the Empirical Bayes (EB) approach using Safety Performance Functions (SPFs). The EB method addresses the issues and limitations of traditional methods and, identifies and prioritizes sites based on the potential for safety improvement (PSI) (Ogle & Alluri, in review). It also provides measures to determine the reliability of predictions.

2.3 Safety Performance Functions:

In comparison to the past, the present and future safety of a roadway is of primary importance (Harwood et al., 2000). Therefore, network screening based on performance of sites in the past alone might not be a true measure of safety. Level Of Service of Safety (LOSS) and Empirical Bayes (EB) approaches are some of the very few site selection and prioritization methods that predict the future expected performance of a site (Kononov & Allery, 2003); (Hauer, Harwood, Council, & Griffith, 2002).

According to the Empirical Bayes approach, prioritization of “problematic sites” is based on their expected safety performance on the roadways which could be calculated by comparing the site’s past and present safety performance with that of sites with similar characteristics. This led to the concept of Safety Performance Functions (SPFs), first introduced by Hauer in 1995, representing a relation between crash frequency and exposure (usually traffic) (Hauer, 1995). “A Safety Performance Function (SPF) is a mathematical function that describes the relationship between the number of crashes per year and the measure of exposure (usually AADT but hourly flow rate by direction is more significant (Qin, Ivan, Ravishanker, & Liu, 2005).” (iTRANS Consulting Ltd & Human Factors North INC, 2003). These are generally used with the Empirical Bayes method to predict the expected safety performance based on the historical and existing trends in crash data. Besides EB analysis, LOSS also uses SPFs in identification of problematic sites. Although it is widely accepted that the use of SPFs aid in “better” identification of problematic sites, it is also understood that they are not readily available to be included in safety analysis requiring statistical expertise, and, reliable and comprehensive data for their development.

The expected crash frequency on a roadway depends on many factors like traffic, functional classification of roadway, area type, number of lanes, lane width, presence and width of median, presence, width and type of shoulder, horizontal and vertical curves etc (American Association of State Highway and Transportation Officials, 2010b)(). Thus, an ideal SPF shall consider all or most of the above mentioned variables while predicting the crash frequency at a site. Collection and maintenance of all the data variables that influence the expected crash frequency is a humungous task for the states which is further aggravated by lack of funding and good data collection infrastructure (Ogle, 2007)(WSDOT, 2010).

Attaining consensus among states about the significant effect of considerably varying roadway characteristics on expected crash frequency, safety researchers across the nation have agreed on the influence of average annual daily traffic (AADT) on the expected crashes on a roadway (Qin, Ivan, & Ravishanker, 2004)(Hauer, 1995). Nevertheless, researchers have also identified and acknowledged the influence of various factors in the safety performance of a roadway. SPFs are therefore divided into two broad categories: fundamental SPFs and all-inclusive SPFs. Fundamental SPFs predict the relation between crashes and traffic for roadway segments with varying characteristics (crashes are normalized on per mile per year basis) and intersections (crashes are normalized on per year basis) (Harwood, Torbic, Richard, & Meyer, July, 2010). All-inclusive SPFs use a base set of conditions and consider the effects of varying roadway characteristics that influence the expected performance of a site through crash modification factors (CMFs) (American Association of State Highway and Transportation Officials, 2010b).

A more statistical term for SPF is a regression equation “that relate crash experience to the traffic and other characteristics of locations” (Persaud, 2001) Highway safety analysis includes development of either crash prediction models (statistical models that “estimate the safety of a location as a function of variables found to be the best predictors”) or crash causation models (models used to relate factors that explain crash causation to crashes). Various statistical techniques are often used in this area for the generation of the above mentioned models of which the following are most frequently used: Log-linear analysis, contingency table analysis, induced exposure/risk estimation, logit models, ordered probit models, logistic models, meta analysis, factor analysis, and data imputation (Persaud, 2001).

Considering the types of models used specifically for accident-frequency studies, Poisson regression models (which consider the dependant variables to be discrete, positive and random) have been shown to be more appropriate than conventional linear regression models (Poch & Mannering, 1996). Poisson distribution, though frequently considered, has a limitation of variance equals mean, which is often not observed with crash data. This is well addressed by negative binomial (NB) regression analysis as it accounts for “extra-Poisson variation due to other variables not included in the model”

(Dean & Lawless, 1989; Vogt & Bared, 1998). Though NB doesn't require variance to be equal to mean, it measures overdispersion (presence of greater than expected variability in predictions), which occurs when variance is greater than mean (Poch & Mannering, 1996), (Shankar, Mannering, & Barfield, 1995),(Kononov, Bailey, & Allery, 2008), (Hauer, 2001; Kononov et al., 2008).

This overdispersion parameter, K, in the negative binomial distribution has been reported in different forms by various researchers. For example, in the report Validation of Accident Models for Intersections by Washington, S et.al, K is associated with variance as:

$$(Var\{m\} = E\{m\} + K * E\{m\}) \quad \text{Equation 1}$$

Where:

Var {m}= the estimated variance of the mean accident rate;

E {m} = the estimated mean accident rate from the model; and

K = the estimated overdispersion constant. ((Washington, Persaud, Lyon, & Oh, 2005))

From the above equation, as the overdispersion gets larger, variance increases, and consequently all of the standard errors of estimates become inflated. As a result, all else being equal, a model with smaller overdispersion (i.e., a smaller value of K) is preferred to a model with larger overdispersion. ((Washington et al., 2005))

Differing slightly from the above discussion, Hauer (Hauer, 2001) stated that when a constant overdispersion parameter is applied to all road sections, the maximum likelihood estimate of parameters will be over-influenced by short segments and thus, leading to inconsistency in EB estimates. Overdispersion per unit length would address to the above noted issues (Hauer, 2001). Reinforcing this conclusion, the overdispersion parameter used in Highway Safety Manual for two-way two-lane rural roads is a function of length ($= \frac{0.236}{segment\ length\ in\ miles}$). (American Association of State Highway and Transportation Officials, 2010b).

2.3.1 Fundamental SPFs:

As mentioned earlier, fundamental SPFs calculate the predicted number of crashes based on AADT for each type of roadway segment and intersection. It is illogical to group all roadway sections into one category and all intersections into another category while performing safety analysis due to their varied characteristics. Therefore, for example, for roadway segments, depending on functional classification, area type, number of lanes, presence/ absence of median, interchanges etc, roadways are broadly classified into various subtypes like rural two-lane roads, multilane undivided and divided urban roadways, etc and a SPF is generated for each subtype and multiple crash severity type.

The general form of the equation used with respect to roadway segments is:

$$N = e^{\alpha} * ADT^{\beta} \quad \text{Equation 2}$$

Where:

N = Predicted number of target crashes per mile per year;

ADT = Average Daily Traffic (veh/day);

α, β = Regression Coefficients.

The general form of the equation used with respect to intersections is:

$$N = e^{\alpha} * MajADT^{\beta 1} * MinADT^{\beta 2} \qquad \text{Equation 3}$$

Where:

N = Predicted number of target crashes per intersection per year;

MajADT = Average Daily Traffic on major road (veh/day);

MinADT = Average Daily Traffic on minor road (veh/day); and

α, β_1, β_2 = Regression Coefficients.

Note: The regression coefficients α, β, β_1, β_2 are different in equations 1 and 2 and are dependent on the specific relation between crashes and traffic for each site subtype.

2.3.2 All-inclusive SPFs:

In fact, traffic is not the sole predictor of roadway safety. Hence, all-inclusive SPFs that consider the effects of various roadway characteristics that influence the expected performance could potentially result in better estimations of the predicted crashes as more variables are used to explain the trend. When a number of factors are considered, the equation to predict the crash frequency on a roadway segment is:

$$N = e^{\alpha} * ADT^{\beta} * V1^{\beta 1} * V2^{\beta 2} * V3^{\beta 3} * V4^{\beta 4} * \ldots * Vn^{\beta n} \qquad \text{Equation 4}$$

Where:

N = Predicted number of target crashes per mile per year;

ADT = Average Daily Traffic (veh/day);

α, β, β_1, $\beta_{2\ldots}$ β_n= Regression Coefficients;

V_1, V_2, … V_n = Independent variables (or roadway characteristics that influence crash frequency).

An example of an all-inclusive SPF is (Zegeer, Reinfurt, Hummer, Herf, & Hunter, 1986):

$$A = 0.0019 * ADT^{0.8824} * (0.8786)^{W} * (0.9192)^{PA} * (0.9316)^{UP} * (1.2365)^{N} * (0.8822)^{TER1} * (1.3221)^{TER2}$$

Where:

A = number of crashes per mile per year;

ADT = two directional average daily traffic;

W = lane width in feet;

PA = Width of paved shoulder in feet;

UP = Width of unpaved shoulder in feet;

H = median roadside hazard rating;

TER1 = 1 for flat terrain, 0 otherwise;

TER2 = 1 for mountainous terrain, 0 otherwise.

Similar to the above model, there are a number of models representing various types of roadways and intersections and also considering various factors and roadway features as influencing variables in predicting target crash frequency. Researchers steered clear of using these inclusive models for national safety analysis programs because they tend to only represent one area well. Many researchers have developed SPFs using various variables for different site subtypes and crash severity levels. For example, (Wang, Hughes, & Stewart, 1995) had developed an equation for predicting annual crashes on rural multi-lane highways based on many roadway characteristics.

A more generalized and complete SPF was generated in the recent past which introduced the concepts of "base conditions" and "crash modification factors (CMFs)". Base conditions are defined as "a specific set of geometric design and traffic control features" (American Association of State Highway and Transportation Officials, 2010b). Base SPFs are generated using a subset of the entire data whose geometric design and traffic control features align with the pre-defined "base conditions". Due to minimum variations within the features, the base SPFs address to greater variability within the crash data. The predicted crash frequency at a site is calculated by adjusting the predicted frequency calculated using base conditions to the site specific and local conditions using CMFs and calibration factors respectively (American Association of State Highway and Transportation Officials, 2010b). For two-lane rural highways, following is the general equation to calculate the predicted crash frequency.

$$Npredicted = Nspf * \ Cr * (CMF1r * \ CMF2r * \ CMF3r * ... * \ CMF12r \qquad \text{Equation 5}$$

Where:

$N_{predicted}$ = predicted average crash frequency for an individual roadway segment for a particular year;

N_{spf} = predicted average crash frequency for base conditions for an individual roadway segment;

C_r = Calibration factor for roadway segments of a specific type developed for a specific agency; and

$CMF_{1r} \ldots CMF_{12r}$ = Crash Modification Factors for roadway segments.

Crash Modification Factors (also known as Crash Reduction Factors) are "used to adjust the SPF estimate of predicted average crash frequency for the effect of individual geometric design and traffic control features" (American Association of State Highway and Transportation Officials, 2010b).

2.4 Advanced safety analysis methods:

Crash frequencies, rates, and safety indices are some of the many site selection methods that are termed as "traditional methods" as they have minimum data requirements and do not address to various serious issues like regression-to-mean effect, bias toward either low volume / high volume roads (depending on the ranking method) and their incapability of predicting the frequency and severity of crashes in the future. As discusses earlier, the advanced methods (primarily Empirical Bayes approach) successfully address these issues and limitations. The Highway Safety Manual, released in July 2010 gives a step-wise guidance to the use of EB methodology.

This approach is automated in a software package called SafetyAnalyst, which is capable of performing various steps in the roadway safety improvement process.

Although HSM includes procedures for both traditional and advanced site selection methods, greater emphasis is given to the Empirical Bayes (EB) approach using Safety Performance Functions (SPFs). The EB method addresses the issues and limitations of traditional methods and, identifies and prioritizes sites based on the potential for safety improvement (PSI) (American Association of State Highway and Transportation Officials, 2010b). It also provides measures to determine the reliability of predictions.

SafetyAnalyst is a state-of-the-art analytical tool for making system wide safety decisions. It has many modules within itself and could act as a complete “safety toolbox” for any safety office. The modules in SafetyAnalyst include (American Association of State Highway and Transportation Officials, 2010c):

Network screening module: It identifies and ranks sites with potential for safety improvements.

Diagnosis and countermeasure selection module: Diagnosis 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 accident frequency and severity at specific sites.

Economic appraisal and priority ranking module: 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.

Countermeasure evaluation module: It provides the capability to conduct before/after evaluations of implemented safety improvements.

SafetyAnalyst software has a data management tool, analytical tool, administration tool and implemented countermeasure tool to perform the complete roadway safety management process. The data management tool is used to import, post process and calibrate data. The analytical tool is used to perform analysis on the data. All the modules of SafetyAnalyst discussed earlier could be performed in this tool. Administrative tool is used to perform a variety of tasks like adding and removing data items (with an exception of mandatory data elements). Data re-coding of various data elements' attributes could also be performed and saved. This tool also gives access to the national default SPFs used in the analysis and could also be replaced with agency specific SPFs, if available.

CHAPTER 3 **METHODOLOGY**

The approach towards this research is taken in six phases. The following flowchart gives an overview of various phases.

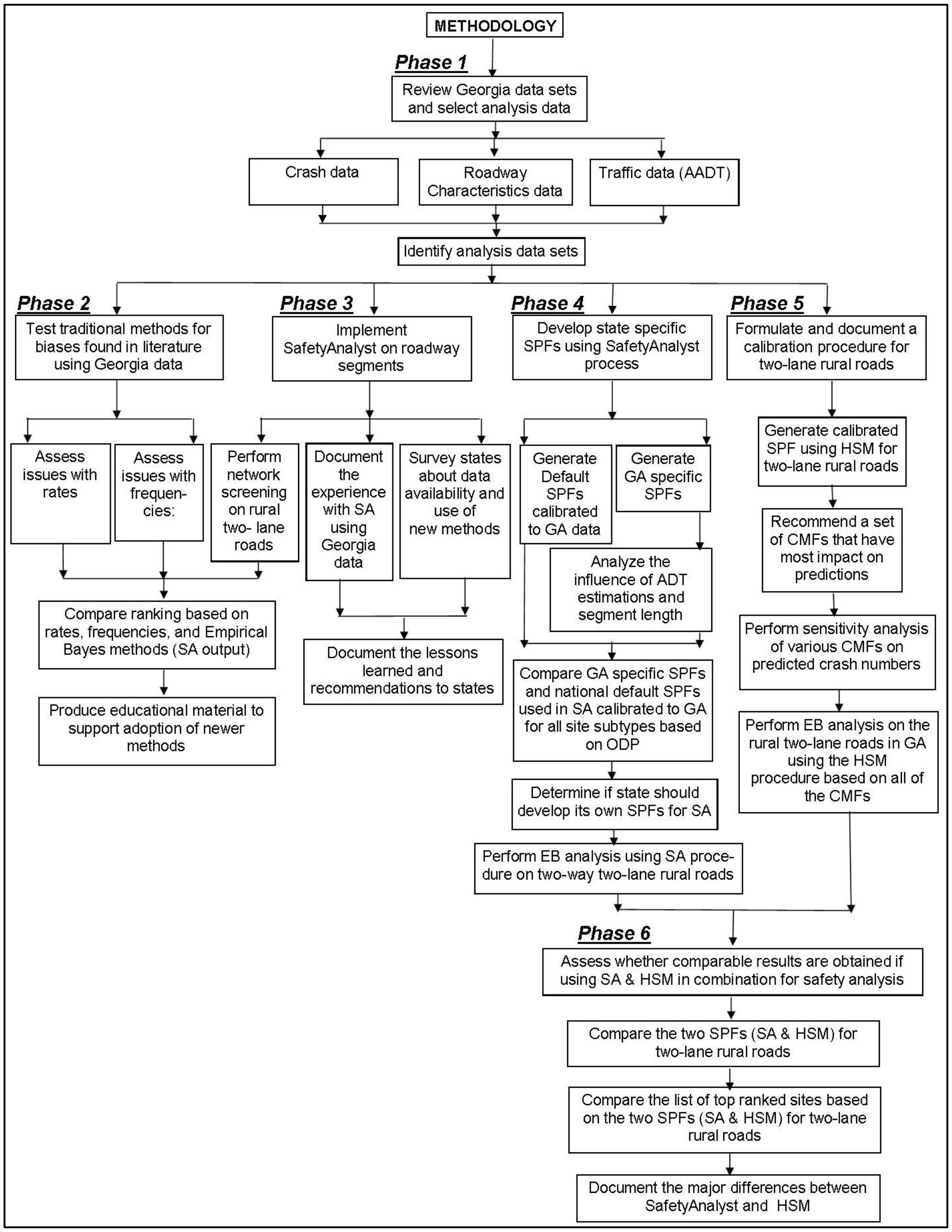

Figure 5: Various phases and steps in research methodology

Following are the six main phases in this research:

1. Review Georgia datasets and select analysis data
2. Test traditional methods (crash frequencies and crash rates) using Georgia data for biases stated in the literature
3. Implement SafetyAnalyst on roadway segments
4. Develop state specific SPFs using the methodology used to develop default SPFs that are used within SafetyAnalyst and determine if state should develop its own SPFs to use with SafetyAnalyst
5. Formulate and document calibration procedure for two-way two-lane rural roads using the HSM approach
6. Assess whether comparable results are obtained if using SafetyAnalyst and Highway Safety Manual in combination for safety analysis

Phase 1:

3.1 Review Georgia datasets:

For the present study, the following datasets were reviewed and analyzed.

- Crash data for the years 2004-2006
- Roadway characteristics data (snap shot from December 2007)
- Spatial reference to the roadway characteristics data (snap shot from 2007)
- Both actual and estimated traffic data for all the roadway segments for 2004-2006

Analysis was carried out only on roadway segments excluding all intersections and ramps. The following sections briefly describe the methods followed to achieve the objectives of this research.

A number of software tools were used in this study. Following is the list and brief description of the tools used:

a. *SafetyAnalyst:* SafetyAnalyst is “a set of state-of-the-art software tools for use in the decision-making process to identify and manage a system-wide program of site-specific improvements to enhance highway safety by cost-effective means” (American Association of State Highway and Transportation Officials, 2010a).

Usage: SafetyAnalyst was used throughout this research to identify the data requirements for advanced methods, and for identifying and prioritizing sites based on Empirical Bayes method using both national default SPFs calibrated to Georgia data and the Georgia specific SPFs. The software was also used to merge shorter segments into longer aggregated segments based on predefined criteria.

b. *SAS:* Statistical Analysis Software is one of the many commercially available statistical software packages that can perform regression analyses.

Usage: In this research, SAS was used to perform negative binomial regression analysis to estimate the regression coefficients for Georgia specific SPF model development for 17 site subtypes. It was also used to assess the fit of SPFs by calculating Freeman Tukey’s R square. T test to compare the expected crashes calculated using SafetyAnalyst and HSM was also performed using this software package.

c. *ArcGIS 10:* A geographic information system (GIS) “integrates hardware, software, and data for capturing, managing, analyzing, and displaying all forms of geographically referenced information” (ESRI,).

Usage: ArcGIS was used in the initial stages of the research to obtain a spatial reference to the roadway characteristics database. Later on, it was used to assess some of the issues with the data which are discussed in the later sections of this dissertation. It was also used extensively to determine the location of the segments identified in plan and profile sheets for HSM calibration.

d. *Microsoft Access:* Microsoft Office Access 2007 is a relational database management system (RDBMS) used to maintain databases and to create simple database solutions.

Usage: MS Access was used in the initial stages of the research to generate the data files required for advanced methods.

e. *Microsoft SQL Server 2008:* SQL is also a RDBMS used to maintain large databases and to perform complex operations on the data sets.

Usage: During the later stages of the project, all the files were transferred from MS Access to SQL server due to the limitations of Access. Various operations like assigning crashes to roadway segments, generating import files for SafetyAnalyst, generating aggregated segments, and identifying segments with base conditions were performed using this software package.

f. *Microsoft Excel:* Excel is a spreadsheet application used to perform basic calculations and to generate graphs.

Usage: Excel was used throughout the project, mainly to display the SPFs in graphical form and to perform various steps in the Empirical Bayes method to identify and prioritize sites. Excel was also used to apply the Crash Modification Factors (CMFs) while calculating calibration factors, used within the HSM procedure.

The following three sections review descriptive statistics and provide explanations of individual data files (crash, roadway characteristics, and traffic) used in this research.

3.1.1 Crash data:

Crash data, for the years 2004-2006, was obtained from GDOT. Two sets of crash data were used: One contains very detailed information of the crash (non spatial database), and the other contains spatial reference to most of the crashes (spatial database). The non-spatial crash database consisted of 1,032,263 crashes while the spatial crash database consisted of 1,032,446 crashes. However, due to data coding and other issues, there was a slight discrepancy between the two databases. It was found that 99.51% of the reported crashes were spatially located for the complete study period. About 0.5% of the reported crashes were missing spatial location. In the similar manner, about 0.5% of the spatially referenced crashes were missing in the non-spatial database. Figure 6 gives the summary of the above discussion.

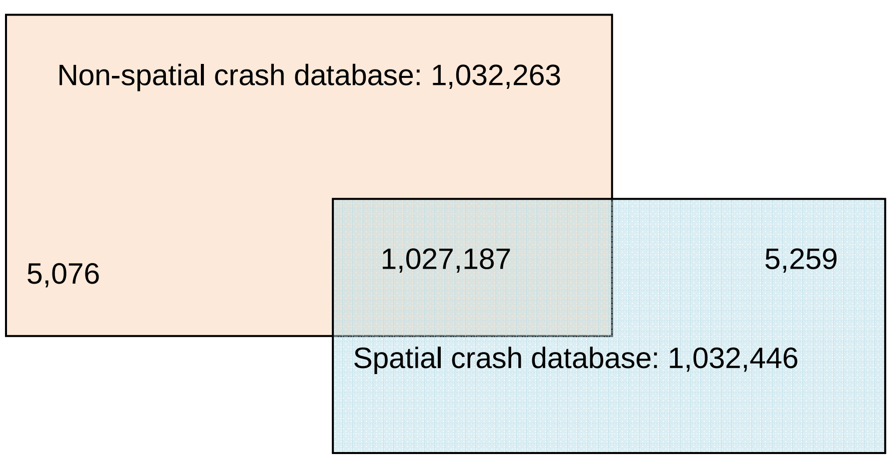

Figure 6: Summary of crashes found in crash database and spatially located for the years 2004-2006 in Georgia

As this study is focusing only on roadway segments, all the crashes that were related to intersections were excluded from further analysis. Based on earlier research, crashes that occur within 200 ft from an intersection were treated as “intersection-related” crashes. Therefore, a 200 ft buffer was created around the intersections in ArcGIS and all the crashes within the buffer were coded as intersection related crashes and excluded. About 56.9% of total crashes were excluded; a total of 442,233 crashes were identified as segment related.

3.1.2 Roadway characteristics and associated GIS data:

Georgia Department of Transportation (GDOT) maintains two different files associated with roadway inventory data. One is the base shape file or Location Referencing System (LRS) file and the other is a Roadway Characteristics file (RC).

The LRS file is a shape file compatible with ArcGIS and has attribute data stored in a .dbf (dbase) format to be used with other database management systems (DBMS). It consisted of 153,308 records. Each record is a specific route and has a unique ID, the “RCLINK”. RCLINK id consists of ten digits. The first three digits represent the county number, followed by one digit representing route type and the last six digits represent the route name. The RCLINK id is used to associate LRS file with detailed roadway characteristics of the RC file. Each record in the LRS file has an RCLINK and the length of the route in addition to various other data variables.

Each route (with a unique RCLINK) in LRS is divided into smaller segments consisting of similar roadway characteristics. 121,915.17 miles of roadway network in Georgia is divided into 884,598 roadway segments with an average length of 0.138 miles. Implying, one or more roadway characteristics change every 0.138 miles. Each roadway segment in the RC file has an RCLINK (id), beginning milepost, and an ending milepost. A unique ID was generated to identify each unique roadway segment. The unique ID consisted of 20 digits starting and ending with an alphabet ‘B’, maintaining the alpha/numeric nature of the ID. It has the RCLINK, followed by the beginning milepost (represented by four digits), and the ending milepost (represented by four digits).

Roadway Characteristics data is an MS Access database and has no spatial reference attached to it. To obtain a spatial dimension to the RC data, a process called “Dynamic Segmentation” was used in ArcGIS. To perform this, a new project in ArcGIS was created and RC text file was imported into ArcGIS. Based on LRS data, a spatial reference was attached to this file by adding route events (by going to Tools → Add Route Events). The segments were added along each RCLINK based on beginning milepost and ending milepost. Following is the screen shot of this step.

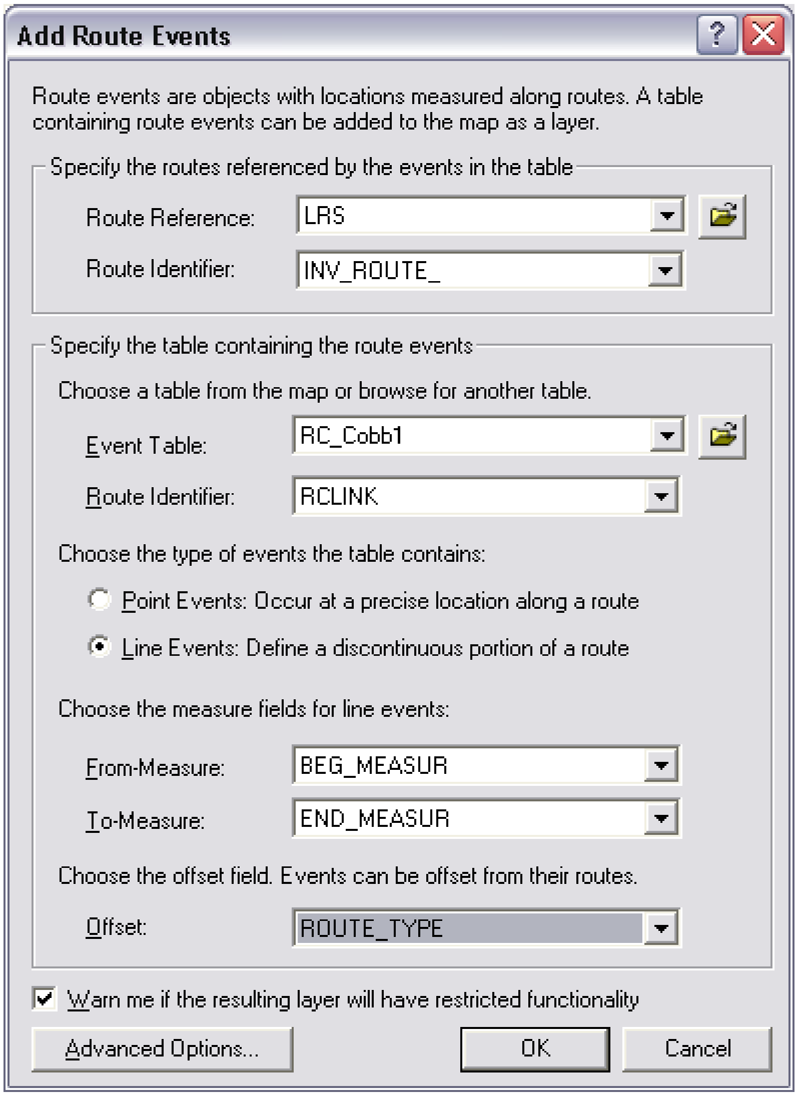

Figure 7: Add Route Events dialogue box in ArcGIS

3.1.3 Traffic data:

In addition to the roadway characteristics and crash data, traffic operations data is also required. At a minimum, average annual daily traffic (AADT) information for all the roadway segments and for the years 2004-2006 is required. AADT information was obtained for the years 1995-2007 from GDOT. Data for the years 2004-2006 was queried and retained for further analysis. A considerable amount of data processing was performed to link the traffic data to the RC data. In this process, it was found that the traffic data file is incomplete in various aspects. There were a considerable number of roadway segments that have no traffic data (20,295 segments summing to ~6,736 miles) and a number of segments that have missing traffic data for a year or two within the analysis period of three years (3253 segments summing to ~4,939.03 miles).

Given the practical limitations of collecting traffic data, on all the roadway segments, Georgia estimates traffic information on substantial number of segments based on the actual count data for proximate and similar road segments. These counts are also adjusted for seasonal variations. To determine the effect of actual versus estimated AADT in SPF development, a specific dataset constituting of actual traffic data was obtained from GDOT. It is observed that, of 121915.17 miles of total roadway network, actual traffic data was collected on only 28,479.97 miles (~23.36%) while the rest was estimated.

Phase 2:

3.2 Test traditional methods for biases stated in the literature:

Network screening based on basic selection methods (crash frequency, crash rate, and critical crash rate) was performed and various issues with these methods were identified.

Approximately 884,598 records representing about 121,915.17 miles of road network with an average segment length of 0.138 miles was used in the analysis. It was observed that about 66% of the segments are shorter than 0.1 miles.

While most of the states use raw segment lengths defined by changes in various roadway characteristic (RC) data elements, such segmentation using Georgia data produces extremely small segments. This is due to the finer levels of detail recorded for various roadway characteristic data elements (that is, lane width and shoulder width recorded at 0.1ft increments). A detailed discussion about the effect of shorter segments on network screening is undertaken in the next chapter.

Researchers attempted to match 442,233 segment related crashes over the three year period to their respective segments. Of these, 435,230 crashes (98.4%) were successfully identified on the road segments. 6,972 crashes were not assigned to any segments due to missing/ incorrect data and 31 crashes were found to have coding errors.

The following analyses were performed to identify the pros and cons of crash frequency, crash rate, and EB approach. This step is important because it is clear from the five-percent reports that many states are still using these tools as their primary analysis methods. Conducting a test of the traditional methods is intended to provide information for educational materials for senior managers to encourage them to shift to more advanced methods.

a. Test for known biases in traditional methods
    i. Frequencies – bias toward high volume roads and longer segments
    ii. Rates – bias toward low volume roads and shorter segments
b. Compare differences in ranking outcomes between crash frequency, crash rate, critical crash rate, and EB approach using SafetyAnalyst for two-lane rural roads

c. Compare differences in ranking outcomes between longer aggregated and shorter disaggregated segments

3.2.1 Description of methods used in the analysis:

For manual analysis, initial analysis table was generated with segment information, traffic data, and number of crashes on each segment. For each segment, crash frequency, crash rate, and critical crash rate were calculated. Two sets of roadway segment files were prepared: shorter unmodified segments and longer aggregated segments (generated with only required data elements and predefined ranges as discussed in section 3.2.2). Each method was implemented twice, once for each set of segments. Following is a detailed discussion of each analysis method, and generation of aggregated segments.

3.2.1.1 Crash Frequency:

Segments were sorted based on crash count per year in descending order and ranked. With this method, the site with highest per year crash count was ranked number 1 and the site with second highest per year crash count was ranked number 2 and so on.

3.2.1.2 Crash Rate:

The ratio between crash count and exposure is termed as "crash rate". Exposure (EXPO) in million vehicle miles of travel (MVMT), was calculated using the formula,

$$EXPO = \frac{AADT * 365 * 3 * total\ segment\ length}{1{,}000{,}000} \qquad \text{Equation 6}$$

Where, 3 is the number of years for which crash data is available.

$$Crash\ rate = \frac{Three\ year\ crash\ count}{EXPO} \quad \text{Equation 7}$$

The calculated crash rate was sorted in descending order. The site with highest crash rate was ranked number 1, and the site with second highest crash rate was ranked number 2 and so on.

3.2.1.3 Critical Crash Rate:

Critical crash rate for a set of sites is calculated using the formula:

$$Rci = RA + Kc * \sqrt{\frac{RA}{EXPO}} + \frac{1}{2 * EXPO} \quad \text{Equation 8}$$

Where:

Rci = Critical crash rate for site i;

RA = Average crash rate for each reference population;

Kc = 1.645 (the probability constant based on the confidence interval of 95%); and

EXPO = Million vehicle miles of travel.

The difference between the crash rate for each site obtained from Equation 7 and the critical crash rate obtained from Equation 8 was calculated and sorted in descending order. The site with highest positive difference was ranked number 1 and the site with second highest positive difference was ranked number 2 and so on. However, sites are ranked only if their observed crash rate is greater than the critical crash rate. It is to be noted that critical crash rate is calculated only for a set of similar sites. Therefore, segments need to be sub-classified into site subtypes prior to performing this analysis.

3.2.2 Generation of longer aggregated segments:

Shorter segments are merged into considerably longer segments, known as aggregated segments while preserving the varying characteristics to the required detail. Aggregated segments are generated in two ways: by considering fewer data elements in defining the segments, and by reducing the sensitivity of data. First, a considerable increase in segment length is achieved by including only the required data elements in defining a roadway segment. Second, it is observed that greater sensitivity in data elements might not be necessary if the thresholds used during the analyses are less sensitive. In the state of Georgia, variables such as lane width, shoulder width, and median width are recorded to the tenth of a foot. However, it is observed that these variables are mostly used in calculating crash modification factors - CMFs (also known as crash reduction factors) to adjust for the base conditions in Empirical Bayes approach (American Association of State Highway and Transportation Officials, 2010b). These CMFs were generated based on 1 ft variations for lane width and 2 ft variations for shoulder width. Therefore, changes of 0.1 ft need not be maintained, as the variable will not be analyzed at this level of detail. For this study, the sensitivity of these variables has been reduced to 1 ft or 2 ft increments to increase the segment length. Therefore, the longer aggregated segments were developed by considering only the required minimum data elements and lesser sensitive data. Results from the longer aggregated segments were compared against traditional rankings using typical segmentation based on all roadway characteristics at finite thresholds.

Phase 3:

3.3 Implement SafetyAnalyst on roadway segments:

As discussed in the earlier chapters of this dissertation, network screening is the fundamental step in highway safety improvement process. For this research, SafetyAnalyst, one of the advanced safety analysis tools, had been used to perform network screening, to identify and prioritize problematic sites based on Empirical Bayes method. The following paragraphs briefly describe the procedure used to generate and import, post process, and calibrate various files in SafetyAnalyst.

3.3.1 Generate import files for SafetyAnalyst:

SafetyAnalyst requires three separate files to be imported in a particular format in order to perform the network screening analysis: AltAccident file, AltRoadwaySegment file, and AltSegmentTraffic file.

a. Accident file:

Microsoft SQL was used to generate the import files. The SQL queries used are included in Appendix A. Only the variables that are required for analysis by SafetyAnalyst were included in the data files. Following are the variables included in the crash database:

Table 3: Data elements that are included in accident file to be imported into SafetyAnalyst

ACCIDENT FILE variable list			
Agency ID	Accident Date	Junction Relationship	v1 Vehicle Configuration
Loc System	Accident Time	Light Condition	v2 Vehicle Configuration
Route Type	Accident Severity1	Weather Condition	v1 Initial Travel Direction
Route Name	Number of Fatalities	Surface Condition	v2 Initial Travel Direction
County	Number of Injuries	Collision Type	v1 Vehicle Maneuver
Loc Offset	Number of Vehicles	Road Condition	v2 Vehicle Maneuver

SafetyAnalyst has a stringent set of enumeration values for each data element. Georgia has a completely different coding structure and therefore, most of the data elements were re-coded either within the Georgia file or within the Administration tool of the software depending on practicality. In addition, some of the mandatory SafetyAnalyst elements required merging data from multiple fields and/or elements in the Georgia datasets. The data mapping guide is shown in Appendix B. The generated AltAccident file was saved in comma separated value (csv) format and it consisted of 442,233 records. The first record in the file had to be the file name and so, 'AltAccident' was added in the first row to indicate file name.

b. Roadway Segment file:

Similar to the Accident file, the road segment file was generated using a set of SQL queries (documented in Appendix A). Only the minimum variables that were required for analysis by SafetyAnalyst were included in the data files. Following are the variables included in the roadway characteristics database:

Table 4: Data elements that are included in roadway segment file to be imported into SafetyAnalyst

ROADWAY SEGMENT FILE variable list		
Agency ID	Roadway Class1	Median Width
Loc System	Num of Thru Lanes in direction1	Start Offset
Route Type	Num of Thru Lanes in direction2	End Offset
Route Name	medianType1	Section Length
County	Access Control	
Area Type	Operation Way	

The file was saved in .csv format and it consisted of 884,598 records. The first row in the file has to be the file name and so, 'AltRoadwaySegment' was added to the first row.

c. Segment Traffic file:

Similar to accident and roadway segment files, SQL queries were run to generate the import file for segment traffic data. The set of queries used are detailed in Appendix A. Agency ID, calendar year, and AADT were the three variables included in this file. Similar to the other files, the file name, 'AltSegmentTraffic' was added to the first row.

<u>3.3.2 Import, post process, and calibrate the input files in SafetyAnalyst:</u>

Various versions of the SafetyAnalyst were used over a period of two years to implement the network screening module of the software (since the work was conducted during continued development of the software). SafetyAnalyst consists of four tools: Data management tool, analytical tool, administration tool, and implemented countermeasure tool. The data management tool is used to import, post process, and calibrate the state's data. The administration tool is used to add/ remove/ change data variables (not all data variables could be changed), recode, add/ remove enumeration values, change SPFs, and perform many other similar functions. The analytical tool is used to run the various modules of the software which are similar to the various steps in safety management process. The implemented countermeasure tool is used to perform the benefit cost analysis of implemented countermeasures.

Once the required data files were generated, the data management tool was opened and the three files were imported. The time required to import the datasets was dependent on the processor speed and the number of applications that were simultaneously run on the system. After importing was performed, a log file was generated with detailed information about the errors and warnings. After completing the import process without major errors, post processing was done. The analysis period had to be defined in this step. For this project, data from the years 2004-2006 was being analyzed. A concept called “Homogeneous Segments” is introduced in this step. SafetyAnalyst has a capability of merging one of more consecutive roadway segments together into one depending on the homogeneity of roadway characteristics. The minimum thresholds for generating these homogeneous segments were set and post processing was started. Figure 8 shows the screenshot of the window for editing and viewing threshold limits for homogeneous segment aggregation.

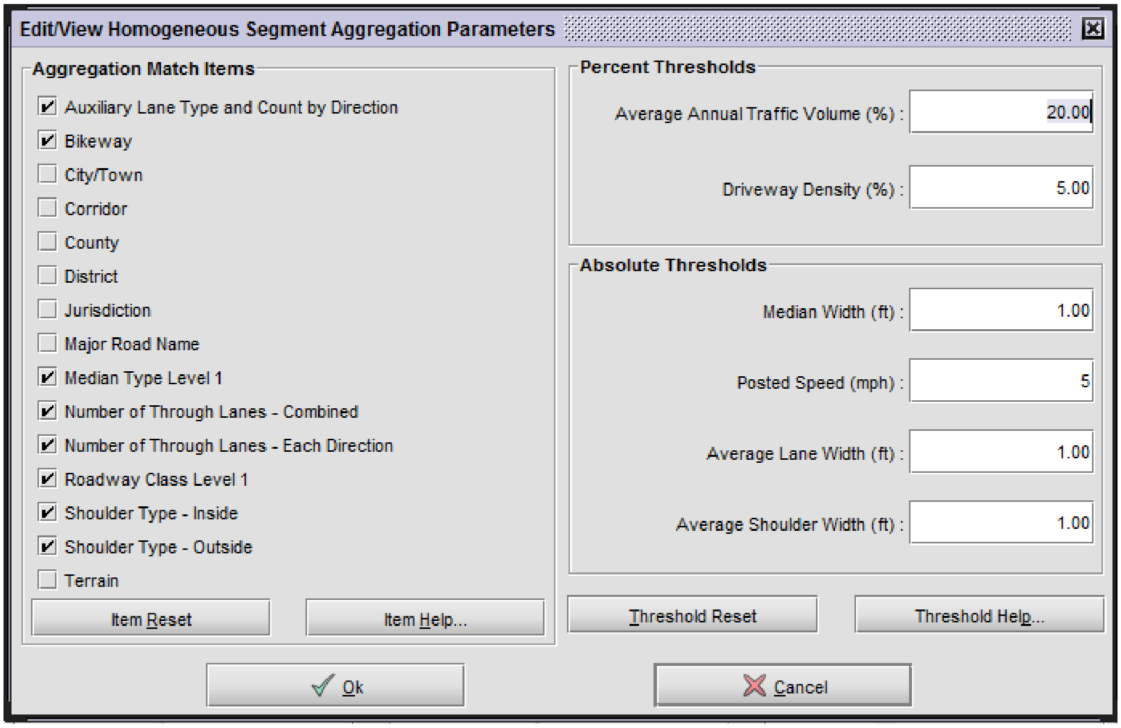

Figure 8: Screenshot of the Edit/View Homogeneous Segment Aggregation Parameters and their threshold limits

After the import and post process steps, calibration needs to be carried out. This step generates calibration factors (the ratio of the observed crashes (actual number of crashes occurred in Georgia) to the predicted crashes (number of crashes predicted using national default SPFs)) for each year to address the variability due to factors like weather, driver population, changes over time, travel behavior etc. Once the import process was completed without major errors, the three files were exported. Each roadway segment was allocated a specific site subtype based on the roadway characteristics. However, there exist a few roadway segments that do not belong to any of the pre-defined site subtypes and therefore, were excluded from further analysis. These excluded segments were found to be special cases such as segments with reversible lanes, segments with one way truck routes, one way during school hours etc.

These segments were not included as specific SPF information for such scenarios is unavailable.

Detailed descriptive statistics about the errors and warnings, and the imported files are given in the results chapter of this dissertation.

3.3.3 Run the administration tool in SafetyAnalyst:

In the administration tool, there are three databases: federal, agency, and system. The federal database is an embedded database that is distributed with the administration tool. It contains federal default site subtype definitions, default values for the countermeasures, diagnostics, and the default national SPF coefficient data along with the national averages for the crash distributions. The agency database is the repository for all agency-specific data and agency modifications to the federal default data. The system database is populated by merging the federal default data and the agency-specified data (American Association of State Highway and Transportation Officials, 2010c). It is almost always required to alter the agency data and therefore the system database needs to be regenerated frequently.

Due to the rigid list of data variables and enumeration values used in SafetyAnalyst, states are required to recode a number of data variables. The process of recoding can be done either within the import files using SQL queries or in the administration tool of the software. In most of the cases, recoding and data manipulation for Georgia was conducted within the administration tool. All data recoding and data mapping is documented in Appendix B.

As mentioned in the earlier chapters, SafetyAnalyst uses Empirical Bayes approach to perform network screening, the first module of roadway safety management process. The safety performance functions (SPFs) used in the various modules of SafetyAnalyst were accessed through the administration tool of the software. By default, SafetyAnalyst uses the national SPFs generated with northern and western states' data, calibrated to Georgia data. The calibration process is automated within the software and is discussed in the results section.

Table 5: States and the years of data used to generate the default national SPFs used in SafetyAnalyst (Harwood et al., July, 2010)

State	Years of data used
California	1997 to 2001
Minnesota	1995 to 1999
Ohio	1997 to 1999
Washington	1993 to 1996

Two versions of default SPFs are available: one for total crashes (spf_{TOT}), and one for fatal and injury crashes (spf_{FI}), for the three types of roadways (segments, intersections, and ramps), and for all subtypes in each roadway type. As this research deals with roadway segments alone, each of the various roadway segment site subtypes is listed here.

Table 6: Site subtype code and description used for roadway segments in SafetyAnalyst

Site Subtype Code	Site subtype description
101	Rural two-lane roads
102	Rural multilane undivided roads
103	Rural multilane divided roads
104	Rural freeways--4 lanes
105	Rural freeways--6+ lanes
106	Rural freeways within interchange area--4 lanes
107	Rural freeways within interchange area--6+ lanes
151	Urban two-lane arterial streets
152	Urban multilane undivided arterial streets
153	Urban multilane divided arterial streets
154	Urban one-way arterial streets
155	Urban freeways - 4 lanes
156	Urban freeways - 6 lanes
157	Urban freeways - 8+ lanes
158	Urban freeways within interchange area - 4 lanes
159	Urban freeways within interchange area - 6 lanes
160	Urban freeways within interchange area - 8+ lanes

In addition to the above mentioned site subtypes, separate agency specific site subtypes could also be generated and used in the analysis. In this case, the state would also need to have a SPF for the same. There is also a possibility of using alternative pre-existing SPFs for new or existing subtypes if it is believed that the data and the safety performance are similar.

If agency specific SPFs are developed, then, the default SPFs could be replaced in the administration tool of the software. In this case, the default SPFs in the administration tool are replaced by the agency specific SPFs, and the system database is populated by merging the federal and agency databases. Thus, states may replace none, one, several, or all of the default SPFs with state specific ones. However, it is not a requirement of SafetyAnalyst that the states provide their own SPFs.

3.3.4 Run the analytical tool in SafetyAnalyst:

Different modules in SafetyAnalyst could be performed in the analytical tool. For this research, only the network screening module was studied in depth. The ‘Getting Started Wizard’ walks users through the tool. When the network screening analysis module was selected, a new workbook was created to store the dataset that was generated in the data management tool. Site lists could be created and saved based on the user requirements. In addition, site lists could also be generated by selecting sites based on queries. For the present project, all the roadway segments were selected for analysis. The types of network screening available include:

- Basic network screening (with peak searching on roadway segments and CV test)
- Basic network screening (with sliding window on roadway segments)
- High proportion of specific accident type
- Sudden increase in mean accident frequency
- Steady increase in mean accident frequency

For this research, “Basic Network Screening with peak searching on roadway segments” method was performed using crash data for the three years. Total (Fatal, injury, and PDO) crashes for all the available years were considered. Potential for safety improvement (PSI) could be calculated based on either expected accident frequency or excess expected accident frequency. For this project, PSI was calculated based on expected accident frequency. Rural and urban areas were weighted equally. To prevent some of the roadway segments that have minimal crashes from being ranked highest, the crash frequency limiting values were set to 5 crashes/mile/year. The coefficient of variation (CV) for the roadway segments determines the number of sites to be included in the output report (the lower the CV limit, the fewer the sites displayed in the output report). CV limit was set to 0.50. The accident screening attribute, such as accident type and manner of collision, vehicle turning movement etc based on which the analysis had to be done was selected and for this analysis, accident type and manner of collision were selected, and all the values were selected within the attribute. This step was performed to make sure that all crashes were analyzed. Appendix C includes the screenshots of all the steps in the analytical module of SafetyAnalyst and a sample report is attached in Appendix D.

3.3.5 Interpret the SafetyAnalyst output:

SafetyAnalyst output consists of a number of columns which require a detailed description. The various columns in the output are explained in the following table:

Table 7: Various columns in the output from SafetyAnalyst

<table>
<tr><td colspan="2">ID</td><td>Roadway Segment ID</td></tr>
<tr><td colspan="2">Site Type</td><td>Whether Segment/ Intersection/ Ramp</td></tr>
<tr><td colspan="2">Site Subtype</td><td>Sub-categories in the site type</td></tr>
<tr><td colspan="2">County</td><td>County where the roadway segment is located</td></tr>
<tr><td colspan="2">Route</td><td>Route number of the roadway segment</td></tr>
<tr><td colspan="2">Site Start Location</td><td>Start location of the roadway segment</td></tr>
<tr><td colspan="2">Site End Location</td><td>End location of the roadway segment</td></tr>
<tr><td colspan="2">Average Observed Accidents for Entire Site*</td><td>Observed crashes for the entire site in crashes/mile/year</td></tr>
<tr><td rowspan="8">Location with Highest Potential for Safety Improvement</td><td>Average Observed Accidents*</td><td>Observed crashes for the roadway sub segment in crashes/mile/year</td></tr>
<tr><td>Predicted Accident Frequency*</td><td>Predicted crash frequency in crashes/mile/year</td></tr>
<tr><td>PSI Expected Accident Frequency*</td><td>PSI Expected accident frequency in crashes/mile/year</td></tr>
<tr><td>Variance**</td><td>Variance in crashes/square mile/ year</td></tr>
<tr><td>Start Location</td><td>Start location of the roadway sub segment where PSI is greater</td></tr>
<tr><td>End Location</td><td>End location of the roadway sub segment where PSI is greater</td></tr>
<tr><td>No. of Expected Fatalities</td><td>Total number of expected fatalities per mile per year</td></tr>
<tr><td>No. of Expected Injuries</td><td>Total number of expected injuries per mile per year</td></tr>
<tr><td colspan="2">Rank</td><td>Overall Rank based on PSI</td></tr>
<tr><td colspan="2">Additional Windows of Interest</td><td>Additional windows whose PSI exceeded the threshold limits, but the expected accident frequencies are between the limiting accident threshold and the highest calculated PSI for the site</td></tr>
</table>

* expressed as crashes/ mile/ year
** expressed as crashes/ mile^2/ year

3.3.6 Conduct survey to states about safety data availability, and use of new methods:

A review of the 2009 five-percent (transparency) reports submitted by the states to FHWA describing at least five percent of highway locations exhibiting the most pressing safety needs had indicated that most DOTs are still using traditional safety analysis measures such as frequency, rate, critical rate, or safety index. Two out of 50 states reported use of EB methods. With this information, the research path called for a survey regarding safety data, present safety analysis methods, use of advanced safety analysis tools, and implementation of newer tools. The survey was prepared and sent to the Safety Director of the Department of Transportation in each state. The questionnaire and the letter accompanying the survey are provided in Appendix H. In summary, the survey was divided into seven major parts:

1. Contact information;

2. General questions about safety data;

3. General questions about safety data analyses;

4. Questions about SafetyAnalyst;

5. Questions about safety performance functions;

6. Questions about SafetyAnalyst implementation; and

7. Questions about Highway Safety Manual implementation.

Only the states that have been working with advanced safety analysis tools like SafetyAnalyst were asked to answer the questions in parts 4-6.

Of the 50 states, 24 states completed the survey in full, and one state answered a portion of the survey. Responses for the answered questions from the incomplete survey were considered in the analysis. Survey responses are discussed in detail in the results chapter.

<u>Phase 4:</u>

<u>3.4 Develop state specific SPFs</u>

The default SPFs used within SafetyAnalyst were generated using northern and western states' data using the years 1993-2001. SafetyAnalyst calibrates the default SPFs to fit to the state's data. But, how well the calibrated SPFs fit the state's data is a point of concern. Therefore, state specific SPFs were generated and compared against the calibrated default SPFs. To maintain consistency and transferability between default national SPFs used in SafetyAnalyst and Georgia specific SPFs, similar logic that was used to generate the default SPFs was used to develop Georgia specific SPFs.

The functional form considered for roadway segments is:

$$k = e^{\alpha} * ADT^{\beta} \quad \text{Equation 9}$$

Where:

k = Predicted number of target crashes per mile per year;

ADT = Average Daily Traffic (veh/day) for roadway segments in both directions of travel; and α, β = Regression constants.

To obtain the predicted crashes per site per year, the formula used is:

$$N = e^{\alpha} * ADT^{\beta} * L \qquad \text{Equation 10}$$

Where:

N = Predicted number of target crashes per site per year; and

L = Length of the roadway segment in miles.

For this project, SPFs are generated for all site subtypes listed in Table 6.

The base equation is:

$$k = e^{\alpha} * ADT^{\beta}$$

Applying natural logarithm on both sides,

$$Ln\ (k) = intercept + coefficient * Ln(ADT)$$

$$k = e^{\{intercept + coefficient * Ln(ADT)\}} \qquad \text{Equation 11}$$

Statistical analysis software, SAS, was used to estimate the intercept and coefficient for 17 site subtypes. Data requirements for running SAS include:

- Roadway segment ID
- Site subtype
- Segment length in miles
- Natural logarithm of Average Annual Daily Traffic

- Offset
- Total number of crashes (TOT) occurring on each roadway segment during the period of analysis
- Total number of Fatal and Injury (FI) crashes occurring on each roadway segment during the analysis period

All the above variables are self explanatory except ‘Offset’ which is calculated as natural logarithm of the product of segment length and the number of years analyzed.

$$Offset = Ln\,(3 * segment\ length)$$

The SAS code used to estimate the regression coefficients is shown in Appendix E. The SAS code used to calculate Freeman Tukey’s R square value is documented in Appendix F.

<u>3.4.1 Analyze the influence of accuracy of traffic data on the development of SPFs:</u>

As shown in Equation 9, the functional form of the SPFs is $k = e^{\alpha} * ADT^{\beta}$. It implies that the crashes are predicted as a function of traffic alone. Therefore, the accuracy of the model, to a great extent relies on the accuracy of traffic data. With Georgia data, it was found that less than 25% of the total traffic data is actually counted in the field while the rest is estimated.

Considerable data cleaning was carried out prior to generating SPFs. Within the 209,636 aggregated segments, it was observed that 20,295 segments (~10%) have no traffic information, and 11,423 segments (~6% of the remaining segments) have missing traffic information for at least one of the three years. To include the segments with a year or two of missing data, the following procedure for estimation, recommended in the HSM was followed (American Association of State Highway and Transportation Officials, 2010b):

- If only one year of traffic data is available, the same value is assumed for the remaining two years.
- If two years of traffic data is available, value for the third year is calculated based on either interpolation or extrapolation. However, segments with unrealistic estimations (negative traffic volumes) were excluded from further analysis.

As the average traffic volume for the analysis period is considered, it is obvious that the yearly variations drastically influence the SPF and its fit. The influence of the variations in traffic data and the effect of traffic estimations were studied and the following methods were adopted.

3.4.1.1 Analyze the influence of AADT and segment length on SPF development:

The influence of variations in traffic data and their effects on SPF generation was studied. As the segment length also plays a vital role in predicting crashes, influence of minimum segment length on the fit of SPFs was also studied. Table 8 gives an idea about the data relating to two-lane rural roads in the context of traffic data and segment length.

Table 8: Statistics about two-lane two-way rural roads data relating to traffic and segment length

Description	Number of segments
Total number of two-way two-lane rural segments	70,167
Total segments with missing AADT value (excluded from analysis)	5,274
Total segments with adjusted AADT (at least AADT for a year is available and)	964
Total segments with unrealistic adjusted AADT (negative values – due to extrapolation of available AADT data)	61
Total segments shorter than 0.05 miles	2,504
Total segments shorter than 0.1 miles	7,283

In phase 1, examination of traffic data revealed extreme variations in the yearly traffic data. Thus, questions were generated about the reliability of the data and the model developed with the estimated AADT data. Due to the lack of significant research in acceptable yearly variations in traffic data, for each roadway segment, the ratio of maximum to minimum AADT value was calculated and thresholds on the acceptable ratio were defined. As discussed above, various datasets were categorized and analyzed to come up with acceptable variations. The goodness of fit of the models (or SPFs) generated from various datasets are discussed in the next chapter.

3.4.1.2 Analyze the influence of actual and estimated AADT data on the fit of the SPFs:

In Georgia, it is found that less than 25% of the traffic data is actually collected while the rest is estimated. To understand the effect of estimation on the fit of SPFs, separate datasets were prepared constituting of segments with measured AADT values and segments with both measured and estimated AADT values. Data cleaning was performed on both the datasets. Segments with null and unrealistic AADTs, and segments shorter than 0.1 miles were excluded from the analysis. The SAS software was run on all the datasets and R^2_{FT} values and overdispersion parameters were calculated for the 17 site subtypes. The results of this study are discussed in the next chapter.

3.4.2 Compare Georgia specific SPFs to the calibrated default SPFs used in SafetyAnalyst:

One of the main objectives of this research is to determine the need to develop agency specific SPFs in order to use advanced safety analysis methods. The basic default national SPFs, the national SPFs calibrated to Georgia data, and the Georgia specific SPFs were compared by assessing their goodness-of-fit. Calibrated SPFs were generated from the default SPFs by using a multiplying factor called calibration factor. The calibration factor is calculated as the ratio of total number of observed crashes to the total number of predicted crashes obtained from the default SPFs.

The following three SPFs were plotted and compared against the observed crash data:

a) Georgia specific SPFs: SPFs generated using Georgia data for all 17 site subtypes

b) Non-calibrated default SPFs used in SafetyAnalyst: Default national SPFs used within SafetyAnalyst *without* calibrating to Georgia data

c) Calibrated default SPFs used in SafetyAnalyst: Default national SPFs used within SafetyAnalyst calibrated to Georgia data

The interpretations and results are discussed in the following chapter. The Freeman Tukey's R^2 coefficient was used to determine the goodness-of-fit (Fridstrom, Ifver, Ingebrigtsen, Kulmala, & Thomsen, 1995). The following formulae were used for calculating Freeman Tukey's R^2 coefficient (R^2_{FT}).

$$\mathrm{R2FT} = 1 - \frac{\sum \hat{\mathrm{e}}\mathrm{i}^2}{\sum(fi - \bar{\mathrm{f}})^2} \qquad \text{Equation 12}$$

Where:

$$fi = \sqrt{(yi)} + \sqrt{(yi+1)} \qquad \text{Equation 13}$$

The statistic is approximately normally distributed with mean,

$$\Phi\mathrm{i} = \sqrt{(4\hat{\mathrm{y}}\mathrm{i}+1)} \qquad \text{Equation 14}$$

The deviation of the Freeman Tukey's Coefficient is estimated by the corresponding residual

$$\hat{\mathrm{e}}\mathrm{i} = \sqrt{(yi)} + \sqrt{(yi+1)} - \sqrt{(4\hat{\mathrm{y}}\mathrm{i}+1)} \qquad \text{Equation 15}$$

In the above equations,

y_i is the observed number of crashes at site i;

$\hat{y}_i$ is the mean of the observed number of crashes at all sites similar to site i;

f_i is the value obtained from Equation 13; and

$\bar{f}$ is the average of all the f for sites considered (Fridstrom et al., 1995).

R^2_{FT} values were calculated for both the calibrated SPFs used in SafetyAnalyst and for the Georgia specific SPFs for all site subtypes. The results are discussed in the next chapter.

3.4.3 Identify base conditions for Georgia data and generate SPFs using base conditions for two-way two-lane rural roads:

The default SPFs used within SafetyAnalyst were generated using data from California, Ohio, Minnesota, and Washington. The analysis datasets were not limited to the base conditions, that is, the complete road network for all site subtypes was considered for performing negative binomial regression and for generating default national SPFs. However, this is not the procedure used within HSM for generating base default SPFs. It is believed that SPFs generated from segments with base conditions result in better fit as the influence of varying roadway characteristics are minimized. To understand the effect of the same, base conditions for two-way two-lane rural roads were identified and the SPF was generated. Its fit was compared to the calibrated SPFs used in SafetyAnalyst and Georgia specific SPFs. The base conditions identified and the SPFs generated using base conditions are discussed in the results chapter.

Phase 5:

3.5 Formulate and document calibration procedure for two-way two-lane rural roads as illustrated in the Highway Safety Manual:

3.5.1 Background of HSM procedure:

The Highway Safety Manual, released by AASHTO in July 2010, provides "analytical tools and techniques for quantifying the potential effects on crashes as a result of decisions made in planning, design, operations, and maintenance" (American Association of State Highway and Transportation Officials, 2010b). HSM explains the step-by-step procedure to perform EB analysis, also known as the predictive method, both at a site and at a project level.

There are three basic elements required to perform the predictive methods:

- Safety performance functions (SPFs): As discussed in earlier sections, a SPF establishes the relation between crashes and exposure, generally, exposure being AADT. These SPFs are called "base SPFs" as they are used to estimate the crash frequency of certain types of roadway with specified base conditions. The base conditions considered within the HSM are discussed later in this section.
- Crash modification factors (CMFs): CMFs are defined as the ratio of the effectiveness of one condition in comparison to the other condition. CMFs need to be calculated for various roadway features, if they deviate from the predefined "base conditions". The safety performance of a roadway is affected by various roadway characteristics like lane width, shoulder width, presence of horizontal and vertical curve, etc. These CMFs when multiplied by the predicted crash frequency obtained using the

base SPFs account for the difference between the existing site conditions and specified base conditions.

- Calibration factor (C): A calibration factor is calculated as the ratio between the total observed crashes and the total predicted crashes. This factor mainly addresses the differences between the jurisdiction and the time period for which the base models were developed to the present jurisdiction and the time period for which they are being applied. A calibration factor greater than 1.0 implies that these roadways, on average, experience more crashes than the roadways used in developing the SPFs. And, a value lower than 1.0 implies that these roadways, on average, experience fewer crashes than the roadways used in developing the SPFs.

Given the three basic elements, it is possible to determine predicted crashes at a site using the following formula:

$$Npredicted = Nspf * \ Cr * (CMF1r * \ CMF2r * \ CMF3r * \ldots * \ CMF12r) \qquad \text{Equation 16}$$

Where:

$N_{predicted}$= Predicted number of crashes in crashes per year;

N_{spf} = Predicted number of crashes in crashes per year determined for base conditions;

CMF_r = Crash Modification Factors for various roadway characteristics; and

Cr = Calibration factor to adjust for differences in jurisdiction and time period.

3.5.2 HSM Calibration procedure:

There are 121,915.17 miles of road network in Georgia, of which, over 65% (~80,000 miles) are two-way two-lane rural roads. According to the HSM, small sections of rural two-way two-lane segments need to be identified randomly and their geometrical information recorded to calculate the calibration factor which is later used to calculate the predicted crashes using the available default SPFs. Prior to using predictions, users are recommended to calibrate model to existing conditions. To perform calibration, the Highway Safety Manual recommends collecting geometric, traffic, and roadway inventory data for at least 30-50 sites having a minimum of 100 crashes/ year.

Data availability is considered to be the toughest hurdle faced with Georgia plan profile information. Plan profile sheets are available only for segments which were revisited and improved at some point in the last 100 years and not for all the roadway segments in the state. So, the amount of available data within GDOT is limited, thus limiting the initial candidate dataset from which the segments for calculating calibration factors are to be randomly selected. Various data sources and tools used for data retrieval include plan profile sheets from GDOT website, Google maps, Google Earth, and ArcGIS. Plan profile sheets were obtained from an internal plan/profile server at GDOT. Plan profile sheets were accessed using a look up form requiring county information.

To ensure the sample would be representative of the state, a random selection process had to be established. The sample selection had to be as random as practical. Using this procedure, care was taken such that a county was given neither lesser nor greater weight depending on the total two-lane two-way rural miles within the county. The cumulative number of miles of rural two lane roads in each county was calculated and about 150 random numbers between 0 and 79,585.52 were generated (where 79,585.52 are the total miles of two-way two-lane rural roads in Georgia). These were considered as the mile numbers and the counties associated with each random mile number were identified. This procedure helps in giving due representation of each county based on the county's total number of rural two lane miles. From the GDOT website, the number of projects GDOT had undertaken in the past in each county was recorded and random numbers were generated. For example, if county 'X' has 123 projects, at least 25 random numbers were generated between numbers 1 and 123. And, if county 'X' has a cumulative mileage of rural two lane roads from 11,592.25 miles to 18,572.87 miles, and if two of the random numbers generated initially were 12,000 and 14,234.98. Then, two projects in the county 'X' will be identified in accordance with the random numbers generated for the county.

For a project to be flagged for further review, the following conditions need to be fulfilled:

- The project needs to have plan profile sheets
- The project location needs to be identified accurately in GIS
- The segment needs to be a two-way two-lane rural road
- The segment needs to be of a considerable length to represent the geometric features of the roadway segment (~2 miles).

Once the project fulfills all the above mentioned criteria, the plan profile sheets were downloaded and the segment was divided into horizontal curves and straight tangent sections. As some of the projects are considerably longer, only a sub section of the complete project length is considered. Approximately, the first two – four mile section of the project was considered.

The following table shows the data variables recommended by HSM and the source from which the information is obtained:

Table 9: Data variables that need to be collected to perform calibration and their data source

Data variable	Data source
Segment length*	Plan and profile sheets
RCLINK+	LRS file and GIS database
Beginning mile post+	Plan and profile sheets
Ending mile post+	Plan and profile sheets
Yearly AADT	AADT database
Length of horizontal curve*	Plan and profile sheets
Length of tangent*	Plan and profile sheets
Radii of horizontal curve*	Plan and profile sheets
Presence of spiral transition for horizontal curves	Plan and profile sheets
Superelevation variance for horizontal curves	Plan and profile sheets
Percent grade	Plan and profile sheets
Total # of crashes*	Accident database
Lane width*	Roadway characteristics (RC) database
Shoulder type*	Roadway characteristics (RC) database
Shoulder width*	Roadway characteristics (RC) database
Presence of lighting	Google Earth and Google maps
Driveway density	Google Earth and Google maps
Presence of passing lane	RC database and Google maps
Presence of short 4-lane section	RC database and Google maps
Presence of center TWLTL*	RC database and Google maps
Presence of centerline rumble strips	Google Earth and Google maps
Roadside hazard rating	Google Earth and Google maps
Use of automated speed enforcement	None

*Variables required by the Highway Safety Manual
+ Variables identified within GDOT database for researcher's convenience and easy mapping

For performing calibration, the Highway Safety Manual recommends collecting data for 30-50 sites with at least 100 crashes/ year. For this project, about 52 segments were identified with a total of 302 crashes over a period of three years. The average segment length was 1.94 miles.

All the required information was gathered from various sources as shown in Table 9. For two-way two-lane rural roadway segments, 12 crash modification factors (CMFs) are required to adjust the base SPFs to account for differences between the base conditions and the local site conditions. The following paragraphs discuss in detail the calculations relating to each CMF ((American Association of State Highway and Transportation Officials, 2010b)). Table 10 gives a list of various CMFs and their base conditions.

Table 10: Various CMFs required to calibrate two-way two-lane rural roads and their base conditions

CMF	CMF variable	Base condition
1	Lane width*	12 feet lanes
2	Shoulder width and type*	6 feet paved shoulders
3	Horizontal curves: Length, Radius and presence. Absence of spiral transitions	None
4	Horizontal curves: Superelevation	None
5	Vertical grades	0%
6	Driveway density	5 driveways/mile
7	Centerline rumble strips	None
8	Passing lanes	None
9	Two-way-left-turn lanes	None
10	Roadside hazard rating	3
11	Lighting	None
12	Automated speed enforcement	None

*Only related crashes are effected and hence required to be adjusted to total crashes

3.5.2.1 CMF1r – Lane width

Research has proven that variations in lane width effect only a certain type of crashes. CMF for lane width first calculates the effect of lane width on related crashes (CMF_{ra}) and later is adjusted to the total crashes based on the proportion of total crashes constituted by related crashes (p_{ra}). CMF_{ra} was obtained from table 10-8 of the HSM and later adjusted to total crashes. CMF_{ra} was calculated using Table 11.

Table 11: CMF for lane width on related crashes (CMF_{ra})

Lane width	AADT (veh/day)		
	< 400	400 to 2000	> 2000
≤ 9 ft	1.05	$1.05+2.81*10^{-4}$(AADT-400)	1.50
10 ft	1.02	$1.02+1.75*10^{-4}$(AADT-400)	1.30
11ft	1.01	$1.01+2.5*10^{-5}$(AADT-400)	1.05
≥ 12ft	1.00	1.00	1.00

Equation 17 was used to calculate the crash modification factor for the effect of lane width on total crashes from the CMF of lane width on related crashes and the proportion of total crashes constituted by related crashes.

$$CMF_{1r} = (CMF_{ra} - 1.0) * P_{ra} + 1.0 \qquad \text{Equation 17}$$

Where:

CMF_{1r} = Crash Modification Factor for the effect of lane width on total crashes;

CMF_{ra} = Crash Modification Factor for the effect of lane width on related crashes. The related crashes include single-vehicle run-off-the-road and multiple-vehicle head on, opposite direction sideswipe and same direction sideswipe crashes.

P_{ra} = Proportion of total crashes constituted by related crashes.

3.5.2.2 CMF_{2r} – Shoulder width and type

The crash modification factor for shoulder width and type has two components: Shoulder width (CMF_{wra}), and shoulder type (CMF_{tra}). The variations in shoulder width and type effect only a certain type of crashes and therefore, needs to be adjusted to total crashes. CMFs for shoulder width and shoulder type were calculated separately and then combined using Equation 18. The values were then used in a formula along with the proportion of total crashes constituted by related crashes (p_{ra}) to obtain the adjusted and combined CMF for total crashes. The crash types related to variations in shoulder width and shoulder type are similar to those related to variations in lane width.

The unadjusted CMF for shoulder width (CMF_{wra}) was obtained from table 10-9 of the HSM (as shown in Table 12) and later used in calculating the final combined CMF for shoulder type and width.

Table 12: CMF for shoulder width on related crashes (CMF_{wra})

Shoulder width	AADT (veh/day)		
	< 400	400 to 2000	> 2000
0-ft	1.10	$1.10+2.5*10^{-4}(AADT-400)$	1.50
2-ft	1.07	$1.07+1.43*10^{-4}(AADT-400)$	1.30
4-ft	1.02	$1.02+8.125*10^{-5}(AADT-400)$	1.15
6-ft	1.00	1.00	1.00
≥ 8-ft	0.98	$0.98+6.875*10^{-5}(AADT-400)$	0.87

The unadjusted CMF for shoulder type (CMF_{tra}) was obtained from table 10-10 of the HSM (as shown in Table 13) and later used in calculating the final combined CMF for shoulder type and width.

Table 13: CMF for shoulder type based on shoulder width on related crashes (CMF_{tra})

Shoulder type	Shoulder width (ft)						
	0	**1**	**2**	**3**	**4**	**6**	**8**
Paved	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Gravel	1.00	1.00	1.01	1.01	1.01	1.02	1.02
Composite	1.00	1.01	1.02	1.02	1.03	1.04	1.06
Turf	1.00	1.01	1.03	1.04	1.05	1.08	1.11

The obtained CMF was adjusted to total crashes using Equation 18.

$$CMF_{2r} = (CMF_{wra} * CMF_{tra} - 1.0) * P_{ra} + 1.0 \quad \text{Equation 18}$$

Where:

CMF_{2r} = Crash Modification Factor for the effect of shoulder width and type on total crashes;

CMF_{wra} = Crash Modification Factor for related crashes based on shoulder width. The related crashes include single-vehicle run-off-the-road and multiple-vehicle head on, opposite direction sideswipe and same direction sideswipe crashes;

CMF_{tra} = Crash Modification Factor for related crashes based on shoulder type. The related crashes include single-vehicle run-off-the-road and multiple-vehicle head on, opposite direction sideswipe and same direction sideswipe crashes; and

P_{ra} = Proportion of total crashes constituted by related crashes.

3.5.2.3 CMF_{3r} – Horizontal curves:

Length, radius, and presence/ absence of spiral transitions play an influential role in calculating the CMF for horizontal curves. The formula used to calculate the CMF for horizontal curves is as follows:

$$CMF_{3r} = \frac{(1.55 * Lc) + \left(\frac{80.2}{R}\right) - (0.012 * S)}{(1.55 * Lc)} \quad \text{Equation 19}$$

Where:

CMF_{3r} = Crash Modification Factor for the effect of horizontal alignment on total crashes;

Lc = Length of horizontal curve (in miles) which includes spiral transitions, if present;

R = Radius of curvature (in feet); and

S = 1 if spiral transition curve is present; 0 if spiral transition curve is not present; 0.5 if spiral transition curve is present at one end of the horizontal curve.

3.5.2.4 CMF_{4r} – Horizontal curves: Superelevation

The value of this CMF is calculated based on the superelevation variance of a horizontal curve. Superelevation variance is the difference between the actual superelevation and the superelevation identified by AASHTO policy.

With Georgia data, it is assumed that the superelevation identified by AASHTO policy is used, resulting in a superelevation variance of 0.0

Therefore, CMF_{4r} = 1.00 for all roadway segments.

3.5.2.5 CMF_{5r} – Vertical grades

The following table is used to determine the value of CMF_{5r}.

Table 14: CMF for vertical grade of roadway segments (CMF_{5r})

Approximate grade (%)		
Level grade (≤ 3%)	Moderate terrain (3% < grade ≤ 6%)	Steep terrain (>6%)
1.00	1.10	1.16

3.5.1.6 CMF_{6r} – Driveway Density

Five or fewer driveways per mile are considered in the base condition. A higher number needs to be adjusted using the following formula:

$$CMF_{6r} = \frac{0.322 + DD * [0.05 - 0.005 * ln(AADT)]}{0.322 + 5 * [0.05 - 0.005 * ln(AADT)]} \qquad \text{Equation 20}$$

Where:

CMF_{6r} = Crash Modification Factor for the effect of driveway density on total crashes;

AADT = Average Annual Daily Traffic in vehicles/day; and

DD = Driveway density considering driveways on both sides of the highway (driveways/mile).

When the driveway density is lower than 5 driveways per mile, CMF_{6r} was considered to be 1.00

3.5.2.7 CMF_{7r} – Presence of centerline rumble strips

None of the roadways were found to have centerline rumble strips and hence, the default CMF value of 1.00 was used.

3.5.2.8 CMF_{8r} – Presence of passing lanes

A CMF of 0.75 for total crashes for a roadway with a passing lane was used. In the absence of a passing lane, the default CMF value of 1.00 was used. When short four-lane sections were present, a CMF of 0.65 was used.

3.5.2.9 CMF_{9r} – Presence of two-way left-turn lanes

The formula used to determine the CMF for the presence of two-way left-turn lane is given below:

$$CMF_{9r} = 1.0 - \left(0.7 * P_{dwy} * P_{LT/D}\right) \qquad \text{Equation 21}$$

Where:

CMF_{9r} = Crash Modification Factor for the effect of two-way left-turn lanes on total crashes;

DD = Driveway density (driveways per mile);

$P_{LT/D}$ = Left-turn crashes susceptible to correction by a TWLTL as a proportion of driveway related crashes. An estimated value of 0.5 is used throughout.

P_{dwy} = Driveway related crashes as a proportion of total crashes which is calculated using the following equation

$$p_{dwy} = \frac{(0.0047 * DD) + (0.0024 * DD^2)}{1.199 + \ (0.0047 * DD) + (0.0024 * DD^2)} \qquad \text{Equation 22}$$

3.5.2.10 CMF_{10r} – Roadside design

“The Roadside Hazard Rating (RHR) system considers the clear zone in conjunction with the roadside slope, roadside surface roughness, recoverability of the roadside and other elements beyond the clear zone such as barriers and trees” (American Association of State Highway and Transportation Officials, 2010b). As the RHR increases from 1 to

7, the crash risk for frequency and/ or severity increases. It is used to determine the level of roadside design. The formula used to determine the CMF for roadside design is given below:

$$CMF_{10r} = \frac{e^{(-0.6869+0.0668*RHR)}}{e^{-0.4865}} \qquad \text{Equation 23}$$

Where:

CMF_{10r} = Crash Modification Factor for roadside design; and

RHR = Roadside hazard rating (A value between 1 and 7).

Sample pictures showing the seven RHR levels which are used as a basis for giving a RHR for Georgia roadways are in Appendix I (American Association of State Highway and Transportation Officials, 2010b).

3.5.2.11 CMF_{11r} – Lighting

None of the roadways were found to have lighting and therefore, the default base CMF value of 1.00 was used.

3.5.2.12 CMF_{12r} – Automated speed enforcement

Due to the limited use of automated speed enforcement in Georgia, automated speed enforcement was assumed to be absent and therefore, the default base CMF value of 1.00 was used for all calibrated segments.

3.5.2.13 Calculation of calibration factor

Once all the twelve CMFs for all the 52 segments were calculated, the calibration factor was computed using the formula:

$$Calibration\ Factor = \frac{\sum Observed\ crashes}{\sum Predicted\ crashes} \qquad \text{Equation 24}$$

The predicted crashes are obtained from the following formula:

$$Nspf = AADT * L * 365 * 10^{-6} * e^{(-0.312)} \qquad \text{Equation 25}$$

Where:

N_{spf} = Predicted total crash frequency per site for roadway segment base conditions;

AADT = Average Annual Daily Traffic in veh/day; and

L = Segment length in miles.

3.5.3 Analyze the effect of various combinations of CMFs:

Of the twelve CMFs, only 4 are mandatory. These mandatory CMFs include lane width, shoulder width and type, presence of horizontal curve, and presence of TWLTL. At this point of time, standard errors for CMFs are unavailable. Therefore, for this research, a standard error of 0.1 is assumed for all CMFs.

Even though, in theory, twelve CMFs need to be applied to address to all the variations between local data and base conditions, multiplication of all the 12 CMFs is not advisable as the standard error increases considerably with increase in the number of

CMFs used. With Georgia data, default values were used for four of the 12 CMFs (super elevation variance, centerline rumble strips, automated speed reinforcement, and lighting). The effect of both the individual and the combination of CMFs is studied by calculating the number of predicted crashes and the calibration factor for each scenario. The results of the analysis are discussed in the results chapter of the document.

3.5.4 Perform sensitivity analysis:

Sensitivity analysis was performed to illustrate the effect of variations of each CMFs on the total number of predicted crashes. Two types of sensitivity analyses were performed to assess the:

a) Effect of variation of AADT on the predicted number of crashes; and

b) Effect of variations of each CMF on the predicted number of crashes if means of all other CMFs are considered, and the effect of variations of each CMF on the predicted number of crashes if all other variables are considered equal to base conditions.

The results of the sensitivity analysis were presented in the next chapter.

3.5.5 Perform EB analysis on two-way two-lane rural roads using the HSM procedure:

The Highway Safety Manual documents the detailed steps to be performed to prioritize sites. Following are the various steps taken to calculate the expected crashes (American Association of State Highway and Transportation Officials, 2010b):

Step 1: Generate analysis dataset: From the roadway characteristics database, two-way two-lane rural roads were identified based on the following criteria.

Table 15: Criteria for identifying two-way two--lane rural roads

Data variable	Condition
Area type	Rural
Number of through lanes	≤ 3
Two-way vs. one-way operation	two-way

Step 2: Assign yearly AADT and crash data to the segments: MySQL was used to assign traffic data and crash data for the years 2004-2006 for all the segments.

Step 3: Determine the calibration factor to be used based on the analysis performed earlier on CMFs: Considering only the required CMFs (lane width, shoulder width and type, horizontal curve, and presence of a TWLTL), a calibration factor of 0.79 was used in the EB analysis.

Step 4: Calculate N_{spf} (predicted crashes for base conditions): N_{spf} was calculated using the following equation.

$$Nspf = AADT * L * 365 * 10^{-6} * e^{(-0.312)} * Calibration\ factor$$ Equation 26

Step 5: Calculate overdispersion parameter: Overdispersion parameter K was determined using the following equation.

$$K = \frac{0.236}{L}$$ Equation 27

Step 6: Calculate weighting to be applied to $N_{predicted}$ and $N_{observed}$ values: Weighting factor was calculated using the following equation.

$$w_{TOT} = \frac{1}{1 + k * (\sum_{y=1}^{Y} Nspf)} \qquad \text{Equation 28}$$

Step 7: Calculate $N_{\text{expected crashes}}$: The number of expected crashes per mile per year was calculated using the following equation.

$$Nexpected \ = Wt * Npredicted\ crashes + (1 - Wt) * Nobserved\ crashes \qquad \text{Equation 29}$$

Phase 6:

3.6 Compare Georgia specific SPFs, national default SPFs used in SafetyAnalyst calibrated to Georgia data, and the calibrated SPFs generated using HSM procedure for two-way two-lane rural roads:

The final phase of this research is to compare the SPFs used in SafetyAnalyst calibrated to Georgia data, and the SPFs used in HSM calibrated using Georgia data for two-way two-lane rural roads.

3.6.1 Compare the list of top ranked sites identified based on the SafetyAnalyst procedure and the HSM procedure:

The procedure used to calculate the number of expected crashes is slightly different in the HSM and SafetyAnalyst. The steps followed within SafetyAnalyst were followed with Georgia specific SPFs. Calculations used to perform EB analysis (and determine the number of expected crashes) using HSM procedure are shown in section 3.5.5

Following are the calculations used to perform EB analysis (and determine the number of expected crashes) using the procedure followed within SafetyAnalyst (Harwood et al., July, 2010).

Step 1: Generate analysis dataset: From the roadway characteristics database, two-way two-lane rural roads were identified.

Step 2: Assign yearly AADT and crash data to the segments: MySQL was used to assign traffic data and crash data for the years 2004-2006 for all the segments.

Step 3: Determine the yearly calibration factor to be used: Yearly calibration factors were calculated as the ratio of yearly observed crashes to yearly predicted crashes.

Step 4: Calculate yearly $N_{\text{predicted crashes}}$: $N_{\text{predicted}}$ for each year was calculated using the year specific AADT information and yearly calibration factor from the following equation.

$$Npredicted\ i = e^{-3.63} * AADTi^{0.53} * calibration\ factor\ i \qquad \text{Equation 30}$$

Where, i is the year for which the predicted crashes was calculated

Step 5: Calculate correction factors: Correction factors are calculated to correct for variations in the yearly predictions. The following equation was used for calculating yearly correction factors.

$$C_y = \frac{ky}{k1} \qquad \text{Equation 31}$$

Where:

C_y = Correction factor for year y relative to year 1;

k_y = Predicted crashes for year y; and

k_1 = Predicted crashes for year 1.

Step 6: Calculate weighting to be applied to $N_{predicted}$ and $N_{observed}$ values: Weighting factor was calculated using the following equation.

$$w_{TOT} = \frac{1}{1 + d_{TOT} \sum_{y=1}^{Y} \kappa_y * L} \qquad \text{Equation 32}$$

Where:

w_{TOT} = Weighting factor;

d_{TOT} = Overdispersion parameter;

κ_y = Predicted crashes for year y;

L = Segment length in miles; and

Y = total number of years in the analysis period

Step 7: Calculate the base EB adjusted expected number of crashes: The base EB adjusted expected number of crashes for year 1 was calculated using the formula:

$$X_1 = w_{TOT}\kappa_1 + \frac{(1 - w_{TOT})}{L} \frac{\sum_{y=1}^{Y} K_y}{\sum_{y=1}^{Y} C_y} \qquad \text{Equation 33}$$

Where:

X_1 = Expected number of crashes in year 1;

w_{TOT} = Weighting factor;

κ_1 = Predicted number of crashes in year 1;

K_y = Observed number of crashes in year y;

C_y = Correction factor for year y;

Y = total number of years in the analysis period; and

L = Segment length in miles.

Step 8: Calculate the EB-adjusted expected number of accidents for the last year Y: The EB-adjusted expected number of crashes for the last year Y was calculated using the following formula:

$$X_Y = X_1 * C_Y \quad \text{Equation 34}$$

Where:

X_Y = Expected number of crashes in year Y (last year of the analysis period); and

C_y = Correction factor for the last year Y.

Step 9: Calculate variance: Variance is used to obtain a measure of the precision of these calculated expected accident frequencies. Variance was calculated using the following formula:

$$\mathrm{Var}(X_Y) = X_Y * (1 - w_{TOT}) * \frac{C_Y}{\sum_{y=1}^{Y} C_y} \qquad \text{Equation 35}$$

Step 10: Ranking of segments: Segments are ranked based on their expected crash frequencies. The higher the expected crash count, the lower the rank (or in other words, the worse the site). Or, segments can also be ranked based on excess expected crash frequencies which is the difference between the observed crash count and the expected crash count. The higher the excess expected crash count, the lower the rank.

3.6.2 Assess whether comparable results are obtained if using SafetyAnalyst and the Highway Safety Manual in combination for safety analysis:

SafetyAnalyst is considered to be companion software to the Highway Safety Manual. Yet, SafetyAnalyst is designed for more system-wide analysis, and HSM is better suited for site specific analysis. It is expected that HSM and SafetyAnalyst working together would constitute a more comprehensive set of safety improvement tools for an agency. Once the high priority sites were identified by SafetyAnalyst, a site specific analysis using HSM procedure would be recommended. Therefore, both the results need to be comparable to gain the confidence of the practitioners.

In this step, results from SafetyAnalyst and HSM were compared and recommendations were made on the usage of the two newer tools.

CHAPTER 4 **ANALYSIS AND RESULTS**

The analysis for this project was completed in six phases. Following are the six phases:

1. Review Georgia data and compile analysis datasets - This phase sets the stage for performing the various analyses required to achieve the objectives of this research. It describes the compilation of the analysis datasets and uncovers potential problems and issues with crash data, roadway characteristics data, and traffic data along with recognizing data cleaning requirements.

Products:

a. Database of aggregated segments with associated yearly traffic data (3 years), observed crash information, and roadway characteristics.

b. Summary list of issues identified with crash, roadway characteristics data, and traffic data in preparation for advanced safety analyses.

2. Test traditional methods for biases found in the literature - This phase deals with the aggravated issues of both traditional and advanced methods when coupled with shorter segments (~0.01 mile segments). It also discusses the variations in rankings between traditional and advanced methods given different methods of compiling aggregated segments.

Products:

a. Definition of factors associated with inclusion of shorter disaggregated segments in the database.

b. Strategic methods for reducing data sensitivity without effecting the analysis and results.

c. Results tables comparing various site selection methods (frequency, rate, and EB method using SafetyAnalyst) by considering both shorter disaggregated and longer aggregated segments.

d. Educational PowerPoint presentation on the results of traditional methods and their biases intended to convince the safety managers of need to move toward newer and more advanced methods.

3. Implement SafetyAnalyst on roadway segments - This phase records the experience with implementing the network screening module of SafetyAnalyst from data collection to data formatting and handling data import errors. It is divided into three sub-phases.

i. Perform network screening on two-way two-lane rural roads in Georgia using SafetyAnalyst.

Products:

a. A list of top ranked two-way two-lane rural sites identified by EB analysis using SafetyAnalyst procedure.

b. Results table comparing various top ranked sites from previous phase with the output from SafetyAnalyst.

ii. Document the SafetyAnalyst implementation experience.

Products:

a. Definition of problems likely to arise while importing files into SafetyAnalyst.

b. A list of errors and warnings in the log files received while importing, post processing, and calibrating Georgia data in SafetyAnalyst and strategic approaches used to fix them.

c. Discussion on a list of issues identified after performing network screening analysis.

iii. Survey states about data availability and use of (or plans of using) newer safety analysis tools. Present results of the survey sent to states to determine their current stand with respect to data needs, data availability, data accuracy, and their willingness to shift to newer safety analysis tools. The section also includes the experience of various states currently working with SafetyAnalyst and/ HSM.

Products:

a. Summary of results of survey responses with specific context for inclusion of the questions. A report on the present stand of the states with respect to their safety data availability and accuracy.

b. A list of observations that encourage and discourage the deployment of the software.

4. Develop state specific SPFs using SafetyAnalyst process and compare with default SafetyAnalyst SPFs - In this phase, state specific SPFs were developed in accordance with the SafetyAnalyst procedure. The non calibrated default SPFs provided in SafetyAnalyst, default SPFs calibrated with Georgia data, and the SPFs generated using state specific data for all 17 site subtypes were compared based on overdispersion parameter. This phase also includes an analysis of the influence of actual and estimated AADT on the fit of SPFs.

Products:

a. Results table containing parameters for national default SPFs and Georgia specific SPFs for both total and Fatal/Injury crashes for the 17 site subtypes.

b. Results table comparing R^2_{FT} values and overdispersion parameters of default, calibrated, and Georgia specific SPFs for total and Fatal/Injury crashes for the 17 site subtypes.

c. Results table assessing the influence of actual and estimated AADT values on the fit of Georgia specific SPFs by comparing R^2_{FT} values and overdispersion parameters.

5. Formulate and document a more articulated calibration procedure for two-way two-lane rural roads than that provided in the Highway Safety Manual - This phase describes a procedure to calculate a calibration factor and various CMFs as per the procedure described in the HSM and analyzing the influence of multiple CMFs on crash predictions. Sensitivity analysis was performed to understand the effect of variations of each CMF on the calibration factor and also on the predicted number of crashes.

Products:

a. Detailed procedure used to randomly select two-way two-lane rural sites for calibration.

b. Results tables showing the sensitivity of predicted number of crashes to individual CMFs when all other variables are set equal to the Georgia conditions.

c. Results tables showing the sensitivity of predicted number of crashes to variations in each CMF when all other variables are assumed to be at base conditions.

d. Recommendation regarding the use of numerous CMFs during calibration process based on results of sensitivity analysis.

6. Assess whether comparable CMFs are obtained if the SPF for two-way two-lane rural roads was generated using HSM procedure - In this phase, the format/ coefficients of SPFs in SafetyAnalyst and HSM are compared along with overdispersion parameter. The EB procedure within SafetyAnalyst and HSM were performed to do the comparisons. The top sites from HSM procedure were compared to those identified by the calibrated SPF used in SafetyAnalyst, and the Georgia specific SPF.

Products:

a. Results table comparing overdispersion parameter of the three SPFs (Georgia specific SPF, calibrated default SPF used in SafetyAnalyst, and calibrated default SPF used in HSM).

b. Results table comparing the list of top ranked sites based on the two SPFs (calibrated default SPF used in SafetyAnalyst, and calibrated default SPF used in HSM).

c. Statistical test results with determination if a significant difference in predictions exist between SafetyAnalyst and HSM.

d. A list of the differences between SafetyAnalyst and HSM.

Phase 1:

4.1 Review Georgia datasets:

As mentioned in the methodology chapter, three datasets- crash data, roadway characteristics data, and traffic data are required to perform safety analyses. The following paragraphs explain various issues and potential problems along with recommended solutions with each of the datasets in detail:

4.1.1 Crash data:

Two crash databases maintained by GDOT include: 1) a crash database with information taken directly from the police crash report, and 2) a crash database with a spatial reference to the GDOT roadway network location reference system. It has been observed that there is a slight variation between the two datasets. However, about 99.5% of the crashes were identified in both databases and therefore the quality of the databases is considered to be acceptable. These two databases are combined to provide all of the details of the crash from the police report along with the spatial reference. The spatial reference can then be used to link the crash data with roadway characteristics data.

Since the crashes were linearly referenced along routes by a third party, it is nearly impossible to cross check whether a crash is correctly located or not. The crash location, in part, depends on the police perception noted in the crash report form. The spatial distribution of crash location appears to be reasonable for the most part. However, researchers found that a large number of crashes were located at 0.1 miles beyond the route start point. Thus, these sites may produce biased results in analysis if the crashes did not actually have occurred at these locations.

In the case of some divided highways, crashes were located only on one side of the roadway. This is mainly due to a missing direction code. The screenshot of an example is in Figure 9. As shown in this figure, all of the crashes were located on the I-75 North link, and none were located on I-75 South. While visually this doesn't look correct, this issue hasn't been addressed since most of the divided highways were identified as a single segment in the roadway characteristics database.

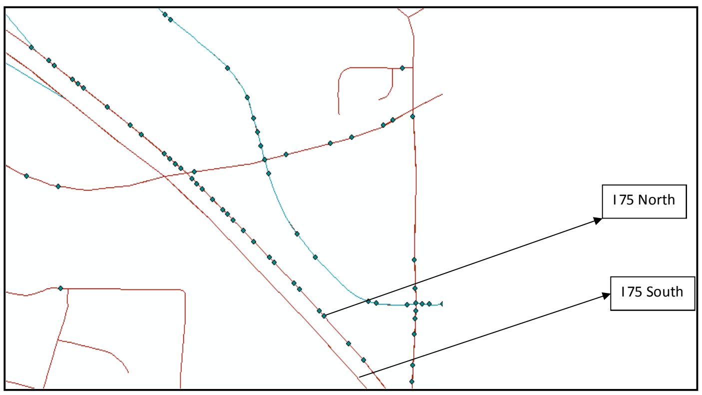

Figure 9: Crash data on divided highways do not have a direction code

4.1.2 Roadway Segment data:

It is unlikely for a state to maintain a single database representing the entire population of roadway inventory/design data for a state. When there are two or more databases pointing to (or identifying) the same roadway in the state, it is also unlikely for the various databases to perfectly overlap. Although this is an ideal situation, it is common to find duplicated links or fields with missing information. The completeness of the databases and interoperability of one database with the other play a vital role in safety analysis.

Like most states, the Georgia Department of Transportation (GDOT) maintains two different files associated with its roadway inventory. One is a base shape file or Location Referencing System (LRS) file and the other is a measure file which provides roadway characteristics info for continuously measured segments along the LRS shape. The LRS file consists of all the routes in Georgia. A spatial reference has been given to all the routes in LRS file based on route and milepost. The LRS file consists of 153,308 records. Some of the records in this file (41,153 or ~27%) have no spatial reference because of a coding error: the measure column (i.e. length of the route) has been noted as “0.00” even though each route has a length. This issue had to be fixed manually by obtaining the correct length from the shape file using ArcGIS. It was also observed that not all LRS links were distinct. Of the total 153,308 records, only 152,500 links were distinct. The redundancy was found to have no pattern. Table 16 illustrates the magnitude of these issues and comments on resolutions.

Table 16: Descriptive statistics for the LRS file

Issue/ description	Number of records	Explanation of the issue/solution
Initial total number of records in LRS file	153,308	
No spatial reference as the length marked is 0.00	41,153	Issue had been fixed by calculating the length of segments in ArcGIS
Redundant RCLINKs	808	No pattern and due to many reasons. These segments were removed from the database
Distinct RCLINKs	152,500	Final LRS file that is used for further analysis

The other database maintained by GDOT is an RC database (roadway characteristics database). Each LRS link is divided into multiple records in the RC file. Each record contains a segment with homogeneous characteristics. Every time there is a change in one of the 75 characteristics recorded in the RC file, a new record is established. A list of variables being collected and recorded in Georgia is given in Table 17.

Table 17: Data variables collected in Georgia

ROADWAY CHARACTERISTICS FILE variable list		
County	Div hwy shoulder type left	Right of way type
Route type	Div hwy surf width	TC number
Route num	Div hwy surf type	Maint. surface description
Beg measure	Div hwy shoulder width right	Sidewalk left
End measure	Div hwy shoulder type right	Sidewalk right
Section length	Div hwy median width	Improve type
Description	Div hwy median type	Truck percent
District	Div hwy barrier type	Truck percent type
Maintenance area	Udiv hwy shoulder width left	Signal
Population	Udiv hwy shoulder type left	AADT old
Inventory date	Udiv hwy surface width	HPMS id
Designated way	Udiv hwy surface type	Paces rating
Truck route	Udiv hwy shoulder width right	AADT
Travel way	Udiv hwy shoulder type right	Intersect road1
Area type	Aux lane width left	Intersect road2
Speed limit	Aux lane type left	S functional class id
FAS route number	Aux lane width right	Dual maint rating
Truck route id	Aux lane type right	Road width
Congress dist	Maintenance year	Divided
State route sequence	Maintenance type	Open to traffic
Access control	Improve year	City code
Operation	Functional classification	Total lanes left
Total lanes	Traffic count type	Total lanes right
Special class	Traffic count year	Land domain
Div hwy shoulder width left	Right of way	Rclink

In total, 152,500 LRS roadway sections were divided into 884,598 link segments in the RC database. Of which, 774,407 links have proper parent LRS records, resulting in 110,191 records in RC database without a spatial reference (i.e. with null LRS data). These records could not be corrected and were therefore not spatially located.

The zero length issue was also a point of concern for the RC database. It was found that, of 884,598 records, 167,703 records (i.e. 18.95%) have a recorded length of zero miles. Approximately 80.6% of the zero length segments were associated with the end of the LRS link. It is logical to assume that some of these records were at intersections. However, it was found that only 785 zero length segments were identified at intersections. The link length of RC was compared to that of LRS and was found that 11,798 records (i.e. 7% of total zero length segments) in RC database were at the end of parent link in the LRS file. These were corrected by equating their length (length of zero length segments in the RC file) to be the length of the sections in the LRS file. However approximately 74 % of the total zero length segments (i.e. 123,398 records) were found to be at the end of the parent link in the LRS file (i.e. the end measure of the segment in RC file is equal to the end measure of the parent link in the LRS file) and could not be corrected. About 15,730 records of the remaining 31,722 zero length segments were corrected by obtaining the correct section length from the LRS file. The remaining 9.5 % of the zero length segments were found to have no pattern and had to be excluded from further analysis. Table 18 gives the descriptive statistics for the roadway characteristics file.

Table 18: Descriptive statistics for the RC file

Issue/ description	Number of records	Explanation of the issue
Initial total number of records in RC file	884,598	
Length marked as '0.00 miles'	167,703	
Zero length segments @ intersections	785	Issue fixed by deleting these records
Zero length segments @ ends that were corrected as the LRS segment length is greater than the point location of RC segment	27,528	Issue fixed by comparing the link to the length of segments in LRS file
Zero length segments @ ends that could NOT be corrected	123,398	Issue couldn't be fixed
Zero length segments with no pattern	15,992	Issue couldn't be fixed
No LRS data	110,191	Either no link in LRS file or the length in LRS file is shorter than that in RC file
RC with proper LRS data	774,407	

During the review of the LRS file, several issues were uncovered and many were found to be associated with data entry. ArcGIS was used to identify some of the issues. In addition, many issues were identified purely from visual observations of the data, thus, there are likely additional issues that have not been identified as of yet. Following are some of the issues that were identified with LRS data using ArcGIS:

- Two records in the LRS file have the same RCLINK and different measures. Upon closer inspection it was noted that the two records make a continuous section. The issue could not be rectified through automation had to be manually verified on each and every link. The continuous sections with same RCLINK could be identified based on traveled way. The traveled way is coded as 1 or 2 depending on the link. The issue is identified in Figure 10 (ex: RCLINK: 0011001500).

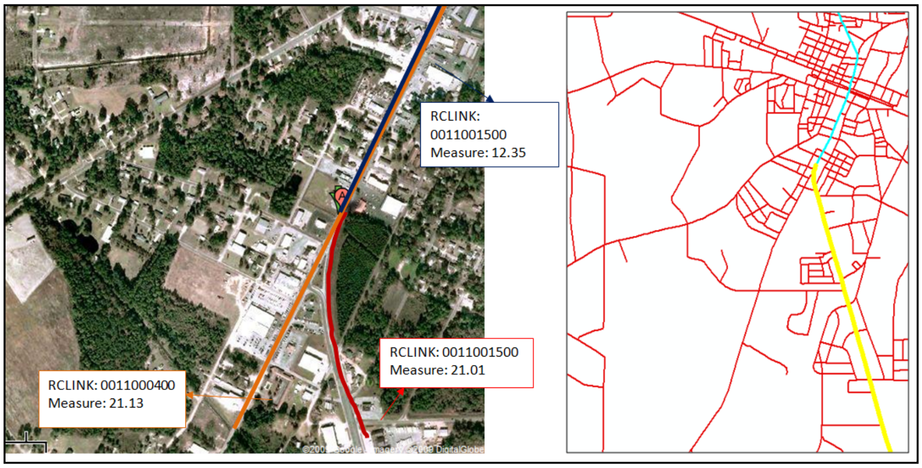

Figure 10: issue with LRS data: Two records have same RCLINK but different measures

- When the continuation of a route is broken due to another route, segments on either side of the break section were given the same RCLINKs. This approach seems logical, yet, assigning crashes to such links was difficult. Figure 11 gives an example of the issue.

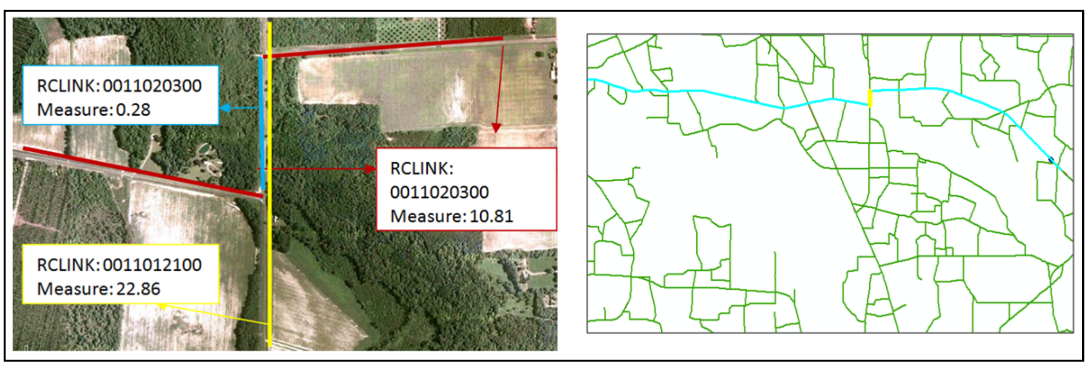

Figure 11: issue with LRS data: Two discontinuous roadway sections have same RCLINK

- At merge and diverge locations, the routes and their corresponding ramps can have same RCLINK (ex: RCLINK: 0051000400). An example of this issue is shown in Figure 12.

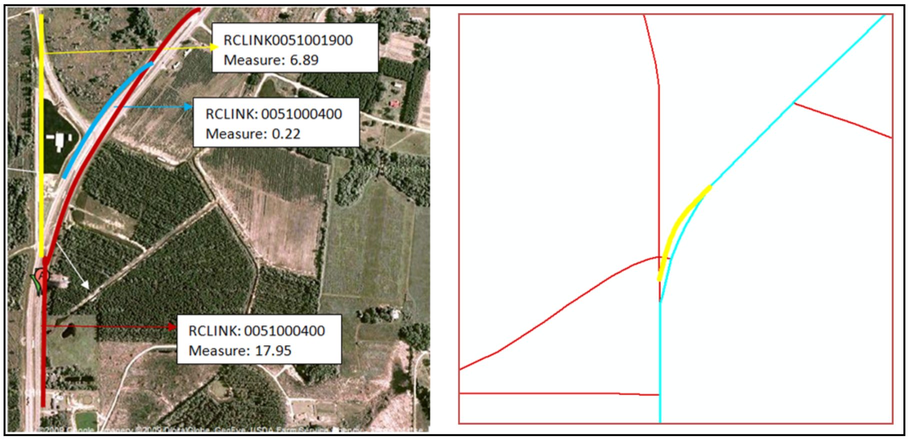

Figure 12: Roadways and their corresponding ramps have same RCLINK

Having spent considerable amount of time with LRS data in ArcGIS, a large portion of the problems and issues with the spatial reference were identified and corrected for use in the safety analysis. However, there is always room for improvement in GIS network coding and processing and there will likely be different issues in other states that may not be present in the Georgia data.

4.1.3 Traffic data:

Due to the strong correlation between traffic volumes and crash occurrence, Annual Average Daily Traffic (AADT) information, is one of the most crucial and fundamental data for any type of roadway safety analysis, be it crash rate, safety index, or Empirical Bayes method. Though the traffic information file obtained from Georgia DOT had data for the years 1995 to 2008, only three years of data was used since the crash data was available for only three years (2004-2006). The main issues associated with traffic data were the completeness of the data (not all segments had traffic data available), and the variations in yearly AADT (some variations were extreme or illogical).

While some roadway segments had no AADT data available, and had to be excluded from the safety analysis, data for a few other segments was found to be incomplete. There were quite a number of segments with only one or two years of traffic data instead of all three. Due to their substantial number, these segments were not excluded from the analysis. The missing AADT values were estimated based on the procedure explained in section 3.4.1. To maintain consistency and to address selection bias, the researchers followed a procedure to flag the segments with unrealistic growth factors. The ratio of the largest AADT to the smallest AADT of the three years for each segment was calculated. After further review and plotting these values, it was found to be acceptable to exclude all segments whose ratio was > 5, as these were considered to be unrealistic. This approach resulted in smaller datasets in each site subtype with fewer outliers. The final dataset is used for generating safety performance functions for various site subtypes. About 7% of the total miles of roadway were excluded due to unrealistic traffic volumes. Table 19 shows an example of segments with extreme growth factors.

Table 19: An example of segments with extreme traffic growth factors

Agency id	segment length (miles)	AADT			ratio of largest to smallest AADT values
		2004	2005	2006	
B057202370000000583B	5.83	740	5130	5630	7.61
B109202020000000131B	1.31	100	150	1270	12.7

Phase 2:

4.2 Test traditional methods for biases found in the literature:

The basic site selection methods including frequencies, rates, and safety indices fail to identify the true deviant sites due to their issues, limitations, and biases. The random nature of crashes needs to be addressed to obtain the true safety performance of a roadway. As it was noted earlier, the roadway characteristics file consists of numerous roadway segments with homogeneous characteristics throughout their length. Typically, many data variables of interest are collected and a roadway is divided into sub segments when at least one of the recorded data elements changes. Georgia is collecting and maintaining information of about 75 different data elements and as a result, the road network is divided into 884,598 shorter segments with an average segment length of 0.1378 miles. However, there are a considerable number of segments that are shorter than 0.1 miles. Table 20 gives descriptive statistics for roadway characteristics database in relation to segment length.

Table 20: Descriptive statistics for RC database with respect to segment length

	RC database
Total # of segments	884, 598
# of miles of roads	121915.17
Avg. segment length in miles	0.1378
# of segments < 0.1 miles	586,653
% of segments < 0.1 miles	66.32%
# of segments = 0.01 miles	216,867
% of segments = 0.01 miles	24.52%

it is observed that even though the average segment length is 0.138 miles, over 65% of the segments are shorter than 0.1 miles and about a quarter of the total number of segments are equal to 0.01 miles (~52.8ft).

Issues with shorter segments are not obviously known or observed in any type of analysis. However, it is noticed that they bias results and often question their reliability irrespective of the type of network screening method used.

When crash rates are considered, shorter segment lengths result in higher crash rates compared to relatively longer segments (Alluri, 2008). Figure 13 helps in understanding the influence of segment length on crash rate. Consider a hypothetical situation, in which one crash has occurred on a 1 mile long segment with an AADT of 1000 veh/day in the year 2004.

Figure 13: One mile segment with one crash

Exposure = AADT * 365 * segment length/ 1 million VMT

= (1000*365*1/1000000)

Crash Rate = (Number of crashes) / (Exposure)

= 2.739 crashes/mile/year

Consider another similar case (as shown in Figure 14) where the previous 1 mile segment has been divided into 10 segments of 0.1 miles each based on the variations in roadway inventory elements, with a single crash in 2004 and an AADT of 1000 veh/day.

Figure 14: One mile segment divided into 10 segments of 0.1 miles each with one crash

Exposure =AADT * 365 * segment length/ 1 million MVMT

= (1000*365*0.1/1000000)

Crash Rate = (Number of crashes) / (Exposure)

= 27.39 crashes/mile/year

In this case, the segment length has a drastic influence on crash rate and also on the criteria for prioritizing sites with greater potential for safety improvement.

When crash frequency is used for site selection, shorter segments are not typically flagged as “problematic sites.” This occurs, because fewer crashes are typically recorded on shorter segments in comparison to their corresponding longer segments.

Given that each roadway segment has different length, advanced methods use normalization in-order to make the crashes comparable across segments. For example, when a crash occurs on a roadway segment of 0.01 miles, the calculated normalized crash frequency will be 100 crashes/ mile. When the same crash occurs on a 0.1 mile segment, the calculated normalized crash frequency will be 10 crashes/ mile and the calculated normalized crash frequency for a similar crash occurring on a 1 mile segment is 1 crash/ mile.

Advanced methods have another advantage – that of a measure of predictive power. The predictive power of the EB method is provided as the variance in expected crashes for each site. This is a key component in assessing the reliability of the method as measures of the predictive powers of rates and frequencies do not exist. Variance measures the square of deviation of the expected crashes from the mean value. Though not uncommon, greater variance might result in an unrealistic number of expected crashes. For example, if at a site, the variance is 100 crashes/mile^2/yr and the number of expected crashes is 5 crashes/mile/yr, it means for that site, the total crashes in the coming year is expected to be between ($5 \pm \sqrt{100}$) = -5 crashes/mile/yr and 15 crashes/mile/yr.

When expected crash predictions cross the zero crash threshold, a reversal of prediction is possible. Thus, sites with tighter variance would be better bets for achieving the most bang for the buck (MBB). Even though, there can be many factors that influence the variance of expected crashes at a site, it is found that segment length has a considerable influence on the variance. Shorter segments are found to have very high variance compared to longer segments. Table 21 gives an example of the influence of segment length on the variance of the expected crash frequency. It could also be noted that the expected crash frequencies are also out of line with predicted crash frequencies for shorter segments.

Table 21: Influence of segment length on variance

Site start loc.	Site end loc.	Seg length (miles)	Location with Highest Potential for Safety Improvement (PSI)				SA Rank
			Average Observed Crashes (crashes/ mile/yr)	(SPF) Predicted Crash Frequency (crashes/ mile/yr)	(PSI) Expected Crash Frequency (crashes/ mile/yr)	Variance (crashes /mile^2/yr)	
5.48	5.49	**0.01**	3,666.67	198.99	3,960.91	**142,601.25**	1
0.36	0.58	0.22	140.73	2.20	35.84	2.99	2
9.35	18.18	8.83	213.44	1.34	35.03	1.98	3
4.10	6.43	2.33	132.03	1.72	28.69	1.96	4

When EB analysis is run on both shorter and longer segments, the variance for the shorter (0.01 mile) segment is unrealistic at about 36 times the expected frequency. However, the variance is reduced for longer segments and valid predictions (i.e. expected frequency 35.84 ± 1.73) are attained. Hence longer segments help improve performance of EB approach.

Having discussed the issues resulting from shorter segment lengths, a closer and detailed study revealed that short segments were commonly associated with variations in roadway characteristics. These variations are caused by two primary factors: coding error, and data sensitivity.

4.2.1 Coding errors:

It was found that data inaccuracy (coding errors) may result in discontinuity of segments. For example, Figure 15 shows the area type of a 0.01 mile long roadway segment being coded as urban, whereas a long multiple mile section on either side are coded as rural sections. This appears to be a coding error and it results in splitting the segment at the beginning and ending of the 0.01 mile urban roadway segment.

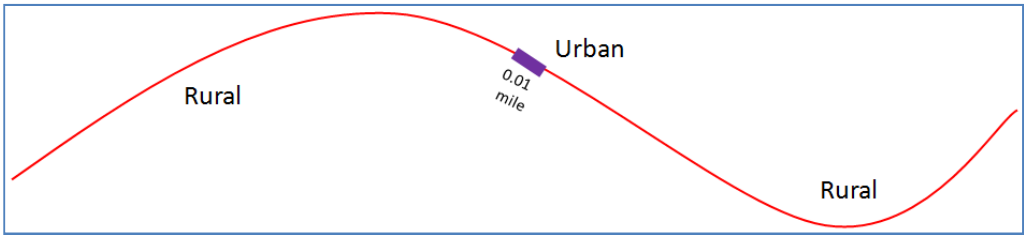

Figure 15: Coding error relating to area type in roadway characteristics file

Coding errors with respect to Annual Average Daily Traffic (AADT) were also frequently identified. Just like the 0.01 mile urban setting, a 0.01 mile roadway segment with an AADT of 2,150, may be contained within a larger segmental AADT of 12,150 on either side. Issues similar to those discussed above result in a number of short segments in the RC database. While some of the discrepancies might be real, there are limited ways to validate the legitimacy of the records. Table 22 shows an example of a coding error relating to AADT data.

Table 22: Coding error relating to AADT data

RCLINK	Beg measure (miles)	End measure (miles)	AADT		
			Year: 2004	Year: 2005	Year: 2006
0872001000	0.00	1.10	31,000	31,000	32,000
0872001000	1.10	1.97	31,000	31,000	32,000
0872001000	1.97	3.50	31,000	31,000	32,000
0872001000	3.50	4.01	31,000	**310,000**	32,000
0872001000	4.01	5.00	31,000	31,000	32,000

4.2.2 Data sensitivity:

With all other data issues set aside, there can still be problems associated with proper data records. The resolution of each of the data elements make the data very sensitive as all the minor changes are recorded. A one foot variation in the shoulder width at every 0.1 mile results in breaking up of the segments into smaller segments that are 0.1 mile long. When the variations in various other characteristics are overlapped, infinite numbers of smaller segments are generated (see Figure 16). Thus, the number of records in the RC file is exponentially increased.

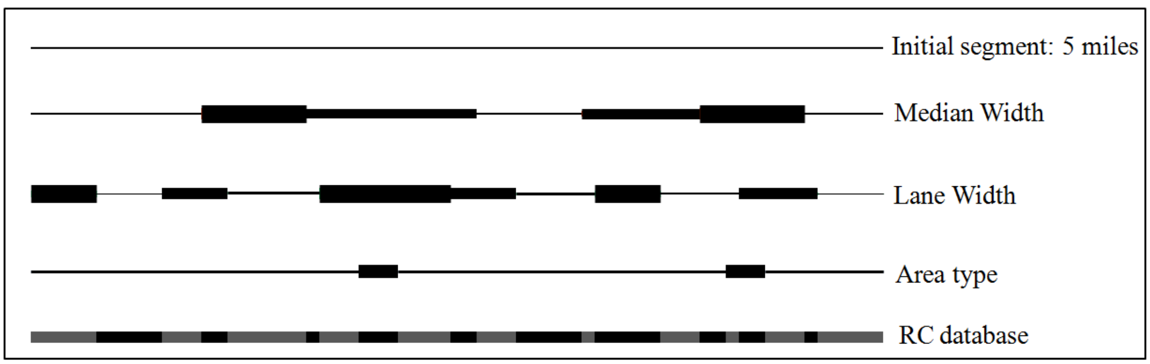

Figure 16: Segmentation of considerably longer segments into shorter segments

In Figure 16, when a roadway segment is considered without any major changes within a 5 mile distance, there will be one record in the roadway characteristics database representing a 5 mile long roadway segment. However, as shown in this example, when the median width, lane width, and area type are considered, the 5 mile segment is divided into 20 records in the RC database; a new record beginning whenever there is a slight change in any of the roadway characteristics. This however, depends on the sensitivity of the data being recorded. The greater the sensitivity, the greater the number of shorter segments resulting in unreliable results. In the above hypothetical situation, even though the average segment length in the roadway characteristics database is 0.2 miles, there are a few sub segments which are less than 0.1 miles that might result in a biased list of problematic sites.

The problem of shorter segments, to some extent, could be addressed to if the shorter segments are somehow merged into considerably longer segments while preserving the varying characteristics to the required detail. These merged segments can be referred to as "*aggregate segments*" and must be generated based on agreed threshold variation limits. Aggregate segments could be generated by considering fewer data elements in defining a segment and/ or by reducing the sensitivity of the data elements being collected.

4.2.3 Aggregated segment generation by considering fewer data elements:

As mentioned earlier, the roadway characteristics file for Georgia has 75 different variables of which some are not very useful in preliminary crash data analysis. Hence, including only the required data elements in defining a roadway segment helps considerably in increasing the segment length.

For this research, only the data elements (given in Table 23) required by SafetyAnalyst were used in generating the homogeneous roadway segments file.

Table 23: The required data elements used by in SafetyAnalyst for Roadway Segments file

Agency ID	roadwayclass1	Operation Way
Route Type	d1numThruLane	v2medianWidth
Route Name	d2numThruLane	Start offset
county	medianType1	End offset
Area type	Access Control	Section length

MySQL was used to generate the aggregate segments from the required variables and the code is attached in Appendix A.

4.2.4 Aggregated segment generation by reducing data sensitivity:

It is also observed that a high level of resolution or sensitivity in data elements might not be as helpful as expected especially if the filter being applied during the analysis is less sensitive than the data themselves. In the state of Georgia, variations of the magnitude of 0.1 ft have been recorded for variables like lane width, shoulder width, and median width. However, it is observed that these variables are mostly used in calculating crash modification factors (also known as crash reduction factors) to adjust for the base conditions, and for countermeasure evaluation. The countermeasure CMFs were generated based on 0.5 ft variations for lane width and 1 ft variations for shoulder width. Thus if data is recorded at 0.1 ft intervals, then intervals between 0.5 ft and 1.0 ft are not being utilized. While generating the import files for roadway segments to be used within SafetyAnalyst, it was observed that median width was recorded within 0.1 ft variations. However, the threshold level for variable change in SafetyAnalyst does not require 0.1 ft increments. The data only needs to be recorded every 0.1 ft if there is a change. Table 24 shows the changes made to the median width data.

Table 24: Reducing the sensitivity of median width data

Minimum width (ft)	Maximum width (ft)	Width changed to (ft)
0	0.5	0
0.6	3.5	2
3.6	6.5	5
6.6	9.5	8
9.6	12.5	11
12.6	15.5	14
15.6	20	17
20.1	30	25
30.1	-	30

The process of including fewer required data variables and reducing the data sensitivity while generating aggregated segments had reduced the number of segments from 884,598 to 209,636. Table 25 briefly compares a few statistics between longer aggregated and shorter unmodified segments.

Table 25: Comparison of shorter disaggregated and longer aggregated segments

	Shorter disaggregated segments	Longer aggregated segments
Total # of segments	884,598	209,636
# of miles of roads	121915.17	121915.17
Avg. segment length in miles	0.14	0.58
# of segments < 0.1 miles	586,653	54,659
% of segments < 0.1 miles	66.32%	26.07%
# of segments = 0.01 miles	216,867	5,858
% of segments = 0.01 miles	24.52%	2.79%

Observing the shorter and longer segments, the average segment length was increased from 0.138 miles to 0.582 miles. It is also observed that the percentage of shorter segments was reduced from 66% to a little over 25%. Less than 3% of longer aggregated segments are shorter than 0.02 miles. These aggregated segments are used for further analysis.

Having discussed about the specific issues related to crash data, roadway characteristics data, and traffic data, the following paragraphs test traditional methods against SafetyAnalyst to ascertain if rankings are comparable.

4.2.5 Issues with traditional methods:

Two-way two-lane rural roads in Georgia were considered for the analysis. Frequency, rate, critical crash rate, and EB approaches using SafetyAnalyst were tested on the data. Aggregated segments were used for EB analysis while both shorter disaggregated and longer aggregated segments were used for other methods. Table 26 describes the number of aggregated and disaggregated segments used in the analysis.

Table 26: Number of aggregated and disaggregated segments

	Georgia statewide roads	Georgia two-lane rural roads
Total number of aggregated segments (AS)	209,636	70,167
Total number of shorter disaggregated segments (DAS)	716,895	328,726

4.2.5.1 Traffic volume: Frequency is biased toward high volume roads and rate is biased toward low volume roads

Previously, statements were made regarding trends and biases for frequencies and rates. It is stated that frequencies identify high volume roads while rates identify low volume roads (Alluri, 2008; iTRANS Consulting Ltd & Human Factors North INC, 2003). To test this bias, the top 100 sites in the state based both on frequency and rate are identified. This information was then used to test relationships between the two methods with volume. To test the volume bias, the top 100 results were grouped by site subtype as shown in Table 27.

Table 27: Total number of segments ranked as top 100 by crash frequency and crash rate by site subtype

Site subtype Description	# of lanes in each direction	Total # of segments identified by Crash Frequency	Total # of segments identified by Crash Rate
Unknown	--	0	1
Rural 2 lane	1	0	19
Rural multilane undivided	2+	0	2
Rural multilane divided	2+	0	3
Rural freeways - 4 lanes	2	2	0
Rural freeways - 6+ lanes	3+	0	0
Rural freeways within interchange area--4 lanes	2	0	0
Rural freeways within interchange area--6+ lanes	3+	0	0
Urban two-lane arterial streets	1	7	42
Urban multilane undivided arterial streets	2+	30	7
Urban multilane divided arterial streets	2+	0	22
Urban one-way arterial streets	1	0	3
Urban freeways - 4 lanes	2	2	0
Urban freeways - 6 lanes	3	4	0
Urban freeways - 8+ lanes	4+	4	0
Urban freeways within interchange area - 4 lanes	2	2	1
Urban freeways within interchange area - 6 lanes	3	10	0
Urban freeways within interchange area - 8+ lanes	4+	39	0
Total segments	**--**	**100**	**100**

Site subtypes were assessed by number of lanes, and thus representing potential AADT. It is assumed that the higher the number of lanes, the greater the AADT. In this analysis, using actual AADT or range of AADT would overly complicate the table.

Site subtypes primarily identified by the frequency method include urban multilane undivided arterials and urban freeways within interchange areas with 8+ lanes. On the other hand, the crash rate method tended to identify urban two-lane arterials and rural two-lane roads. The frequency method ranked 57% of sites with greater than 3 lanes in each direction, whereas, crash rate method ranked 64% of sites with only one lane in each direction. Thus higher functional class is also associated with higher volume.

4.2.5.2 Segment Length: Frequency is biased toward longer segments and rate is biased toward shorter segments

As for segment length, it is stated that frequencies identify longer segments while rates identify shorter segments (iTRANS Consulting Ltd & Human Factors North INC, 2003). To test this bias, the top 100 sites in the state based both on frequency and rate are identified. It was found that the average segment length for the top 100 frequency and top 100 rate sites to be 12.35 miles (std dev = 6.62 miles) and 0.1 miles (std dev = 0.22 miles) respectively.

From the results above, it can be concluded that frequency is biased toward high volume roads and longer segments, while crash rate is biased toward low volume roads and shorter segments. Most DOTs have tried to overcome these biases by combining frequency and rate methods (sometimes with severity) into an index approach. However, the inherent biases are still present, and appropriate site selections will not be made using these methods.

Phase 3:

4.3 Implement SafetyAnalyst on roadway segments:

Although the prior analysis was completed with data that had been parsed into various site subtypes, this is commonly not the norm. As the survey results will show later, most states dump all segments together regardless of whether they are urban or rural, multilane or single lane. Given the practical implication, it is unacceptable to consider all the roadway segments alike for safety analyses. As discussed in the earlier chapter, SafetyAnalyst divides the roadway segments into various site subtypes based on functional classification, area type, roadway class etc. The following paragraphs discuss the roadway characteristics database by site subtype. Table 28 shows the number of records and the total number of miles of roadway in each site subtype.

Table 28: Site subtypes with number of records and total miles

Site subtype	Site subtype Description	# of records	Total miles
(null)		5,210	428.34
101	Rural 2 lane	70,167	79,585.52
102	Rural multilane undivided	347	474.59
103	Rural multilane divided	4,490	1,432.55
104	Rural freeways - 4 lanes	187	393.28
105	Rural freeways - 6+ lanes	86	120.73
106	Rural freeways within interchange area--4 lanes	127	159.11
107	Rural freeways within interchange area--6+ lanes	60	59.38
151	Urban two-lane arterial streets	110,720	34,650.62
152	Urban multilane undivided arterial streets	2,873	1,534.19
153	Urban multilane divided arterial streets	9,341	1,396.63
154	Urban one-way arterial streets	4,092	683.87
155	Urban freeways - 4 lanes	676	285.49
156	Urban freeways - 6 lanes	166	121.35
157	Urban freeways - 8+ lanes	67	24.85
158	Urban freeways within interchange area - 4 lanes	561	245.18
159	Urban freeways within interchange area - 6 lanes	189	130.53
160	Urban freeways within interchange area - 8+ lanes	277	188.96
Total		**209,636**	**121,915.17**

Although all the aggregated segments were imported into SafetyAnalyst, not all were used for the calibration process. As a default threshold, segments shorter than 0.1 miles were not used for calibration. However, this threshold could be set to a different threshold level within the data management tool of the software. Table 29 shows the number and percent of segments used for calibration by site subtype in addition to other statistics.

Table 29: Descriptive statistics for the segments imported into SafetyAnalyst by site subtype

Site Subtype	# of segments imported	# of segments NOT used for calibration	# of segments used for calibration	% of segments used	# of accidents associated with segments used for calibration	avg. segment length (miles)
101	64,893	7,287	57,606	88.77	75,242	1.33
102	334	67	267	79.94	3,199	1.75
103	4,428	1,240	3,188	72.00	5,308	0.38
104	182	32	150	82.42	5,982	2.60
105	86	7	79	91.86	3,608	1.52
106	127	9	118	92.91	2,841	1.34
107	60	7	53	88.33	2,604	1.11
151	98201	27,186	71,015	72.32	118,162	0.42
152	2,628	924	1,704	64.84	64,338	0.82
153	8,233	4,700	3,533	42.91	31,899	0.30
154	3,666	1,379	2,287	62.38	2,603	0.23
155	522	198	324	62.07	7,007	0.71
156	164	47	117	71.34	6,743	0.94
157	67	20	47	70.15	5,359	0.51
158	558	184	374	67.03	5,476	0.47
159	189	40	149	78.84	14,691	0.86
160	277	83	194	70.04	37,520	0.95
Total	**184,615**	**68,431**	**141,205**	**76.49**	**392,582**	

Approximately 12% of the segments were not imported into the software and were excluded from the analysis. The reasons for these exceptions will be discussed in section 4.3.1.2.

4.3.1 Problems that arose while generating import files for SafetyAnalyst:

4.3.1.1 Coding mismatch: SafetyAnalyst requires a restrictive coding structure for each data variable. Because most of the Georgia's coding structures are different, coding mismatch existed for most of the data variables. Table 30 identifies one of the more severe cases of coding mismatch using Georgia data. Data recoding could be done either by changing the enumeration values of data variables within SafetyAnalyst in the administration tool or within the dataset using a database management system. It is advisable to determine whether data recoding at the state level is more appropriate or the recoding of the attributes of data elements within SafetyAnalyst. If a state is setting out to collect data from the very beginning to use in more advanced methods, then changing and recoding the state data variables to work in SafetyAnalyst would be more feasible. However, on the other hand, if the state's data has already been collected and updated, it would be easier and less time consuming to change the attributes of data variables within SafetyAnalyst.

Table 30 An example of coding mismatch between SafetyAnalyst data atttributes and Georgia data

MEDIAN TYPE1	
SafetyAnalyst	**GDOT**
Field Name: medianType1	**Field Name: Median Type**
1-rigid barrier system (i.e., concrete)	0-No barrier
2-Semi-rigid barrier system (i.e., box beam, W - beam strong post, etc.)	1-Curb
3-Flexible barrier system (i.e., cable, W - beam weak post, etc.)	2-Guardrail
4-Raised median with curb	3-Curb and Guardrail
5- Depressed median	4-Fence
6-Flush paved median [at least 4 ft in width]	5-New Jersey Concrete Barrier
7-HOV lane(s)	6-Cable
8-Railroad or rapid transit	7-Other
9-Other divided	
0-Undivided	
98-Not applicable	
99-Unknown	

Approach to resolve the issue: A majority of the variables were recoded in the administration tool of the software. However, for some of the variables, the GDOT coding structure was entirely different from the values used in SafetyAnalyst. As well, a few variables required combination of multiple variables within the GDOT database. For these, SQL queries were run on Georgia database to restructure data prior to import. Further, there are a few required variables in SafetyAnalyst which cannot be altered within the software. These mandatory variables (for example, operation way), are used to sub-categorize roadway segments into site subtypes and to generate calibration factors. These cases were also recoded within the GDOT database prior to the import process.

4.3.1.2 Not all segments and crashes were imported: Although all the 209,636 aggregated segments were imported into SafetyAnalyst, only 88% of the segments were considered to be valid and the remaining segments were excluded from further analysis. Reasons for exclusion ranged from missing traffic info to limitations on site subtype in which the segments can be placed. Traffic information was missing for a substantial number of segments and SafetyAnalyst flagged all the segments with no traffic data. Loop roads and segments with zero length (i.e. the beginning and end milepost of the segment is the same) were also flagged and excluded in the import process. Previewing these in the GIS roadway inventory database, most represented point features and not segments. Only the loop roads were problematic as these actually exist, but are being excluded from analysis. Some sections had missing lane information (number of lanes), and these could not be imported either. Finally, some combinations of roadway characteristics in the Georgia database did not fit within any of the pre-defined site subtypes. For example, Georgia has a few reversible roadways, since SafetyAnalyst has not defined a site subtype for this configuration, the site cannot be analyzed. If there were many of these sites in Georgia, it would be possible to generate a SPF and create a new subtype in the administration tool to allow import and analysis. Table 31 provides the specific reasons for exclusions in the segment file.

Table 31: Descriptive statistics for the segments imported into SafetyAnalyst

Issue	# of segments
Total # of aggregated segments	209,636
Segments with loops and zero length	3,629
Segments with missing lane information	9
Segments that were not assigned to any site subtype	1,088
Segments with missing traffic information	20,295
Total number of valid segments to be used within SafetyAnalyst	**184,615**

Approach to the issue: As traffic data is a required input, all the segments with no traffic information were excluded from the analysis. Segments with missing location and missing lane information were also excluded. Segments without an assigned site subtype were queried out and were found to be special cases such as segments with reversible lanes, segments with one way truck routes, one way during school hours etc. These segments were not included as specific SPF information for such scenarios is unavailable. Segments such as these could be manually compared to their closest site subtype to ensure that they are being monitored for safety performance as it is likely that there would never be enough data to generate a proper SPF for these special roads.

Over 98% of the crashes were considered to be valid with only 7,003 out of 442,233 crashes not being used in the analysis for a couple of reasons. Very few crashes (31) were not included because they did not match the standard types of crashes included in SafetyAnalyst analysis. This could be associated with miscoding of crash info from police reports. Further analysis may be needed, but they are small in number comparatively.

The majority of the crashes were not included because they could not be assigned to a segment. This is partially a carry-over effect from segments not being imported into SafetyAnalyst as identified in Table 31. When the segment is not imported, crashes belonging to that segment cannot be associated to it, and thus, they are excluded from analysis. Table 32 summarizes the specific reasons for exclusions in the crash data file along with the effected number of crashes.

Table 32 Descriptive Statistics for the crashes imported into SafetyAnalyst

Crashes that are valid versus invalid	
Total number of crashes imported	442,233
Total number of valid crashes	435,230
Crashes of non-standard type	31
Crashes that are not assigned to any segment	6,972
Total number of invalid crashes	7,003

4.3.1.3 Miscoded data: Data quality checks need to be performed prior to importing the files into SafetyAnalyst to reduce the number of warnings in the import process. Various coding errors, possibly resulting from manual data entry need to be identified and flagged. These errors could be identified by verifying the location using spatial reference software like Google Earth. Though not fool proof, coding errors could be identified by observing the previous and the next segments and noting the variations in roadway characteristics. If the variation is insignificant (based on the researcher's opinion), the shorter segment could be merged with either the previous segment or the next segment. However, consistency needs to be maintained in this approach and these decisions should be documented for possible use in more detailed stages of analysis. Various examples of miscoded data were discussed in section 4.2.1

4.3.2 Errors and warnings in the log files while importing, post processing and calibrating Georgia data in SafetyAnalyst:

During the import, post process, and calibration stages, the user can specify in SafetyAnalyst to generate a log file in each step to record all the errors and warnings. The following section shows some of the errors and warnings that were most frequently identified with Georgia data.

i. *'Accidents are not located on any roadway segment':* Not all crashes could be located on roadway segments. Most likely candidate for this error is related to segment import error listed in Table 31.

ii. *'There is no traffic data associated with the segments':* Segment traffic information is considered to be the most important attribute to perform roadway safety analysis and therefore, all segments without traffic information are excluded from further analysis.

iii. *'The traffic data and/or the growth factor is unrealistic':* As most of the traffic data is estimated and/ or entered manually, there is a greater probability of error. For each roadway segment, SafetyAnalyst conducts a comparison of yearly variations in traffic data. The segments with unrealistic growth factors (±20 %) were flagged. The growth factors from year to year may help agencies identify potential problematic sections that require follow-up. However, smaller changes on sites with lower AADT values can appear as unrealistic growth. This is an issue that each state will have to tackle. Some operational rules to deal with these warnings would likely be required. This issue, to an extent, was addressed by increasing the acceptable traffic growth factor in the data management tool. By default, the annual growth factor is set at 20 % and it was increased to 50% (the maximum allowable growth rate within the software). This may even change over time within a particular area or state.

iv. *‘Segments do not fall under any of the predefined site subtypes’:* SafetyAnalyst divides the roadway segments into subtypes based on area type, functional classification, number of lanes, presence of median etc. But, sometimes, there are a few segments which do not fall under any predefined site subtype and therefore are flagged and excluded from further analyses. In the Georgia dataset, there are some segments such as High Occupancy Vehicle lanes and reversible lanes that do not fall under any of the site subtype and therefore are excluded from the analysis. In such situations, if there are a sufficient number of similar roadways, they could be identified as a separate site subtype. Specific SPFs would have to be developed for their analysis. Else, these segments could be assigned the most closely fit subtype category with the main goal of including them in the analysis. This step could be performed in the administration tool of the software.

4.3.3 Issues identified after performing network screening:

After successfully importing files into SafetyAnalyst, network screening could be performed in the analytical tool of the software. The output from the analysis needs to be reviewed to identify possible anomalies and issues. When the initial output from Georgia data was observed, the variance was found to be extremely high questioning the reliability of the results. Shorter segment length (as discussed in the earlier sections) was found to be the main reason for unacceptably high variance.

Merging of exiting segments into longer segments to increase the segment length (also known as generation of homogeneous segments) could be performed within the post process step of the data management tool of SafetyAnalyst. Additionally, the sensitivity of the data elements can be controlled in SafetyAnalyst as well using thresholds defined in the post process step of the data management tool. These methods were previously addressed in section 3.3.2

4.3.4 Comparison of differences in ranking outcomes between crash frequency, crash rate, critical crash rate and EB approach using SafetyAnalyst for two-lane rural roads

The following analysis was conducted for two sets of segments.

1) Shorter disaggregate segments such as those generated by raw roadway inventory data
2) Longer aggregated segments generated by reducing sensitivity of data elements and limiting segmentation to only required elements

For each set of elements, a ranking was produced using traditional methods (crash frequency and crash rate). However, since the generation of longer aggregated segments is inherent to the SafetyAnalyst process, the SafetyAnalyst rankings are all based on second dataset of aggregated segments.

The top ten shorter sites based on frequency and their corresponding ranks by other methods are shown in Table 33.

Table 33: Ranking of two-lane rural roadways in Georgia based on different selection criteria sorted according to the rank based on crash frequency (disaggregate segments)

	Ranking Method					
Agency ID	Crash Freq[1]	Crash Rate	Critical Crash Rate	Safety Analyst (SA)	Longer SA seg length (mi)	Shorter Seg Len (mi)
B227100530013201780B	**1**	214,425	not ranked[3]	129	30.08	4.60
B02310087BU01680406B	**2**	14,737	6,399	59	4.06	2.38
B085100530015741577B	**3**	93	73	2	8.83	0.03
B121101540004971609B	**4**	40,238	20,165	123	14.71	11.12
B015206330005400608B	**5**	20,950	8,529	3	2.33	0.68
B015206360000000468B	**6**	22,583	9,643	not ranked[2]	4.68	4.68
B015206330006370643B	**7**	2,416	952	3	2.33	0.06
B015206330006080637B	**8**	14,743	6,461	3	2.33	0.29
B151101550005830597B	**9**	4,538	1,957	not ranked[2]	0.65	0.14
B241100150010101014B	**10**	1,389	335	5	10.22	0.04

[1]Each SA (homogeneous segment) link might have multiple links in shorter segment database
[2]Does not meet the minimum of 5 crashes/mile/year threshold for SafetyAnalyst homogeneous segments
[3]Observed crash rate is lower than the critical crash rate for that particular site and hence is not flagged as a site with potential for safety improvement based on critical crash rate criteria

Five of the top ten sites by frequency are identified by SafetyAnalyst as problematic sites. However, the 5 shorter sites correspond to only 3 longer aggregated roadway sections generated by SafetyAnalyst, thus reducing the actual number of “problematic sites”.

Table 34 shows the top ranked shorter disaggregated sites identified by crash rate and their corresponding ranks by other methods.

Table 34: Ranking of two-lane rural roadways in Georgia based on different selection criteria sorted according to the rank based on crash rate (shorter disaggregated segments)

	Ranking Method					
Agency ID	Crash Rate[1]	Crash Freq	Critical Crash Rate	Safety Analyst (SA)	Longer SA seg length (mi)	Shorter segment length (mi)
B151305480500000002B	**1**	225	1	not ranked[2]	0.02	0.02
B035204200000000004B	**2**	728	2	not ranked[2]	0.04	0.04
B035204200000080011B	**3**	2,213	3	not ranked[2]	0.16	0.03
B103305930500000006B	**4**	91	4	not ranked[2]	1.06	0.06
B103305930500800081B	**5**	1,489	6	not ranked[2]	1.06	0.01
B291103480000000002B	**6**	6,654	7	not ranked[2]	6.95	0.02
B233201820006720677B	**7**	3,053	8	not ranked[2]	3.07	0.05
B035204200002200024B	**8**	2,198	9	not ranked[2]	0.04	0.04
B005305120100410043B	**9**	2,023	10	not ranked[2]	0.96	0.02
B055103370002990304B	**10**	6,385	11	not ranked[2]	12.85	0.05

[1]Each SA (homogeneous segment) link might have multiple links in shorter segment database
[2]Does not meet the minimum of 5 crashes/mile/year threshold for SafetyAnalyst homogeneous segments

None of the top ten sites identified by crash rate were ranked by SafetyAnalyst since they do not meet the minimum threshold of 5 crashes/mile/year to be included in the SafetyAnalyst list. Also, the top ranked sites by rate method are all very short, with all the segments considerably shorter than 0.1 miles. SafetyAnalyst recommends segment lengths above 0.1 miles, thus in SafetyAnalyst these shorter segments are aggregated to longer adjacent homogeneous segments. Else, if short segments are used in SafetyAnalyst, the high variances associated with these short segments would disqualify them as potential study sites. SafetyAnalyst has the ability to highlight problem locations within a longer segment, thus allowing these sites (if truly deviant) to be identified.

When longer aggregated segments are considered while ranking sites based on both traditional and EB methods, only two of the top 10 sites based on crash frequency were also ranked in top 10 by SafetyAnalyst. Table 35 shows the top ranked longer aggregated sites identified by crash frequency and their corresponding ranks by other methods.

Crash rate had no significant improvement in its top ranked sites when aggregated segments are considered as there are still a substantial number of shorter segments which influence rates resulting in false identification of shorter segments as problematic sites. It is observed that even with aggregated segments, none of the top ranked sites identified using crash rate were flagged by SafetyAnalyst since they do not meet the minimum required criteria of 5 crashes/mile/year. Table 36 shows the top ranked longer aggregated sites identified by crash rate and their corresponding ranks by other methods.

Table 35: Ranking of two-lane rural roadways in Georgia based on different selection criteria sorted according to the rank based on crash frequency considering longer aggregated segments

	Ranking Method				
Agency Id	Crash Frequency	Crash Rate	Critical Crash Rate	Safety Analyst (SA)	Aggregated segment Length (mi)
B085100530009351818B	**1**	5,744	1,612	2	8.83
B117103690000001187B	**2**	9,951	not ranked[3]	31	11.87
B151100420000000929B	**3**	9,858	not ranked[3]	55	9.29
B255101550007231305B	**4**	4,433	1,251	37	5.82
B151101550000000532B	**5**	5,822	1,668	9	5.32
B255100160000001037B	**6**	10,396	not ranked[3]	38	10.37
B311101150000001556B	**7**	9,354	3,044	98	15.56
B187100090000001231B	**8**	6,421	1,896	19	12.31
B221100100000001856B	**9**	10,314	not ranked[3]	126	18.56
B103101190000002199B	**10**	6,762	2,022	not ranked[2]	21.99

[2]Does not meet the minimum of 5 crashes/mile/year threshold for SafetyAnalyst homogeneous segments
[3]Observed crash rate is lower than the critical crash rate for that particular site and hence is not flagged as a site with potential for safety improvement based on critical crash rate criteria

Considering the top ten ranked sites based on EB approach, it is seen in Table 37 that crash rates fail to identify at least one; while crash frequency performed better by identifying two of the ten sites.

Table 36: Ranking of two-lane rural roadways in Georgia based on different selection criteria sorted according to the rank based on crash rate considering loner aggregated segments

	Ranking method				
Agency Id	Crash Rate	Crash Frequency	Critical Crash Rate	Safety Analyst (SA)	Aggregated segment Length (mi)
B01510020SP04240425B	**1**	2,232	1	not ranked[2]	0.01
B151305480500000002B	**2**	1,461	2	not ranked[2]	0.02
B151305480500020003B	**3**	5,788	4	not ranked[2]	0.01
B035204200000000004B	**4**	2,235	3	not ranked[2]	0.04
B069201000003890392B	**5**	7,283	14	not ranked[2]	0.03
B185204690000000003B	**6**	7,494	15	not ranked[2]	0.03
B077205030000360037B	**7**	6,757	7	not ranked[2]	0.01
B151305480500090013B	**8**	2,273	5	not ranked[2]	0.04
B151217390001380140B	**9**	4,136	6	not ranked[2]	0.02
B045202160002970299B	**10**	9,945	20	not ranked[2]	0.02

[2]Does not meet the minimum of 5 crashes/mile/year threshold for SafetyAnalyst homogeneous segments

Table 37: Ranking of two-lane rural roadways in Georgia based on different selection criteria sorted according to the rank based on EB approach using SafetyAnalyst

	Ranking Method				
Agency Id	Safety Analyst (SA)	Crash Frequency	Crash Rate	Critical Crash Rate	HS Length (mi)
B255100030000360058B	**1**	238	2,154	640	0.22
B085100530009351818B	**2**	1	5,744	1,612	8.83
B015206330004100643B	**3**	21	5,392	1,556	2.33
B151101550005550563B	**4**	938	517	171	0.08
B151101550005720597B	**5**	161	516	156	0.25
B035100160001671011B	**6**	12	9,025	2,920	8.44
B211100120011171343B	**7**	181	7,229	2,380	2.26
B021100190000000337B	**8**	260	9,889	not ranked[3]	3.37
B151101550000000532B	**9**	5	5,822	1,668	5.32
B151101550005670572B	**10**	649	128	45	0.05

[3]Observed crash rate is lower than the critical crash rate for that particular site and hence is not flagged as a site with potential for safety improvement based on critical crash rate criteria

4.3.5 Comparison of differences in ranking outcomes for longer aggregated and shorter disaggregated segments

Considering the aggregated and disaggregated segments, the five top deviant sites identified by SafetyAnalyst along with their corresponding conventional ranks are shown in Table 38.

Table 38: Ranking of two-lane rural roadways in Georgia based on different selection criteria sorted according to the rank based on SafetyAnalyst

Agency ID	**Ranking Method**					
	Crash Freq[1]	Crash Rate	Critical Crash Rate	Safety Analyst (SA)	HS Length (mi)	Shorter segment length (mi)
B255100030000360058B	298,100	39,029	not ranked[3]	**1**	0.22	0.06
	19	8,158	3,474			0.07
	298,101	30,112	not ranked[3]			0.02
	2,995	21,433	13,072			0.03
	116	8,033	3,442			0.04
B085100530009351818B	**3 – 171,366	**93 – 150,011	**73 - not ranked[3]	**2**	8.83	0.067*
B015206330004100643B	32,578	229,993	not ranked[3]	**3**	2.33	0.22
	407	39,883	22,045			0.7
	32,587	229,978	not ranked[3]			0.03
	405	33,592	17,058			0.35
	5	20,950	8,529			0.68
	8	14,743	6,461			0.29
	7	2,416	952			0.06
B151101550005550563B	211,597	11,196	16,664	**4**	0.08	0.01
	53	4,557	1,995			0.07
B241100150000401062B	**10 – 289,238	**1,389 – 65,822	**335 - not ranked[3]	**5**	10.22	0.083*

*Average length of multiple segments that make up equivalent homogeneous segments
**Range of ranks associated with multiple segments that make up homogeneous segments

It is observed that the top 5 homogeneous sites are represented by about **269** shorter segments, thus, making “shorter segments” a greater issue in identifying SWiP. The top ten shorter segments could actually be one long segment. Hence, a significant number of deviant sites may not be included in top lists when shorter segments are considered. Of the traditional methods, frequency tends to identify more of the same sites identified by EB method while none of the crash rate sites were identified as top sites by SafetyAnalyst. Crash rates tend to identify shorter segments with few crashes. SafetyAnalyst uses SPFs and EB approach to overcome these biases.

4.3.6 Survey to states

A review of the 2009 five-percent (transparency) reports submitted by the states to FHWA describing at least five percent of highway locations exhibiting the most pressing safety needs had indicated that most DOTs are still using traditional safety analysis measures such as frequency, rate, critical rate, or crash index. Two out of 50 states reported use of EB methods. With this information, the research path called for a survey regarding safety data, present safety analysis methods, use of advanced safety analysis tools, and implementation. The survey was prepared and sent to the Safety Director of the Department of Transportation in each state. The questionnaire and the letter accompanying the survey are provided in Appendix H. In summary, the survey was divided into seven major parts:

1. Contact information;
2. General questions about safety data;
3. General questions about safety data analyses;
4. Questions about SafetyAnalyst;

5. Questions about Safety Performance Functions;
6. Questions about SafetyAnalyst implementation; and
7. Questions about Highway Safety Manual implementation.

Only the states that have been working with advanced safety analysis tools like SafetyAnalyst were asked to answer the questions in parts 4-6.

Of the 50 states, 24 states completed the survey in full, and one state answered a portion of the survey. Responses for the answered questions from the incomplete survey were considered in the analysis. Thirteen of the 25 states mentioned that they have been working with new highway safety analysis tools (SafetyAnalyst, IHSDM, or HSM). These states were asked to answer the questions related to SafetyAnalyst, safety performance functions, and Highway Safety Manual. Table 39 gives a list of states that have completed the survey and also identifies the responded states that have been working with the new safety analysis tools.

Table 39: States that have completed the survey on safety data, road safety analyses methods and tools

Alaska	Missouri*
Arkansas	Nevada*
California	New Hampshire*
Delaware	New Jersey
Florida*	New York
Georgia*	North Carolina*
Hawaii	Ohio*
Idaho	Oregon
Kansas*	Pennsylvania*
Maine	South Carolina*
Maryland	South Dakota*
Massachusetts*	Washington*
Minnesota	

*States that have been working with new safety analysis tools

Figure 17 shows the geographic distribution of responding states. In the figure, states that have been working with either SafetyAnalyst or any other new advanced safety analysis tools are shaded.

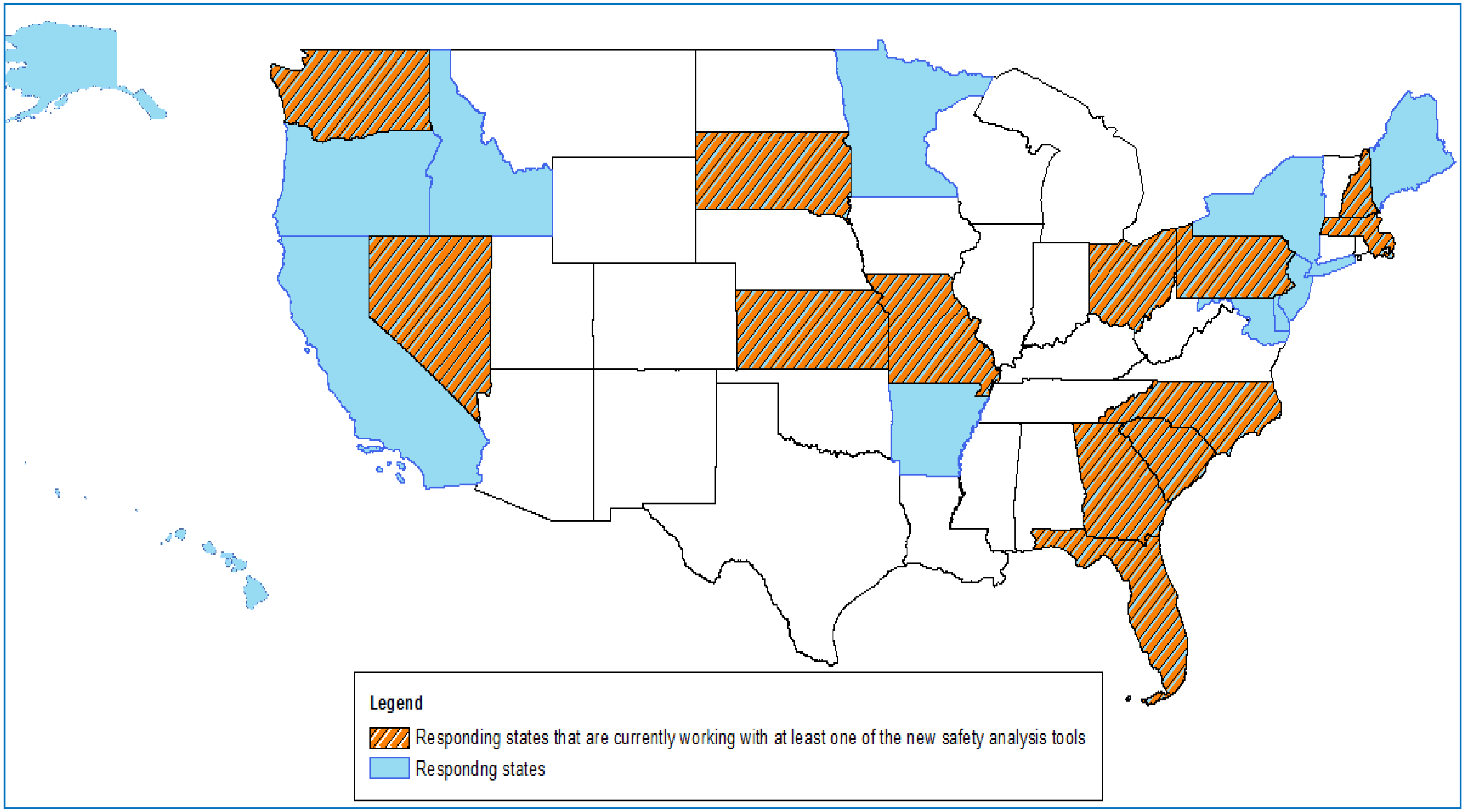

Figure 17: Geographic distribution of states responding to survey

4.3.6.1 Crash data: As mentioned in the literature review, safety data (crash, roadway characteristics, and traffic) plays a vital role in highway safety analyses. To an extent, the availability and accuracy of data governs site selection and prioritization methods. Understanding the importance of safety data, various questions on data availability and maintenance were asked in the survey, to which most of the states responded.

Use of longer periods (3 to 5 years) of crash data for safety analysis is recommended to account for the random nature of crashes and to address regression-to-the-mean effect. On the contrary, longer analysis periods might fail to give accurate results if changes in the roadway characteristics are not accounted for in the analysis. Research has shown that fewer years of data coupled with traditional methods fail to rank the “true-deviant sites” within the roadway network. In this context, of 25 responding states, 1 state uses two years of crash data, 13 states use three years of crash data, 7 states use five years of crash data, 1 state uses seven years of data and 2 states use ten years of crash data. When the availability of historical crash data is considered, 84% of responding states (21 out of 25) reported that they maintain at least ten years of crash data while two states reported that they maintain five years of historical crash data.

Most of the site selection methods (except project based Empirical Bayes approach) require specific location of crashes for identifying and prioritizing segments and intersections for safety improvements. A majority of the surveyed states maintain specific location information for crashes. Greater than a half of responding states (13 out of 25) are able to identify at least 90% of their crashes spatially. Six of 25 responding states are successful at locating 80-90% of their crashes, while less than a quarter (5 states), are spatially identifying fewer than 80% of their crashes. A few states, including

New Jersey and California, give priority to fatal and injury crash locations with almost all severe crashes being located where as a lower percent of PDO crashes are located.

In summary, most of the states use 3-5 years of crash data which is as expected given traditional data analysis methods. The overall performance of the states with respect to the maintenance of specific location information of crashes is adequate (as most of the states are able to spatially locate a minimum of 80% of their crashes) for comprehensive safety analyses.

A roadway network is comprised of segments, intersections and ramps with various classifications such as rural two-lane and urban multilane. Research proves that each sub-category of roadway network behaves differently and as a result, their safety performance needs to be evaluated separately. Therefore, it is important for the state's safety office to be able to identify crashes on segments, intersections and ramps separately and accurately. From the survey, all of the survey respondents are able to locate crashes on segments and over ninety percent (23 of 25 states) can locate crashes at intersections. As expected, crashes on ramps are the most difficult to locate precisely and only 17 states can identify the precise location of ramp related crashes. Among the responding states, Alaska and Missouri had mentioned that they are currently working on adding ramp information to the crash database. While dealing with ramp related crashes, various approaches are followed across states: for example, treating crashes on ramps as crashes at an interchange influence area, or assigning ramp related crashes to the gore area, or a more comprehensive alternative of visually identifying and inspecting ramp related crashes.

None of the responding states identified categorizing intersection-related crashes as problematic. On a separate note, a couple of states do not identify crashes on local roads as they are typically not mile posted. This approach results in not including the local roads (that constitute a considerable miles of road network within a state) in safety analysis. Since, SAFETEA-LU states all public roads, states will have to collect additional data or partner with local agencies.

Crash patterns, crash severity and performance measures vary with the type of a roadway – segments, intersections and ramps. From the survey, it could be inferred that most of the states are able to perform safety analyses on segments and intersections, but a comprehensive analysis of ramps may require additional work to locate ramp related crashes more accurately.

4.3.6.2 Roadway characteristics data: Roadway characteristics are not static. Many characteristics can change continuously along the roadway. Ideally, the road network database needs to be dynamic to allow for changes such as shoulder width to be recorded whenever there is a slight variation in any of the elements being collected. This approach is considered to be important because safety analysis depends on the roadway characteristics which need to be as up-to-date and accurate as practically feasible. In practice, eight of the 21 responding states (38% of the responding states) update their databases on a yearly basis, while about 10 states (50% of the responding states) update continuously whenever there is a change, two states update irregularly whenever there is a change and one state updates every 3 months. Idaho is planning on continuously updating the roadway characteristics database once their asset management system is implemented beginning in January 2011.

In addition to the information on changes in roadway characteristics, information about the implementation date of the individual changes is also important. This information is critical for conducting before and after studies, performing countermeasure evaluation and assigning roadway segments, intersections, and ramps to various site subtypes. Less than a quarter of the responding states (6 out of 25) do collect and maintain change date information while approximately the same number of states (5) do not maintain the same. From a safety analysis point of view, it may be acceptable to not maintain information about all the variables, as some might not be useful. About forty percent of the responding states (10) collect and maintain date information for only a few variables including, but not limited to lane width, shoulder width and type, and median width. However, many other elements have been shown to be associated with crash increases/ reductions. Changes in these elements should also be noted. Finally, other elements that we may not readily monitor or track date changes for may have significance in safety analysis, but, we don't know if we can't track them.

Changes in any roadway characteristics, typically defines the length of a homogeneous segment. Segment length plays a vital role in identifying and prioritizing sites with potential for safety improvement. Depending on the type of site selection method used, segment length can negatively impact the results. Crash rates tend to identify extremely short segments while frequencies identify longer segments. Empirical Bayes method even results in greater variance for shorter versus longer segments, thus the reliability of crash predictions remains questionable. Earlier sections of this document mentioned that Georgia divides the roadway segments into as small as 0.01 miles if there is any change in its roadway characteristics. Similar to Georgia, 0.01 miles is the smallest segment length used in 13 of 20 states for recording changes in the roadway characteristics.

About 25% of responding states (5 of 20 states) consider the minimum segment length as 0.1 miles, while one state records roadway characteristics every 0.25 miles. Florida is the only state that records roadway characteristics data at a higher resolution of 5 ft (0.001 miles). With over 50% of the responding states collecting data every 0.01 miles (if there is a change in roadway characteristics), segment length will likely be an issue to be addressed prior to site selection.

Intersection data is another area where additional work is likely to be needed. It is tedious and extremely data intensive to collect and maintain intersection data (constituting of lane configuration, signal plan, traffic control type and turning volumes). Even though 90% of the states identify crashes on intersections, far fewer states are found to have specific datasets for intersection characteristics on which they can perform intersection specific safety analyses using EB methods. Five states mentioned that they do not maintain intersection data. At the least, many states collect traffic control type and volumes which are considered to be the basic requirements for any type of site selection and prioritization methods (other than crash frequencies). Nevertheless, there is an agreed consensus among various states about the need for more detailed intersection data including turn volumes and lane configurations in addition to the minimum required data elements. It is observed that incomprehensive datasets are being maintained by many states. A few states maintain graphic files with intersection data that could be obtained when needed. Washington State Department of Transportation is working on these grounds by collecting and building an intersection database with all the required variables to be compatible with SafetyAnalyst software.

Thirteen of the 23 responding states maintain specific datasets for ramps while 9 do not. However, South Carolina and Ohio are currently working on developing their ramp datasets. For the states to shift to newer and advanced tools like SafetyAnalyst, Highway Safety Manual and IHSDM and to perform a complete road safety analysis, comprehensive datasets are required and from the current stand of the states, it can be concluded that most states need to start to collect and maintain the required data elements.

4.3.6.3 Traffic data: Excluding crash frequencies, all of the other traditional and advanced methods of site selection and prioritization require traffic data (either actual or estimated values) for the complete analysis period (time period for which crash data is available). In this context, several questions related to traffic data were asked in the survey. Fifteen of the 24 responding states mentioned that they maintain traffic data for at least 10 years while six states maintain data for the past 5 years and one state maintains 3 years of traffic data. About a half of the responding states (11 of 23) do maintain a comprehensive state-wide traffic database while 11 states do not. 17 of 21 responding states do maintain a comprehensive traffic data on all of the state maintained and federal maintained roadways, while the data on local roads is sporadic. The state highway administration (SHA) of Maryland maintains approximately 17% of the roadways in the state, and consequently has accurate traffic information for the same sections.

Considering the importance of accurate and comprehensive traffic data in road safety analyses, a few survey questions have been designed to understand the availability of actual and estimated traffic counts. In the survey, the roadways were grouped into interstates, state routes, secondary routes, county routes, city routes, other and low volume roads for better understanding of patterns in traffic data availability. From the survey, it is found that most of the states do measure traffic counts on roadways with higher functional classification, that is, interstates and state routes. Eighteen and fourteen of the 23 responding states mentioned that over 75% of their interstates and state routes have actual traffic count information respectively, and the percentage of roadway miles with actual traffic counts declined consistently with the decrease in the functional classification of the roadways. The least amount of the actual traffic data is collected on local, city and low volume roads. From the survey, it is found that about nine out of 18 responding states collect actual traffic data on less than 25% of their low volume roads. A few states indicated that default/estimated traffic volumes might be out of date.

Supplementing actual traffic data, state Departments of Transportation often estimate volumes for roads lacking actual counts. It is found that about three-fourths of the responding states estimate traffic on the roadways for which counts are not available. A minority of the responding states (4 out of 23) stated that they don't estimate AADT volumes. Washington State has about 7,000 miles of state highways and Interstates, and collects traffic counts at 4,000 locations to get a representative sample. In the state of California, as a continuing process, approximately a third of the roadways are actually counted every year while the rest of the segments are estimated.

Nine of the 24 responding states have a documented procedure for estimating traffic counts. Of the remaining states, seven states do not have a documented estimating procedure and eight other states are unsure. Over 85% of the responding states (21 out of 24) have traffic volume data (combining actual and estimated data) on over three quarters of the total interstate and state roadway network. About 30% of the responding states also have over three quarters of the city and county roadway network. As expected, roadways with lower functional classification (low volume roads and local roads) have the least amount of volume data (both actual and estimated data combined).

4.3.6.4 Currently used safety analysis methods: Research in the field of safety over the past few decades primarily focused on understanding the advantages, issues and limitations of traditional site selection methods and developing more statistically sound methods for selection and prioritization of “unsafe or problematic sites”. With the release of SafetyAnalyst and Highway Safety Manual, the requirement to understand the current stand of the states with regard to their safety analysis methods has gained immense interest among the researchers, practitioners and the administrators. In light of this, states were surveyed on their current safety analysis methods and their future plans regarding the adoption of newer tools.

Of the 24 responding states, twenty-one states mentioned that they perform their safety analysis in-house while three states do some of the analysis and contract the remainder. Management of projects and allocation of safety funds depends extensively on the administrative process followed within a state. Forty percent of the responding states (10 out of 24) noted that for management and allocation of safety funds, they follow both centralized and decentralized distribution. Nearly 50% of the responding states (11 out of 24) are using centralized procedure for both management and allocation of safety funds while three states use decentralized procedure for both. Two states use both centralized and decentralized procedures for project identification, and decentralized system for safety fund disbursement. On the similar grounds, Ohio uses centralized and decentralized procedures for safety analysis but, distributes funds centrally. Washington State identifies and funds projects centrally, but, its low cost enhancement program funding is decentralized. Massachusetts and Oregon distribute funds based on a formula which is decided centrally. The structure of management and allocation of funds will play an important role in the implementation of newer safety analysis tools.

Roadway safety analysis, whether basic site selection or advanced methods, require a considerable amount of expertise in the field of Transportation Engineering and Statistics. In this scenario, it is found that eight of the responding 10 states have at least one person with a Masters degree working with SafetyAnalyst. The remaining two states have a person with a Bachelor degree.

Segments and intersections need to be re-grouped into site subtypes based on various categories like area type, functional classification, number of lanes etc for safety analyses as different types of roadways behave differently. SafetyAnalyst divides the roadway segments into 17 site subtypes, intersections into 12 sub groups, and ramps into 16 subtypes. Such sub-classification, though not to that extent, is recommended while using traditional site selection methods. This is because the site characteristics significantly influence the relation between crashes and exposure (for example AADT).

With this background, several questions were targeted to understand the sub-classification scheme used by the states in relation to safety analysis of segments and intersections. Nine of the 22 responding states reported that they run safety analysis on segments on the complete state's data as a whole. It is also observed that the same states run safety analysis on the entire state's intersection data as a whole. This approach is particularly not favored within the research community because urban multilane intersections have very different safety considerations than intersections at two-lane rural roads. Another more partially feasible and better approach of sub-grouping segments/ intersections is to use a couple of variables for reclassification. Reclassification of segments based on a couple of variables is practical in comparison to the similar task with intersections. Of the thirteen remaining states, it is not surprising to find two states reclassifying segments based on a couple of variables, but, performing safety analysis on the complete state's intersection data as a whole. Three states reclassify both segments and intersections based on a couple of variables before performing safety analysis. One state classifies segments based on multiple variables and intersections on a couple of variables. Seven other states reclassify both segments and intersections based on multiple variables prior to performing safety analysis. It is interesting to note that, irrespective of the states' method of sub-categorization of segments and intersections, a majority of states believe that rates require sub-classification of roadway segments to obtain better results.

With the release of new safety analysis tools in the form of SafetyAnalyst and Highway Safety Manual, the knowledge of tools/ methods used by states for safety analysis earned greater significance than anticipated. In this scenario, it is found that thirteen of the 24 responding states are currently using a combination of traditional and advanced methods while six states haven't switched methods in the past five years. Four of the remaining 5 states have completely switched methods in the last 2-5 years. Florida, Missouri, New York State, Ohio, and South Dakota are currently working with Highway Safety Manual. And at present, Florida, Georgia, Kansas, Missouri, New Hampshire, and

Ohio are working with SafetyAnalyst. However, all the twenty-four responding states are found to use traditional methods for their safety analysis. The most commonly used traditional methods include crash frequency (20), crash rate (18), equivalent property damage only (8), high proportion of crash types (8), relative severity index (8), and rate quality control (6). Pennsylvania and South Carolina are currently using EB methodology and Nevada is using CARE (Critical Analysis Reporting Environment) for their site selection. As a supplement to the traditional methods, Washington State is using 0.25 mile sliding window method based on GIS. Survey results show that substantial number of states are using a combination of traditional methods to determine a value - safety index which is used for project selection and prioritization. More emphasis is being given to the locations with high severity crashes and many states are incorporating severity in their analysis. Hawaii is looking at “corridor analysis” instead of “black spots” with more emphasis on low cost safety improvement. Three fourths of the responding states (18 out of 24) are planning to use new highway safety analysis tools (IHSDM, SafetyAnalyst, and HSM) while three states are not considering the option as of now. Some states are hesitant about SafetyAnalyst due to its high initial cost and extensive data requirements. Thirteen out of 24 responding states are currently working with at least one of the new safety analysis tools which include HSM, SafetyAnalyst and IHSDM. Table 40 gives a brief summary of the states that are either currently working/ planning to work with SafetyAnalyst and HSM.

Table 40: States that are currently using and/or planning to use SafetyAnalyst and HSM in the future

State	SafetyAnalyst	HSM
Alaska	No	No
Arkansas	NA	NA
California	No	Yes
Delaware	No	Yes
Florida	Yes	Yes
Georgia	Yes	Not sure
Hawaii	Not sure	Not sure
Idaho	No	Yes
Kansas	Yes	Yes
Maine	No	Yes
Maryland	No	Yes
Massachusetts	No	Yes
Minnesota	No	No
Missouri	Yes	Yes
North Carolina	Yes	Yes
Nevada	Yes	Yes
New Hampshire	Yes	Yes
New Jersey	No	No
New York	No	Yes
Ohio	Yes	Yes
Oregon	Not sure	Yes
Pennsylvania	Yes	Yes
South Carolina	No	Not sure
South Dakota	No	Yes
Washington	Yes	Yes

<u>4.3.6.5 SafetyAnalyst and SPFs:</u> Understanding the data availability, data issues, current safety analysis methods in practice, and the present stand of the states with regard to HSM and SafetyAnalyst, the rest of the survey aims at the states that have been using SafetyAnalyst. As identified earlier, of the 24 responding states, eleven have some experience with SafetyAnalyst. Georgia, Massachusetts, Nevada, North Carolina, and

Ohio have been involved with SafetyAnalyst for over two years now. Kansas, Missouri, New Hampshire, and Washington have been working with the software for 1-2 years now. Florida and Pennsylvania have just started to work with SafetyAnalyst (less than 6 months ago). Ohio, Pennsylvania and Washington State are currently working with all four modules in SafetyAnalyst (network screening, diagnosis and countermeasure selection, economic appraisal and priority ranking, and countermeasure evaluation). Florida, Kansas, New Hampshire and Georgia have yet worked with countermeasure evaluation.

For a comprehensive safety analysis, the software requires several files to be imported which include crash, traffic, segments, intersections and ramps; however none of the states except Florida are able to import ramp data into SafetyAnalyst. Florida is able to import all the five files into SafetyAnalyst while New Hampshire, Ohio and Washington are able to import all the required files except ramps. Georgia and Kansas have successfully imported crash, traffic and segment files. Kansas, New Hampshire and Ohio are using SafetyAnalyst to prioritize sites. Although familiar with SafetyAnalyst, Massachusetts is not actually using the software due to data limitations, and Washington State is still evaluating the software and has just completed importing the files.

As discussed earlier, SafetyAnalyst uses the Empirical Bayes method for performing safety analyses and requires SPFs. Default SPFs for various site subtypes (for segments, intersections and ramps) were generated using data from California, Minnesota, Ohio and Washington. These SPFs are calibrated to reflect the state's road network by multiplying the default SPFs with a calibration factor (which is calculated as the ratio of observed crashes to predicted crashes).

States are encouraged to generate their own SPFs (with their data) as they are believed to produce more reliable results than the default SPFs. However, many states might not have the resources and expertise to generate state specific SPFs. Therefore, the fit of national SPFs to state data is a point of concern.

Of the eleven states currently working with SafetyAnalyst - Georgia, Kansas and Missouri are using their own state specific SPFs (replacing the default SPFs within the software). Florida is also following the same path by having state specific SPFs for some site subtypes. SPFs for Kansas were found to be completely different from the default SPFs used within SafetyAnalyst, and the calibrated default SPFs did not fit the state's SPFs. Whereas, Georgia, Ohio, and Washington have found that some of their state specific site subtypes are well represented by the calibrated version of default SPFs within the software. Missouri has been using its own SPFs but it hasn't looked at their fit versus the fit of default and calibrated SPFs within SafetyAnalyst.

New Hampshire, Ohio and Washington are using the default SPFs within the software. New Hampshire believes that the default SPFs represent the state's SPFs pretty well and therefore has no intentions of generating state specific SPFs. The other two states consider the states' SPFs to be well represented by the default SPFs only for a few site subtypes. Kansas and Ohio are not planning on developing their state specific SPFs while Florida, Minnesota, Pennsylvania and Washington are planning on developing the same. Due to the lack of state specific SPFs, none of the responding states except Georgia have compared the fit of calibrated default SPFs used in SafetyAnalyst against the state specific SPFs.

Seven out of the 8 responding states have identified the process of generation of import files for SafetyAnalyst as difficult, at a minimum. Of the five states that had completed the import process, one state reported less than 4 man-months, two states reported 6-8 man-months, and two states have taken 8-12 man months for importing files into the software. One state has not completed the import process and it is currently using data supplied by AASHTO to evaluate the software.

Importing files into SafetyAnalyst is not a one-time job and needs to be repeated, at a minimum, annually and/or whenever there are major changes to the roadway characteristics database. The time it takes to repeat the import process depends a lot on the magnitude of changes that are made within the database. Six states have reported the time to repeat the process to be between one day and 2-3months.

While importing data into SafetyAnalyst, five of the 6 responding states received numerous errors. A couple of states were able to fix all the errors but ignored warnings, whereas others ignored the errors altogether. Most states had a portion of sites that failed to be processed. Memory issues and data coding were other problems identified by states, Issues with data coding due to data elements not conforming to the requirements of SafetyAnalyst is considered to be another issue for many states.

A majority of states are expected to have shorter segments, as they reported that they identify segments every 0.01 miles if there is a significant change in at least one of their roadway characteristics. Therefore, emphasizing the importance of longer segments in safety analysis, it is found that four out of the responding 8 states generate homogeneous segments using SafetyAnalyst while the rest are unsure about the process. With the Georgia data, a significant difference is observed in the output report in terms of variance when shorter segments and longer aggregated segments are considered. Longer segments produced by the homogeneous segment generation process gave more reliable results and therefore, aggregated segments are used consistently by Georgia.

SafetyAnalyst is not a free software and states are required to buy the annual license from AASHTO. The annual fee for a single workstation use is $11,000 and for multiple workstations is $22,500 for the states that had initially participated in the pooled fund study and $44,000 for the other states. Each license includes 24 hours of engineering support (American Association of State Highway and Transportation Officials, 2010a). Nine states that are currently working with SafetyAnalyst have either already bought the software or are planning to buy it from AASHTO and one state is currently evaluating the software. While a couple of states are unsure, the remaining six states (of the 8 responding states) do recommend SafetyAnalyst to other states. When surveyed about the top 5 most difficult hurdles faced in the process of implementing SafetyAnalyst, the following hurdles are identified where the number of states identifying a hurdle is indicated in parenthesis.

a. Data importing (8)
b. Initial set-up cost (5)
c. Data requirements and intersection data in particular (4)
d. Learning curve (4)
e. Interpreting the results and understanding the defaults (2)
f. IT compatibility issues (3)
g. Switch-over of analysis methodologies, and, processes and procedures (1)

h. Physical memory issues (1)

Based on the states' experience, following are the tips that the states offer to other states, universities and research institutes planning to use the software:

a. Start with a subset of data on a local machine
b. Involve the IT department early on in the process
c. Cross walk state data to SafetyAnalyst data
d. Know what you plan on using it for
e. Understand that it takes considerable resources and time to start-up and spend time accordingly
f. Train users on the capabilities, outputs and validate data to ensure buy-in at various levels
g. Take advantage of the consultant's expertise
h. Understand that expertise must be developed and maintained
i. Factor in time required for implementation
j. Work with management

As the data needs for implementing SafetyAnalyst are intense, funding for collection and processing of data is of importance. Four of the 9 responding states have received federal funds for implementing the new data needs to work with SafetyAnalyst, and others are using research funding and a 408 grant to improve data. Only three states are using SafetyAnalyst to generate priorities for SHSP (Strategic Highway Safety Plan)

<u>4.3.8.6 HSM implementation:</u> In addition to SafetyAnalyst, AASHTO has also released Highway Safety Manual which is considered to "provide the best factual information and

proven analysis tools for crash frequency prediction" (American Association of State Highway and Transportation Officials, 2010b). With proven benefits for using advanced safety analysis methods against traditional frequencies and rates, the implementation and adoption of HSM by states is crucial for safety improvements across the states. Of the 19 responding states, about 18 states have received copies of the 1st edition of HSM with a majority of states receiving multiple copies. A majority of the states have distributed the manuals: a) among its districts and b) to various sections of DOT such as traffic engineering, safety, planning, operations etc. Substantial number of responding states (9) have given the manuals to their respective safety offices. When asked about an implementation plan for the deployment of HSM, three states have an implementation plan in place while seven other states are currently working on developing one. Nine of the 19 responding states have no specific/ formal implementation plan in place. More than fifty percent of the responding states (10 out of 18), have a specific person in charge for HSM implementation, of which two states have a team responsible for the same.

In the view of the complexity of shifting from the current safety analysis methods to the more comprehensive advanced methods, states are asked to determine a time frame for the complete transfer and deployment of HSM. Table 41 gives a brief summary of the time frame the states are looking at for complete deployment of the HSM

Table 41: Summary of the time frame of the states for complete deployment of the HSM

Time frame for complete deployment of the HSM	# of states
not sure	5
several years	2
1-2yrs	5
2-5yrs	2
not likely	3

As expected, a majority of responding states (12 out of 19) are looking towards using HSM as a supplement to their current practices and one state is looking at a complete conversion.

4.3.7 Observations that encourage the deployment of SafetyAnalyst:

Having worked with SafetyAnalyst for substantial amount of time importing Georgia's data and performing network screening module, the following observations that support the deployment of SafetyAnalyst are made:

- SafetyAnalyst uses Empirical Bayes method and addresses the issues, biases and limitations of traditional methods.
- SafetyAnalyst divides roadway network into site subtypes and merges segments into longer homogeneous segments automatically. This process increases the speed and accuracy of safety analysis. The files exported from data management tool of the software are clean with specific information on the site's subtype, number of crashes that occurred at the site and whether the site is valid or not along with the reason for the site's exclusion. Even if using other methods of site selection, the exported files could be used to perform the analyses.

- SafetyAnalyst performs all the steps in roadway safety management process and is mostly automated.
- SafetyAnalyst doesn't require extensive statistical expertise. As the required SPFs for performing EB analysis were already in SafetyAnalyst, the agencies do not require extensive statistical expertise to perform the analyses. However, if available, the default SPFs could be replaced with the agency specific SPFs.
- SafetyAnalyst performs basic data quality checks and logs a list of errors, warnings and potential issues with the data. During the import, post process and calibration steps of the data management tool, SafetyAnalyst identifies and flags segments with unrealistic growth factors, shorter segments, and segments with missing roadway characteristics information.
- The software has the ability to perform sliding window analysis. For example, a 5 mile long segment could be flagged as a site with potential for safety improvement. However, safety might not be an issue on the complete segment. There could be shorter sub segments within the longer 5 mile segment which need to be flagged. SafetyAnalyst identifies the shorter sub-segments as additional windows of interest which may be a cause for concern. Detailed analysis on these shorter sub-segments could be beneficial to the safety analysis.

4.3.8 Observations that discourage the deployment of SafetyAnalyst:

Following are the observations that hinder the SafetyAnalyst deployment:

- *Cost:* SafetyAnalyst is an AASHTOWare product costing an agency about $22,500 annually with additional charges if technical support is required (American Association of State Highway and Transportation Officials, 2010a).
- *Initial generation of import files is tedious:* As discussed in the earlier sections, SafetyAnalyst requires a number of import files to be generated in line with the data requirements and format recommended by the software. This process is often tedious since the data may need to be merged from various files and significant amount of data recoding may need to occur both within the software and within the agency database. However, once the initial files are generated, with proper documentation, repeating the process is quite easy.
- *Data requirements are stringent:* The software requires the agencies to collect certain number of data elements with predefined enumeration values. These requirements could be worked around by changing the array of values collected within the administration tool of SafetyAnalyst. Though, there will be some required elements within the software that could not be changed increasing the complexity of data recoding.
- *SafetyAnalyst could be a 'black box':* Empirical Bayes method of performing network screening requires a significant statistical expertise. This process is automated within the software often making it difficult for the end user to understand the internal steps performed to obtain the output and the output itself.
- *Prescriptive error handling:* SafetyAnalyst has the capability of flagging sites with unrealistic AADT growth factors and miscoded information. The software fails to distinguish between the actual data from errors and thus, might result in flagging normal/ acceptable sites.

In summary, for the states to shift to newer and advanced tools like SafetyAnalyst and Highway Safety Manual, and to perform a complete road safety analysis, comprehensive datasets (crash, roadway characteristics and traffic data) are required and from the current stand of the states, it can be concluded that most states need to start to collect and maintain the required data elements. Most of the states are still using traditional methods like frequency, rate, and safety index. About thirteen states are familiar with newer tools like SafetyAnalyst and HSM. States working with SafetyAnalyst are finding data requirements and data compatibility as issues. A majority of states are looking toward using the HSM as a supplement to their current practices.

Phase 4:

4.4 Develop state specific SPFs using the SafetyAnalyst procedure:

The fourth phase of the project dealt with generating safety performance functions for various site subtypes using Georgia data and comparing their fit to the default national and calibrated SPFs used in SafetyAnalyst (for all site subtypes). Base conditions were identified for two-way two-lane rural roads based on Georgia data and SPFs were generated using the base conditions. Negative binomial regression analysis was carried out to generate SPFs and overdispersion parameter was used to assess the SPFs' goodness-of-fit.

4.4.1 Process of SPF generation:

The first step of SPF generation was the compilation of a dataset with all three pieces of information – segment characteristics, crash data, and traffic data. The dataset included all aggregated segments excluding: a) segments shorter than 0.1 miles; b) segments with null AADTs, and c) segments with unrealistic AADT growth factors. Table 42 shows the number of homogeneous segments excluded in each category by site subtype.

Table 42: Number of segments excluded for generating SPFs by site subtype

Site Subtype	Total # of segments in each site subtype	segments shorter than 0.1 miles	segments with null AADTs	segments with unrealistic AADT growth factors	total number of excluded segments based on segment length and AADT
(null)	5,210	4,113	872	6	4,991
101	70,167	8,611	5,340	2,573	16,524
102	347	72	51	17	140
103	4,490	1,261	303	330	1,894
104	187	32	5	0	37
105	86	7	0	0	7
106	127	9	0	1	10
107	60	7	0	0	7
151	110,720	31,988	9,959	1,539	43,486
152	2,873	1,019	287	158	1,464
153	9,341	5,374	769	246	6,389
154	4,092	1,536	1,605	32	3,173
155	676	256	238	23	517
156	166	47	4	0	51
157	67	20	0	0	20
158	561	184	266	2	452
159	189	40	2	0	42
160	277	83	0	0	83
Total	**209,636**	**54,659**	**19,701**	**4,927**	**79,287**

Once the outliers within each site subtype were identified and excluded, negative binomial regression analysis was run in SAS (see code in Appendix E) to obtain the regression coefficients and over-dispersion parameters. R^2_{FT} was also calculated in SAS for both Georgia specific SPFs and default national SPFs calibrated to Georgia data. Analysis was performed on both total crashes, and Fatal and Injury (FI) crashes. Table 43 and Table 44 provide the results from the analysis. The number of segments used along with the total number of miles of roadway in each site subtype were also shown in addition to the regression coefficients, overdispersion parameters and R^2_{FT} for both Georgia specific SPFs and national default SPFs for both total and FI crashes.

Freeman Tukey's R square is not considered to be a good measure of goodness-of-fit since the model used is negative binomial regression. It was also found that the variations in R square values were insignificant and R^2_{FT} values were also negative for some site subtypes. Therefore, greater emphasis is placed on the overdispersion parameter.

Table 43: National default SPFs and Georgia specific SPFs for TOTAL CRASHES for various site subtypes

			Georgia specific SPFs				National default SPFs used in SafetyAnalyst calibrated to Georgia data				
site subtype	# of segs used	length of segments (miles)	Intercept (α)	Ln AADT (β)	ODP	R^2_{FT}	Intercept (α)	Ln AADT (β)	ODP	calibration factor	R^2_{FT}
101	53,643	79,585.52	-7.660	0.950	1.38	0.604	-3.63	0.53	0.50	0.268	0.581
102	207	474.59	-2.352	0.388	1.31	0.141	-3.17	0.49	0.53	0.997	0.075
103	2,596	1,432.55	-6.601	0.781	1.26	0.198	-5.05	0.66	0.32	0.698	0.121
104	150	393.28	-7.910	0.925	0.23	0.867	-6.82	0.81	0.17	1.162	0.858
105	79	120.73	-10.592	1.173	0.14	0.790	-8.28	0.94	0.09	1.372	0.755
106	117	159.11	-7.493	0.892	0.18	0.687	-7.76	0.97	0.15	0.573	0.682
107	53	59.38	-10.350	1.166	0.22	0.552	-9.63	1.06	0.21	1.653	0.567
151	67,234	34,650.62	-7.694	1.018	1.51	0.595	-7.16	0.84	4.40	2.300	0.596
152	1,409	1,534.19	-3.586	0.685	1.32	0.381	-10.24	1.29	0.85	2.121	0.275
153	2,952	1,396.63	-3.605	0.636	1.57	-0.043	-11.85	1.34	5.91	3.293	-0.023
154	919	683.87	-6.871	0.975	1.77	0.267	-3.53	0.60	1.38	0.147	0.127
155	159	285.49	-4.461	0.664	0.91	0.581	-7.85	1.00	0.99	0.752	0.550
156	115	121.35	-6.918	0.906	0.76	0.656	-5.96	0.78	0.48	1.638	0.635
157	47	24.85	-7.166	0.952	0.70	0.783	-16.24	1.67	0.45	1.450	0.777
158	109	245.18	-5.600	0.786	0.88	0.287	-11.23	1.30	0.81	0.815	0.077
159	147	130.53	-13.401	1.482	0.70	0.456	-11.25	1.28	0.60	1.259	0.470
160	194	188.96	-20.593	2.085	0.96	-0.630	-26.76	2.58	0.52	1.087	-0.540

Table 44: National default SPFs and Georgia specific SPFs for FATAL AND INJURY CRASHES for various site subtypes

			Georgia specific SPFs				National default SPFs used in SafetyAnalyst calibrated to Georgia data				
site subtype	# of seg used	length of segments (miles)	Intercept (α)	Ln AADT (β)	ODP	R^2_{FT}	Intercept (α)	Ln AADT (β)	ODP	Calibration factor	R^2_{FT}
101	53,643	79,585.52	-8.400	0.919	1.053	0.592	-4.860	0.530	0.670	0.295	0.534
102	207	474.59	-4.594	0.495	0.839	0.431	-4.200	0.500	0.530	0.729	0.365
103	2,596	1,432.55	-6.615	0.667	1.021	0.145	-7.460	0.720	0.090	1.553	0.116
104	150	393.28	-6.492	0.683	0.221	0.849	-8.820	0.890	0.160	1.255	0.848
105	79	120.73	-9.080	0.926	0.066	0.836	-10.250	1.030	0.090	1.083	0.824
106	117	159.11	-8.168	0.850	0.205	0.604	-8.860	0.960	0.240	0.613	0.598
107	53	59.38	-6.227	0.694	0.261	0.614	-10.480	1.040	0.200	1.480	0.533
151	67,234	34,650.62	-10.430	1.177	1.570	0.563	-8.840	0.890	4.540	1.623	0.556
152	1,409	1,534.19	-5.138	0.695	1.238	0.459	-12.070	1.390	0.810	1.149	0.343
153	2,952	1,396.63	-5.825	0.720	1.558	-0.005	-14.870	1.520	5.810	2.714	-0.024
154	919	683.87	-10.445	1.162	1.274	0.248	-5.150	0.650	1.450	0.418	0.318
155	159	285.49	-4.385	0.529	0.658	0.600	-8.820	1.020	1.150	0.408	0.551
156	115	121.35	-6.480	0.739	0.476	0.771	-7.600	0.850	0.540	0.885	0.771
157	47	24.85	-13.651	1.367	0.507	0.805	-19.160	1.850	0.520	0.703	0.795
158	109	245.18	-6.594	0.746	0.623	0.448	-12.890	1.380	0.790	0.600	-0.354
159	147	130.53	-13.189	1.347	0.701	0.493	-13.620	1.420	0.550	0.651	0.485
160	194	188.96	-20.390	1.948	0.851	-0.428	-25.630	2.420	0.530	0.576	-0.394

As shown previously in Table 5, the national default SPFs used in SafetyAnalyst were generated from northern and western states' data. A comparison of the fit of default SPFs using state specific data that was originally used to develop the national models and the fit of Georgia's SPFs to Georgia data yielded interesting results. Table 45 compares the two overdispersion parameters and R^2_{FT} values. It is observed that Georgia had significantly more miles of roadway (for 14 out of 17 site subtypes) to generate its SPFs in comparison to the miles of roadway used to generate default SPFs.

Table 45: Comparison of R^2_{FT} and ODP of national default SPFs used in SafetyAnalyst and Georgia specific SPFs for TOTAL CRASHES

	Default National SPFs used in SafetyAnalyst				Georgia specific SPFs		
Site subtype	State	ODP	R^2_{Ft} (%)	Total length of segments (mi)	ODP	R^2_{FT} (%)	Total length of segments (mi)
101	OH	0.50	72.5	12,412	1.38	60.4	79,585.52
102	NC	0.53	46.5	308	1.31	14.1	474.59
103	MN	0.32	49.8	467	1.26	19.8	1,432.55
104	MN	0.17	88.0	379	0.23	86.7	393.28
105	CA	0.09	84.3	201	0.14	79.0	120.73
106	MN	0.15	65.0	90	0.18	68.7	159.11
107	CA	0.21	46.1	238	0.22	55.2	59.38
151	OH	4.40	13.6	1,504	1.51	59.5	34,650.62
152	WA	0.85	23.5	194	1.32	38.1	1,534.19
153	OH	5.91	1.4	327	1.57	-4.3	1,396.63
154	MN	1.38	4.1	170	1.77	26.7	683.87
155	WA	0.99	9.2	126	0.91	58.1	285.49
156	WA	0.48	53.5	35	0.76	65.6	121.35
157	WA	0.45	43.1	15	0.70	78.3	24.85
158	WA	0.81	40.9	156	0.88	28.7	245.18
159	WA	0.60	56.1	83	0.70	45.6	130.53
160	WA	0.52	51.6	31	0.96	-63.0	188.96

Table 46: Comparison of R^2_{FT} and ODP of national default SPFs used in SafetyAnalyst and Georgia specific SPFs for FATAL & INJURY CRASHES

	Default National SPFs used in SafetyAnalyst				Georgia specific SPFs		
Site subtype	State	ODP	R^2_{Ft} (%)	Total length of segments (mi)	ODP	R^2_{Ft} (%)	Total length of segments (mi)
101	OH	0.670	59.9	12,412	1.053	59.2	79,585.52
102	NC	0.530	45.9	308	0.839	43.1	474.59
103	MN	0.090	37.2	467	1.021	14.5	1,432.55
104	MN	0.160	82.2	379	0.221	84.9	393.28
105	CA	0.090	82.8	201	0.066	83.6	120.73
106	MN	0.240	53.1	90	0.205	60.4	159.11
107	CA	0.200	45.3	238	0.261	61.4	59.38
151	OH	4.540	14.0	1,504	1.570	56.3	34,650.62
152	WA	0.810	25.8	194	1.238	45.9	1,534.19
153	OH	5.810	2.2	327	1.558	-0.5	1,396.63
154	MN	1.450	11.1	170	1.274	24.8	683.87
155	WA	1.150	12.8	126	0.658	60.0	285.49
156	WA	0.540	46.4	35	0.476	77.1	121.35
157	WA	0.520	39.9	15	0.507	80.5	24.85
158	WA	0.790	38.1	156	0.623	44.8	245.18
159	WA	0.550	56.0	83	0.701	49.3	130.53
160	WA	0.530	48.9	31	0.851	-42.8	188.96

The graphs of safety performance functions for various site subtypes for total and fatal injury crashes are shown in Appendix G. Figure 18 and Figure 19 shows a set of SPFs (SPFs for two-way two-lane rural roads for total, and fatal and injury crashes) as an example.

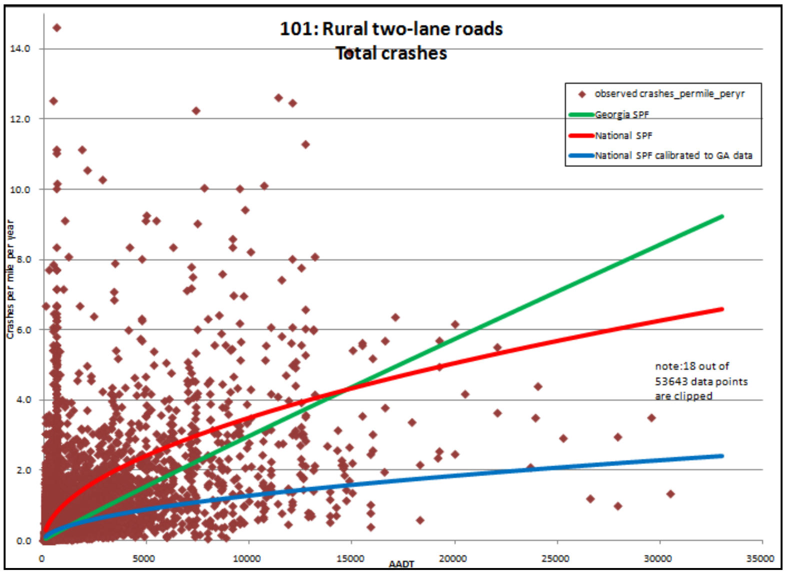

Figure 18: SPFs for site subtype 101 considering total crashes

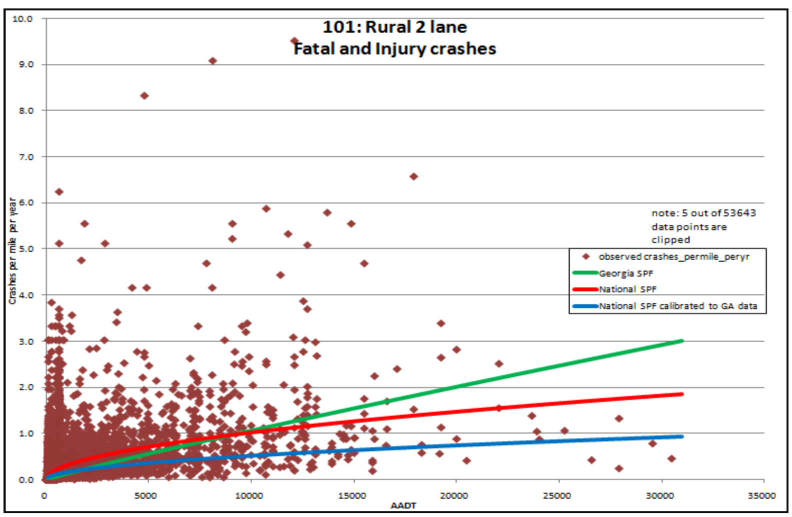

Figure 19: SPFs for site subtype 101 considering Fatal and Injury crashes

4.4.2 Influence of actual measured AADT on the fit of SPFs:

In Georgia, it is found that less than 25% of the traffic data is actually collected while the rest is estimated. Therefore, the fit of SPFs depend a lot on the reliability of the traffic data. The influence of actual AADT on the fit of SPFs is therefore studied.

For all site subtypes, segments with actual versus estimated AADT information were parsed out into a separate dataset and SPFs were generated using the new datasets. Prior to generating SPFs, data cleaning was performed as discussed in section 3.4.1.

With Georgia data, it is observed that traffic counts are estimated on over 75% of the total roadways, therefore bringing the reliability of predictions into question as they are solely based on traffic data. Predictably, it was found that the number of segments with actual measured traffic counts are not distributed evenly across site subtypes. Rural segments had a higher proportion of actual traffic count data, whereas urban segments had a lower proportion in general. Site subtypes 101 (rural two-lane), 151 (urban two-lane arterial streets), and 154 (urban one-lane arterial streets) had actual counts for less than a quarter of mileage.

Table 47 gives a split of the number of miles of roadway with actual and estimated traffic counts.

Table 47: Comparison of total miles of roadway segments with actual and total data by site subtypes

Site Subtype	Total number of segments		Segments with actual traffic data		% of total length with actual traffic data
	# of segments	total segment length in miles	# of segments	total segment length in miles	
(null)	5,210	428.34	144	31.39	7.33
101	70,167	79,585.52	3,629	18,675.32	23.47
102	347	474.59	273	450.24	94.87
103	4,490	1,432.55	3,926	1,195.01	83.42
104	187	393.28	171	389.75	99.10
105	86	120.73	74	110.08	91.18
106	127	159.11	124	157.76	99.15
107	60	59.38	46	48.94	82.42
151	110,720	34,650.62	4,371	4,455.37	12.86
152	2,873	1,534.19	1,565	1,178.23	76.80
153	9,341	1,396.63	5,502	970.28	69.47
154	4,092	683.87	245	70.84	10.36
155	676	285.49	242	189.75	66.46
156	166	121.35	152	110.79	91.30
157	67	24.85	62	24.48	98.51
158	561	245.18	114	109.70	44.74
159	189	130.53	166	126.20	96.68
160	277	188.96	245	185.84	98.35
Total	**209,636**	**121,915.17**	**21,051**	**28,479.97**	**23.36**

Table 48 gives the SPFs for total crashes for the 17 site subtypes generated using segments with actual AADT counts. It also gives the R^2_{FT} values and overdispersion parameters.

The low R^2_{FT} value can be caused by a number of issues:

1) The limited ability to explain the crash relationship with AADT in urban settings

2) The limited number of segments and/or mileage used to generate state specific SPFs

3) The use of incorrect functional form of the model

Table 48: Georgia specific SPFs for TOTAL CRASHES using segments with actual AADTs for various site subtypes

TOTAL CRASHES: Using segments with actual AADTs						
site subtype	Intercept (α)	Ln AADT (β)	ODP	R^2_{Ft}	# of segments	length of segments (miles)
101	-6.402	0.812	0.736	0.476	2945	16579.32
102	-1.996	0.351	1.266	0.127	213	425.62
103	-7.218	0.849	1.190	0.209	2634	1031.55
104	-7.466	0.881	0.213	0.869	147	388.65
105	-11.284	1.233	0.135	0.802	69	109.74
106	-7.681	0.910	0.170	0.690	116	157.32
107	-8.712	1.022	0.222	0.500	41	48.67
151	-6.560	0.914	0.829	0.543	2857	3993.23
152	-7.344	1.073	1.282	0.237	1019	1029.97
153	-5.769	0.855	1.491	-0.137	2591	770.31
154	-2.762	0.544	1.266	-0.459	131	59.45
155	-3.330	0.561	0.796	0.563	140	172.00
156	-7.495	0.958	0.845	0.613	113	108.61
157	-7.043	0.942	0.717	0.783	45	23.61
158	-9.700	1.157	0.696	-0.037	94	108.66
159	-12.857	1.438	0.695	0.431	139	124.85
160	-21.992	2.200	0.943	-0.696	185	183.06

When the R^2_{FT} values and overdispersion parameters are compared between the SPFs generated for total crashes using segments with only actual AADTs and segments with both actual and estimated AADTs, it is observed that the reliability of the functions depend on the percent of miles of segments with actual traffic data. The overdispersion parameter was significantly improved when the segments with actual traffic data were used in model generation. However, this is not true in two instances (site subtypes 156 and 157: Urban freeways with 6 and 8+ lanes), there has been a slight increase in the overdispersion parameter value. It is believed that the reliability of a SPF does not solely depend on the R^2_{FT} and the R^2_{FT} values of SPFs with just AADT as an independent variable are expected to be lower as there are many factors that influence the frequency and severity of crashes in addition to traffic. R^2_{Ft} values are not considered as a good measure of the goodness-of-fit when the negative binomial regression models are considered. The overdispersion parameter is considered to be an important factor since it is used to weigh the reliability of the function against the observed crashes in the EB method while calculating the expected crash frequency at a site.

Table 49 compares the overdispersion parameters and R^2_{FT} values of SPFs for total crashes generated using segments with actual AADT, and segments with both actual and estimated AADT values for all site subtypes. It is observed that, when the percent change in overdispersion parameter and increase in R^2_{FT} values are compared for SPFs generated using segments with both actual and estimated AADTs, and segments with just actual AADTs, it could be concluded that in most of the cases, there is no significant change in R^2_{FT} values while the overdispersion parameters are significantly lowered on SPFs generated using segments with actual AADT counts. It is interesting to observe that greatest percent reduction in overdispersion parameter occurred on site subtypes which have fewer percent of total segment length with actual traffic counts. This observation is acceptable as traffic estimations, which, depending on the estimating method might not resemble the actual traffic counts.

Table 49: Comparison of overdispersion parameters and R square values of SPFs for TOTAL crashes generated using segments with actual AADT and segments with both actual and estimated AADT values for all site subtypes

site subtype	% of total length with actual traffic data	Using segments with actual + estimated AADTs (CASE 1)		Using segments with actual AADTs (CASE 2)		% reduction in Dispersion factor from case 1 to case 2	change in R square value from case 1 to case 2
		ODP	R^2_{FT}	ODP	R^2_{FT}		
101	23.47	1.377	0.604	0.736	0.476	46.55	-0.13
102	94.87	1.307	0.141	1.266	0.127	3.14	-0.01
103	83.42	1.262	0.198	1.190	0.209	5.71	0.01
104	99.10	0.233	0.867	0.213	0.869	8.58	0.00
105	91.18	0.145	0.790	0.135	0.802	6.90	0.01
106	99.15	0.177	0.687	0.170	0.690	3.95	0.00
107	82.42	0.224	0.552	0.222	0.500	0.89	-0.05
151	12.86	1.512	0.595	0.829	0.543	45.17	-0.05
152	76.80	1.323	0.381	1.282	0.237	3.10	-0.14
153	69.47	1.570	-0.043	1.491	-0.137	5.03	-0.09
154	10.36	1.767	0.267	1.266	-0.459	28.35	-0.73
155	66.46	0.907	0.581	0.796	0.563	12.24	-0.02
156	91.30	0.762	0.656	0.845	0.613	-10.89	-0.04
157	98.51	0.696	0.783	0.717	0.783	-3.02	0.00
158	44.74	0.884	0.287	0.696	-0.037	21.27	-0.32
159	96.68	0.701	0.456	0.695	0.431	0.86	-0.03
160	98.35	0.960	-0.630	0.943	-0.696	1.77	-0.07

Similar observations were registered while considering the SPFs for Fatal and Injury crashes. Table 50 gives the SPFs for FI crashes for the 17 site subtypes generated using segments with actual AADT counts.

Table 50: Georgia specific SPFs for FATAL & INJURY CRASHES using segments with actual AADTs for various site subtypes

FATAL AND INJURY CRASHES: Using segments with actual AADTs						
site subtype	Intercept (α)	Ln AADT (β)	ODP	R^2_{Ft}	# of segments	length of segments (miles)
101	-6.519	0.705	0.561	0.550	2945	16579.32
102	-3.696	0.397	0.810	0.450	213	425.62
103	-7.251	0.738	1.016	0.140	2634	1031.55
104	-6.232	0.658	0.216	0.848	147	388.65
105	-9.973	1.005	0.060	0.849	69	109.74
106	-8.303	0.863	0.202	0.605	116	157.32
107	-3.406	0.445	0.232	0.607	41	48.67
151	-7.163	0.829	0.658	0.580	2857	3993.23
152	-8.645	1.058	1.122	0.338	1019	1029.97
153	-7.446	0.884	1.491	-0.094	2591	770.31
154	-3.072	0.391	1.335	-0.357	131	59.45
155	-3.449	0.445	0.625	0.568	140	172.00
156	-7.358	0.819	0.550	0.728	113	108.61
157	-13.458	1.351	0.529	0.805	45	23.61
158	-8.929	0.961	0.540	0.262	94	108.66
159	-11.922	1.242	0.725	0.471	139	124.85
160	-21.683	2.054	0.803	-0.524	185	183.06

When the R square values and the overdispersion parameters of SPFs for Fatal and Injury crashes generated using segments with actual AADT and segments with both actual and estimated AADT values are compared, it is found that there is no significant difference in the R square values of the two cases. However, though not for all site subtypes, overdispersion parameters improved considerably when segments with actual traffic counts were used for SPF development. Supporting the above discussion, Table 51 compares the overdispersion parameters and R square values of SPFs generated from segments with actual and estimated AADTs, and segments with just actual AADT values for fatal and all injury crashes for all site subtypes.

Table 51: Comparison of overdispersion parameters and R square values of SPFs for Fatal & Injury crashes generated using segments with actual AADT and segments with both actual and estimated AADT values for all site subtypes

site subtype	% of total length with actual traffic data	Using segments with actual + estimated AADTs (CASE 1)		Using segments with actual AADTs (CASE 2)		% reduction in Dispersion factor from case 1 to case 2	change in R square from case 1 to case 2
		ODP	R^2_{FT}	ODP	R^2_{FT}		
101	23.47	1.053	0.592	0.561	0.550	46.72	-0.04
102	94.87	0.839	0.431	0.810	0.450	3.42	0.02
103	83.42	1.021	0.145	1.016	0.140	0.52	0.00
104	99.10	0.221	0.849	0.216	0.848	2.31	0.00
105	91.18	0.066	0.836	0.060	0.849	9.85	0.01
106	99.15	0.205	0.604	0.202	0.605	1.61	0.00
107	82.42	0.261	0.614	0.232	0.607	11.23	-0.01
151	12.86	1.570	0.563	0.658	0.580	58.1	0.02
152	76.80	1.238	0.459	1.122	0.338	9.4	-0.12
153	69.47	1.558	-0.005	1.491	-0.094	4.31	-0.09
154	10.36	1.274	0.248	1.335	-0.357	-4.77	-0.61
155	66.46	0.658	0.600	0.625	0.568	5.08	-0.03
156	91.30	0.476	0.771	0.550	0.728	-15.53	-0.04
157	98.51	0.507	0.805	0.529	0.805	-4.24	0.00
158	44.74	0.623	0.448	0.540	0.262	13.37	-0.19
159	96.68	0.701	0.493	0.725	0.471	-3.41	-0.02
160	98.35	0.851	-0.428	0.803	-0.524	5.64	-0.10

In summary, it is found that the overdispersion parameter improved significantly for the SPFs generated using segments with just the actual traffic data. It implies that traffic data estimations increase the variations within the data resulting in larger overdispersion factors. A larger overdispersion parameter results in giving less weight to the developed model and more weight to the observed crash counts. Therefore, the use of models with lower ODP values giver better estimations in the Empirical Bayes analysis.

4.4.3 Identify base conditions for two-way two-lane rural roads for Georgia data and generate SPFs using base conditions:

In order to compare SafetyAnalyst SPF with HSM SPF (generated in phase 6), another version of Georgia specific SPF was developed using only base conditions for two-way two-lane rural roads as required by the Highway Safety Manual.

Following are the assumptions made while identifying base conditions for two-way two-lane rural roads:

- Horizontal curves are assumed to be absent
- There is no super elevation
- Vertical grades are absent
- Driveway density is 5 driveways/ mile
- There are no centerline rumble strips and no TWLTL
- Roadside hazard rating is 3.0
- There is no lighting and no automated speed enforcement

Several combinations of variables and their corresponding array of values are compared to identify the “base conditions” that constitute a majority of the roadways.

Longer aggregated segments were generated from the segments with base conditions using the following variables: Rclink, area type, functional classification, road width, access control, operation way, total lanes on each direction, auxiliary lane type on each direction, and undivided shoulder width and type on each direction.

Table 52 gives the criteria used for identifying base conditions on two-way two-lane rural roads. Table 53 shows different criteria used to identify base conditions for two-lane two-way rural roads in Georgia.

Table 52: Criteria used for identifying base conditions for two-way two-lane rural roads in Georgia

Variable	Condition
Area type	Rural
Operation way	Two-way
Total number of lanes	≤ 3
Total lane width	= 24 ft
Undivided Shoulder width left and right	= 2 ft
Undivided Shoulder type left and right	Paved
Auxiliary lane left and right	neither a passing lane nor a climbing lane

Table 53: Criteria used to identify base conditions for two-lane two-way rural roads

Area type	Operation way	Total lanes	Road width (ft)	Left Shldr width (ft)	Right Shldr width (ft)	Aux lane type left	Aux lane type right	Total # of segs.	total length (miles)
Rural	Two way	≤ 3	24	6	6			73	5.11
Rural	Two way	≤ 3	23	6	6			1	0.05
Rural	Two way	≤ 3	24	5	5			172	19.99
Rural	Two way	≤ 3	24	4	4			1583	276.31
Rural	Two way	≤ 3	24	3	3			1031	214.17
Rural	Two way	≤ 3	24	2	2	!=E*	!=E*	9682	2103.89
Rural	Two way	≤ 3	23	2	2			1114	268.38
Rural	Two way	≤ 3	23	3	3			590	163.14
Rural	Two way	≤ 3	23	4	4			33	5.02
Rural	Two way	≤ 3	22	3	3			305	100.26
Rural	Two way	≤ 3	22	2	2			1130	285.57

*E corresponds to either a passing lane or a climbing lane

In total, there are 2,458 two-way two-lane rural aggregated segments (or homogeneous segments) for use in development of base condition SPF with SafetyAnalyst functional form. Of these 2458 segments, 60 segments have no traffic information and about 66 segments have estimated traffic counts. Therefore, actual traffic counts were available for about 95% of the segments. However, as previously identified, about 45% of the aggregated segments (1,074) are shorter than 0.1 miles.

Different analysis datasets were created and SAS was run to determine the SPF that best fits Georgia data. Table 54 gives the SPFs for total crashes developed using different analysis datasets and their corresponding R^2_{FT} values.

Table 54: SPFs of various analysis datasets for total crashes generated using base conditions and their corresponding R^2_{FT} values

Criteria	# of segs	Total length of segments (miles)	Alpha	Beta	ODP	R^2_{FT}
All HS	2398	1997.35	-5.8680	0.7617	0.7361	0.395
All HS > 0.1miles	1323	1949.71	-5.3163	0.6818	0.6416	0.330
All HS with actual traffic data & > 0.1miles	1258	1879.70	-5.3576	0.6865	0.6206	0.350

Based on the overdispersion parameter of different analysis datasets, it is recommended to use the longer aggregated segments (segments longer than 0.1 miles) with actual traffic data.

Phase 5:

4.5 Formulate and document a calibration procedure for two-way two-lane rural roads using HSM procedure:

Up until this point, all of the analysis has focused on SafetyAnalyst. However, another tool, the Highway Safety Manual is also available for similar analysis. While SafetyAnalyst has been dubbed as a suite of tools for state wide safety analysis, it is expected that many states may use SafetyAnalyst to screen for sites and HSM to conduct more detailed site specific analysis. Thus, the process of implementing the HSM for two-way two-lane rural roadway segments was conducted.

The national SPFs published in the Highway Safety Manual, which are recommended to be used by other states need to be calibrated to address the variations in geography, travel patterns, climate etc across states and also within states (if there are identifiable differences).

Phase 5 includes the development of a calibration factor for two-lane rural roads. As the HSM documentation provides only an overview of the procedure, more detailed procedures are explained herein.

The procedure calls for a random sample of 30-50 two-lane rural segments within the area of interest on which a minimum of 100 crashes are recorded within a year time. For this research, 52 rural two-way two-lane roadway segments with a total of 302 observed crashes over a span of 3 years were randomly selected from across the state of Georgia. Site-specific and detailed geometric information was obtained from plan profile sheets and the segments were constituted of both horizontal and tangent sections. The

52 random segments were divided into 399 smaller segments separating segments with horizontal curves from segments on tangent sections. Roadside information, such as the presence of a passing lane, roadside hazard rating, driveway density etc. were obtained from Google maps. The traffic and crash information was obtained from databases described previously in phase 1. For each of these 399 segments, CMFs were identified for all characteristics that did not match the following base conditions:

- Lane width: 12 ft
- Shoulder width: 6 ft
- Shoulder type: Paved
- Roadside hazard rating: 3
- Driveway density: 5 driveways per mile
- Horizontal curve: none
- Vertical grade: 0%
- Presence of a passing lane: none
- Presence of a TWLTL: none

The different CMFs were then multiplied with the predicted crashes (calculated using the HSM procedure) obtained using base SPFs to obtain the total predicted crashes adjusted to site specific conditions. The calibration factor was calculated as the ratio of total observed crashes to total adjusted predicted crashes.

4.5.1 Crash Modification Factors:

Eight of the 12 CMFs were calculated with Georgia data. Table 55 gives a list of the CMFs calculated. The procedure described in the HSM was followed in calculating CMFs for the 52 randomly identified segments.

Table 55: CMFs calculated with Georgia data

CMF	CMF variable
1*	Lane width*
2*	Shoulder width and type*
3*	Horizontal curves: Length, Radius and presence. Absence of spiral transitions*
5	Vertical grades
6	Driveway density
8	Passing lanes
9*	Two-way-left-turn lanes*
10	Roadside design

*Required CMFs

As expected, a majority of data variables in Georgia deviate from the pre-defined base conditions that were used to generate CMFs and calibration factors. Therefore, it is observed that almost all segments have a number of variables to be adjusted to base conditions. Table 56 shows the descriptive statistics for CMFs calculated with Georgia data.

Table 56: Descriptive statistics for CMFs calculated with Georgia data

Variable and its criteria	% of miles with conditions worse than base conditions	Average CMF value	Min CMF value	Max CMF value
lane width < 12 ft	30.24	1.022	1.000	1.146
shoulder is not paved	71.85	1.080	0.937	1.250
Shoulder width < 6ft	62.22			
horizontal curve is present	28.56	1.193	1.000	10.471
vertical grade > 3%	51.69	1.041	1.000	1.160
Driveway density > 5 driveways/mile	67.61	1.214	1.000	1.928
Passing lane is present	6.65	0.993	0.650	1.000
TWLTL is present	4.94	0.996	0.813	1.000
Roadside hazard rating > 3	18.91	1.007	0.818	1.143

Given the predominance of non base conditions, each of the 399 smaller segments had one or more of the CMFs multiplied by the predicted crashes to obtain the calibration factor. In practice, the HSM recommends that no more than 2-3 CMFs be multiplied together to determine crash reductions associated with implementing countermeasures. Thus, there were concerns about doing the same in developing the calibration factor.

Each CMF is associated with a standard error, most of which are unknown in the first version of the HSM. For this research, a standard error of 0.1 was assumed for all the CMFs. When a standard error is considered for a particular CMF, if its range includes 1.00, it implies that the CMF might increase or decrease the predictions and the CMF is considered to have a potential “reversal phenomenon”.

Table 57 gives the mean, standard deviation, minimum, and maximum values for required variables as well as for CMFs associated with each of the 399 segments.

Table 57: Descriptive statistics for the variables and CMFs used in the calibration process

Variable/ CMF	Mean	Std. dev	Min value	Max value
segment length in miles	0.236	0.307	0.010	2.100
Average yearly AADT (veh/day)	3562.490	3313.152	136.667	15923.333
length of horizontal curves in ft	756.788	399.498	35.000	2466.700
Tangent length in ft	397.840	208.129	67.800	1252.800
Radius of horizontal curves in ft	3311.00	3092.03	116.800	17189.000*
Vertical grade in %	0.567	1.553	0.000	8.000
road width in ft	23.737	2.782	20.000	38.000
shoulder width in ft	4.619	1.957	0.000	12.000
driveway density (driveways/mile)	12.110	10.274	0.000	37.000
roadside hazard rating	3.093	0.709	1.000	5.000
CMF for lane width	1.022	0.042	1.000	1.146
CMF for shoulder width and type	1.080	0.055	0.937	1.250
CMF for horizontal curve	1.193	0.740	1.000	10.471
CMF for vertical grade	1.041	0.058	1.000	1.160
CMF for driveway density	1.214	0.278	1.000	1.928
CMF for passing lane	0.993	0.043	0.650	1.000
CMF for TWLTL	0.996	0.021	0.813	1.000
CMF for roadside hazard rating	1.007	0.049	0.818	1.143

*Eight values were considered as outliers due to extremely large values

4.5.2 Sensitivity analysis:

Three types of sensitivity analyses were performed to illustrate the effect of various variables on the predicted number of crashes.

c) Effect of variation of AADT on the predicted number of crashes,

d) Effect of variations of each CMF on the predicted number of crashes if means of all other CMFs are considered, and effect of variations of each CMF on the predicted number of crashes if all other variables are considered equal to base conditions, and

e) Effect of each CMF individually and a combination of CMFs on the calibration factor

4.5.2.1 Effect of variation of AADT on the predicted number of crashes:

The following tables show the variations in predicted crash frequency with AADT when all CMFs were equal to their mean values. From Table 58 and Table 59, it is observed that the predicted crashes in Georgia are about 1.6 times greater than the default predicted crash numbers that were generated using base conditions. This is because the base conditions vary considerably with Georgia data.

Table 58: Sensitivity of predicted crashes to AADT when existing values of variables are used

AADT (veh/day)	Predicted crashes (crashes/mile year)
400	0.18
1000	0.44
3000	1.33
5000	2.21
10000	4.43

Table 59: Sensitivity of predicted crashes to AADT when all other variables are base conditions

AADT (veh/day)	Predicted crashes (crashes/mile year)
400	0.11
1000	0.27
3000	0.80
5000	1.34
10000	2.67

4.5.2.2 Effect of variations of each CMF on the predicted number of crashes:

Next, variations in predicted crash frequency with specific changes to each CMF are shown.

Lane width:

Table 60 shows the sensitivity of predicted total crashes to lane width when all other variables are kept equal to the site conditions. It is observed that the effect of lane width increases considerably with the increase in AADT. For example, for an AADT of 10,000 veh/day, a 9 ft lane results in ~25% increase in the predicted total crashes when compared to the base condition of 12 ft lanes when all other variables remain unchanged.

Table 60: Sensitivity of predicted total crashes to lane width when all other variables are kept constant equal to their average value

AADT	Predicted crashes	Lane width (ft)			
		9	10	11	12
400	0.17	0.18	0.17	0.17	0.17
1000	0.43	0.48	0.46	0.44	0.43
3000	1.30	1.62	1.49	1.33	1.30
5000	2.17	2.69	2.48	2.22	2.17
10000	4.33	5.39	4.96	4.44	4.33

Similarly, when the effect of changes in lane width on the number of predicted total crashes considering all other variables to be base conditions is considered, it is observed that for an AADT of 10,000 veh/day, there is approximately 25% increase in predicted total crashes in comparison to the base lane width of 12 ft. Table 61 shows the sensitivity of predicted total crashes to lane width when all other variables are at base conditions.

Table 61: Sensitivity of predicted total crashes to lane width when all other variables are at base conditions

AADT	Predicted crashes	Lane width (ft)			
		9	10	11	12
400	0.11	0.11	0.11	0.11	0.11
1000	0.27	0.30	0.28	0.27	0.27
3000	0.80	1.00	0.92	0.82	0.80
5000	1.34	1.66	1.53	1.37	1.34
10000	2.67	3.32	3.06	2.74	2.67

Shoulder width and type:

Table 62 and Table 63 show the sensitivity of predicted total crashes to shoulder width and type when all other variables are kept equal to the site conditions. The worst possible situation is to have a 2 ft turf shoulder with an AADT of 10,000 veh/day. This scenario results in an increase of about 16.5% crashes when compared to the ideal base conditions of 6 ft paved shoulder. Independent of the type of shoulder, a 2 ft shoulder has a minimum of 14.8% increase in crashes when compared to the ideal condition.

Table 62: Sensitivity of predicted total crashes to shoulder type and width when all other variables are kept constant

AADT	Predicted crashes	Shoulder type and width (ft)							
		paved				Gravel			
		2	4	6	8	2	4	6	8
400	0.16	0.17	0.17	0.16	0.16	0.17	0.17	0.17	0.16
1000	0.41	0.44	0.42	0.41	0.41	0.44	0.43	0.41	0.41
3000	1.23	1.41	1.32	1.23	1.15	1.42	1.33	1.24	1.16
5000	2.05	2.35	2.20	2.05	1.92	2.36	2.21	2.07	1.94
10000	4.10	4.70	4.40	4.10	3.84	4.72	4.42	4.14	3.88

Table 63: Sensitivity of predicted total crashes to shoulder type and width when all other variables are kept constant cont

AADT	Predicted crashes	Shoulder type and width (ft)							
		Composite				Turf			
		2	4	6	8	2	4	6	8
400	0.16	0.17	0.17	0.17	0.17	0.17	0.17	0.17	0.17
1000	0.41	0.45	0.43	0.42	0.42	0.45	0.43	0.43	0.43
3000	1.23	1.42	1.34	1.25	1.18	1.43	1.35	1.28	1.21
5000	2.05	2.37	2.23	2.09	1.97	2.39	2.26	2.13	2.02
10000	4.10	4.75	4.47	4.18	3.95	4.78	4.51	4.26	4.03

As expected, when the sensitivity of predicted total crashes to shoulder type and width when all others variables are considered to be base conditions observed, as shown in Table 64 and Table 65, the increase in the number of predicted total crashes with variations in shoulder width and type is up to 16.47%. The adverse effects of shoulder width and shoulder type are severe when larger AADT values are considered.

Table 64: Sensitivity of predicted total crashes to shoulder width and type when all other variables are base conditions

AADT	Predicted crashes	Shoulder type and width (ft)							
		paved				Gravel			
		2	4	6	8	2	4	6	8
400	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11
1000	0.27	0.29	0.28	0.27	0.26	0.29	0.28	0.27	0.27
3000	0.80	0.92	0.86	0.80	0.75	0.92	0.86	0.81	0.76
5000	1.34	1.53	1.43	1.34	1.25	1.54	1.44	1.35	1.26
10000	2.67	3.06	2.87	2.67	2.50	3.08	2.88	2.70	2.53

Table 65: Sensitivity of predicted total crashes to shoulder width and type when all other variables are base conditions cont

AADT	Predicted crashes	Shoulder type and width (ft)							
		Composite				Turf			
		2	4	6	8	2	4	6	8
400	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11
1000	0.27	0.29	0.28	0.27	0.27	0.29	0.28	0.28	0.28
3000	0.80	0.93	0.87	0.82	0.77	0.93	0.88	0.83	0.79
5000	1.34	1.55	1.46	1.36	1.29	1.56	1.47	1.39	1.31
10000	2.67	3.10	2.91	2.72	2.57	3.11	2.94	2.78	2.63

Horizontal curve:

The third CMF is for the presence of horizontal curve with and without a spiral transition. The presence of a spiral transition slightly improves safety. Table 66 to Table 69 illustrates the sensitivity of predicted crashes to the presence of a horizontal curve with/ without a spiral transition when all other variables are kept constant and when all other variables are considered to be of base conditions. Similar to earlier discussed CMFs, AADT has a significant effect on the predicted crashes on a horizontal curve. The predicted crash number depends on the length and radius of the horizontal curve. The worst situation, a horizontal curve of 100 ft length and a 100 ft radius without a spiral transition, result in 105 crashes for an AADT of 10,000 veh/day. This is about a 27% increase in predictions from the base conditions of no horizontal curve. A similar trend is observed when the sensitivity of the presence of horizontal curve is calculated when all other variables are base conditions. Similar to the predictions considering the site conditions, there is approximately 27 % increase in predicted crashes while base conditions are considered.

Table 66: Sensitivity of predicted crashes to the presence of horizontal curve without spiral transition when all other variables are kept constant

		Horizontal curve - Curves without spiral transition												
		Curve length = 100 ft			Curve length = 500 ft			Curve length = 1,000 ft			Curve length = 2,000 ft			
		Radius (ft)			Radius (ft)			Radius (ft)			Radius (ft)			
AADT	**Pred. crashes**	**100**	**200**	**500**	**500**	**1000**	**2000**	**1000**	**2000**	**5000**	**1000**	**2000**	**5000**	**Tangent**
400	0.15	4.20	2.18	0.96	0.31	0.23	0.19	0.19	0.17	0.16	0.17	0.16	0.15	0.15
1000	0.37	10.51	5.44	2.40	0.78	0.57	0.47	0.47	0.42	0.39	0.42	0.40	0.38	0.37
3000	1.11	31.53	16.32	7.20	2.33	1.72	1.42	1.42	1.27	1.17	1.27	1.19	1.14	1.11
5000	1.86	52.56	27.21	12.00	3.88	2.87	2.36	2.36	2.11	1.96	2.11	1.98	1.91	1.86
10000	3.71	105.11	54.41	23.99	7.77	5.74	4.73	4.73	4.22	3.91	4.22	3.97	3.81	3.71

Table 67: Sensitivity of predicted crashes to the presence of horizontal curve with spiral transition when all other variables are kept constant

		Horizontal curve - Curves with spiral transition												
		Curve length = 100 ft			Curve length = 500 ft			Curve length = 1,000 ft			Curve length = 2,000 ft			
		Radius (ft)			Radius (ft)			Radius (ft)			Radius (ft)			
AADT	**Pred. crashes**	**100**	**200**	**500**	**500**	**1000**	**2000**	**1000**	**2000**	**5000**	**1000**	**2000**	**5000**	**Tangent**
400	0.15	4.14	2.12	0.90	0.30	0.22	0.18	0.18	0.16	0.15	0.17	0.16	0.15	0.15
1,000	0.37	10.36	5.29	2.25	0.75	0.54	0.44	0.46	0.41	0.38	0.41	0.39	0.37	0.37
3,000	1.11	31.08	15.87	6.74	2.24	1.63	1.33	1.37	1.22	1.13	1.24	1.17	1.12	1.11
5,000	1.86	51.80	26.45	11.24	3.73	2.72	2.21	2.29	2.03	1.88	2.07	1.94	1.87	1.86
10,000	3.71	103.60	52.90	22.48	7.46	5.44	4.42	4.57	4.07	3.76	4.14	3.89	3.74	3.71

Table 68: Sensitivity of predicted crashes to the presence of horizontal curve without spiral transition when all other variables are base conditions

		Horizontal curve - Curves without spiral transition												
		Curve length = 100 ft			Curve length = 500 ft			Curve length = 1,000 ft			Curve length = 2,000 ft			
		Radius (ft)			Radius (ft)			Radius (ft)			Radius (ft)			
AADT	**Pred. crashes**	**100**	**200**	**500**	**500**	**1000**	**2000**	**1000**	**2000**	**5000**	**1000**	**2000**	**5000**	**Tangent**
400	0.11	3.03	1.57	0.69	0.22	0.17	0.14	0.14	0.12	0.11	0.12	0.11	0.11	0.11
1000	0.27	7.57	3.92	1.73	0.56	0.41	0.34	0.34	0.30	0.28	0.30	0.29	0.27	0.27
3000	0.80	22.70	11.75	5.18	1.68	1.24	1.02	1.02	0.91	0.85	0.91	0.86	0.82	0.80
5000	1.34	37.83	19.58	8.63	2.80	2.07	1.70	1.70	1.52	1.41	1.52	1.43	1.37	1.34
10000	2.67	75.66	39.17	17.27	5.59	4.13	3.40	3.40	3.04	2.82	3.04	2.85	2.74	2.67

Table 69: Sensitivity of predicted Sensitivity of predicted crashes to the presence of horizontal curve with spiral transition when all other variables are base conditions

		Horizontal curve - Curves with spiral transition												
		Curve length = 100 ft			Curve length = 500 ft			Curve length = 1,000 ft			Curve length = 2,000 ft			
		Radius (ft)			Radius (ft)			Radius (ft)			Radius (ft)			
AADT	**Pred. crashes**	**100**	**200**	**500**	**500**	**1000**	**2000**	**1000**	**2000**	**5000**	**1000**	**2000**	**5000**	**Tangent**
400	0.11	2.98	1.52	0.65	0.21	0.16	0.13	0.13	0.12	0.11	0.12	0.11	0.11	0.11
1,000	0.27	7.46	3.81	1.62	0.54	0.39	0.32	0.33	0.29	0.27	0.30	0.28	0.27	0.27
3,000	0.80	22.37	11.42	4.85	1.61	1.17	0.95	0.99	0.88	0.81	0.89	0.84	0.81	0.80
5,000	1.34	37.29	19.04	8.09	2.69	1.96	1.59	1.65	1.46	1.35	1.49	1.40	1.35	1.34
10,000	2.67	74.57	38.08	16.18	5.37	3.91	3.18	3.29	2.93	2.71	2.98	2.80	2.69	2.67

Vertical grade:

Table 70 and Table 71 address the sensitivity of the predicted crash numbers to the changes in percent vertical grade while considering Georgia variables and base conditions respectively. The variations are exactly equal to the CMF values for vertical grade. This is because, the CMF value is constant for a given specific grade. Similar numbers are observed while considering the sensitivity of predicted crashes to the percent vertical grade when all other variables are base conditions.

Table 70: Sensitivity of predicted crashes to the percent vertical grade when all other variables are kept constant

AADT	Predicted crashes	Percent vertical grade				
		0	2	4	6	8
400	0.17	0.17	0.17	0.19	0.19	0.20
1000	0.43	0.43	0.43	0.47	0.47	0.49
3000	1.28	1.28	1.28	1.40	1.40	1.48
5000	2.13	2.13	2.13	2.34	2.34	2.47
10000	4.25	4.25	4.25	4.68	4.68	4.93

Table 71: Sensitivity of predicted crashes to the percent vertical grade when all other variables are base conditions

AADT	Predicted crashes	Percent vertical grade				
		0	2	4	6	8
400	0.11	0.11	0.11	0.12	0.12	0.12
1000	0.27	0.27	0.27	0.29	0.29	0.31
3000	0.80	0.80	0.80	0.88	0.88	0.93
5000	1.34	1.34	1.34	1.47	1.47	1.55
10000	2.67	2.67	2.67	2.94	2.94	3.10

Driveway density:

Table 72 and Table 73 present the sensitivity of predicted crash numbers to driveway density when all variables and base conditions are respectively considered. By observing the variations in CMF value for driveway density, it is found that a segment with 40 driveways/ mile experience 2.5 times more crashes than a similar segment with ≤ 5 driveways/ mile.

Table 72: Sensitivity of predicted crashes to driveway density when all other variables are kept constant

AADT	Predicted crashes	Driveway density in driveways/mile							
		5	10	15	20	25	30	35	40
400	0.15	0.15	0.18	0.22	0.25	0.28	0.32	0.35	0.39
1000	0.36	0.36	0.44	0.51	0.58	0.65	0.72	0.79	0.86
3000	1.09	1.09	1.24	1.39	1.53	1.68	1.83	1.97	2.12
5000	1.82	1.82	2.01	2.20	2.39	2.58	2.77	2.95	3.14
10000	3.65	3.65	3.86	4.07	4.28	4.49	4.70	4.91	5.12

Table 73: Sensitivity of predicted crashes to driveway density when all other variables are base conditions

AADT	Predicted crashes	Driveway density in driveways/mile							
		5	10	15	20	25	30	35	40
400	0.11	0.11	0.13	0.16	0.18	0.21	0.23	0.26	0.28
1000	0.27	0.27	0.32	0.37	0.42	0.47	0.53	0.58	0.63
3000	0.80	0.80	0.91	1.02	1.12	1.23	1.34	1.45	1.55
5000	1.34	1.34	1.47	1.61	1.75	1.89	2.03	2.16	2.30
10000	2.67	2.67	2.83	2.98	3.13	3.29	3.44	3.60	3.75

Presence of a TWLTL:

Table 74 and Table 75 present the sensitivity of predicted crash numbers to the presence of a TWLTL with varying driveway density (driveways/mile) when all variables and base conditions are respectively considered. It is observed that fewer crashes are predicted when a TWLTL exists for any specific number of driveways.

Table 74: Sensitivity of predicted crashes to the presence of TWLTL when all other variables are kept constant

AADT	Predicted crashes with no driveways	Driveway density in driveways/mile							
		5	10	15	20	25	30	35	40
400	0.15	0.14	0.17	0.19	0.21	0.23	0.25	0.26	0.28
1000	0.36	0.36	0.41	0.45	0.48	0.52	0.55	0.59	0.63
3000	1.09	1.07	1.16	1.22	1.28	1.34	1.41	1.48	1.55
5000	1.82	1.78	1.88	1.94	2.00	2.06	2.13	2.21	2.29
10000	3.65	3.56	3.60	3.59	3.58	3.59	3.62	3.67	3.74

Table 75: Sensitivity of predicted crashes to the presence of TWLTL when all other variables are base conditions

AADT	Predicted crashes with no driveways	Driveway density in driveways/mile							
		5	10	15	20	25	30	35	40
400	0.11	0.10	0.12	0.14	0.15	0.17	0.18	0.19	0.21
1000	0.27	0.26	0.30	0.33	0.35	0.38	0.40	0.43	0.46
3000	0.80	0.78	0.85	0.90	0.94	0.98	1.03	1.08	1.13
5000	1.34	1.31	1.37	1.42	1.46	1.51	1.56	1.62	1.68
10000	2.67	2.61	2.64	2.63	2.62	2.63	2.65	2.69	2.74

Presence of a passing lane or a short four-lane section:

Independent of AADT, the presence of a passing lane and a short four-lane section reduce the crashes by 25% and 35% respectively. Table 76 and Table 77 show similar trend.

Table 76: Sensitivity of predicted crashes to the presence of either a passing lane or a short four-lane section when all other variables are kept constant

AADT	Predicted crashes	Presence of passing lane		Presence of short four-lane section	
		No	Yes	No	Yes
400	0.18	0.18	0.13	0.18	0.12
1000	0.45	0.45	0.33	0.45	0.29
3000	1.34	1.34	1.00	1.34	0.87
5000	2.23	2.23	1.67	2.23	1.45
10000	4.46	4.46	3.34	4.46	2.90

Table 77: Sensitivity of predicted crashes to the presence of either a passing lane or a short four-lane section when all other variables are base conditions

AADT	Predicted crashes	Presence of passing lane		Presence of short four-lane section	
		No	Yes	No	Yes
400	0.11	0.11	0.08	0.11	0.07
1000	0.27	0.27	0.20	0.27	0.17
3000	0.80	0.80	0.60	0.80	0.52
5000	1.34	1.34	1.00	1.34	0.87
10000	2.67	2.67	2.00	2.67	1.74

Roadside hazard rating:

A segment is rated on a scale of 1 to 7, based on its roadside features. With an increase in AADT, the predicted crash numbers increase slightly. While considering the sensitivity of predicted crashes to roadside hazard rating when all other variables are kept constant, for an AADT of 400 veh/day, the number of predicted crashes increases by 28% and for an AADT of 10,000 veh/day, the predicted crash numbers increase by 30% for the worst RHR of 7. Similar percentage differences are observed when the base conditions are considered.

Table 78: Sensitivity of predicted crashes to roadside hazard rating when all other variables are kept constant

AADT	Predicted crashes	Roadside Hazard Rating						
		1	2	3	4	5	6	7
400	0.18	0.15	0.16	0.18	0.19	0.20	0.21	0.23
1000	0.44	0.38	0.41	0.44	0.47	0.50	0.54	0.57
3000	1.32	1.15	1.23	1.32	1.41	1.51	1.61	1.72
5000	2.20	1.92	2.06	2.20	2.35	2.51	2.69	2.87
10000	4.40	3.85	4.11	4.40	4.70	5.03	5.37	5.74

Table 79: Sensitivity of predicted crashes to roadside hazard rating when all other variables are base conditions

AADT	Predicted crashes	Roadside Hazard Rating						
		1	2	3	4	5	6	7
400	0.11	0.09	0.10	0.11	0.11	0.12	0.13	0.14
1000	0.27	0.23	0.25	0.27	0.29	0.31	0.33	0.35
3000	0.80	0.70	0.75	0.80	0.86	0.92	0.98	1.05
5000	1.34	1.17	1.25	1.34	1.43	1.53	1.63	1.75
10000	2.67	2.34	2.50	2.67	2.86	3.05	3.26	3.49

4.5.2.3 Effect of each individual CMF and a combination of CMFs on the calibration factor:

The final analysis considers the effect of individual CMFs on the calibration factor. As shown in Table 80, assuming all roadway characteristics are equal to base conditions, the total predicted crashes for a three-year period is 284. This is compared to 302 observed crashes in the same time period. Thus, the resulting calibration factor (observed crashes/ predicted crashes) is 1.064

When all CMFs for each sub segment are used in calculations, predicted crashes are 382 and resulting calibration factor is 0.791. In contrast, if only the required CMFs are used for calibration, the calibration factor is 0.937. When the effect of individual CMFs are compared, it is observed that the CMF for driveway density resulted in a total of 323 predicted crashes. It is found that the CMF for driveway density is substantially influencing the calibration factor. This observation is strengthened when the calibration factor considering all the CMFs excluding driveway density is found to be 0.912 which is very close to the calibration factor calculated using only the required CMFs (0.937).

Table 80 : Effect of each individual CMF and a combination of CMFs on the calibration factor

Condition	total predicted crashes for 3 year period	total observed crashes for 3 year period	calibration factor
Base condition	284	302	1.064
Multiple CMFs			
All CMFs	382	302	0.791
Required CMFs	322	302	0.937
All CMFs excluding CMF for driveway density	331	302	0.912
Single CMFs			
Lane width	286	302	1.057
Shoulder width and type	306	302	0.987
Horizontal curve	301	302	1.005
Vertical grade	298	302	1.012
Driveway density	323	302	0.936
Passing lane	276	302	1.093
TWLTL	280	302	1.077
RHR	284	302	1.062

Looking at the variations in calibration factor considering individual CMFs and a combination of CMFs, it is recommended to use all CMFs excluding the CMF for driveway density. As discussed earlier, CMF for driveway density is skewing the predictions in the opposite direction, resulting in a cancelling effect.

Phase 6:

4.6 Assess whether comparable results are obtained if using SA & HSM in combination for safety analysis:

SafetyAnalyst and the Highway Safety Manual are the two advanced safety analysis tools that use Empirical Bayes approach. SafetyAnalyst, companion software to the HSM, is recommended for state-wide analysis and for performing various steps in the roadway safety management process. The HSM is more geared toward site-specific improvements, even though, project based EB methodology is also illustrated in the manual. Ideally, both the tools together constitute a comprehensive safety analysis. Both the safety analysis tools are expected to give similar results when used in combination for safety analysis. Given this expectation, the main objective of this phase is to compare the results obtained from the two tools.

4.6.1 Compare different SPFs for two-lane rural roads based on R^2_{FT} and overdispersion parameter:

When two-way two-lane rural roads are considered, this research has looked at 5 separate SPFs. Following are the different SPFs considered:

a. Georgia specific SPF with all AADT (estimated and actual values)
b. Default SPF used within SafetyAnalyst
c. Default SPF used within SafetyAnalyst calibrated to Georgia data (with a calibration factor of 0.37)
d. Default SPF published in the HSM
e. Default SPF published in the HSM calibrated to Georgia data (with a calibration

factor of 0.93)

All the five SPFs were plotted against the observed crashes. Figure 20 shows the graph and Table 81 gives the R^2_{FT} and overdispersion values for each SPF. As hypothesized, it is observed that the SPFs used in the Highway Safety Manual, when calibrated to Georgia data considering all the required CMFs gives a better fit to Georgia data. However, there is not a lot of difference between the calibrated and base SPFs of HSM. The calibrated SPFs used in SafetyAnalyst also provide an acceptable fit with an R^2_{FT} value of 0.58. Excluding the non calibrated SPF used within SafetyAnalyst, all the other SPFs have an R square value within a range of 0.6 +/- 0.02.

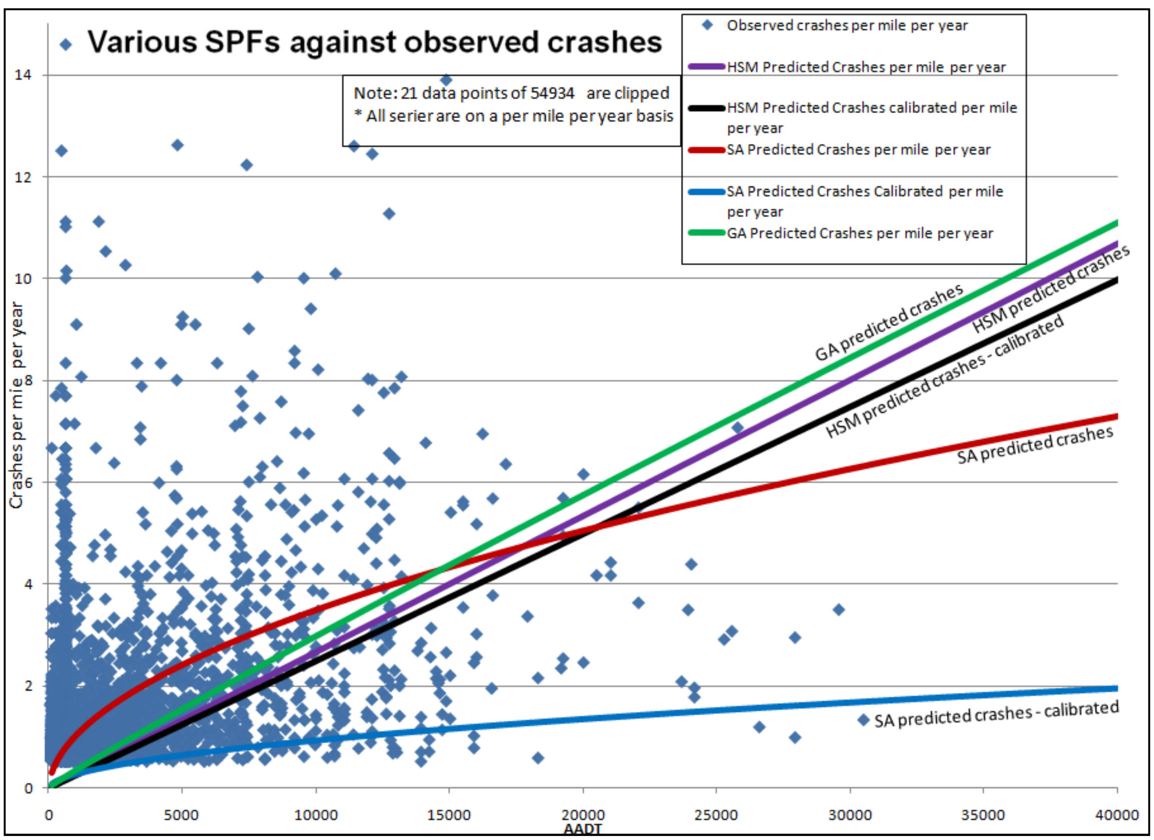

Figure 20: Various SPFs plotted against observed crashes

Table 81: R square values of various SPFs for two-lane rural roads

SPF	R^2_{FT}	ODP
SPF generated using Georgia data	0.604	1.377
Base SPF used in SafetyAnalyst	-0.220	0.5
Base SPF used in SafetyAnalyst calibrated to Georgia data with a calibration factor of 0.268	0.581	0.5
Base SPF used in the Highway Safety Manual	0.612	$\frac{0.236}{segment\ length\ in\ miles}$
Base SPF used in the Highway Safety Manual calibrated to Georgia data with a calibration factor of 0.934 (using required CMFs)	0.619	$\frac{0.236}{segment\ length\ in\ miles}$

The overdispersion parameter helps in assessing the reliability of SPFs. The non-calibrated and calibrated SPFs used within SafetyAnalyst, and the SPF generated using Georgia data use a constant overdispersion parameter, while, the HSM considers overdispersion factor as a function of segment length. Lower overdispersion parameter, used to weigh the predicted crashes, consequences more reliable models. ODP is used to weigh the predicted crashes in the EB analysis. Lower OD factor gives greater weight to the predictive model and lesser weight to the observed crashes.

4.6.2 Compare the list of top ranked sites based on two SPFs (default SPFs used in SafetyAnalyst calibrated to Georgia data, and default SPFs used in HSM calibrated to Georgia data) for two-way two-lane rural roads:

Considering one year of crash data and traffic data, EB analysis was performed to calculate the expected crashes based on the HSM procedure and also based on the SafetyAnalyst procedure using their respective default SPFs calibrated to Georgia data. Table 82 gives the descriptive statistics for the segment length of the top 50 sites based on HSM procedure and SafetyAnalyst procedure.

Table 82: Descriptive statistics for the segment length of the top 50 sites based on HSM procedure and SafetyAnalyst procedure

Procedure used	**Segment length (miles)**			
	Mean	Std. dev	Max	Min
HSM	13.23	5.57	30.49	4.60
SafetyAnalyst	3.71	3.18	10.37	0.08

Table 83 and Table 84 shows the list of top 25 ranked sites based on the default SPFs used in HSM calibrated to Georgia data with HSM calculations, and the default SPFs used in SafetyAnalyst calibrated to Georgia data with SafetyAnalyst calculations respectively. From the tables, it is observed that HSM procedure tends to identify longer segments while SafetyAnalyst procedure identifies shorter segments.

Table 83: Ranking of two-lane rural roadways in Georgia based on the procedures illustrated in the HSM and SafetyAnalyst sorted according to the rank based on HSM

Agency ID	**Segment Length (miles)**	**Obs. crashes in 2004**	**HSM calculations**		**SA calculations**	
			$N_{expected}$	HSM RANK	X_{2004}	SA RANK
B031100260021294263B	21.34	36	54.08	**1**	1.68	438
B169100490002861553B	12.67	38	48.49	**2**	2.92	132
B221100100000001856B	18.56	61	46.16	**3**	3.13	112
B255100160000001037B	10.37	69	43.71	**4**	6.05	**17**
B169100110009862415B	14.29	20	42.41	**5**	1.44	587
B027101330000922016B	19.24	25	41.59	**6**	1.31	697
B015100200017752407B	6.32	46	40.66	**7**	6.61	**12**
B133100120000002289B	22.89	42	40.17	**8**	1.79	384
B299100380000001480B	14.8	27	40.09	**9**	1.81	377
B311100750000001782B	17.82	47	39.10	**10**	2.52	169
B045100160018732787B	9.14	40	38.78	**11**	4.11	61
B073102320000000666B	6.66	32	38.39	**12**	4.56	44
B115100530000001487B	14.87	18	37.81	**13**	1.25	777
B069100310010302912B	18.82	21	37.64	**14**	1.14	919
B033100560000002187B	21.87	18	37.05	**15**	0.85	1479
B311100110006531501B	8.48	39	36.15	**16**	4.28	55
B285100010000000852B	8.52	27	35.53	**17**	3.07	121
B117103690000001187B	11.87	54	35.12	**18**	4.18	57
B311101150000001556B	15.56	41	33.33	**19**	2.50	171
B267100230006303679B	30.49	25	33.10	**20**	0.83	1565
B217100120009191646B	7.27	26	32.53	**21**	3.42	91
B193100260000002463B	24.63	28	32.29	**22**	1.13	933
B151100420000000929B	9.29	53	31.54	**23**	5.09	29
B073101040008171494B	6.77	28	31.50	**24**	3.89	66
B077100140018162773B	9.57	48	31.44	**25**	4.53	46

Table 84: Ranking of two-lane rural roadways in Georgia based on the procedures illustrated in the HSM and SafetyAnalyst sorted according to the rank based on SafetyAnalyst

Agency ID	Segment Length (miles)	Obs. crashes in 2004	HSM calculations		SA calculations	
			$N_{expected}$	HSM RANK	X_{2004}	SA RANK
B2551000300003600058B	0.22	14	7.98	446	13.18	**1**
B015206330004100643B	2.33	38	22.42	75	11.51	**2**
B151101550005720597B	0.25	17	5.68	651	10.18	**3**
B077100140010721100B	0.28	14	5.39	683	9.16	**4**
B255101550007231305B	5.82	68	22.76	73	9.02	**5**
B077101540008070858B	0.51	13	6.48	561	8.53	**6**
B139100110019932209B	2.16	27	13.06	227	8.04	**7**
B151101550005550563B	0.08	11	3.57	964	7.66	**8**
B085100530009351818B	8.83	75	30.93	28	7.25	**9**
B151101550000000532B	5.32	46	21.56	87	6.92	**10**
B117204580004500461B	0.11	10	3.22	1054	6.85	**11**
B015100200017752407B	6.32	46	40.66	**7**	6.61	**12**
B111100050004251245B	8.2	63	27.68	50	6.53	**13**
B187100600006350699B	0.64	13	4.73	779	6.49	**14**
B013100080005210707B	1.86	18	9.92	331	6.18	**15**
B139100110022092275B	0.66	11	4.98	743	6.17	**16**
B255100160000001037B	10.37	69	43.71	**4**	6.05	**17**
B187100600002660620B	3.54	28	12.94	235	5.76	**18**
B035100160001671011B	8.44	55	27.90	47	5.65	**19**
B02111104TA00000023B	0.23	7	3.04	1097	5.57	**20**
B111100050012451438B	1.93	18	7.41	481	5.44	**21**
B113100850000000549B	5.49	35	21.63	85	5.36	**22**
B137103850000560514B	4.58	29	20.00	98	5.28	**23**
B187100600006990778B	0.79	11	4.32	836	5.28	**24**
B057101400022232268B	0.45	8	3.51	978	5.21	**25**

4.6.3 Statistical test to determine if a significant difference in predictions exists between the HSM and SafetyAnalyst procedures:

A paired T test was performed on the resulting SafetyAnalyst/ HSM predicted crashes using SAS to determine if there is significant difference between the expected crashes calculated using HSM and SafetyAnalyst procedure. Table 85 shows the SAS output of the paired T test to determine if there is a significant difference in predictions between HSM and SafetyAnalyst procedures.

Hypothesis:

H_0: The difference in means between the two samples = 0 → $\mu_d = 0$

H_a: The difference in means between the two samples ≠ 0 → $\mu_d \neq 0$

Decision: Reject H_0 since the p- value of 0.0305 is less than alphas of 0.05

Conclusion: At a 5% level of significance, there is sufficient evidence to conclude that there is a significant difference in means between the two samples.

Table 85: SAS output of the paired T test to determine if a significant difference in predictions exist between HSM and SafetyAnalyst procedures.

```
                    Considering one year of data.
               Hypothesis testing when samples are paired
                  17:02 Wednesday, November 10, 2010

                         The TTEST Procedure
                         Difference:  HSM - SA
     N        Mean     Std Dev     Std Err     Minimum     Maximum
 64832      0.0121      1.4283     0.00561     -5.2000     52.4010

  Mean        95% CL Mean        Std Dev       95% CL Std Dev
0.0121     0.00114   0.0231      1.4283       1.4206   1.4361

                                     DF     t Value    Pr > |t|
                                  64831        2.16      0.0305
```

Given the results of the paired T test, a significant difference is noted between the expected crashes based on the two methods.

4.6.4 Document the major differences between SafetyAnalyst and HSM:

Even though SafetyAnalyst and HSM use SPFs and EB approach for identifying and prioritizing sites, it is found that the underlying calculations performed in both the tools are slightly different. Table 86 shows the sample calculations using both SafetyAnalyst procedure and HSM procedure for a 0.1 mile segment.

Table 86: Sample calculations based on the procedure described by SafetyAnalyst and HSM

SafetyAnalyst Calculations			
Agency id	B085100530009351818B		
Start mile point	15.65		
End mile point	15.75		
Segment length	0.10 miles		
Analysis year	**2004**	**2005**	**2006**
AADT	7610	10480	11380
# of crashes	13	17	29
Calibration factor (SA)	0.37	0.366	0.358
$N_{predicted}$ SA (crashes/ Mile/ year)	1.12	1.31	1.34
Correction factor (SA)	1.00	1.17	1.20
Wt (SA)			0.84
X_{2004} = Expected crashes in 2004			28.72
X_{2006} = expected crashes in 2006			**34.39**

Highway Safety Manual Calculations				
Agency id	B085100530009351818B			
Start mile point	15.65			
End mile point	15.75			
Segment length	0.10 miles			
Analysis year	**2004**	**2005**	**2006**	**totals**
AADT	7610	10480	11380	
# of crashes	13	17	29	59
$N_{predicted}$ HSM (crashes/ Mile/ year)	0.161	0.222	0.240	0.623
Over-dispersion factor	2.36	2.36	2.36	
Wt (HSM)				0.4049
N expected crashes				**35.36**

Following are the various equations used to perform EB analysis using SafetyAnalyst procedure and HSM procedure:

<u>Equations used for SafetyAnalyst Procedure:</u>

$$Calibration\ factor = \frac{Observed\ crashes\ in\ year\ i}{Predicted\ crashes\ in\ year\ i} \quad \text{Equation 36}$$

$$\text{corr?ction factor} = \frac{Npredicted\ in\ year\ i}{Npredicted\ in\ year\ 1}$$ Equation 37

$$Npredicted\ \ SA = e^{-3.63} * AADT^{0.53} * \ calibration\ factor$$ Equation 38

$$Wt\ SA = \frac{1}{1 + 0.5 * \sum(Npredicted\ \ * Length)}$$ Equation 39

$$X2004 = Wt\ SA\ \ * Npredicted2004 + \frac{(1 - Wt\ SA)}{Segment\ length} * \frac{total\ observed\ crashes}{\sum correction\ factors}$$ Equation 40

$$X2006 = Correction\ factor\ for\ 2006\ \ * X2006$$ Equation 41

Equations used for HSM Procedure:

$$Npredicted\ \ HSM = AADT\ * \ L\ * 365\ * \ 10^{-6}\ * \ e^{-0.312} * calibration\ factor$$ Equation 42

$$Overdispersion\ parameter = \frac{0.236}{Length}$$ Equation 43

$$Wt\ HSM = \frac{1}{1 + overdispersion\ parameter * \sum(Npredicted)}$$ Equation 44

$$Nexpected\ Wt\ HSM\ \ * Npredicted + (1 - Wt\ HSM) * \ total\ observed\ crashes$$ Equation 45

Having worked with SafetyAnalyst and the HSM, it is observed that there are a number of significant differences between the two tools which play a defining role in choosing one tool over the other. Figure 21 shows the functional form of SafetyAnalyst SPF and HSM SPF for two-way two-lane rural roads. For two-way two-lane rural roads, the HSM considers the relation between crashes and traffic to be almost linear, while it is not the case with the functional form of the SPF used within SafetyAnalyst.

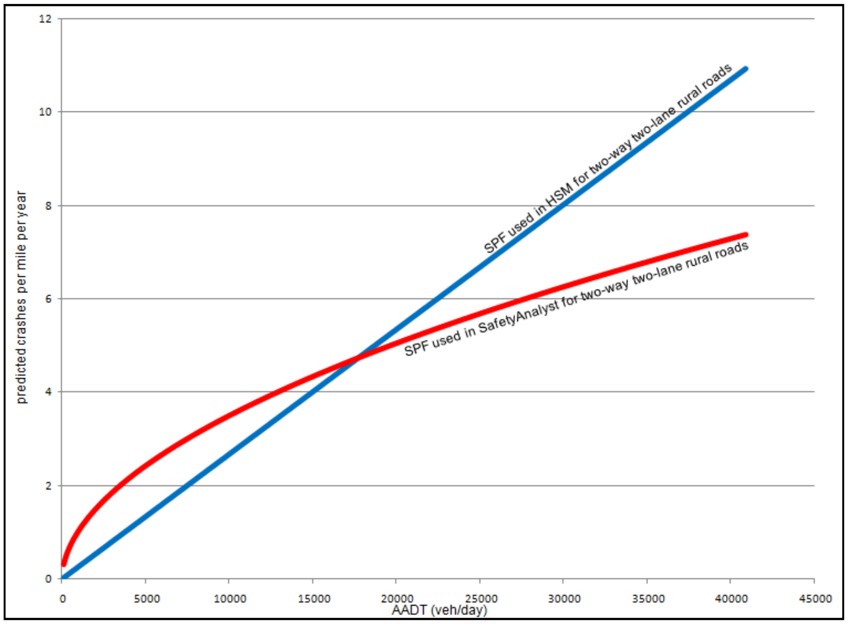

Figure 21 : Functional forms of SafetyAnalyst SPF and HSM SPF for two-way two-lane rural roads

Table 87 briefly summarizes the major differences between SafetyAnalyst and HSM as found in the prior sections.

Table 87: Major differences between SafetyAnalyst and HSM

SafetyAnalyst (SA)	Highway Safety Manual (HSM)
Site selection by EB analysis only	Site selection can be done using a variety of traditional or other EB methods
SA is designed more for system-wide analysis	HSM is designed more for site specific analysis
Cost: $11,000 for single user license	Cost: $390 per manual (for AASHTO members)
Data requirements are less intense compared to HSM requirements	Has intense data requirements for calculating the calibration factor and for each site analyzed
Import process may involve a lot of manual work yearly	Data acquisition could be tedious
Learning curve is steep	Learning curve is manageable
EB method is available for all site subtypes for segments, intersections and ramps	EB method is available for only 3 site subtypes: Rural two lane roads, urban multilane highways and suburban arterials
Base functional form of the SPF used for all types of segments is: $N = e^{\alpha} * AADT^{\beta}$	Base functional form of the SPF used for two-way two-lane rural roads is: $N = \text{coefficient} * AADT * e^{\gamma}$
All segments (irrespective of base conditions) were used to develop default SPFs	Segments with base conditions only were used to develop base SPFs
CMFs are used only for countermeasure selection and evaluation	CMFs are used to address to the variations in base conditions, and for countermeasure selection and evaluation
Weighting factor varies with segment length: Npredicted is given more weigh for shorter segments	Weighting factor is independent of segment length
The end result of EB method is expected crashes per mile per year in the last year of the analysis period	The end result of EB method is average expected crashes per mile per year
SA generates a log with errors and warnings during import, post process and calibration steps	A log file with errors and warnings is not available
SA cannot perform network screening when crashes are not assigned to specific segments	HSM can be used to perform project based EB analysis (when crashes are not assigned to specific segments)

There are significant differences in predicted and expected crashes between the two tools for two-way two-lane rural roads. A large portion is likely due to variations in the functional form and the definition of overdispersion parameter.

CHAPTER 5 **CONCLUSIONS AND RECOMMENDATIONS**

5.1 Conclusions:

From reviewing the literature and the past work that is carried out in the area of safety analysis, it is clear that the conventional methods of selecting “sites with potential for safety improvement” have their own drawbacks and limitations. However, most of the DOTs (that is, all the 24 states that have responded to the survey) use conventional methods like crash frequency, crash rate, or a safety index to identify and prioritize SWiP resulting in improper site selection and lesser safety effect for the money spent. Empirical Bayes approach, in addition to addressing all the limitations of traditional methods, also gives the predictive capability of safety along with the reliability measure of the safety predictions (variance). This research project aims at developing guidance for states transitioning to advanced safety analysis tools like SafetyAnalyst and Highway Safety Manual.

Following are the conclusions developed from the present research which was divided into six broad phases.

Phase 1:

5.1.1 Review Georgia data and identify analysis datasets:

Crash data, roadway characteristics data, and traffic data were obtained from GDOT for the years 2004-2006. GDOT maintains two files for roadway inventory data: Location referencing system (LRS file), and roadway characteristics file (RC file). Many issues were found with the roadway characteristics data. Even though some of the issues were fixed, many were beyond the scope of this research. Following are some of the issues identified with respect to the segments file:

- Incorrect segment lengths in LRS file
- Redundant segments with varying lengths in LRS file
- Presence of zero length segments in RC file
- Absence of spatial reference to some segments in RC file

Traffic data was found to be incomplete with quite a number of segments with just one or two years of traffic data. Missing AADT values were estimated based on the procedure explained in the HSM. Considering that the traffic growth factor significantly influences SPF generation and EB analysis, segments with unrealistic growth factors were excluded from the analysis.

<u>Phase 2:</u>

<u>5.1.2 Test traditional methods for biases found in the literature:</u>

As a result of ~75 data variables being collected in Georgia, the minimum length of a segment is 0.01 miles. The average segment length is 0.138 miles with a significant number of segments (> 65 %) shorter than 0.1 miles. Issues with shorter segments are not obviously known or observed in any type of analysis. However, it is noticed that they bias results and often question their reliability irrespective of the type of network screening method used. As discussed in Table 33 through Table 38, it can be concluded that crash frequencies identify longer segments and segments with higher AADT values while crash rates flag shorter segments and those with lower traffic volumes. Irrespective of the site selection method used, normalization of crashes occurring on shorter segments result in unrealistically extrapolated numbers. Additionally, the variance of expected crashes on shorter segments calculated by the EB method is usually extremely high questioning the reliability of predictions.

Coding errors and data sensitivity are considered to be the two main reasons for shorter segments. Large number of data variables coupled with greater sensitivity of the collected variables result in extremely small segments. Longer aggregated segments were generated by using fewer required data elements and also by reducing the sensitivity of data. The average length of aggregated segments was found to be 0.58 miles with about 26 % of segments shorter than 0.1 miles.

Phase 3:

5.1.3 Implement SafetyAnalyst on roadway segments:

The three important components of safety analysis - crash data, roadway characteristics data, and traffic data were imported into SafetyAnalyst and network screening was performed. Following are the various problems identified while working with the software.

- Coding mismatch: Almost all of the data elements in Georgia database were coded differently from SafetyAnalyst requirements. Therefore, a considerable amount of time was spent in recoding and matching the two coding structures.
- Not all segments and crashes were imported into SafetyAnalyst: Many segments were not imported into the software due to missing traffic and location information. Some segments were not assigned to any site subtype and hence, were not used in the analysis.

- Miscoded data: Various coding errors, possibly resulting from manual data entry were identified and flagged.
- Unrealistic traffic growth factors: Estimations and coding errors in traffic data resulted in unrealistic yearly variations and extremely high growth factors.
- Shorter segments: Shorter segments result in extremely high variance values questioning the predictions made by the EB methodology. This issue was attended to by merging two or more shorter unmodified segments into longer aggregated segments and therefore increasing the segment length.

Understanding the present stand of the states with regard to their safety practices is of immense interest to this project. 24 states have completed the survey, of which 13 states have some experience with either SafetyAnalyst and/or HSM. Following are the observations made.

- A majority of responding states (13 states) use three years of crash data and are capable of successfully locating more than 90% of the crashes spatially.
- All the responding states are successfully identifying crashes on segments. Ninety percent (23 of 25 states) can locate crashes on intersections. As expected, crashes on ramps are more difficult to be located precisely making them difficult for analysis.
- Roadway characteristics database is updated yearly by about 8 states while ten states update it continuously whenever there is a change. Yet, not many states record the date of changes which is important while performing before-after studies. About forty percent of the responding states (10) collect and maintain date information about a few required variables.

- The present research identifies the negative influence of shorter segments on the entire safety analysis and it is surprising to find that, similar to Georgia, 0.01 miles is the smallest segment length that is typically recorded if there is a change in the roadway characteristics in about 13 (of 20) states.
- SafetyAnalyst emphasizes on sub classification of segments, intersections and ramps into site subtypes. But, in practice, a majority of states broadly sub classify segments and intersections based on only two variables, resulting in a more generalized analysis datasets.
- All the responding states do maintain traffic data for as many years as the crash data is available.
- Actual traffic data is collected only on segments with higher functional classification. And the percent of available actual traffic data reduces with the decrease in functional classification of the roadways. A similar trend is followed with the amount of total available traffic data (actual + estimated traffic data). Though this indication is not out-of-normal, it might result in a serious issue of misrepresentation of a state's roadway network in various safety analyses as significant miles of roadway network is excluded from the preliminary analysis.
- About 50% of the responding states stated that they are currently using a combination of traditional and advanced methods. Most commonly used methods include crash frequency (20 states), crash rate (18 states), equivalent property damage only (8 states), high proportion of crash types (8 states), relative severity index (8 states), rate quality control (6 states), and EB methodology (2 states).
- Three fourths of the responding states (18 out of 24) are planning to use new highway safety analysis tools (IHSDM, SafetyAnalyst, and HSM).

- The default SPFs used within SafetyAnalyst might or might not truly represent the state's data. Therefore, assessment of the fit of default SPFs and the development of state specific SPFs are recommended.
- States responding to the survey have identified the following as major hurdles while working with SafetyAnalyst:
    - Data importing (8)
    - Initial set-up cost (5)
    - Data requirements and intersection data in particular (4)
    - Learning curve (4)
    - Interpreting the results and understanding the defaults (2)
    - IT compatibility issues (3)
    - Switch-over of analysis methodologies, and, processes and procedures (1)
    - Physical memory issues (1)
- Based on the states' experience, following are the tips that the states offer to other states, universities and research institutes planning to use the software:
    - Start with a subset of data on a local machine
    - Involve the IT (Information Technology) department early on in the process
    - Cross walk state data to SafetyAnalyst data
    - Know what you (end user) plan on using the software for
    - Understand that it (implementing SafetyAnalyst) takes considerable resources and time to start-up and spend time accordingly

- Train users on the capabilities, outputs ,and validate data to ensure buy-in at various levels
- Take advantage of the consultant's expertise
- Understand that expertise must be developed and maintained
- Factor in time required for implementation of the software
- Work with management throughout the process of SafetyAnalyst implementation

Based on the experience with Georgia data, following are the observations that encourage the deployment of SafetyAnalyst:

- SafetyAnalyst uses Empirical Bayes method and addresses the issues, biases, and limitations of traditional methods
- The software has the ability to perform sliding window analysis
- SafetyAnalyst divides roadway network into site subtypes and merges segments into longer homogeneous segments automatically
- SafetyAnalyst doesn't require extensive statistical expertise
- SafetyAnalyst performs basic data quality checks and logs a list of errors, warnings and potential issues with the data
- SafetyAnalyst performs all the steps in roadway safety management process and is mostly automated

Based on the experience with Georgia data, following are the observations that discourage the deployment of SafetyAnalyst:

- The software is very costly

- Initial generation of import files is tedious
- Data requirements are stringent
- SafetyAnalyst could be a “black box”
- SafetyAnalyst has the capability of flagging sites with unrealistic AADT growth factors and miscoded information as problematic sites.

In addition to SafetyAnalyst, AASHTO has also released the HSM. Its implementation and adoption is crucial for safety improvements across the states. Five states are looking at a time frame of 1-2 years for the complete deployment of the manual, and about 4 states are looking at several years. Several states are looking toward using the HSM as a supplement to their current practices.

Phase 4:

5.1.4 Develop state specific SPFs using SafetyAnalyst procedure:

SafetyAnalyst uses EB approach which requires SPFs. The national default SPFs were generated using northern and western states’ data for the years 1993-2002. The software calibrates the default SPFs to the agency data. But, most of the factors like traffic trends, accident patterns, climate, population, geography, etc change considerably among different regions. Hence, default SPFs (either calibrated or non-calibrated) may or may not very well represent the agency’s data.

Georgia specific SPFs were generated for the 17 site subtypes for both total, and fatal and injury (FI) crashes. The default calibrated and non-calibrated SPFs, and Georgia specific SPFs were compared based overdispersion parameter and R^2_{FT}. Georgia SPFs fit the data well (for most of the site subtypes) compared to the calibrated default SPFs used within SafetyAnalyst. Site subtypes 153 and 160 (urban multilane divided arterial streets and urban freeways within interchange area – 8+ lanes) were not well represented by either of the SPFs. This could be because AADT might not be the only factor influencing crashes on urban roadways and there could be many other distractions and contributing factors. It could also be that the functional form of the SPF used in SafetyAnalyst may not represent the true safety trend in Georgia. Further, the default SPFs used within SafetyAnalyst were developed using fewer miles of segments than the Georgia SPFs. The fit of Georgia specific SPFs to Georgia data was better than the fit of default SPFs to its original data for eight and eleven of the 17 site subtypes for total and FI crashes respectively.

The default SPFs used within SafetyAnalyst were not generated using the base conditions. Assuming that the use of base conditions might improve the fit, base conditions were identified for two-way two-lane rural roads in Georgia. They were found to be rural two way roads with ≤ 3 lanes (total), 24 ft of total lane width, 2 ft paved shoulders on both sides of the undivided road and with neither a passing nor a climbing lane. It was found that SPF for total crashes generated using segments with base conditions has a slightly lower overdispersion parameter.

Traffic is measured on less than 25% of roadway miles in Georgia while estimated traffic counts are used on the rest of the 75% of roadway network. In this context, the influence of actual and total traffic data on the fit of SPFs was analyzed. It was found that the number of segments with measured traffic counts was not distributed evenly across site subtypes. Most of the actual traffic data on rural segments (> 80%), except rural two lane roads was captured while a fewer percent of miles of roadway segments of a few urban site subtypes have actual traffic counts. Urban freeways with over 5 lanes have actual AADT counts on over 90% of their network. When the R^2_{FT} and overdispersion parameters were compared between SPFs generated using segments with actual traffic data and complete dataset, it was found that there was no significant difference in the R^2_{FT} values (except for urban one-way arterial streets). However, when the overdispersion parameters were considered, there was a significant reduction for a majority of the site subtypes (except for urban freeways with more than 5 lanes). Also, due to the increased reliability of traffic counts, a significant reduction in overdispersion factor was observed. A similar trend was observed when SPFs for fatal and injury crashes were compared.

An interesting observation is that the overdispersion parameter was improved significantly only for site subtypes whose total length of segments with actual traffic data was less. Therefore, it could be concluded that estimations in traffic data increases the variability and reduces the reliability of SPFs while performing EB analysis.

Phase 5:

5.1.5 Formulate and document a calibration procedure for two-lane rural roads to be used with the HSM:

Procedure described in the HSM was used to generate calibration factors for two-way two-lane rural roads. 52 segments with an average length of 1.93 miles and a total of 302 crashes over a 3 year analysis period were randomly selected for calculating the calibration factor. Eight CMFs (lane width, shoulder width and type, horizontal curve, vertical grade, presence of passing lane, presence of TWLTL, roadside hazard rating, and driveway density) were calculated.

5.1.5.1 The effect of individual CMFs on calibration factor:

Calibration factors while considering all CMFs and required CMFs (lane width, shoulder width and type, horizontal curve, and presence of TWLTL) were found to be 0.791 and 0.93 respectively. The effect of individual CMFs on the calibration factor was studied and observed that CMF for driveway was the most influential of all the CMFs. The calibration factor considering all the CMFs excluding CMF for driveway density was found to be 0.912 which is close to the calibration factor calculated by considering the required CMFs alone.

5.1.5.2 The sensitivity of predicted crashes to individual CMFs:

The sensitivity of predicted crash numbers to the individual eight CMFs was analyzed when: 1) all other variables are set equal to the Georgia conditions, and 2) all other variables are assumed to be at base conditions.

It is observed that the safety effect by considering variations within individual CMFs when all other variables were assumed to be at base conditions was similar to the corresponding safety effect when all other variables were set equal to the Georgia conditions. From this observation, it is safe to conclude that the acceptable sensitivity of predicted crashes to the CMFs is independent of the variations in other variables.

Phase 6:

5.1.6 Assess whether comparable results were obtained with SafetyAnalyst and HSM:

The HSM and SafetyAnalyst are the two advanced safety analysis tools released by AASHTO. SafetyAnalyst is recommended for state-wide analysis and considered as companion software to the HSM, which is recommended for site specific analysis. Comparable results between the two tools are essential to earn the trust of the safety officers and the researchers. In this context, EB analysis was performed on two-way two-lane rural roads using both the tools and the results were compared.

Paired t-test was performed on the complete dataset and it was found that, there was a significant difference in means between the expected crash numbers calculated using the HSM and the SafetyAnalyst procedure. As shown in Figure 20, the functional form (or shape) of the SPFs used within SafetyAnalyst is different from the functional form of the base SPF considered in the Highway Safety Manual. The significant differences in the expected number of crashes between the two methods could due to the differences in the functional form and the differences in the definition of the overdispersion parameter considered in the analysis. Also, the difference between the two calculated values was found to depend on the segment length; the shorter the segment length, the lesser the difference between the expected crash frequencies (calculated using the two procedures).

The HSM tends to identify longer segments while SafetyAnalyst tends to identify shorter segments. The average segment length of the top 50 sites identified by the HSM and SafetyAnalyst were 13.23 and 3.71 miles respectively.

Even though both the HSM and SafetyAnalyst are AASHTO tools aimed at using Empirical Bayes methods to assess and improve safety, many noteworthy differences were found to exist between the two tools. A few important differences are given below:

- The base functional form of the SPFs used by HSM is different from that used within SafetyAnalyst.
- The SPFs used in the HSM were generated using sites with base conditions and no base conditions were considered while generating default SPFs used within SafetyAnalyst.

In summary, from this research, it could be concluded that the states are ready to shift to the newer safety analysis tools provided they have sufficient comprehensive and accurate data. The knowledge gained through this research helps states in transferring to the newer and more advanced tools. From the 2009 five-percent reports and the nationwide survey, it is found that, for their safety analysis, most of the states are still using traditional methods like crash frequencies, crash rates, or safety indices coupled with shorter segments. Therefore, research on the generation of longer aggregated segments, and the documentation of the proven issues with rates and frequencies would be extremely helpful for the states that are willing to transfer to the newer tools. With about 13 (of the 24 responding states) states currently working with either SafetyAnalyst or the Highway Safety Manual, the documented experience with SafetyAnalyst using Georgia data could be highly beneficial to the states that are starting to work with the software. Assuming Georgia data to be similar to the other states' data, the documentation on SafetyAnalyst implementation would smoothen the learning curve for the states working (or planning to work) with the software. For a few site subtypes, the national default SPFs calibrated to GA data used within SafetyAnalyst did not represent the GA data well enough and therefore Georgia specific SPFs are recommended for those site subtypes. With this experience, generation of state specific SPFs are recommended. However, the overdispersion parameter helps in deciding which SPF to use for the EB analysis (the lower the ODP, the greater the reliability of the model). Further, the critical comparison of the two newer safety tools (SafetyAnalyst and the HSM) would guide the states in deciding which tools to use for their safety analyses.

However, this research is not completely transferable to the states. The Georgia specific SPFs are not transferable and the states need to generate their own SPFs and assess their fit. With regard to the SafetyAnalyst implementation, most of the issues and constraints are documented. By no means, this list of issues is comprehensive. Each state needs to implement the software to identify and address any possible issues specific to the state. Also, the procedure used to randomly select sites (two-way two-lane rural roads) for the HSM calibration is unique to Georgia as sufficient number of plan profile sheets are unavailable. The randomization procedure is specific to the states and is dependent on data availability.

5.2 Future Recommendations:

This research project has a lot of scope for future work. The following follow-up work is recommended to expand the research presented in this dissertation:

- The present work analyzed roadway segments. Similar research can be performed on intersections and ramps. The data requirements, availability and accuracy for intersection and ramp data can be identified from advanced tools point of view.
- Only network screening module of SafetyAnalyst was studied in this research. The software is capable of performing various steps in road safety improvement process (which includes diagnosis and countermeasure selection, economic

appraisal and priority ranking, and countermeasure evaluation). Research on these modules would be highly beneficial to researchers and practitioners.

- In this research, SPFs were generated manually considering the functional form of default SPFs used in SafetyAnalyst as a basis. This might not be the best way to develop SPFs as the relation between traffic and crashes is confined in this approach. The future research can include a study on the relation between AADT and traffic without confining to a specific functional form.
- The positive influence of actual traffic volumes on SPFs was reinforced by the lower overdispersion parameters. However, estimation of traffic volumes on some roadways (low volume roads, city roads, and rural roads) is inevitable. Therefore, research on various estimation methods and their effect on R square values and the overdispersion parameter is valuable.
- Although CMFs for various data variables are available, their standard errors are unavailable at this point. As most of the CMFs are around 1.00, their standard errors gain a lot of importance as the effect of CMFs could be reversible when their standard errors are considered. Therefore, development of standard errors is important and is considered as a significant contribution to the highway safety research.
- EB method based on HSM procedure was applied to only two-way two-lane rural segments. Similar research can be applied to rural multilane highways and suburban arterials and the results between SafetyAnalyst and HSM can be compared.

REFERENCES

Alluri, P. (2008). *Assessment of potential site selection methods for use in prioritizing safety improvements on Georgia roadways* Retrieved from http://etd.lib.clemson.edu/documents/1219852297/

American Association of State Highway and Transportation Officials. (2010a). *AASHTOWare 2010 catalog.* Retrieved July/ 28, 2010, from http://www.aashtoware.org/Documents/FY2011_Catalog_Final-6-4-10.pdf

American Association of State Highway and Transportation Officials. (2010b). *Highway safety manual* (1st ed.). Washington DC: American Association of State Highway and Transportation Officials.

American Association of State Highway and Transportation Officials. (2010c). *SafetyAnalyst.* Retrieved July/ 28, 2010, from http://www.safetyanalyst.org/

CH2MHill. (2009). *Five percent report illinois department of transportation.* Retrieved July 28, 2010, from http://safety.fhwa.dot.gov/hsip/fivepercent/2009/09il.pdf

Chen, S. R. (2009). *Interactive highway safety design model (IHSDM)*

Clifford, C. (2008). *U.S. car accident cost: $164.2 billion.* Retrieved 05/30, 2010, from http://money.cnn.com/2008/03/05/news/economy/AAA_study/

Council, F. M., & Harkey, D. L. (2006). *Traffic safety information systems international scan: Strategy implementation white paper.* No. FHWA-HRT-06-099). McLean, VA: Federal Highway Administration.

Council, F. M., Harkey, D. L., Carter, D. L., & White, B. (2007). *Model minimum inventory of roadway Elements—MMIRE.* No. FHWA-HRT-07-046). McLean, VA: Federal Highway Administration.

Dean, C., & Lawless, J. F. (1989). Tests for detecting overdispersion in poisson regression models. *Journal of the American Statistical Association, 84*(406), 467-472.

Delucia, B., & Scopatz, R. (2005). Crash records systems: A synthesis of practice. national cooperative highway research program synthesis 350. *Transportation Research Board, Washington DC,*

ESRI.*What is GIS?* Retrieved May, 31, 2010, from http://www.gis.com/content/what-gis

FDOT. (2006). *Strategic highway safety plan.* Retrieved October 13, 2010, from http://www.dot.state.fl.us/safety/posted documents/Florida Comp SHSP for print.pdf

Federal Highway Administration.*Model inventory of roadway elements.* Retrieved June 01, 2010, from http://www.mireinfo.org/

Federal Highway Administration. (2005). *A summary of highway provisions in SAFETEA-LU.* Retrieved May, 31, 2010, from http://www.fhwa.dot.gov/safetealu/summary.htm

Federal Highway Administration. (2010a). *Highway safety improvement program "5 percent report" requirement.* Retrieved 05/15, 2010, from http://safety.fhwa.dot.gov/hsip/fivepercent/

Federal Highway Administration. (2010b). *Interactive highway safety design model (IHSDM): Making safety a priority in roadway design.* Retrieved 08 28, 2010, from http://www.tfhrc.gov/safety/ihsdm/ihsdm.htm

Fridstrom, L., Ifver, J., Ingebrigtsen, S., Kulmala, R., & Thomsen, L. K. (1995). Measuring the contribution of randomness, exposure, weather, and daylight to the various in road accident counts. *Accident Analysis and Prevention, 27*(1), 1.

Harwood, D. W., Council, F. M., Hauer, E., Hughes, W. E., & Vogt, A. (2000). *Prediction of the expected safety performance of rural two-lane highways.* No. FHWA-RD-99-207). McLean, VA: Federal Highway Administration.

Harwood, D. W., Torbic, D. J., Richard, K. R., & Meyer, M. M. (July, 2010). *SafetyAnalyst: Software tools for safety management of specific highway sites.* No. FHWA-HRT-10-063).Federal Highway Administration.

Hauer, E. (1995). On exposure and accident rate. *Traffic Engineering Control, 36*(3), 134-138.

Hauer, E. (1996). Identification of sites with promise. *Transportation Research Record: Journal of the Transportation Research Board, 1542*(-1), 54-60.

Hauer, E. (1997). *Observational before-after studies in road safety* Pergamon Press Inc.

Hauer, E. (2001). Overdispersion in modelling accidents on road sections and in empirical bayes estimation. *Accident Analysis & Prevention, 33*(6), 799-808.

Hauer, E., Allery, B. K., Kononov, J., & Griffith, M. S. (2004). How best to rank sites with promise. *Transportation Research Record: Journal of the Transportation Research Board, 1897*(-1), 48-54.

Hauer, E., Harwood, D. W., Council, F. M., & Griffith, M. S. (2002). Estimating safety by the empirical bayes method: A tutorial. *Transportation Research Record: Journal of the Transportation Research Board, 1784*(-1), 126-131.

Hauer, E., Kononov, J., Allery, B., & Griffith, M. S. (2002). Screening the road network for sites with promise. *Transportation Research Record: Journal of the Transportation Research Board, 1784*(1), 27-32.

INDOT. (2006). *Strategic highway safety plan.* Retrieved October 13, 2010, from http://www.in.gov/indot/files/shsp.pdf

Iowa DOT. (2007). *Strategic highway safety plan.* Retrieved October 13, 2010, from http://www.iowadot.gov/traffic/chsp/index.htm

iTRANS Consulting Ltd, & Human Factors North INC. (2003). *New approaches to highway safety analysis* US Department of Transportation Federal Highway Administration.

Kononov, J., & Allery, B. (2003). Level of service of safety A conceptual blueprint and the analytical framework. *Transportation Research Record: Journal of the Transportation Research Board, 1840*, 57-66.

Kononov, J., Bailey, B., & Allery, B. K. (2008). Exploratory examination of the functional form of safety performance functions of urban freeways. *Transportation Research Board,*

Kraft, W. H. (Ed.). (2009). *Traffic engineering handbook* (6th ed.) Institute of Transportation Engineers.

Ludwig, D. A. (2007). *Safety management systems for airports: Overview* Transportation Research Board National Research.

MDOT. (2008). *Strategic highway safety plan.* Retrieved October 13, 2010, from http://www.michigan.gov/documents/MI_CHSP_110103_7.pdf

National Highway Traffic Safety Administration. (2003). *Traffic records: A highway safety program advisory.* Retrieved June 11, 2010, from http://www.nhtsa.gov/people/perform/pdfs/NHTSA_TR_Advisories.pdf

National Highway Traffic Safety Administration. (2008). *Model minimum uniform crash criteria.* Retrieved July 17, 2010, from http://www.mmucc.us/2008MMUCCGuideline.pdf

National Highway Traffic Safety Administration. (2010). *Fatality analysis reporting system, general estimates system-2008 data summary.* No. DOT HS 811 171).

National Highway Traffic Safety Administration. (NA). *National statistics.* Retrieved May 30, 2010, from http://www-fars.nhtsa.dot.gov/Main/index.aspx

O'Day, J. (1993). NCHRP synthesis of highway practice 192: Accident data quality. *Transportation Research Board, National Research Council, Washington, DC,*

Ogle, J. H. (2007). *Technologies for improving safety data: A synthesis of highway practice.* No. 367). Washington DC: Transportation Research Board. Retrieved from http://onlinepubs.trb.org/onlinepubs/nchrp/nchrp_syn_367.pdf

Ogle, J. H., & Alluri, P. (in review). Comparison of traditional and advanced methods in identifying roadway sections for safety improvements. *Journal of Transportation Engineering,*

Persaud, B. (2001). NCHRP synthesis of highway practice 295: Statistical methods in highway safety analysis. *Transportation Research Board, National Research Council, Washington, DC,*

Pfefer, R. C., Neuman, T. R., & Raub, R. A. (1999). *Improved safety information to support highway design* Transportation Research Board.

Poch, M., & Mannering, F. (1996). Negative binomial analysis of intersection-accident frequencies. *Journal of Transportation Engineering, 122*(2), 105-113.

Qin, X., Ivan, J. N., & Ravishanker, N. (2004). Selecting exposure measures in crash rate prediction for two-lane highway segments. *Accident Analysis & Prevention, 36*(2), 183-191.

Qin, X., Ivan, J. N., Ravishanker, N., & Liu, J. (2005). Hierarchical bayesian estimation of safety performance functions for two lane highways using markov chain monte carlo modeling. *Journal of Transportation Engineering, 131*(5), 345.

Shankar, V., Mannering, F., & Barfield, W. (1995). Effect of roadway geometrics and environmental factors on rural freeway accident frequencies. *Accident Analysis & Prevention, 27*(3), 371-389.

Vogt, A., & Bared, J. G. (1998). *Accident models for two lane rural roads: Segments and intersections.* No. FHWA-RD-98-133). McLean, VA: Federal Highway Administration.

Wang, J., Hughes, W. E., & Stewart, R. (1995). Safety effects of cross-section design on rural multi-lane highways.

Washington, S., Persaud, B., Lyon, C., & Oh, J. (2005). *Validation of accident models for intersections.* No. FHWA-RD-03-037). McLean, VA: Federal Highway Administration.

Weingroff, R. F.*Federal aid highway act of 1956: Creating the interstate system .* Retrieved 05/26, 2010, from http://www.fhwa.dot.gov/infrastructure/rw96e.cfm

WSDOT. (2010). *Target zero strategic highway safety plan 2010 review draft.* Retrieved July 17, 2010, from http://wsdot.wa.gov/planning/SHSP.htm

Zegeer, C. V., Reinfurt, D. W., Hummer, J., Herf, L., & Hunter, W. (1986). *Safety effects of cross-section design for two-lane roads.* No. FHWA-RD-87-008). Washington DC: Federal Highway Administration.

APPENDICES

APPENDIX A: IMPORT FILES FOR SAFETYANALYST: SQL QUERIES

SafetyAnalyst requires three files to be imported for analysis of roadway segments: Roadway Characteristics, Accident and Traffic.

AltAccident:

Step1: Open SQL server and create a new database "GDOT04-06 Crash data"

Step2: Import Accident_tbl and Location_tbl into GDOT04-06 Crash data database.

Step3: select * into Georgia_Crashes from accident_tbl left outer join location_tbl on acc_id=Loc_acc_id

This will add all the records from accident_tbl and only those from location_tbl which have loc_acc_id similar to acc_id

Step4: Rename acc_id to agencyID

This will rename acc_id as agencyID

Step5: Alter table Georgia_crashes

add RTE_NAME nvarchar(12)

This will add a column RTE_NAME with a datatype nvarchar

Step6: UPDATE Georgia_crashes

SET RTE_NAME =

LOC_ROUTE_TYPE+LOC_ROUTE_IDENTIFIER+LOC_ROUTE_SUFFIX

This will set RTE_NAME as

LOC_ROUTE_TYPE+LOC_ROUTE_IDENTIFIER+LOC_ROUTE_SUFFIX

Step7: Alter table Georgia_crashes

add LocSystem nvarchar(2)

This will add a column LocSystem with a datatype nvarchar

Step8: UPDATE Georgia_crashes

SET LocSystem = 'B'

This will set LocSystem as 'B'

Step9: Alter table Georgia_crashes

add routeType nvarchar(12)

This will add a column routeType with a datatype nvarchar

Step10: UPDATE Georgia_crashes

```
SET routeType = CASE
        when LOC_FUNCTIONALCLASS_TYPE = 1 then '99'
        when LOC_FUNCTIONALCLASS_TYPE = 11 then '99'
```

else LOC_ROUTE_TYPE
end
This will set SA_RTE_TYPE as ‘99’ for interstates else LOC_ROUTE_TYPE

Step11: Rename Loc_ACC_Milelog as LocOffset
This will rename Loc_ACC_Milelog as LocOffset
Step12: Alter table Georgia_crashes
add SA_ACC_DATE nvarchar(12), SA_ACC_TIME nvarchar(255)
This will add a column SA_ACC_DATE, SA_ACC_TIME with a datatype nvarchar

Step13: You need to have a date time function to separate date and time

```
ALTER FUNCTION [dbo].[udf_GetTimeOnly]
(
        -- Add the parameters for the function here
        @InputDate datetime
)
RETURNS varchar(5)
AS
BEGIN
        Return ( CONVERT(varchar(5),@InputDate,108) )
END
```

This will parse out time from the datetime field

Step14: You need to have a date time function to separate date and time

```
ALTER FUNCTION [dbo].[udf_GetDateOnly]
(
        -- Add the parameters for the function here
        @InputDate datetime
)
RETURNS varchar(5)
AS
BEGIN
        Return ( CONVERT(varchar(5),@InputDate,101) )
END
```

This will parse out date from the datetime field

Step15: update Georgia_crashes

```
set sa_acc_time = dbo.udf_gettimeonly(ACC_ATIME)
update Georgia_crashes
set sa_acc_date = dbo.udf_getdateonly(ACC_ATIME)
```

This will save date and time in the required fields

The following steps are used to obtain accident severity

Step16 (a): Alter table Accident_tbl
add drvr_injc nvarchar(2), pssgr_injc nvarchar(2), ped_injc nvarchar(2),
This will add drvr_inj, pssgr_inj and ped_inj columns with datatype nvarchar

Step16 (b): Alter table Occdrvr_tbl
add occ_inj_pdo5 nvarchar(2)
This will add occ_inj_pdo5 column with datatype nvarchar

Step16 (c): Update Occdrvr_tbl
```
Set occ_inj_pdo5  = CASE
                when Occ_injc_type = 0 then '5'
                else Occ_injc_type
                end
```
This will set occ_inj_pdo5 as '5' for pdos and occ_injc_type for other severities

Step16 (d): Alter table Passngr_tbl
add pssgr_inj_pdo5 nvarchar(2)
This will add pssgr_inj_pdo5 column with datatype nvarchar

Step16 (e): Update Passngr_tbl
```
Set pssgr_inj_pdo5  = CASE
                when pssgr_injc_type = 0 then '5'
                else pssgr _injc_type
                end
```
This will set pssgr_inj_pdo5 as '5' for pdos and pssgr_injc_type for other severities

Step16 (f): Alter table Ped_tbl
add ped_inj_pdo5 nvarchar(2)
This will add ped_inj_pdo5 column with datatype nvarchar

Step16 (g): Update Ped_tbl
```
Set ped_inj_pdo5  = CASE
                when ped_injc_type = 0 then '5'
                else ped_injc_type
                end
```
This will set ped_inj_pdo5 as '5' for pdos and ped_injc_type for other severities

Step16 (h): Update Accident_tbl
```
Set     accident_tbl.drvr_injc = 5
accident_tbl.pssgr_injc = 5
```

accident_tbl.ped_injc = 5
This will set drvr_injc, pssgr_injc and ped_injc as ‘5’ in Accident_tbl

To obtain max severity, a function named “Least” needs to be created. The following steps are required

Step16 (i):
```
create function least
(
@a int,
@b int,
@c int
)
returns int
Begin
	if(@a>@b)
		Begin
			set @a=@b
		End
	if(@a>@c)
		Begin
			set @a=@c
		End

	return @a
End
```
This will create “least” function

Step16 (j):
```
update georgia_crashes
set max_sev = dbo.least((select min(convert(int,occ_inj_pdo5)) as temp from
occdriver_tbl where occ_acc_id=georgia_crashes.agencyid),
(select min(convert(int,pssgr_inj_pdo5)) as temp from passenger_tbl where
occ_acc_id=georgia_crashes.agencyid),
(select min(convert(int,ped_inj_pdo5)) as temp from pedestrian_tbl where
	ped_acc_id=georgia_crashes.agencyid))
```
This is calculate the maximum severity in occdrvr_tbl, passenger_tbl and pedestrian_tbl and assign it to the max_sev column in Georgia_Crashes

Step17: Rename ACC_TNV to numVehicles
This will rename ACC_TNV as numVehicles

Step18: Rename ACC_TNI to numOfInjuries
This will rename ACC_TNI as numOfInjuries

Step19: Rename ACC_TNF to numOfFatalities
This will rename ACC_TNF as numOfFatalities

Step20:

```
Update Georgia_crashes
set Georgia_crashes.junctionRelationship = case
when (ramp.ACC_ID = Georgia_crashes.agencyID) then '5'
when (Georgia_RRX_tbl.rrx_ACC_ID = v.agencyID) then '7'
when (Location_tbl.loc_interroute_type is not null and
        Location_tbl.acc_id = Georgia_crashes.agencyID) then '2'
when (Location_tbl.loc_interroute_type is null and
        Location_tbl.acc_id = Georgia_crashes.agencyID) then '1'
end
from Georgia_crashes, fulton_ramp, Georgia_RRX_tbl, Location_tbl
```

This will identify ramps, rail road crossings, intersections and roadway segments

Step21:

```
Alter table Georgia_crashes
add lightCondition nvarchar(2),
weatherCondition nvarchar(2),
surfaceCondition nvarchar(2),
roadCondition nvarchar(2),
collisionType nvarchar(255)
```

This will add columns lightCondition, weatherCondition, surfaceCondition, roadCondition and collisionType with the abovementioned datatype.

Step22:

```
Update Georgia_crashes
Set weatherCondition = ACC_WEAT_TYPE
Where Georgia_crashes.agencyID = accident_tbl.acc_id
```

This will set weatherCondition as ACC_WEAT_TYPE based on accident ids.

Step23:

```
Update Georgia_crashes
Set lightCondition = ACC_LITE_TYPE
Where Georgia_crashes.agencyID = accident_tbl.acc_id
```

This will set lightCondition as ACC_LITE_TYPE based on accident ids.

Step24:

```
Update Georgia_crashes
Set surfaceCondition = ACC_SURF_TYPE
Where Georgia_crashes.agencyID = accident_tbl.acc_id
```

This will set surfaceCondition as ACC_SURF_TYPE based on accident ids.

Step25:

```
Update Georgia_crashes
Set roadCondition = ACC_RDD_TYPE
```

Where Georgia_crashes.agencyID = accident_tbl.acc_id
This will set roadCondition as ACC_RDD_TYPE based on accident ids.

Step26: Update Georgia_crashes

```
set Collisiontype = case
when (acc_mnrc_type < '6') then acc_mnrc_type
else acc_he1_type
end
```

This will initially set Collisiontype to acc_mnrc_type when acc_mnrc_type <6 else it will set to acc_he1_type

Step27: Update Georgia_crashes

```
set Collisiontype = 'ped'
where Georgia_crashes.agencyID in (Select pedestrian_tbl.PED_ACC_ID from
pedestrian_tbl)
```

This will set Collisiontype to ‘ped’ based on the records in pedestrian table.

Step28: Update Georgia_crashes

```
set Collisiontype = 'bike'
where Georgia_crashes.agencyID in (Select Vehicle_tbl.VEH_ACC_ID from Vehicle_tbl
where Vehicle_tbl.veh_type_type = '19')
```

This will set Collisiontype to ‘bike’ when veh_type_type = '19' in vehicle table.

Step29: Update Georgia_crashes

```
set Collisiontype = case
when (georgia_crashes.acc_mnrc_type = '1') then 'angle'
when (georgia_crashes.acc_mnrc_type = '2') then 'headon'
when (georgia_crashes.acc_mnrc_type = '3') then 'rearend'
when (georgia_crashes.acc_mnrc_type = '4') then 'sssamedir'
when (georgia_crashes.acc_mnrc_type = '5') then 'ssoppdir'
else collisiontype
end
from Georgia_crashes
```

This will change the coding of Collisiontype based on acc_mnrc_type value.

You need to now create 6 columns in the AltAccident table displaying the characteristics of the vehilce: vehicle configuration, vehicle direction and vehicle maneuver for the first two vehicles that are involved in a crash.

Step30: Alter table Georgia_crashes

Add veh_num1 nvarchar(2), veh_dir1 nvarchar(2), veh_type1 nvarchar(2), veh_manv1 nvarchar(2), veh_num2 nvarchar(2), veh_dir2 nvarchar(2), veh_ type2 nvarchar(2), veh_manv2 nvarchar(2)

This will add the above mentioned columns with their respective datatypes.

To accomplish this task, you need to create two separate tables from vehicle_tbl, one for each vehicle.

Step31: create table veh12 (veh_acc_id nvarchar(15), Veh_num nvarchar(2), veh_dir1 nvarchar(2), veh_ type1 nvarchar(2), veh_manv1 nvarchar(2))

This will create a table veh12 with the above mentioned columns.

Step32: Insert into veh12 (veh_num2,veh_dir2,veh_ type2,veh_manv2)
(select veh_no,veh_dirt_type,veh_type_type,
veh_manv_type from vehicle_tbl where vehicle_tbl.veh_no = '01'
and vehicle_tbl.veh_acc_id = veh12.veh_acc_id)

This will insert values into table veh12 for the first vehicle involved in the crash.

Step33: create table veh45 (veh_acc_id nvarchar(15), Veh_num nvarchar(2), veh_dir1 nvarchar(2), veh_ type1 nvarchar(2), veh_manv1 nvarchar(2))

This will create a table veh45 with the above mentioned columns.

Step34: Insert into veh45 (veh_num2,veh_dir2,veh_ type2,veh_manv2)
(select veh_no,veh_dirt_type,veh_type_type,
veh_manv_type from vehicle_tbl where vehicle_tbl.veh_no = '02'
and vehicle_tbl.veh_acc_id = veh45.veh_acc_id)

This will insert values into table veh45 for the first vehicle involved in the crash.

Step35: update georgia_crashes
set georgia_crashes.veh_num1= veh12.veh_num1
set georgia_crashes.veh_dir1= veh12.veh_dir1
set georgia_crashes.veh_type1= veh12.veh_type1
set georgia_crashes.veh_manv1= veh12.veh_manv1
set georgia_crashes.veh_num2= veh45.veh_num2
set georgia_crashes.veh_dir2= veh45.veh_dir2
set georgia_crashes.veh_ type2 = veh45.veh_ type2
set georgia_crashes.veh_manv2= veh45.veh_manv2
from veh12, veh45
where georgia_crashes.agencyID = veh12.veh_acc_id = veh45.veh_acc_id

This will update Georgia_crashes table from veh12 and veh45 tables based on accident Id.

Step36: **Rename columns** veh_type1, veh_type2, veh_dir1, veh_dir2, veh_manv1, veh_manv2 as v1vehicleConfiguration, v2vehicleConfiguration, v1initialTravelDirection, v2initialTravelDirection,v1vehicleManeuver, v2vehicleManeuver
This will rename columns as described above.

Step37: Change the coding of v1vehicleConfiguration and v2vehicleConfiguration as follows

```
                UPDATE Georgia_crashes
SET v1vehicleConfiguration = CASE
                when v1vehicleConfiguration = 01 then '1'
                when v1vehicleConfiguration = 02 then '2'
                when v1vehicleConfiguration = 03 then '9'
                when v1vehicleConfiguration = 04 then '8'
                when v1vehicleConfiguration = 05 then '11'
                when v1vehicleConfiguration = 06 then '13'
                when v1vehicleConfiguration = 07 then '13'
                when v1vehicleConfiguration = 08 then '6'
                when v1vehicleConfiguration = 09 then '17'
                when v1vehicleConfiguration = 10 then '14'
                when v1vehicleConfiguration = 11 then '1'
                when v1vehicleConfiguration = 12 then '5'
                when v1vehicleConfiguration = 13 then '15'
                when v1vehicleConfiguration = 14 then '13'
                when v1vehicleConfiguration = 15 then '16'
                when v1vehicleConfiguration = 16 then '5'
                when v1vehicleConfiguration = 17 then '4'
                when v1vehicleConfiguration = 18 then '4'
                when v1vehicleConfiguration = 19 then '4'
                when v1vehicleConfiguration = 20 then '17'
                when v1vehicleConfiguration = 21 then '17'
                when v1vehicleConfiguration = 22 then '17'
                end

UPDATE Georgia_crashes
SET v2vehicleConfiguration = CASE
                when v2vehicleConfiguration = 01 then '1'
                when v2vehicleConfiguration = 02 then '2'
                when v2vehicleConfiguration = 03 then '9'
                when v2vehicleConfiguration = 04 then '8'
                when v2vehicleConfiguration = 05 then '11'
                when v2vehicleConfiguration = 06 then '13'
```

```
        when v2vehicleConfiguration = 07 then '13'
        when v2vehicleConfiguration = 08 then '6'
        when v2vehicleConfiguration = 09 then '17'
        when v2vehicleConfiguration = 10 then '14'
        when v2vehicleConfiguration = 11 then '1'
        when v2vehicleConfiguration = 12 then '5'
        when v2vehicleConfiguration = 13 then '15'
        when v2vehicleConfiguration = 14 then '13'
        when v2vehicleConfiguration = 15 then '16'
        when v2vehicleConfiguration = 16 then '5'
        when v2vehicleConfiguration = 17 then '4'
        when v2vehicleConfiguration = 18 then '4'
        when v2vehicleConfiguration = 19 then '4'
        when v2vehicleConfiguration = 20 then '17'
        when v2vehicleConfiguration = 21 then '17'
        when v2vehicleConfiguration = 22 then '17'
        end
```

This will change the coding of v1vehicleConfiguration and v2vehicleConfiguration.

Step38: Change the coding of v1vehicleManeuver and v2vehicleManeuver as follows

```
UPDATE Georgia_crashes
SET v1vehicleManeuver = CASE
        when v1vehicleManeuver = 01 then '1'
        when v1vehicleManeuver = 02 then '2'
        when v1vehicleManeuver = 03 then '3'
        when v1vehicleManeuver = 04 then '4'
        when v1vehicleManeuver = 05 then '5'
        when v1vehicleManeuver = 06 then '6'
        when v1vehicleManeuver = 07 then '7'
        when v1vehicleManeuver = 08 then '8'
        when v1vehicleManeuver = 09 then '9'
        else v1vehicleManeuver
end

UPDATE Georgia_crashes
SET v2vehicleManeuver = CASE
        when v2vehicleManeuver = 01 then '1'
        when v2vehicleManeuver = 02 then '2'
        when v2vehicleManeuver = 03 then '3'
```

```
                when v2vehicleManeuver = 04 then '4'
                when v2vehicleManeuver = 05 then '5'
                when v2vehicleManeuver = 06 then '6'
                when v2vehicleManeuver = 07 then '7'
                when v2vehicleManeuver = 08 then '8'
                when v2vehicleManeuver = 09 then '9'
                else v2vehicleManeuver
end
```

This will change the coding of v1vehicleManeuver and v2vehicleManeuver.

Step39: Change the coding of v1initialTravelDirection and v2initialTravelDirection as follows

```
UPDATE Georgia_crashes
SET v1initialTravelDirection = CASE
                        when v1initialTravelDirection = '1' then 'NB'
                        when v1initialTravelDirection = '2' then 'SB'
                        when v1initialTravelDirection = '3' then 'EB'
                        when v1initialTravelDirection = '4' then 'WB'
                else v1initialTravelDirection
end

UPDATE Georgia_crashes
SET v2initialTravelDirection = CASE
                        when v2initialTravelDirection = '1' then 'NB'
                        when v2initialTravelDirection = '2' then 'SB'
                        when v2initialTravelDirection = '3' then 'EB'
                        when v2initialTravelDirection = '4' then 'WB'
                else v2initialTravelDirection
end
```

This will change the coding of v1initialTravelDirection and v2initialTravelDirection.

Step40: Change the coding of collisiontype as follows

```
UPDATE Georgia_crashes
SET collisiontype = CASE
                when collisiontype= '01' then '1'
                when collisiontype= '02' then '2'
                when collisiontype= '03' then '3'
                when collisiontype= '04' then '4'
                when collisiontype= '05' then '5'
                when collisiontype= '06' then '6'
                when collisiontype= '07' then '7'
```

when collisiontype= '08' then '8'
when collisiontype= '09' then '9'
else collisiontype
end

This will change the coding of collisiontype.

Step41: Alter table Georgia_crashes
drop column ACC_ICO_TYPE, ACC_EMSN, ACC_EMSA, ACC_HOSA, ACC_INVS, ACC_CIT, ACC_HE1_TYPE, ACC_MNRC_TYPE, ACC_LOI_TYPE, ACC_RCOMP_TYPE, ACC_RCHAR_TYPE, ACC_DAYOFWEEK_TYPE, DMVS_LAST_UPDATE, DMVSDOT_LAST_UPDATE, LOC_ACC_ID, LOC_ACC_JULDT, LOC_RCLINK_IDENTIFIER, LOC_CITY_IDENTIFIER, LOC_COUNTY_IDENTIFIER, LOC_ROUTE_IDENTIFIER, LOC_ROUTE_SUFFIX, LOC_ACC_MILELOGCUM, LOC_INTERROUTE_TYPE, LOC_INTERROUTE_IDENTIFIER, LOC_INTERROUTE_SUFFIX, LOC_ACCESSCONTROL_TYPE, LOC_AADT_COUNT, LOC_AUXLANELEFT_TYPE, LOC_AUXLANERIGHT_TYPE, LOC_AUXLANELEFT_WIDTH, LOC_AUXLANERIGHT_WIDTH, LOC_DIVHWYBARRIER_TYPE, LOC_DIVHWYMEDIAN_TYPE, LOC_FEDELIG_TYPE, LOC_FUNCTIONALCLASS_TYPE, LOC_RURALURBAN_TYPE, LOC_SIGNAL_TYPE, LOC_SPEEDLIMIT_NUMBER, LOC_LANESLEFT_COUNT, LOC_LANESRIGHT_COUNT, LOC_LOCATE_DATE, LOC_LOCATOR_IDENTIFIER, LOC_X, LOC_Y, ACC_ACCNO, ACC_NCICNO, veh_num1, veh_num2

This will drop all the other columns in Georgia_crashes table that are not required for SafetyAnalyst

Step42: Export the table Georgia_crashes into a text file (.txt) as comma separated values

Step43: Open the file in Wordpad and add a row – AltAccident and save it.

AltRoadwayCharacteristics

Step1: Open RC file in Access database and save it. We need to create an agencyID which is unique to each roadway segment. This is an 18 digit alphanumeric value. It is county number (3 digits) followed by routetype (1 digit) followed by route number (6 digits) followed by beginning milepost without decimals (4 digits) and ending milepost without decimals (4 digits)

To address to some of the issues while opening files in .csv format, we prefer to make the agencyID a 20 digit alphanumeric value. Its the 18 digit value which starts and ends with a **'B'**.

Step2: Write a query in access to obtain just the four digits from beginning milepost and ending milepost.

BegM: 10000+[BEG_MEASURE]*100

EndM: 10000+[END_MEASURE]*100

This will create two columns (BegM and EndM) in the RC file with the beginning and ending mileposts as numbers without decimals.

Step3: Write a query in access to generate the ID.

ID: 'B' & [COUNTY]&[ROUTE_TYPE]&[ROUTE_NUM]&(Right([BegM],4))&(Right([EndM],4)) & 'B'

This will generate the ID column as required.

Step4: Save the query as a table RC_ID1 and add it to the acces database.

Step5: Open SQL server and create a new database "GDOT_RC"

Step6: ImportRC_ID1 from access database into GDOT_RC database.

Step7:

```
create table AltRC
(agencyID nvarchar(255), locSystem varchar(2), routeType nvarchar(6),
 routeName nvarchar(6), county nvarchar(6), startOffset decimal(4, 2),
 endOffset decimal(4,2), segmentLength decimal(4, 2), areaType nvarchar(2),
roadwayClass1 nvarchar(6), d1numThruLane nvarchar(6), d2numThruLane nvarchar(6),
medianType1 nvarchar(6),medianWidth nvarchar(6), postedSpeed nvarchar(6),
accessControl nvarchar(6), operationWay nvarchar(6))
```

This will create a table AltRC with all the required columns

Step8:

```
insert into AltRC (agencyID, routeType, county, startoffset,
endoffset, segmentlength, areatype, d1numthrulane, d2numthrulane, medianwidth,
postedspeed, accesscontrol, operationway)
select ID, ROUTE_TYPE, COUNTY, BEG_MEASURE, END_MEASURE,
SECTION_LENGTH,
RURAL_URAN, T_LANES_LEFT, T_LANES_RIGHT, DIV_HWY_MEDIAN_WIDTH,
SPEED_LIMIT, ACCESS_CONTROL, OPERATION from RC_ID1
```

This will insert values into the above mentioned columns in AltRC table from RC_ID1 table.

Step9:

```
UPDATE AltRC
SET LocSystem = 'B'
```

This will set LocSystem as 'B'

Step10:

```
UPDATE AltRC
```

```
SET routeName = rc_id1.ROUTE_TYPE+rc_id1.ROUTE_NUM
from rc_id1, altrc where RC_ID1.ID = AltRC.agencyID
```
This will set routeName as a combination of route type and route number based on agencyID

Step11:
```
UPDATE AltRC
SET routeType = CASE
        when RC_ID1.Func_Class = 1 then '99'
        when RC_ID1.Func_Class = 11 then '99'
        else routeType
    end
from rc_id1, altrc where RC_ID1.ID = AltRC.agencyID
```
This will set routeType as '99' for interstates else does not change

Step12:
```
update AltRC
set roadwayclass1 = case
when (RC_ID1.func_class= '01' or RC_ID1.func_class = '11') then '1'
when (RC_ID1.func_class= '12' ) then '2'
when (RC_ID1.func_class= '02' or RC_ID1.func_class = '14') then '3'
when (RC_ID1.func_class= '06' or RC_ID1.func_class = '16') then '4'
when (RC_ID1.func_class= '07' or RC_ID1.func_class = '17') then '5'
when (RC_ID1.func_class= '08')        then '6'
when (RC_ID1.func_class= '09' or RC_ID1.func_class = '19') then '7'
else '99'
end
from AltRC, RC_ID1
where AltRC.agencyid = RC_ID1.id
```
This will recode column roadwayclass1 based on functional classification of roads

Step13:
```
alter table AltRC
add  d1shoulderTypeOut nvarchar(1),d1shoulderTypeIn nvarchar(1),
     d2shoulderTypeOut nvarchar(1), d2shoulderTypeIn nvarchar(1)
```
This will add the above mentioned columns to the table AltRC.

Step14:
```
insert into AltRC (d1shoulderTypeOut,d1shoulderTypeIn,d2
shoulderTypeOut, d2shoulderTypeIn)
select udiv_hwy_shldr_type_lft, div_hwy_shldr_type_lft,
     udiv_hwy_shldr_type_rt, div_hwy_shldr_type_rt
from RC_ID1, altrc where RC_ID1.ID = AltRC.agencyID
```
This will insert values into the above mentioned columns in AltRC table from RC_ID1 table.

Step15: UPDATE AltRC
SET mediantype1 = rc_id1.div_hwy_median_type
from rc_id1, altrc where RC_ID1.ID = AltRC.agencyID
This will insert div_hwy_median_type values into medianType1 in AltRC table from RC_ID1 table based on agencyID

Step16: update AltRC
set areaType = case
when (RC_ID1.func_class= '01' or RC_ID1.func_class = '02'
or RC_ID1.func_class = '06' or RC_ID1.func_class = '07'
or RC_ID1.func_class = '08' or RC_ID1.func_class = '09') then '7'
when (RC_ID1.func_class= '11' or RC_ID1.func_class = '12'
or RC_ID1.func_class = '14' or RC_ID1.func_class = '16'
or RC_ID1.func_class = '17' or RC_ID1.func_class = '19') then '8'
else '99'
end
from AltRC, RC_ID1
where AltRC.agencyid = RC_ID1.id
This will recode areaType column based on agencyID and functional class.

Step 17: update AltRC
set operationWay = case
when (operationWay = ‘5’) then ‘1’
when (operationWay = ‘6’) then ‘2’
else operationWay
end
This will recode operationway column.

Step18: Export the table AltRC into a text file (.txt) as comma separated values

Step19: Open the file in Wordpad and add a row – AltRoadwaySegment and save it.

AltSegmentTraffic

Step1: Open AADT file in Access database and save it. We need to create an agencyID which is unique to each roadway segment. This is an 18 digit alphanumeric value. It is county number (3 digits) followed by routetype (1 digit) followed by route number (6 digits) followed by beginning milepost without decimals (4 digits) and ending milepost without decimals (4 digits)

Step2: Write a query in access to obtain just the four digits from beginning milepost and ending milepost.

BegM: 10000+[BEG_MEASURE]*100
EndM: 10000+[END_MEASURE]*100

This will create two columns (BegM and EndM) in the AADT file with the beginning and ending mileposts as numbers without decimals.

Step3: Write a query in access to generate the ID.

ID:
'B'&[COUNTY]&[ROUTE_TYPE]&[ROUTE_NUM]&(Right([BegM],4))&(Right([EndM],4))& 'B'

This will generate the ID column as required.

Step4: Create a crosstab query if required with agencyID as row heading and year (2000, 2001, 2002, 2003, 2004, 2005 etc..) as column headings and save it as agencyID_adt.

Step5: In SQL server open the database "GDOT_RC"

Step6: Import agencyID_adt from access database into GDOT_RC database.

Step7:

```
create table Altadt
(agencyID nvarchar(255), [year] int, adt decimal(6, 0))
```

This will create a table Altadt with the above mentioned columns.

Step8:

```
Insert into Altadt (ID,[year],adt)
(
select d.ID,1995,d.[1995] from agencyid_aadt_table as d where d.[1995] is not null
union
select d.ID,1996,d.[1996] from agencyid_aadt_table as d where d.[1996] is not null
union
select d.ID,1997,d.[1997] from agencyid_aadt_table as d where d.[1997] is not null
union
select d.ID,1998 ,d.[1998] from agencyid_aadt_table as d where d.[1998] is not null
union
```

```
select d.ID,1999,d.[1999] from agencyid_aadt_table as d where d.[1999] is not null
union
select d.ID,2000 ,d.[2000] from agencyid_aadt_table as d where d.[2000] is not null
union
select d.ID,2001,d.[2001] from agencyid_aadt_table as d where d.[2001] is not null
union
select d.ID,2002,d.[2002] from agencyid_aadt_table as d where d.[2002] is not null
union
select d.ID,2003,d.[2003] from agencyid_aadt_table as d where d.[2003] is not null
union
select d.ID,2004,d.[2004] from agencyid_aadt_table as d where d.[2004] is not null
union
select d.ID,2005 ,d.[2005] from agencyid_aadt_table as d where d.[2005] is not null
union
select d.ID,2006 ,d.[2006] from agencyid_aadt_table as d where d.[2006] is not null
union
select d.ID,2007 ,d.[2007] from agencyid_aadt_table as d where d.[2007] is not null )
delete from Altadt where adt=0
```

This will insert values into table Altadt in the required format.

Step9: Export the table Altadt into a text file (.txt) as comma separated values.

Step10: Open the file in Wordpad and add a row – AltSegmentTraffic and save it.

Step11: Open the file in Wordpad and change the column headings to agencyID,calendarYear,aadtVPD,percentHeavyVehicles, peakHourlyVolume,comment

From research, it is understood that generating and using longer aggregated segments yield better and more reliable results. The steps below explain the process of generation of aggregated segments.

AltRChomo:

Step1: Open SQL server and export the table "AltRC" (the table tats generated earlier based on SafetyAnalyst requirements).

Step2:

```
Alter table AltRC
add v2medianwidth nvarchar(2)
```

This will add v2medianwidth column with datatype nvarchar

Step3:

```
update AltRC
set v2medianwidth = case
when (medianWidth  between '0' and '0.5') then '0'
```

```
when (medianWidth between '0.6' and '3.5') then '2'
when (medianWidth between '3.6' and '6.5') then '5'
when (medianWidth between '6.6' and '9.5') then '8'
when (medianWidth between '9.6' and '12.5') then '11'
when (medianWidth between '12.6' and '15.5') then '14'
when (medianWidth between '15.6' and '20') then '17'
when (medianWidth between '20.1' and '30') then '25'
when (medianWidth >= '30') then '30'
end
```

This will add values to column 'v2medianwidth' based on the above mentioned criteria.

Step4: Add 2004, 2005 and 2006 adt values to each roadway segment

```
Alter table Georgia_crashes
add     adt04 decimal(4,2),
        adt05 decimal(4,2),
        adt06 decimal(4,2)

update AltRC
set adt04 = AltSegmentTraffic.adt from AltSegmentTraffic
where AltRC.agencyID = AltSegmentTraffic. agencyID and
AltSegmentTraffic.[year] = '2004'

update AltRC
set adt05 = AltSegmentTraffic.adt from AltSegmentTraffic
where AltRC.agencyID = AltSegmentTraffic. agencyID and
AltSegmentTraffic.[year] = '2005'

update AltRC
set adt06 = AltSegmentTraffic.adt from AltSegmentTraffic
where AltRC.agencyID = AltSegmentTraffic. agencyID and
AltSegmentTraffic.[year] = '2006'
```

This will add 2004, 2005 and 2006 adt values to each roadway segment.

Step5: A cursor is initially required to generate aggregated segments.

```
Declare @temp int
set @temp =0
Declare @curstartOffset decimal(4,2)
Declare @curendOffset decimal(4,2)
Declare @prevstartOffset decimal(4,2)
Declare @prevendOffset decimal(4,2)
Declare @curlocSystem varchar(2)
Declare @currouteType nvarchar(6)
```

```
Declare @currouteName nvarchar(12)
Declare @curcounty nvarchar(6)
Declare @curfunc_areatype decimal(2,0)
Declare @curroadwayclass1 nvarchar(6)
Declare @curd1numThruLane nvarchar(6)
Declare @curd2numThruLane nvarchar(6)
Declare @curmedianType1 nvarchar(6)
Declare @curaccessControl nvarchar(6)
Declare @curoperationWay nvarchar(6)
Declare @curv2medianWidth float
Declare @agencyID nvarchar(255)
Declare @prevlocSystem varchar(2)
Declare @prevrouteType nvarchar(6)
Declare @prevrouteName nvarchar(12)
Declare @prevcounty nvarchar(6)
Declare @prevfunc_areatype decimal(2,0)
Declare @prevroadwayclass1 nvarchar(6)
Declare @prevd1numThruLane nvarchar(6)
Declare @prevd2numThruLane nvarchar(6)
Declare @prevmedianType1 nvarchar(6)
Declare @prevaccessControl nvarchar(6)
Declare @prevoperationWay nvarchar(6)
Declare @prevv2medianWidth float

Declare gb_Cursor cursor for
(select locSystem,routeType,routeName,county,func_areatype, roadwayclass1,
d1numThruLane, d2numThruLane, medianType1, accessControl,
operationWay,v2medianWidth, startOffset,
endOffset,agencyID
from dbo.AltRC061009)
order by county,routeName,routeType,startOffset asc

open gb_Cursor
fetch next from gb_Cursor
into @curlocSystem, @currouteType, @currouteName,
@curcounty, @curfunc_areatype, @curroadwayclass1,
@curd1numThruLane, @curd2numThruLane, @curmedianType1,
@curaccessControl, @curoperationWay, @curv2medianWidth,
@curstartOffset,@curendOffset,@agencyID
while(@@fetch_status=0)
Begin
        if(@prevlocSystem <> @curlocSystem
```

```
        or @prevrouteType <> @currouteType
        or @prevrouteName <> @curroadwayclass1
        or @prevcounty <> @curcounty
        or @prevfunc_areatype <>@curfunc_areatype
        or @prevroadwayclass1 <>@curroadwayclass1
        or @prevd1numThruLane <>@curd1numThruLane
        or @prevd2numThruLane <>@curd2numThruLane
        or @prevmedianType1 <>@curmedianType1
        or @prevaccessControl <>@curaccessControl
        or @prevoperationWay <>@curoperationWay
        or @prevv2medianWidth <> @curv2medianWidth)
        Begin
                set @temp = @temp + 1
        End
        Update dbo.AltRC set new=@temp
                where agencyID = @agencyID

        set @prevlocSystem = @curlocSystem
        set @prevrouteType =@currouteType
        set @prevrouteName =@curroadwayclass1
        set @prevcounty = @curcounty
        set @prevfunc_areatype = @curfunc_areatype
        set @prevroadwayclass1 = @curroadwayclass1
        set @prevd1numThruLane = @curd1numThruLane
        set @prevd2numThruLane = @curd2numThruLane
        set @prevmedianType1 =@curmedianType1
        set @prevaccessControl =@curaccessControl
        set @prevoperationWay =@curoperationWay
        set @prevv2medianWidth = @curv2medianWidth

fetch next from gb_Cursor
into @curlocSystem, @currouteType, @currouteName,
@curcounty, @curfunc_areatype, @curroadwayclass1,
@curd1numThruLane, @curd2numThruLane, @curmedianType1,
@curaccessControl, @curoperationWay, @curv2medianWidth,
@curstartOffset,@curendOffset,@agencyID
End

close gb_Cursor
deallocate gb_Cursor
```

This will be used in the next step to generate aggregated segments.

Step6: select locSystem, routeType, routeName,county, func_areatype,roadwayclass1,d1numThruLane, d2numThruLane, medianType1, accessControl, operationWay,v2medianwidth, new, Min(startOffset)as begst,max(endOffset) as endst,
max(adt04) as [2004], max(adt05) as [2005], max(adt06) as [2006]
from dbo.AltRC
group by locSystem,routeType,routeName,county,func_areatype, roadwayclass1,d1numThruLane, d2numThruLane, medianType1, accessControl, operationWay, v2medianwidth, new
order by routeName,county,begst asc
This will generate aggregated segments.

A cursor might take a very long time (upto 20 hours) to execute. There is also another way to generate aggregated segments without using a cursor.
Following are the steps used to generate the aggregated segments:
Step1: select * into query1
from Altrc
order by rclink, startoffset asc
This will create a new table called query1 with records in ascending order based on rclink and startoffset.

Step2: Add 3 new columns (rcount1, rcount2 and new) to the table query1. The data types for the three new columns is bigint.

Step3: Make Rcount1 as an increment value.

Step4: Update query1
Set Rcount2 = Rcount1 + 1

Step5: select * into query
from query1
This will create a new table called query with records in query1.

Step6: update Query1 set new = 0
This will set the value of new from null to ‘0’ in query1.

Step7: update Query1 set new = 9
from Query1 inner join Query on
(Query1.rcount1 = Query.rcount2)
where Query1.locSystem <> Query.locSystem or
Query1.routeType <> Query.routeType or
Query1.routeName <> Query.routeName or

```
Query1.county <> Query.county or
Query1.areatype <> Query.areatype or
Query1.roadwayclass1 <> Query.roadwayclass1 or
Query1.d1numThruLane <> Query.d1numThruLane or
Query1.d2numThruLane <> Query.d2numThruLane or
Query1.medianType1 <> Query.medianType1 or
Query1.accessControl <> Query.accessControl or
Query1.operationWay <> Query.operationWay or
Query1.v2medianWidth <> Query.v2medianWidth
```

This will set the value of new from ‘0’ to ‘9’ by comparing the first record with its immediate next record and when the value in one of the many fields to be considered is different.

Step8:

```
update dbo.Query1 set new = rcount1
where new=9
```

This will set the value of new to rcount1 when the value of new is ‘9’

Step9:

```
declare @temp int
set @temp = 10
while(@temp<>0)
Begin
update one set one.new = (select new from Query1 where rcount1  = one.RCount1-1)
from Query1 as one
where one.new = 0
set @temp = @@ROWCOUNT
PRINT @temp
End
```

Step10:

```
update dbo.Query1 set new = rcount1
where new=9
```

This will set the value of new to rcount1 when the value of new is ‘9’

Step11:

```
Select locSystem, routeType,
routeName,county,areatype,roadwayclass1,
d1numThruLane, d2numThruLane, medianType1,
accessControl,
 operationWay,v2medianWidth, Min(startOffset)as begst,max(endOffset) as endst,new,
max([2004]) as [2004], max([2005]) as [2005], max([2006]) as [2006]
into AltRChomo from Query1
group by locSystem,routeType,routeName,county,areatype, roadwayclass1,
d1numThruLane, d2numThruLane, medianType1, accessControl,
operationWay,v2medianWidth,new
```

```
order by county,routeName,routeType,begst asc
```

This will generate the aggregated segments.

APPENDIX B: DATA MAPPING GUIDE FOR SAFETYANALYST

Attribute in SafetyAnalyst	Enumeration	Code changed to
Collision type	Parked vehicle	10
	Collision with railroad train	8
	Collision with bicyclist	bike
	Collision with pedestrian	Ped
	Collision with animal	9
	Collision with fixed object	34
	Collision with other object	13
	Overturn	1
	Fire or explosion	2
	Other single-vehicle non-collision	5
	Rear-end	rearend
	Head-on	headon
	Angle	angle
	Sideswipe, same direction	sssamedir
	Sideswipe, opposite direction	ssoppdir
	Unknown	99
	Immersion	3
	Jackknife	4
	Motor Vehicle in Motion	11
	Motor Vehicle in Motion – in other Roadway	12
	Deer	14
	Impact Attenuator	15
	Bridge Pier/Abutment	16
	Bridge Parapet End	17
	Bridge Rail	18
	Guardrail Face	19
	Guardrail End	20
	Median Barrier	21
	Highway Traffic Sign Post	22
	Overhead Sign Support	23
	Luminaries /Light Support	24
	Utility Pole	25
	Other Post	26
	Culvert	27
	Curb	28
	Ditch	29
	Embankment	30
	Fence	31
	Mailbox	32
	Tree	33

Attribute in SafetyAnalyst	Enumeration	Code changed to
accidentSeverity1	Fatal Injury	1
	Incapacitating	2
	Non-Incapacitating Injury	3
	Possible Injury	4
	Property-Damage-Only	5
	Unknown	X
Alcohol/Drug involvement	deployment deleted	
Bicycle indicator	deployment deleted	
Contributing Circumstances, Environment	deployment deleted	
Divided Highway Flag-side of road	deployment deleted	
Driveway Indicator	deployment deleted	
Pedestrian indicator	deployment deleted	
Run off road indicator	deployment deleted	
school bus related	deployment deleted	
tow-away indicator	deployment deleted	
work zone related	deployment deleted	
Contributing circumstances, road	none	1
	surface condition	surf
	Debris	4
	Rut, holes, bumps	3
	Work zone	6
	Worn, travel-polished surface	worn
	Obstruction in roadway	obstruction
	Control device	control
	Shoulders	2
	Non-highway work	delete
	Other	8
	water standing	5
	running water	7
	Unknown	99
Area Type	Urban	8
	rural	7
	unknown	X

Attribute in SafetyAnalyst	Enumeration	Code changed to
Light condition	Daylight	1
	Dawn	3
	Dusk	2
	Dark-lighted	4
	Dark-not lighted	5
	Dark-unknown lighting	deleted
	other	deleted
	unknown	deleted
Relationship to junction	non-junction	1
	At intersection	2
	Intersection-related	3
	At driveway or driveway-related	4
	Entrance/exit ramp	5
	Other part of interchange	6
	Railroad/highway grade crossing	7
	Crossover related	8
	Other	9
	Unknown	99
Roadway surface condition	dry	1
	wet	2
	snow	3
	slush	8
	ice/frost	4
	water	deleted
	sand	7
	mud/dirt/gravel	6
	oil	9
	other	5
	unknown	99
weather condition	clear	1
	cloudy	2
	fog , smog, smoke	6
	rain	3
	sleet/hail	5
	snow	4
	blowing snow	deleted
	severe crosswinds	deleted
	blowing sand, soil, dirt	deleted
	other	7
	unknown	99

Attribute in SafetyAnalyst	Enumeration	Code changed to
Route type	Interstate	99
	US route	deleted
	state route	1
	business route	deleted
	business loop	deleted
	spur route	deleted
	county road	2
	township road	7
	local road	3
	other	0
	Ramp	6
	Public road	8
	Collector- Distributor	9
	Col road	4
	Unofficial road	5
	unknown	X
Access control	Full access control	F
	Partial access control	P
	no access control	U
	unknown	99
Direction of Travel	Northbound	1
	Southbound	2
	Eastbound	3
	Westbound	4
	Not applicable	NA
	Unknown	X
Operation	One way road	1
	Two way road	2
	Reversible lanes	3
	One way during school hours	4
	One direction of travel for a divided highway	9
	Unknown	99
Jurisdiction	Federal maintained	1
	State maintained	2
	County maintained	3
	Township maintained	6
	Local maintained	4
	Other maintained	5
	Unknown	99

APPENDIX C: SAFETYANALYST ANALYTICAL TOOL: SCREENSHOT OF THE STEPS

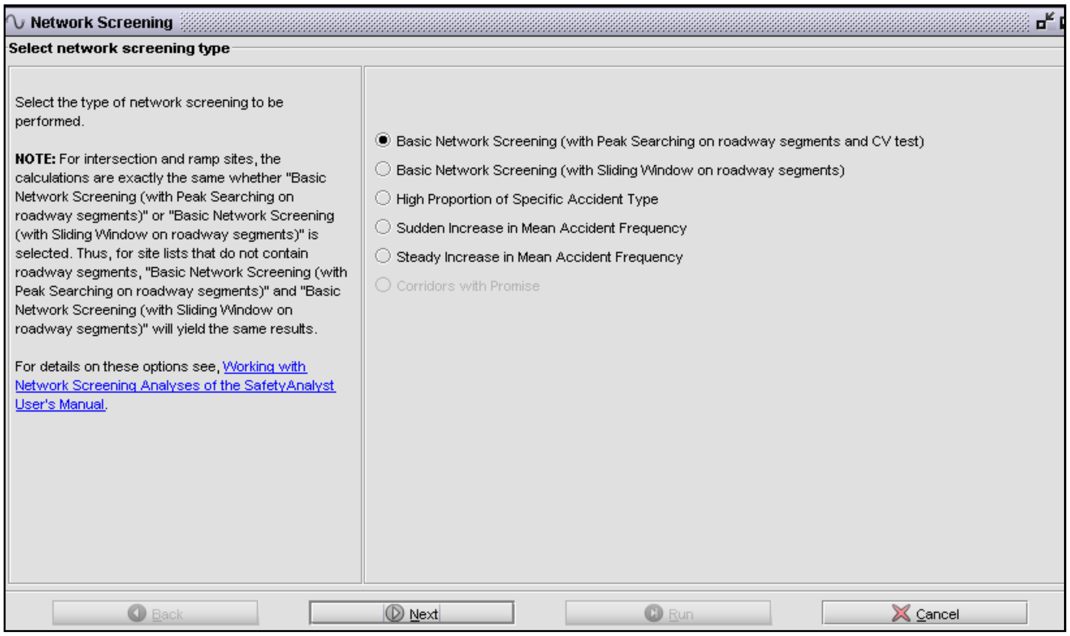

Figure 22: Select Network screening method

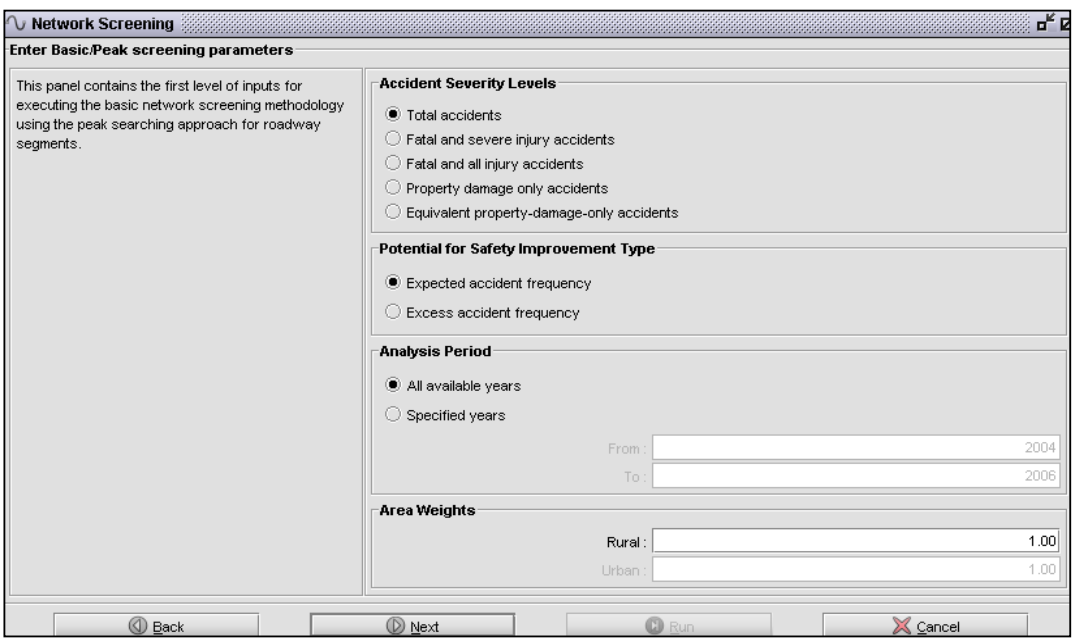

Figure 23: Select Accident Severity Level, PSI type, Analysis period and Area weights

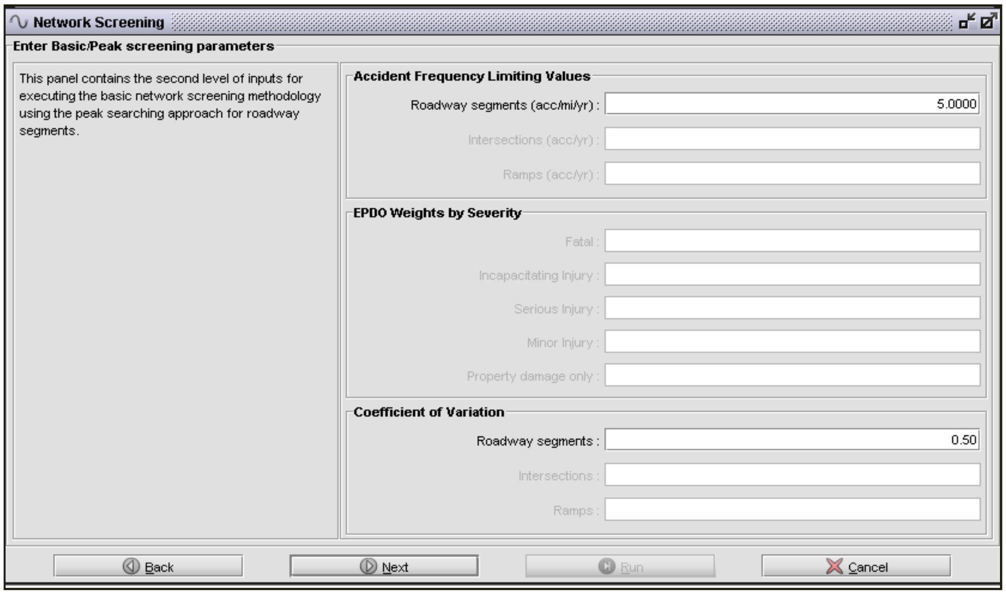

Figure 24: Select limiting value for accident frequency and the coefficient of variation

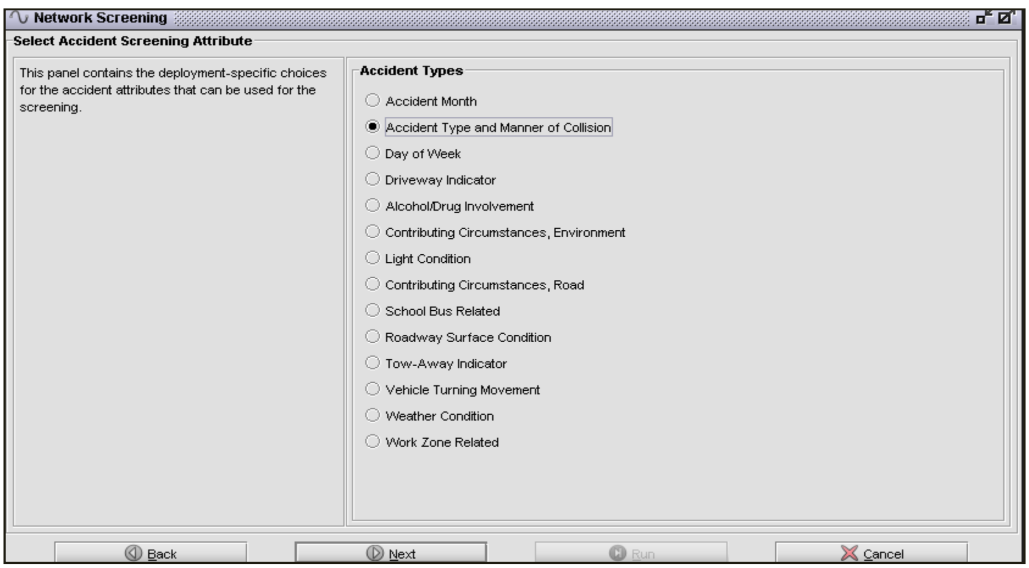

Figure 25: Select the accident type to be analyzed

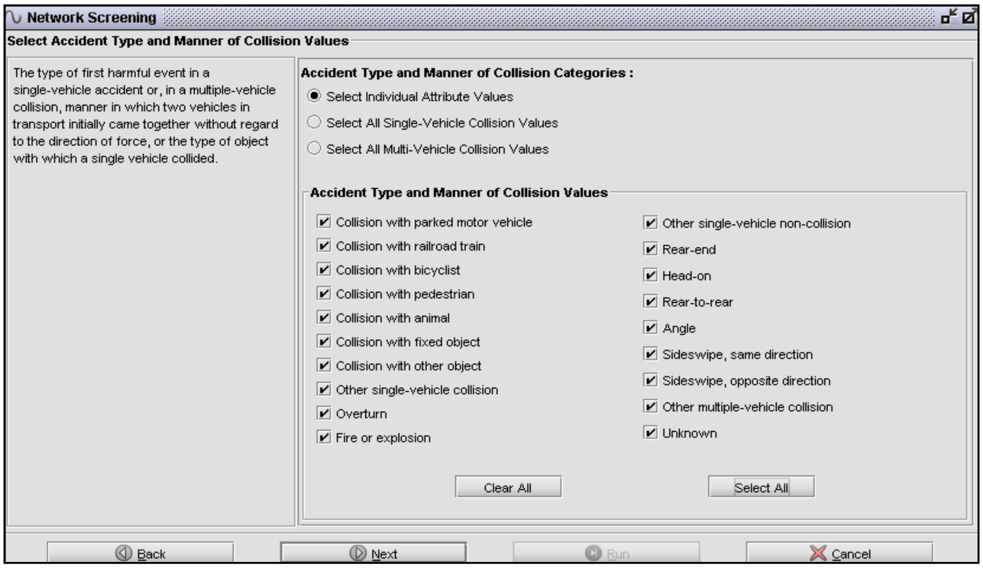

Figure 26: Select attributes for Accident type and manner of collision

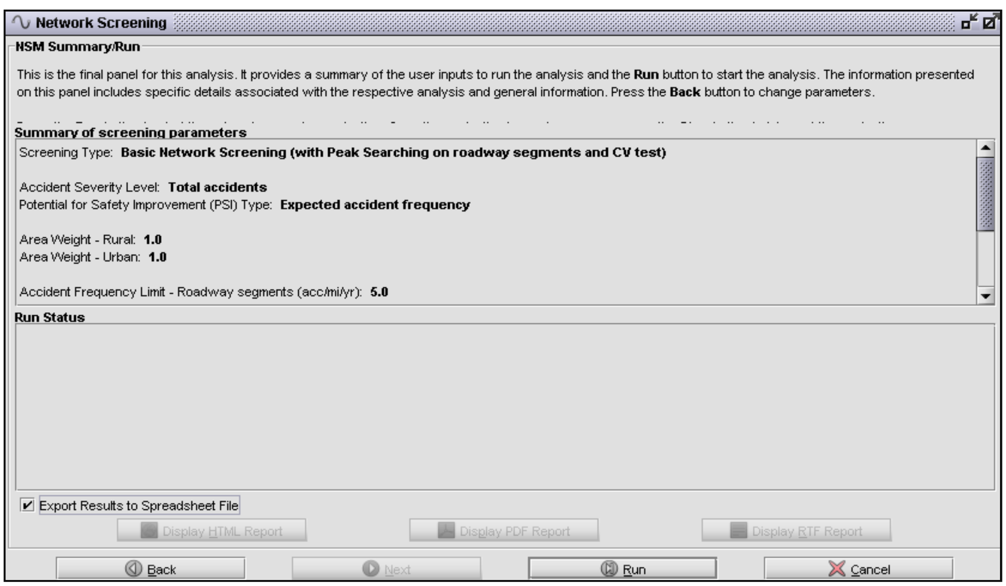

Figure 27: Final step in the “Network Screening” module

APPENDIX D: SAFETYANALYST OUTPUT

SafetyAnalyst

Network Screening Report

Oct 20, 2010

Notice

This Software Product is owned by the American Association of State Highway and Transportation Officials, Inc. ("AASHTO") and is being Licensed to your agency or employer under the terms of a Master and Supplemental agreement ("License Agreements") with AASHTO. The License Agreements contain important terms and conditions relating to the use of the Software Products which are binding on your agency or employer and you, including: limitations on AASHTO warranties; prohibitions on reverse engineering and unauthorized distribution or copying; protections of AASHTO patent, copyright, trademark and other proprietary rights; procedures for addressing defects; and termination and return of the Software Product. Violations of any of the terms of the License Agreements may result in the termination of your right to use this Software Product. Copies of the Licensee Agreements are available from your agency or employer or from AASHTO.

By installing and/or using this Software Product, you acknowledge that you have read and understood this NOTICE and agree to comply with the terms and conditions of the License Agreements.

1. Network Screening Report

Basic Network Screening

SafetyAnalyst: v4.0.8, packaged: Oct 1, 2010 11:02 PM on transvr1.aes.de.ittind.com

Data set title: 101310

Data set comment: segments with just operation way 1 or 2

Data set created: Wed, Oct 13, 09:21PM

Roadway Segments: Peak Searching

Accident Severity Level: Total accidents

Site Types: Segments

Screening Attribute: Accident Month = January; February; March; April; May; June; July; August; September; October; November; December

Potential for Safety Improvement Using: Expected accident frequency

Analysis Period: From 2004 To 2006

Major Reconstruction: No major reconstruction occurred at any sites during the analysis period

CV limit (roadway segments): 0.5

Area Weights (Rural): 1.0

Area Weights (Urban): 1.0

Limiting Value (Roadway Segments): 5.0 crashes/mi/yr

Number of sites in the site list: 64508

Number of sites evaluated: 64508

Number of segments evaluated: 64508

Total length of segments evaluated: 76844.090

Number of intersections evaluated: 0

Number of ramps evaluated: 0

Number of sites flagged: 148

Table: Basic Network Screening (with Peak Searching on roadway segments and CV test)

ID	Site Type	Site Subtype	County	Route	Site Start Loc	Site End Loc	Average Observed Accidents for Entire Site*	Location with Highest Potential for Safety Improvement								Rank	Addtl Windows of Interest
								Average Observed Acc*	Predicted Acc Freq*	Expected Acc Freq*	Var**	Start Loc	End Loc	No. of Exp Fats	No. of Exp Injs		
B255100030 000360058B	Segment	Seg/Rur; 2-lane	255	110003 00	0.36	0.58	77.93	140.58	2.20	35.78	2.98	0.48	0.58	-	-	1	0.36 - 0.46 0.46 - 0.56
B085100530 009351818B	Segment	Seg/Rur; 2-lane	085	110053 00	9.35	18.18	10.10	213.22	1.34	34.97	1.97	15.65	15.75	-	-	2	15.45 - 15.55 15.55 - 15.65 15.75 - 15.85 15.85 - 15.95 17.65 - 17.75
B015206330 004100643B	Segment	Seg/Rur; 2-lane	015	220633 00	4.1	6.43	18.26	131.89	1.72	28.64	1.95	6.3	6.4	-	-	3	5.6 - 5.7 5.7 - 5.8 5.8 - 5.9 5.9 - 6.0 6.0 - 6.1 6.1 - 6.2 6.2 - 6.3 6.33 - 6.43
B151101550 005550563B	Segment	Seg/Rur; 2-lane	151	110155 00	5.55	5.63	170.69	170.69	1.35	24.55	1.15	5.55	5.63	-	-	4	
B241100150 000401062B	Segment	Seg/Rur; 2-lane	241	110015 00	0.4	10.62	4.65	107.46	1.52	20.56	1.28	10.0	10.1	-	-	5	10.1 - 10.2

ID	Site Type	Site Subtype	County	Route	Site Start Loc	Site End Loc	Average Observed Accidents for Entire Site*	Location with Highest Potential for Safety Improvement								Rank	Addtl Windows of Interest
								Average Observed Acc*	Predicted Acc Freq*	Expected Acc Freq*	Var**	Start Loc	End Loc	No. of Exp Fats	No. of Exp Injs		
B151101550 005720597B	Segment	Seg/Rur; 2-lane	151	110155 00	5.72	5.97	62.81	99.00	1.35	17.52	0.99	5.87	5.97	-	-	6	5.72 - 5.82 5.82 - 5.92
B035100160 001671011B	Segment	Seg/Rur; 2-lane	035	110016 00	1.67	10.11	6.40	77.65	1.55	15.23	0.97	9.67	9.77	-	-	7	6.47 - 6.57 8.97 - 9.07 9.17 - 9.27 9.37 - 9.47 9.57 - 9.67 9.77 - 9.87 9.87 - 9.97 10.01 - 10.11
B211100120 011171343B	Segment	Seg/Rur; 2-lane	211	110012 00	11.17	13.43	9.03	89.26	1.55	15.09	0.99	11.97	12.07	-	-	8	
B021100190 000000337B	Segment	Seg/Rur; 2-lane	021	110019 00	0.0	3.37	6.89	87.13	2.04	14.76	1.28	0.8	0.9	-	-	9	3.2 - 3.3 3.27 - 3.37

APPENDIX E: SAS CODE TO GENERATE SPFs

```
DM
'LOG;CLEAR;OUT;CLEAR;';
OPTIONS
NODATE NONUMBER LS=90 PS=80;
DATA
alluri;
INFILE
'U:\profile.cu\My Documents\My SAS Files\GDOT\052110sas_tot.csv'
delimiter=',' firstobs=2;
INPUT agencyID SiteSubtype $ segmentlength avgAADT lnaadt  accidentcount lnlenyrs
;
PROC
GENMOD; BY SiteSubtype;
MODEL accidentcount =lnaadt /
LINK = Log DIST = NEGBIN OFFSET = lnlenyrs;
run;quit;
```

APPENDIX F: SAS CODE TO CALCULATE FREEMAN TUKEY'S R SQUARE

```
PROC IMPORT OUT= WORK.try
        DATAFILE= "C:\Users\labuser\Documents\My SAS Files\GDOT\try101tot.
csv"
        DBMS=CSV REPLACE;
    GETNAMES=YES;
    DATAROW=2;
RUN;

PROC GENMOD DATA = try;
MODEL accidentcount=lnAADT
     /offset=lnlenyrs  LINK=LOG  TYPE1  TYPE3  ALPHA=0.10  WALDCI  DIST=NEGBIN
SCALE=PEARSON
      MAXIT=300 OBSTATS;
output out=ResAll
      pred=PredAcc
      resraw=resraw
      STDRESDEV=STDRESDEV;
ODS OUTPUT ModelInfo=Info ModelFit=Fit ConvergenceStatus=Converge
        ParameterEstimates=ParmEst Type1=Type1 Type3=Type3 obstats=allout;

quit;


DATA AllTOTW; MERGE try (KEEP=accidentcount SiteSubtype) AllOUT;
*by SiteSubtype;
F=SQRT(accidentcount)+SQRT(accidentcount+1);
E=F-SQRT(4*PRed+1);
G=1/SQRT(PRed);
H=RESCHI+SQRT(PRed);
J=accidentcount/PRed;

proc means data=alltotw n css uss;
var f e;
output out=tot_w n=nf ne css=cf ce uss=uf ue;
data tot_w;
set tot_w;
rft2=1-ue/cf;
proc print;
run; quit;
```

APPENDIX G: SAFETY PERFORMANCE FUNCTIONS FOR VARIOUS SITE SUBTYES FOR TOTAL, AND FATAL INJURY CRASHES

The following graphs explain how well each SPF fits the Georgia data. The graphs also show the default national SPFs used in SafetyAnalyst calibrated to Georgia data in against the observed crashes.

All the graphs are plotted with AADT on the X-axis and, predicted and observed crashes (in crashes per mile per year) on the Y-axis. Table 88 describes the colors used to plot various SPFs.

Table 88: Color-codes used in graphs

Color	Code	SPF
Red	SA	Default national SPFs used in SafetyAnalyst - non calibrated
Green	GA	SPFs generated with Georgia data
Blue	SA calibrated to GA	Default national SPFs used in SafetyAnalyst – calibrated to Georgia data

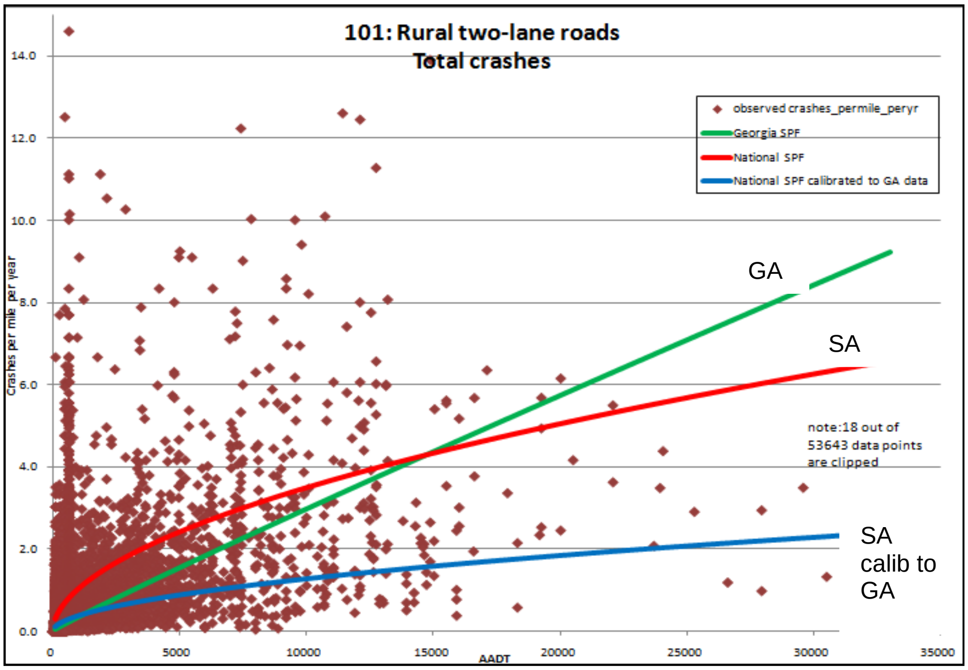

Figure 28: SPFs for site subtype 101 considering total crashes

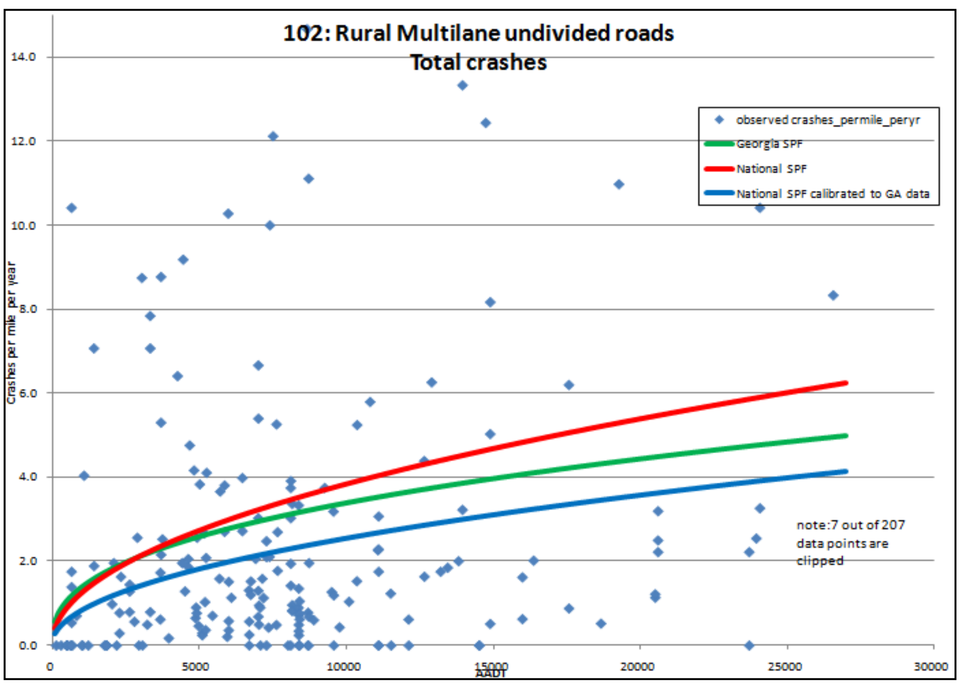

Figure 29: SPFs for site subtype 102 considering total crashes

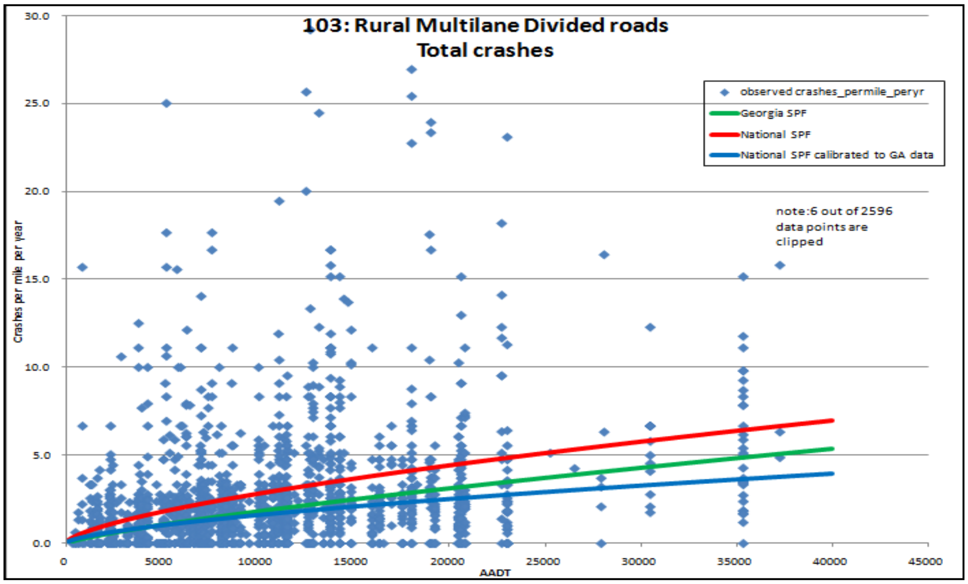

Figure 30: SPFs for site subtype 103 considering total crashes

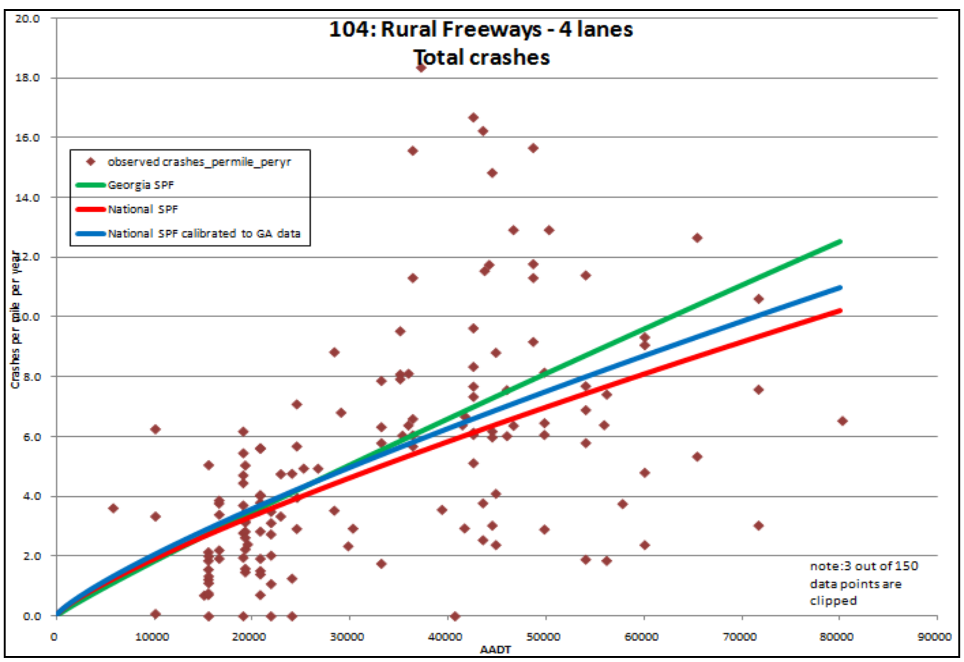

Figure 31: SPFs for site subtype 104 considering total crashes

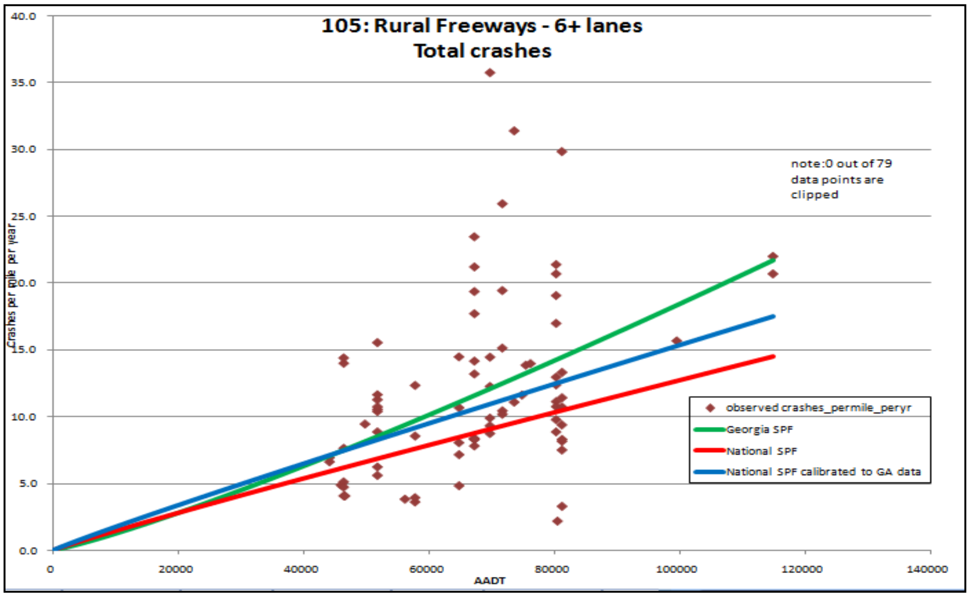

Figure 32: SPFs for site subtype 105 considering total crashes

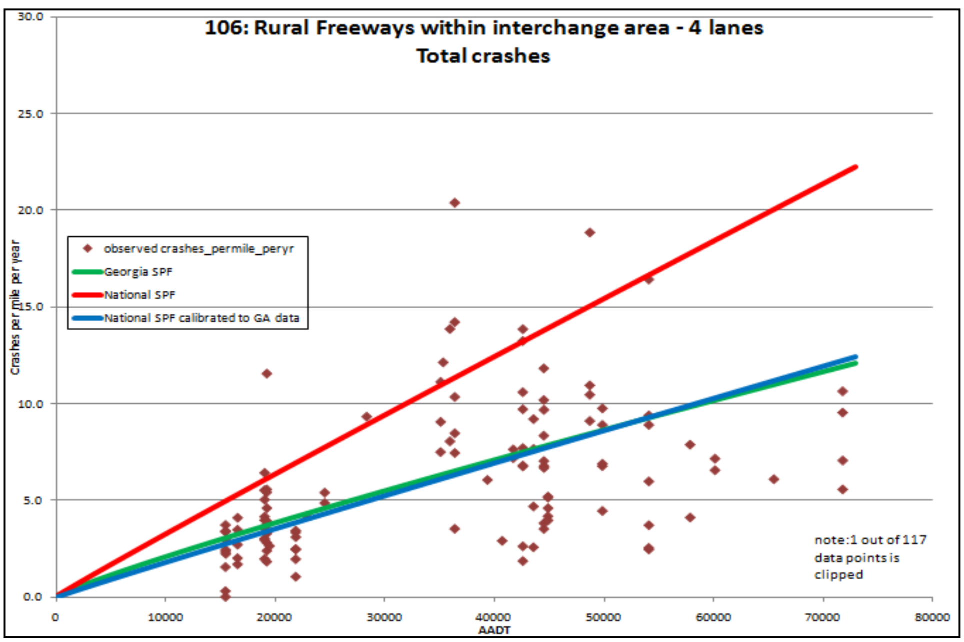

Figure 33: SPFs for site subtype 106 considering total crashes

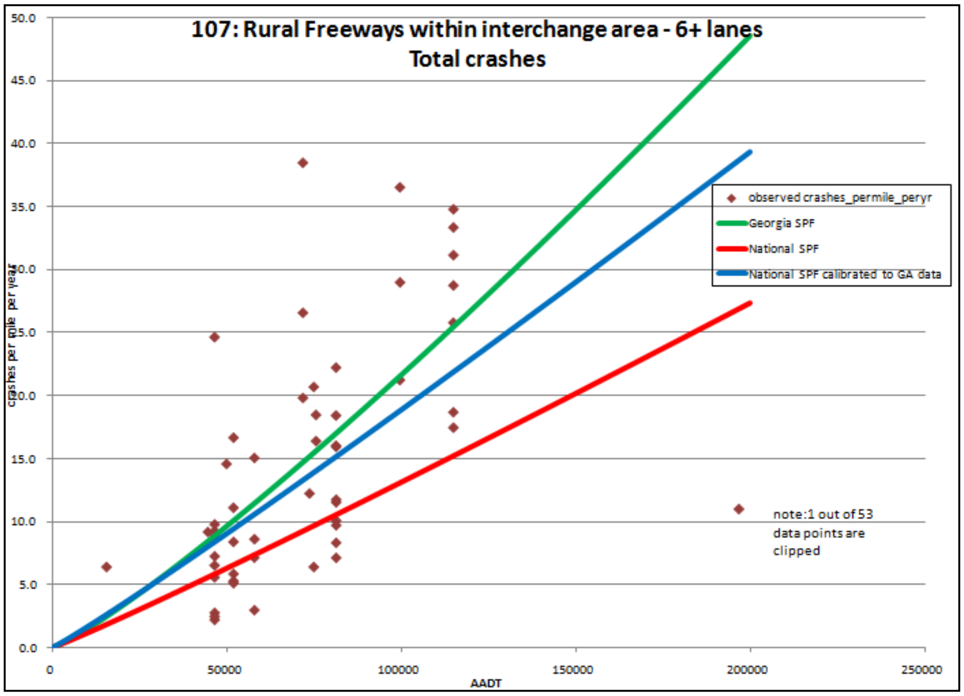

Figure 34: SPFs for site subtype 107 considering total crashes

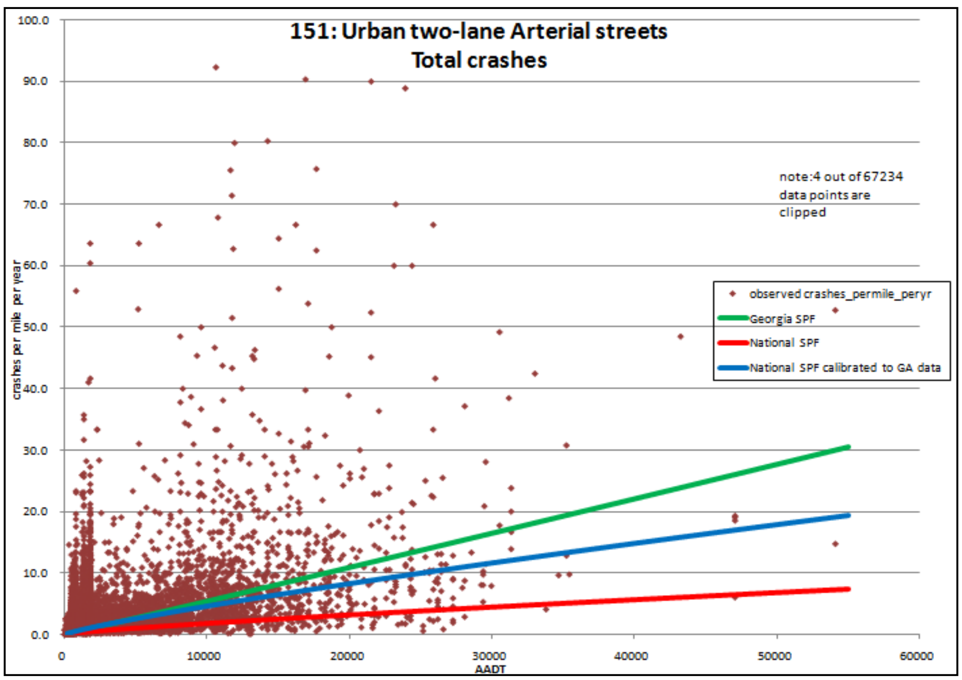

Figure 35: SPFs for site subtype 151 considering total crashes

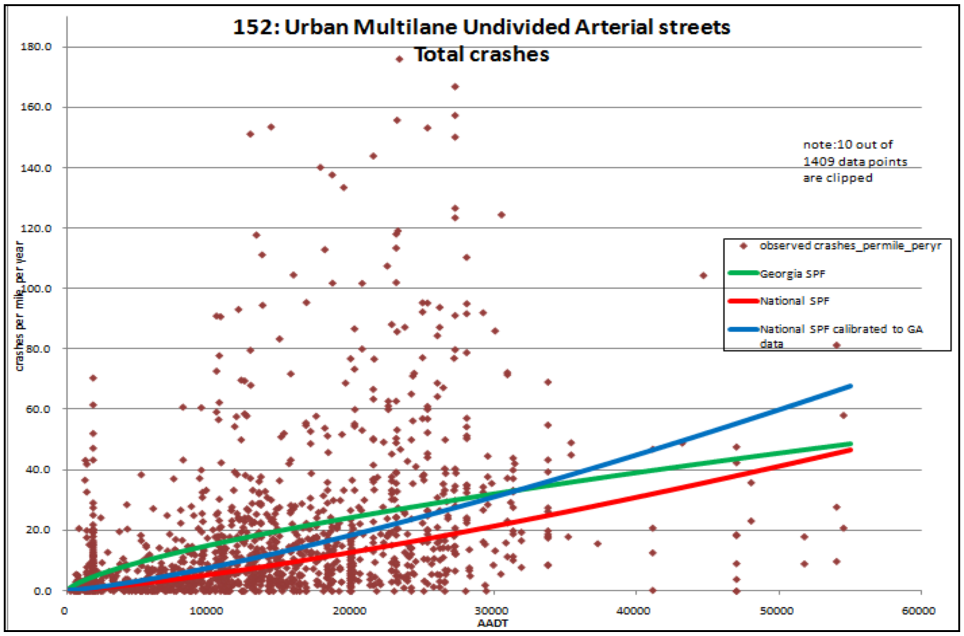

Figure 36: SPFs for site subtype 152 considering total crashes

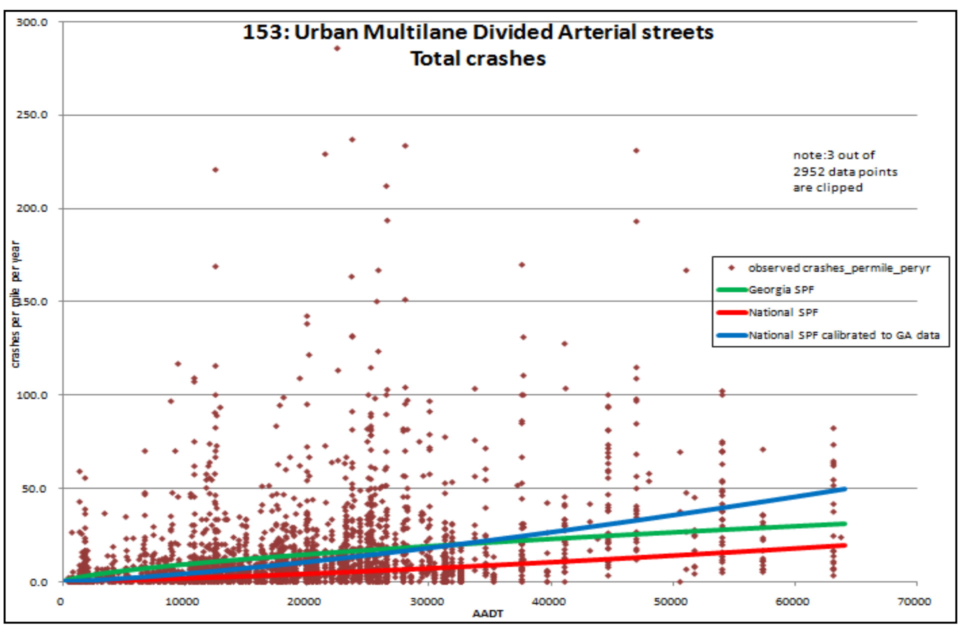

Figure 37: SPFs for site subtype 153 considering total crashes

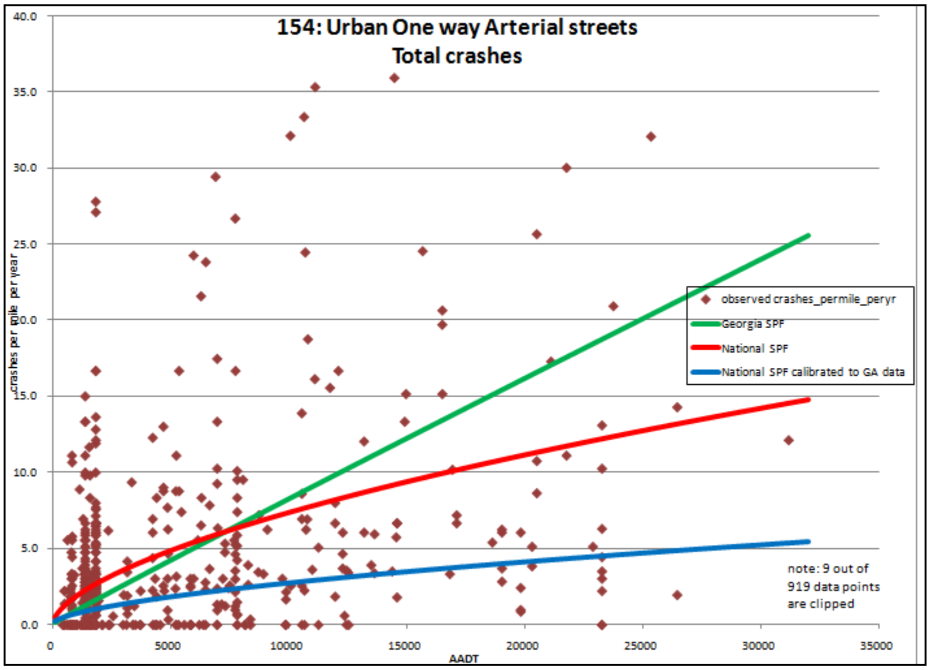

Figure 38: SPFs for site subtype 154 considering total crashes

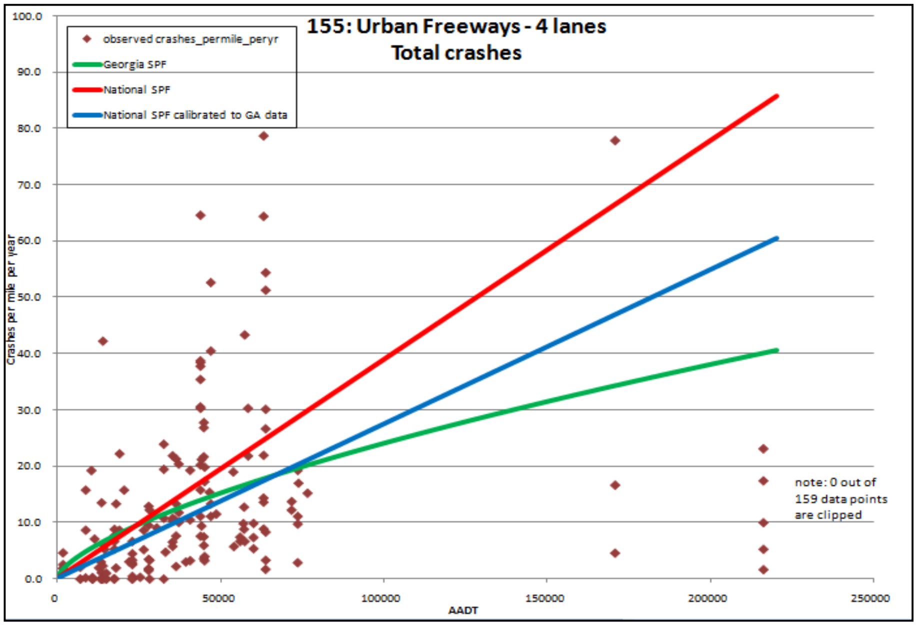

Figure 39: SPFs for site subtype 155 considering total crashes

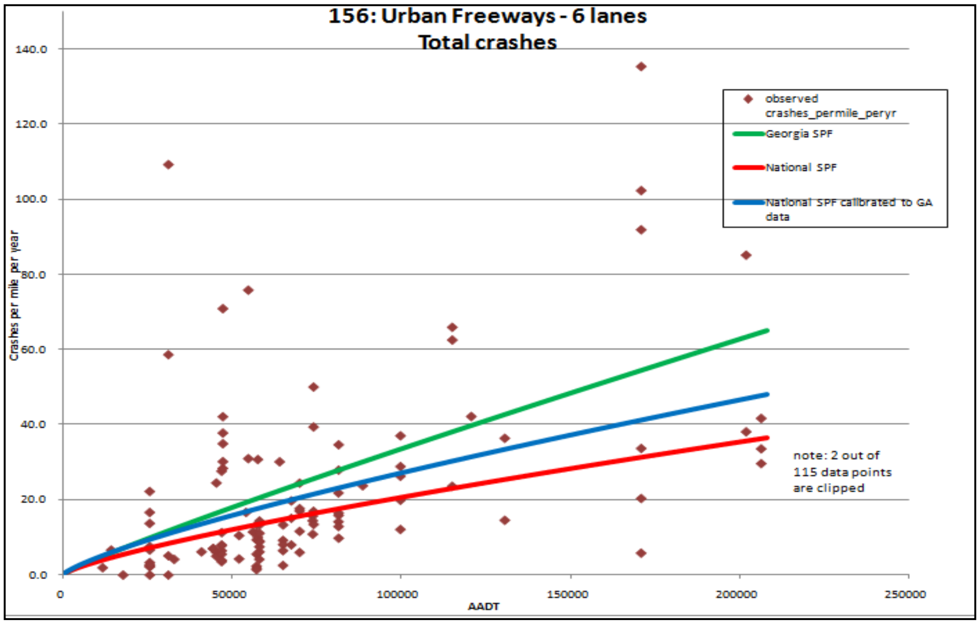

Figure 40: SPFs for site subtype 156 considering total crashes

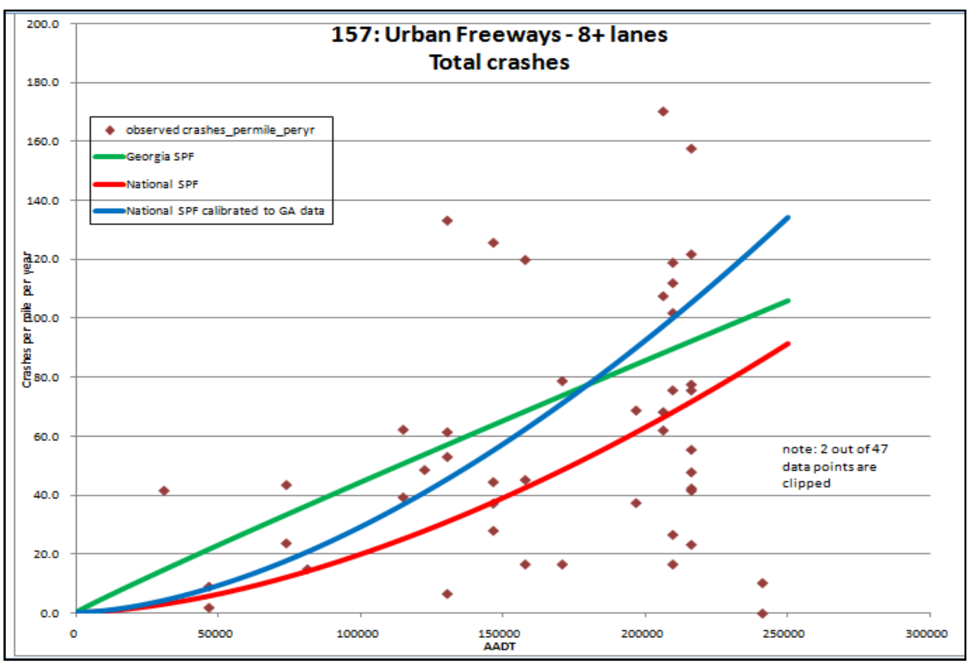

Figure 41: SPFs for site subtype 157 considering total crashes

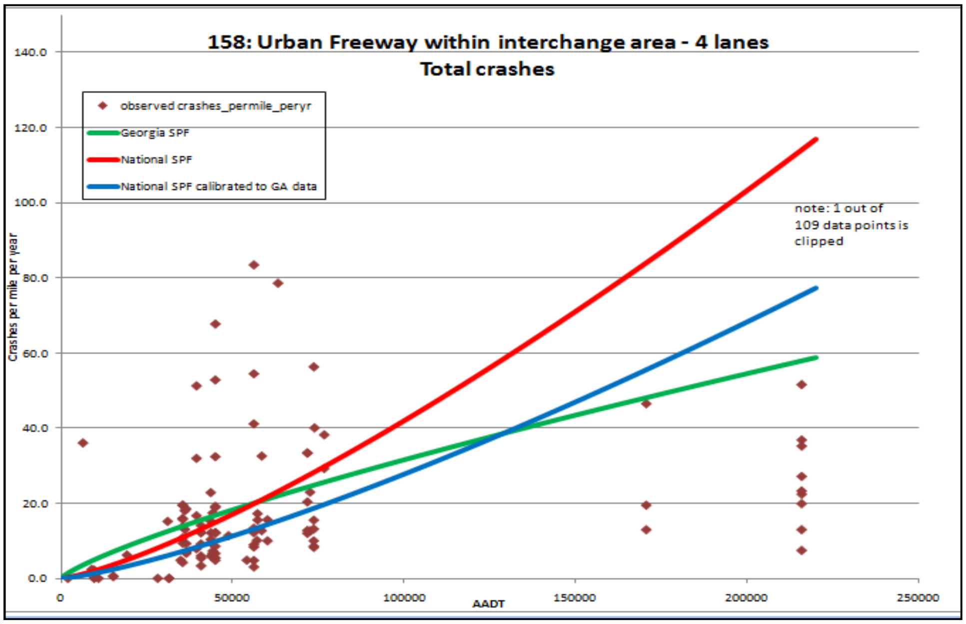

Figure 42: SPFs for site subtype 158 considering total crashes

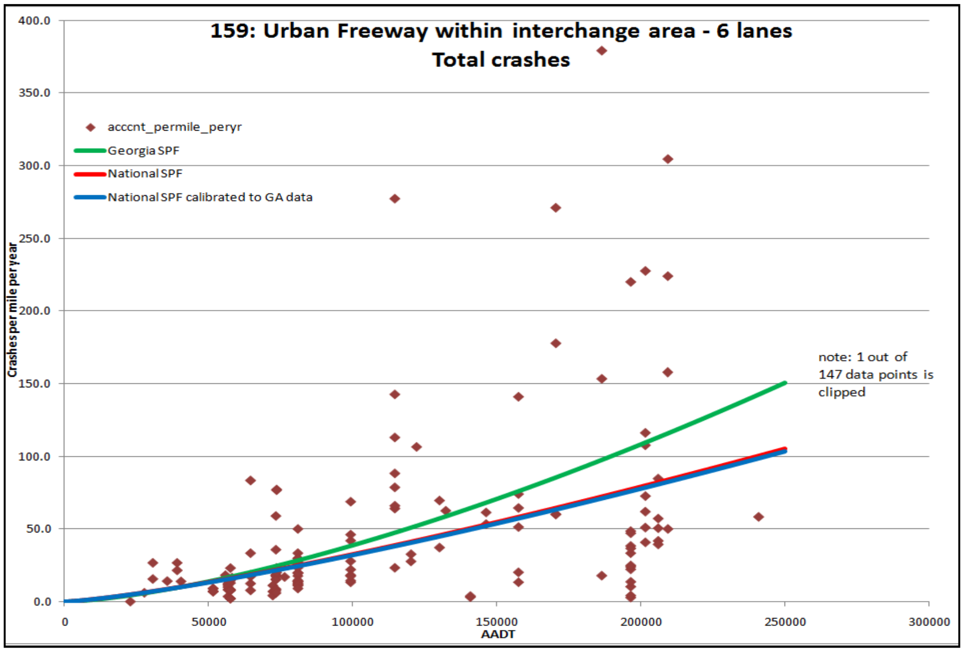

Figure 43: SPFs for site subtype 159 considering total crashes

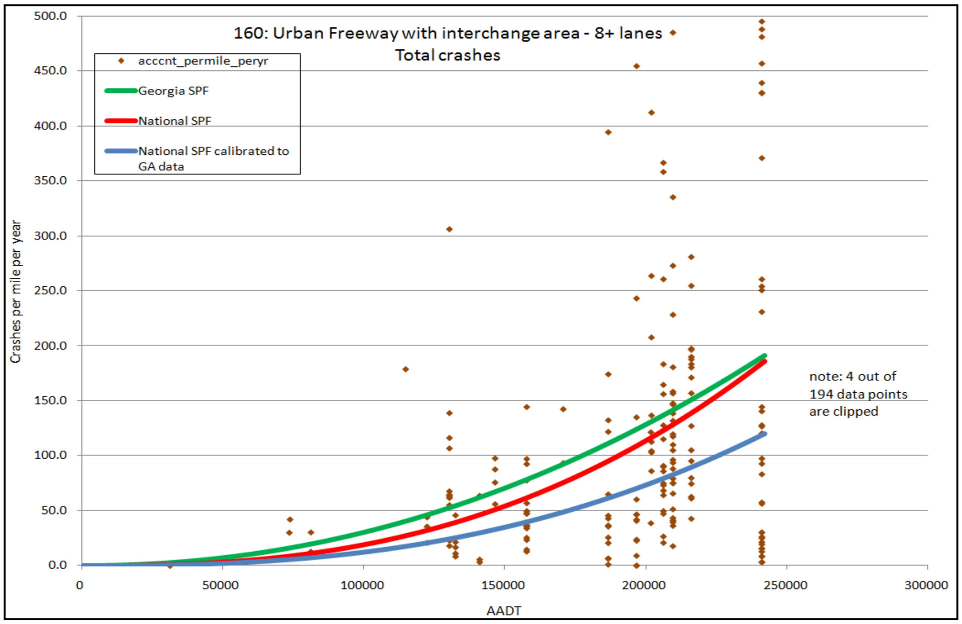

Figure 44: SPFs for site subtype 160 considering total crashes

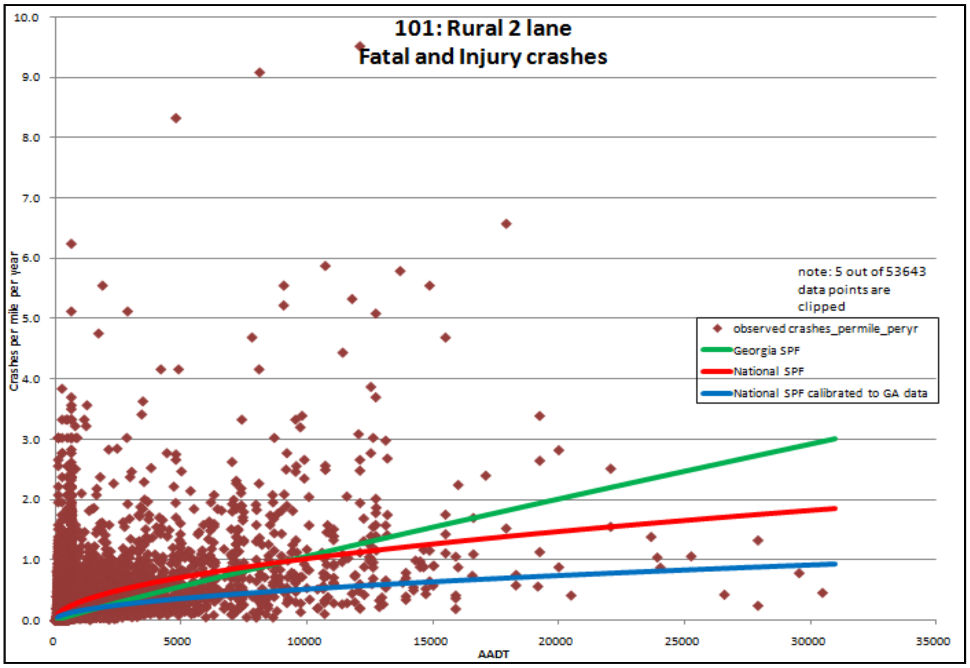

Figure 45: SPFs for site subtype 101 considering Fatal and Injury crashes

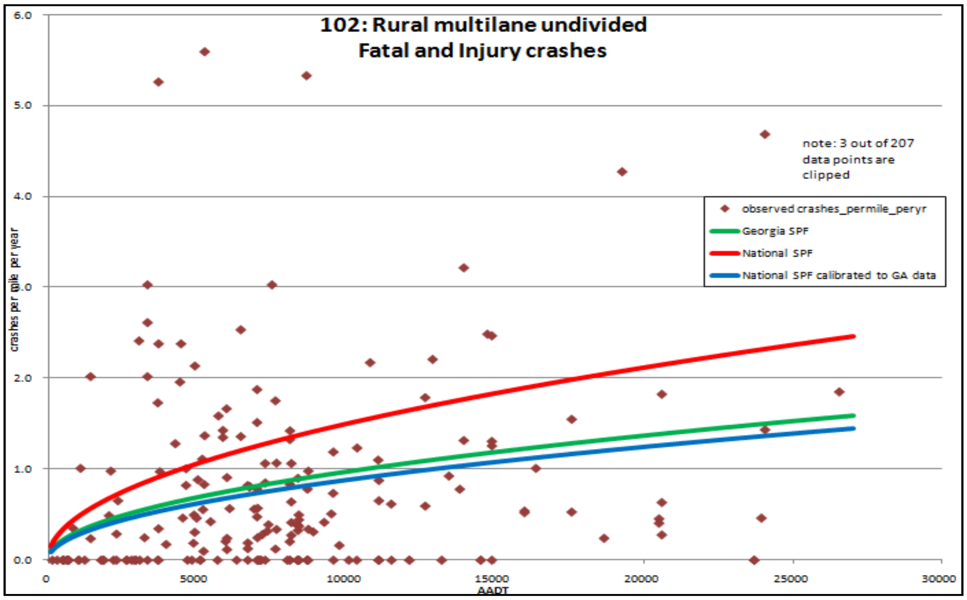

Figure 46: SPFs for site subtype 102 considering Fatal and Injury crashes

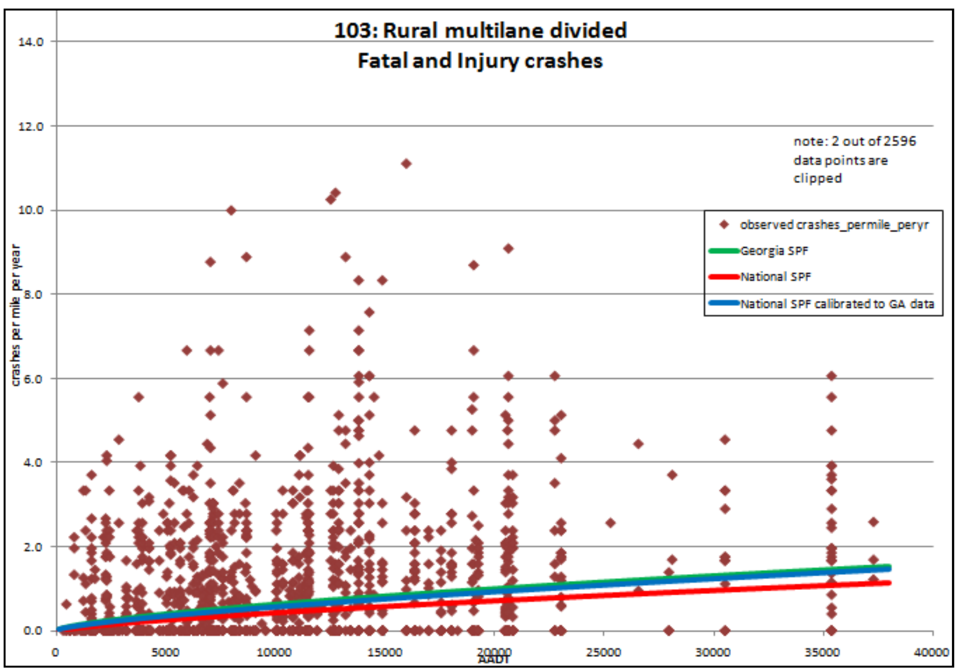

Figure 47: SPFs for site subtype 103 considering Fatal and Injury crashes

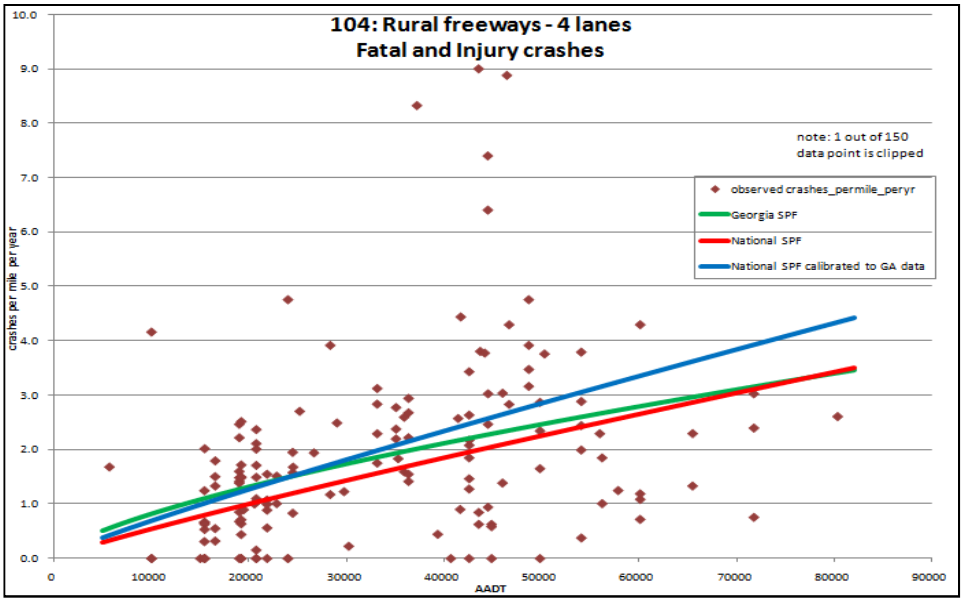

Figure 48: SPFs for site subtype 104 considering Fatal and Injury crashes

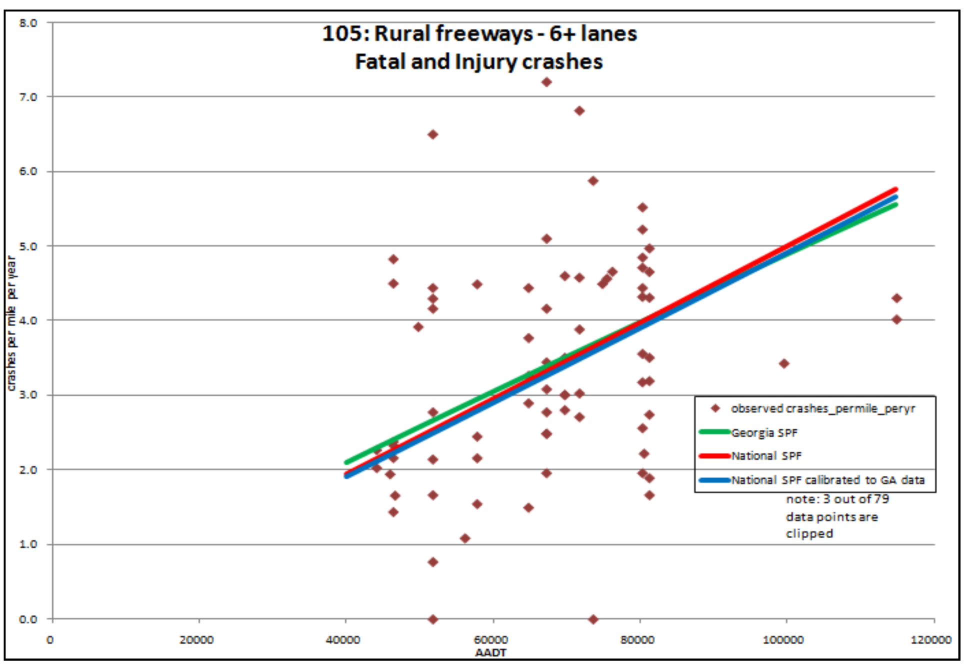

Figure 49: SPFs for site subtype 105 considering Fatal and Injury crashes

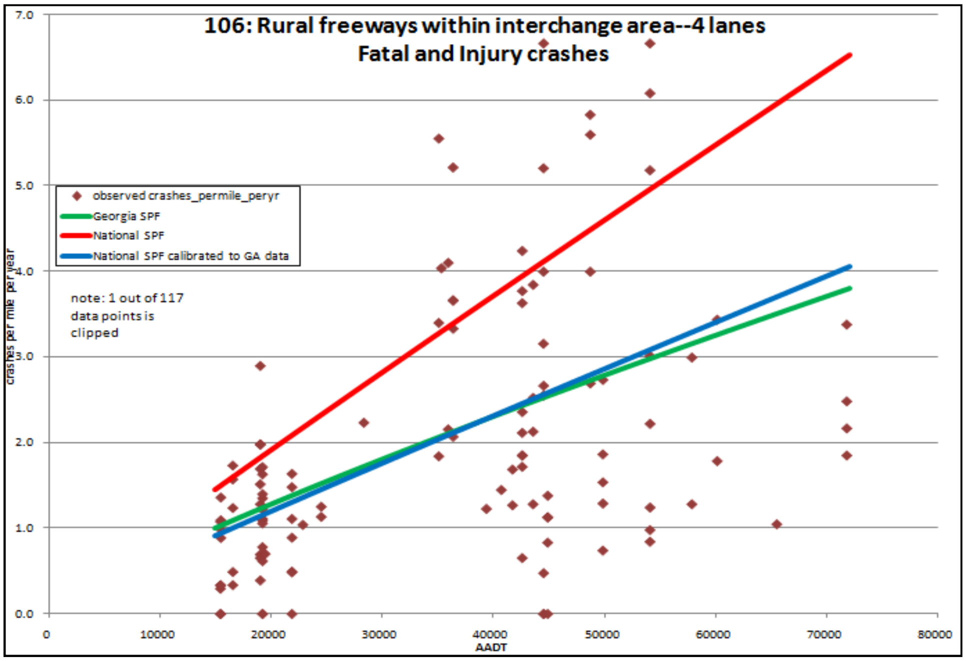

Figure 50: SPFs for site subtype 106 considering Fatal and Injury crashes

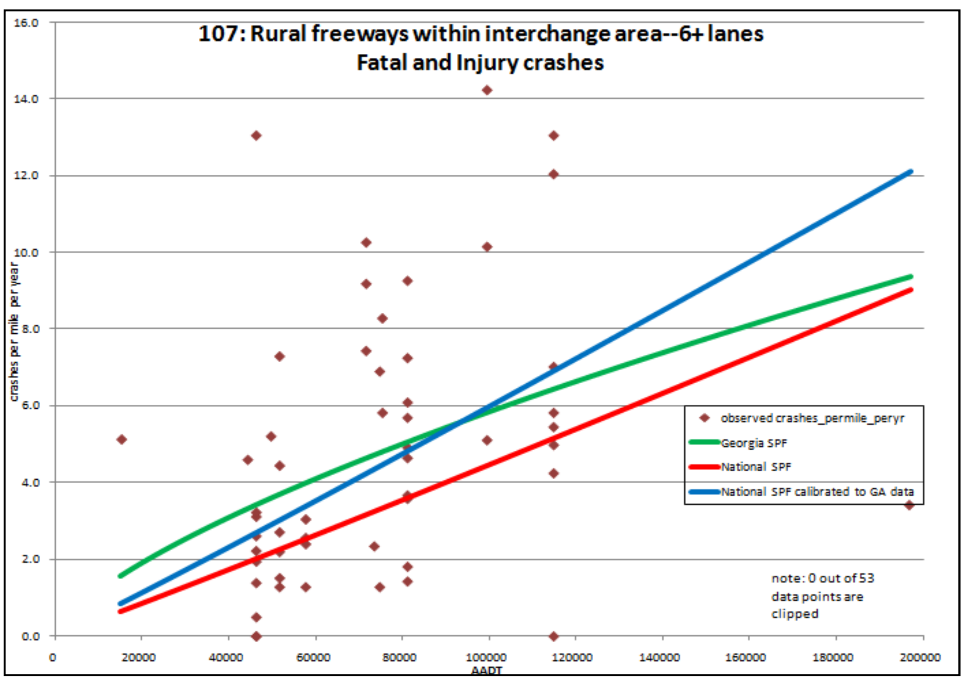

Figure 51: SPFs for site subtype 107 considering Fatal and Injury crashes

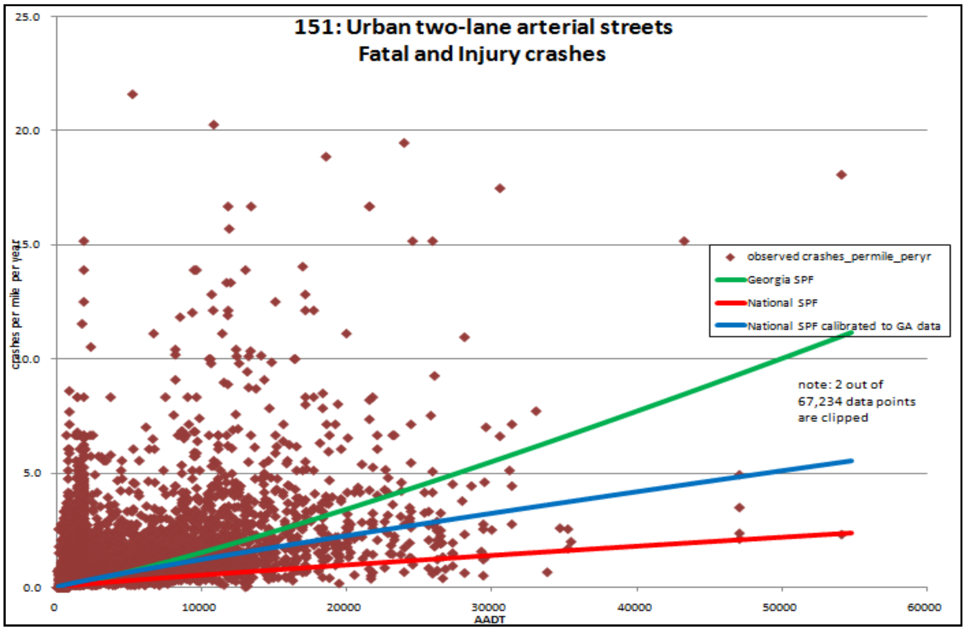

Figure 52: SPFs for site subtype 151 considering Fatal and Injury crashes

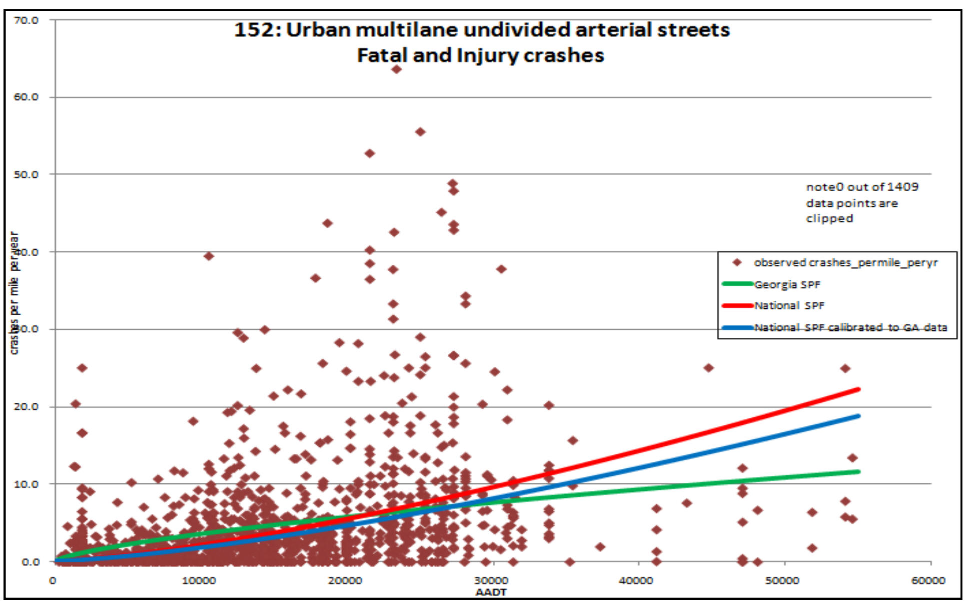

Figure 53: SPFs for site subtype 152 considering Fatal and Injury crashes

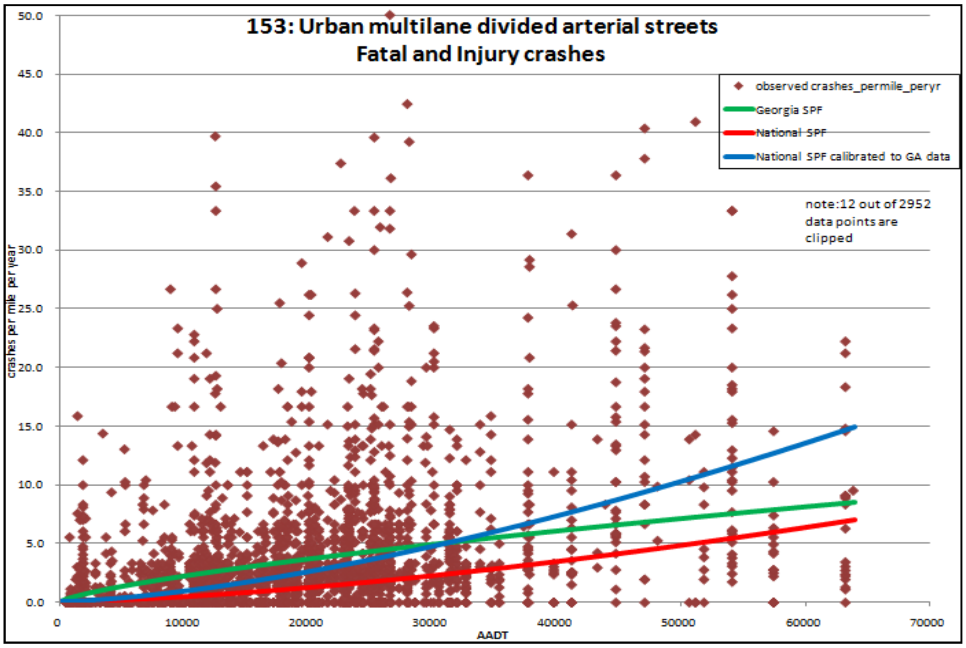

Figure 54: SPFs for site subtype 153 considering Fatal and Injury crashes

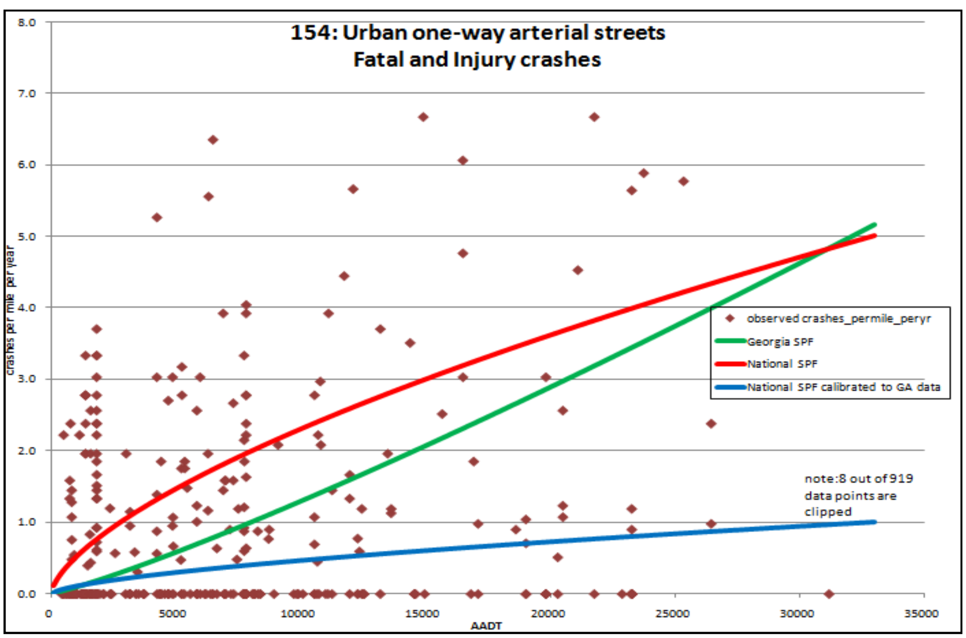

Figure 55: SPFs for site subtype 154 considering Fatal and Injury crashes

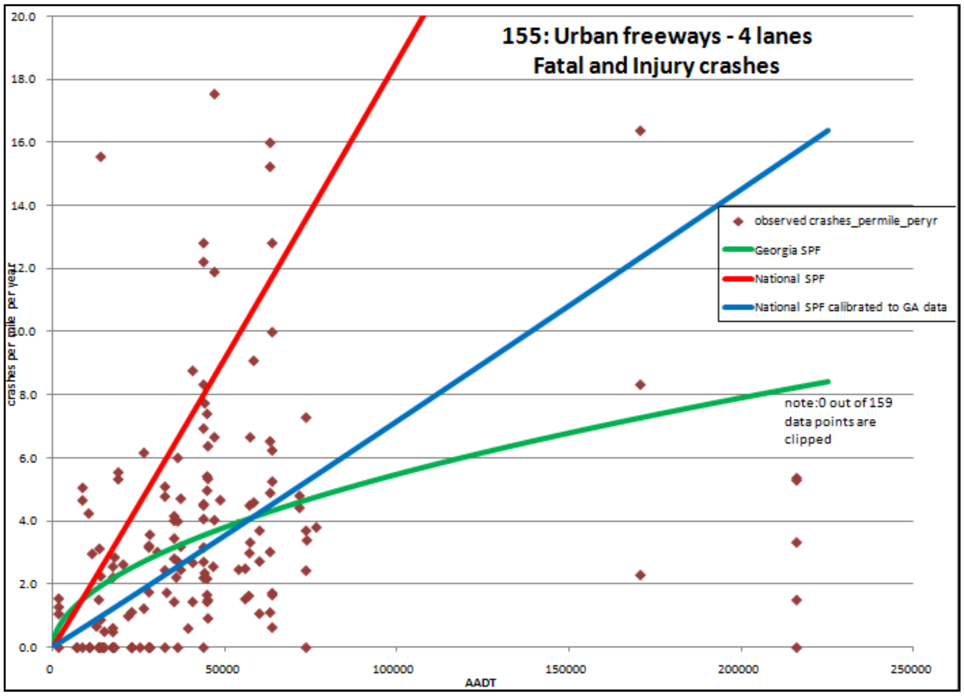

Figure 56: SPFs for site subtype 155 considering Fatal and Injury crashes

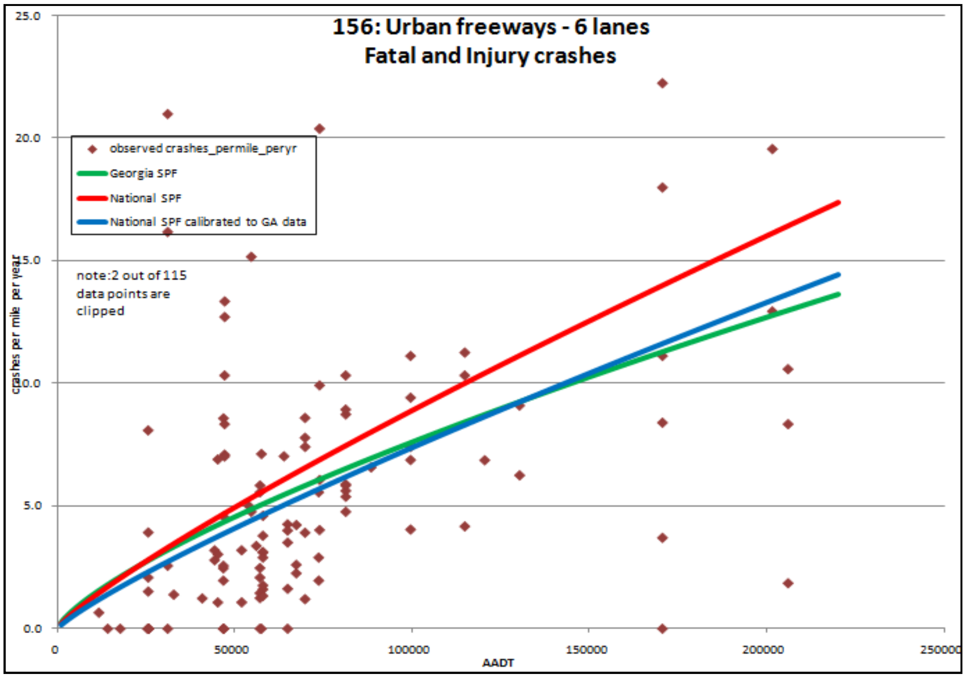

Figure 57: SPFs for site subtype 156 considering Fatal and Injury crashes

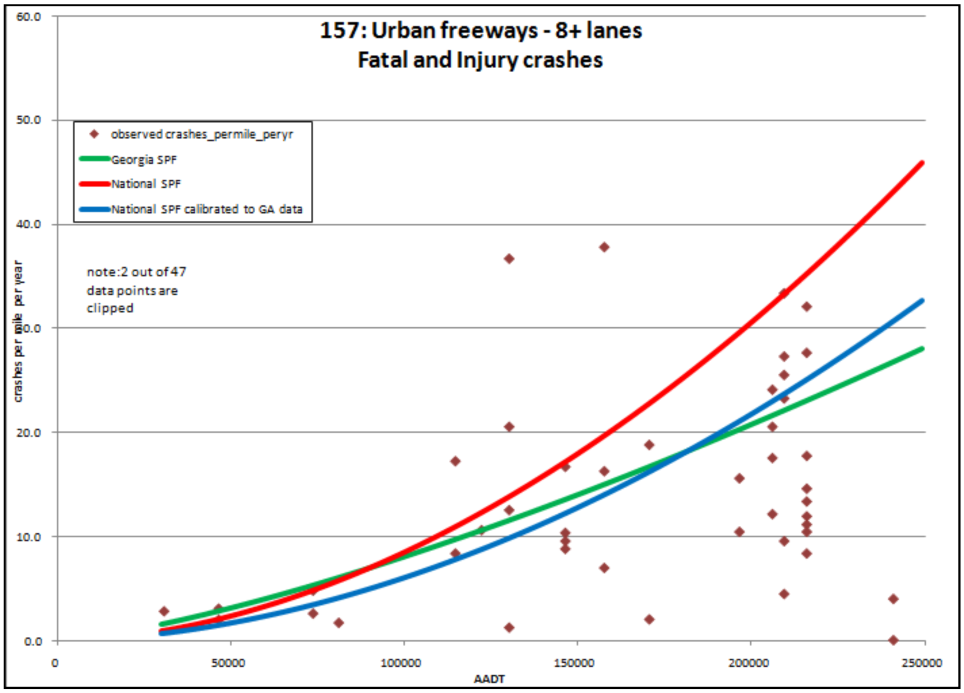

Figure 58: SPFs for site subtype 157 considering Fatal and Injury crashes

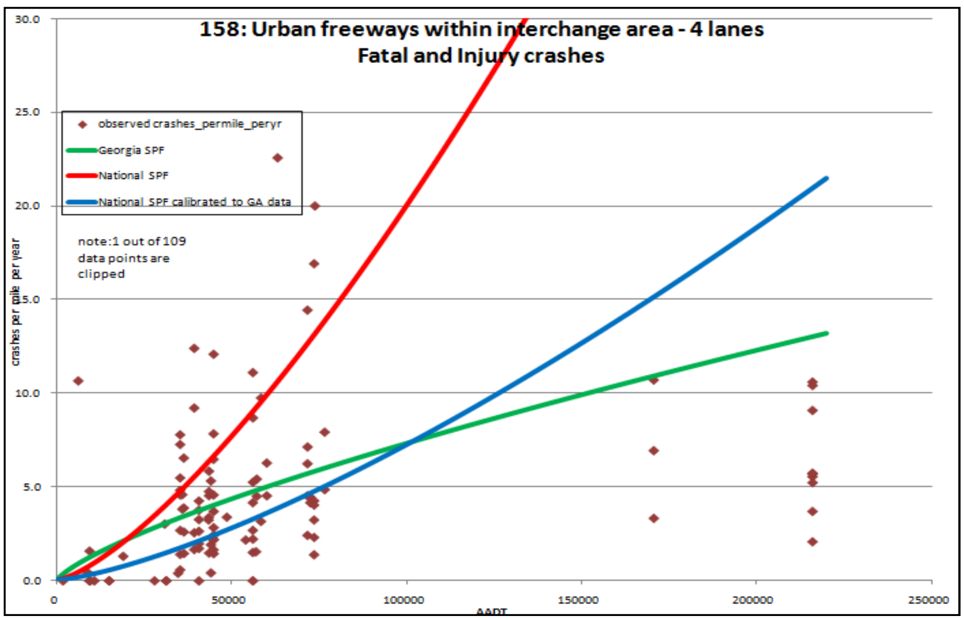

Figure 59: SPFs for site subtype 158 considering Fatal and Injury crashes

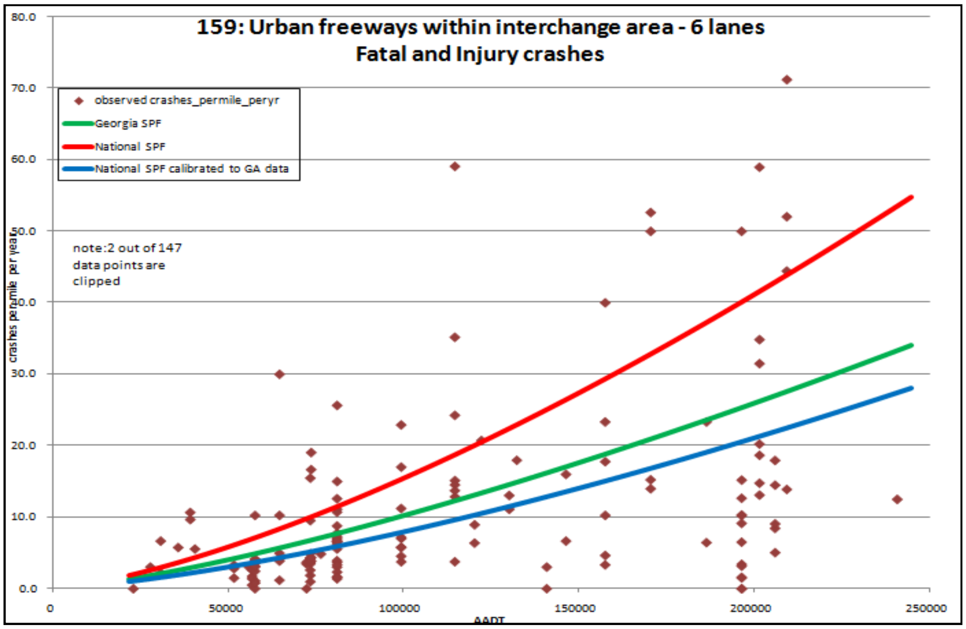

Figure 60: SPFs for site subtype 159 considering Fatal and Injury crashes

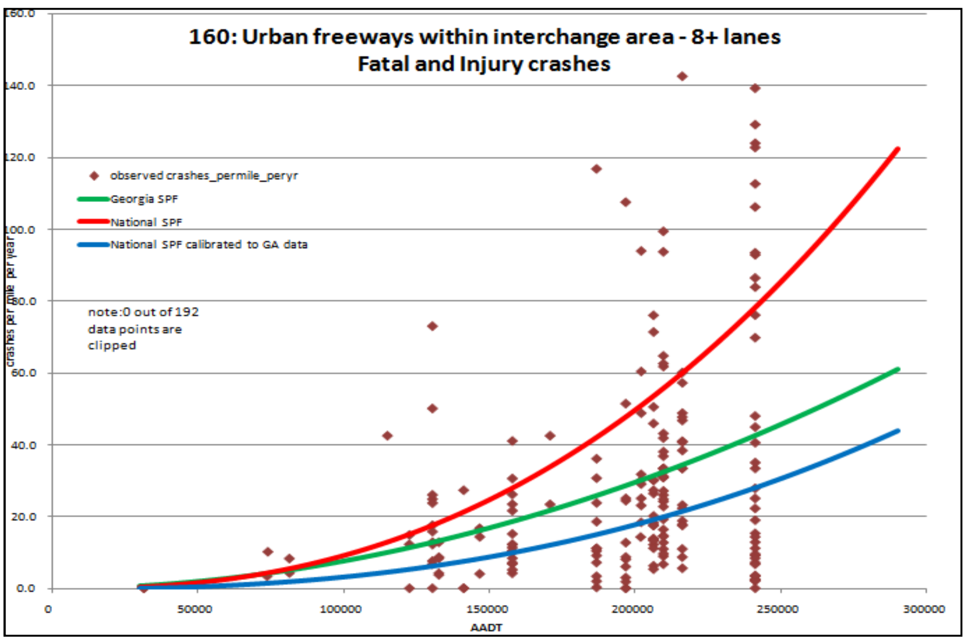

Figure 61: SPFs for site subtype 160 considering Fatal and Injury crashes

APPENDIX H: SURVEY ON ROAD SAFETY ANALYSIS METHODS, TOOLS, AND DATA

Survey on road safety analysis

Dear DOT Safety Director/ Staff Member,

With the passage of the SAFETEA-LU legislation in 2005, all states are required to prepare and adhere to a Strategic Highway Safety Plan (SHSP) that identifies data-driven approaches to prioritize and evaluate program outcomes in order to continue receiving federal money. Central to these plans are crash data analysis methods for identifying and prioritizing safety improvements. While most DOTs are still using traditional safety analysis measures such as frequency, rate, critical rate, or crash index, a few have indicated in their 5% reports that they are moving toward using new software packages and methods developed to overcome some of the biases and errors found in the traditional analysis approaches, and many others have plans for implementation.

In the past few years, several pooled fund studies and other federal funding mechanisms have been used to develop Interactive Highway Safety Design Model (IHSDM), SafetyAnalyst and Highway Safety Manual (HSM). These software and analysis tools are comparatively more advanced in statistical theory and level of accuracy, and have a tendency to be more data intensive. In this context, the researchers at the Clemson University are interested in determining availability of data and current road safety analysis practices in each state DOT. The research team has developed a survey to aid in capturing this information. The survey mainly helps the team to understand the various safety analysis methods used across states, knowledge of new safety analysis tools and the availability of data for use with newer methods. Ultimately this information can be used to help determine training needs and data gaps which may inhibit adoption of new safety analysis methods.

Survey on road safety analysis

The survey is split into 7 major parts:

1. Contact information
2. General questions about data
3. General questions about safety data analyses
4. Questions about SafetyAnalyst
5. Questions about Safety Performance Functions
6. Questions about SafetyAnalyst implementation
7. Questions about Highway Safety Manual Implementation

Questions in parts 4-7 on SafetyAnalyst and Highway Safety Manual are revealed only if the respondent indicates use of these tools in prior sections of the survey. If you have problems completing any question or section, we have tried to offer you a space to answer other. If you are not the right person for this question, please use other to give us contact information (name, email or phone) for someone who may could answer this question for you and we will follow up with them. For example, if you don't have information on traffic data, you may want to send us to someone in another department to retrieve that information. In addition, if you would prefer to complete the survey in paper/pen format, we would be glad to send a paper version to you, just respond to this email and let us know.

We will be happy to share our findings with you at the end of the study. So, please provide your complete contact information in the first section of the survey.

Survey on road safety analysis

We thank you in advance for your participation in this very important survey. If you have any questions, please feel free to contact me.

Following is the link to the survey.

http://www.surveymonkey.com/s/H53HF2W

Sincerely,

Priyanka Alluri

Doctoral candidate
Transportation Engineering
Department of Civil Engineering
18 Lowry Hall
Clemson University
Clemson, South Carolina,29634
(864)-650-7078, palluri@clemson.edu

Survey on road safety analysis

Page1: Contact Information

1. Name:______________________________
2. Title: ______________________________
3. Organization: ______________________________
4. Office/Department: ______________________________
5. Address: ______________________________
6. Contact phone number: ______________________________
7. Contact Email address: ______________________________
8. Best time to contact: ______________________________
9. Time zone where you are located:
   a. Eastern Time Zone
   b. Central Time Zone
   c. Mountain Time Zone
   d. Pacific Time Zone
   e. Other

Page2: General questions about data: Crash data

1. How many years of historical crash data do you use when conducting safety data analysis?
   a. Not sure
   b. 1 year
   c. 2 years
   d. 3 years
   e. 4 years
   f. 5 years
   g. 6 years
   h. 7 years
   i. 8 years
   j. 9 years
   k. 10 years
   l. > 10 years
   m. Other

2. How many years of historical crash data do you maintain in your database?
   a. Not sure
   b. 1 year
   c. 2 years
   d. 3 years
   e. 4 years
   f. 5 years
   g. 6 years

h. 7 years
i. 8 years
j. 9 years
k. 10 years
l. > 10 years
m. Other

3. On average, how many crashes occur in your state each year?
   ____
4. What percent of your crashes do you have specific location information for?
   ____
5. Can you identify crashes separately on:
   Segments yes no I don't know
   Intersections yes no I don't know
   Ramps yes no I don't know
   Comment: ___________________________

Page 3: General questions about data: roadway characteristics data:

1. How frequently do you update your roadway characteristics database?
   a. Every 6 months
   b. Every 1 year
   c. Every 2 years
   d. Every 3 years
   e. Every 4 years
   f. Every 5 years
   g. 5 years
   h. Continuously whenever there is a change
   i. Irregularly whenever there is a change
   j. Not sure
   k. Other (please specify): ___________________________
2. Do you maintain information about date of changes made to roadway characteristics file such that you could easily identify individual changes like addition of lane or addition of signing or markings?
   a. Yes
   b. No
   c. For some data elements
   d. Not sure
   e. Comment: ___________________________
3. What is the smallest segment length you typically use to record a change in the roadway characteristics?
   a. 0.01 miles
   b. 0.05 miles

c. 0.1 miles
d. 0.25 miles
e. Other (please specify) ______________________

4. Do you maintain specific dataset for intersection characteristics including traffic control and lane configuration?
   a. Yes
   b. No
   c. I don't know
   d. Other (please specify) ___________________________

5. Do you maintain specific dataset for ramps?
   a. Yes
   b. No
   c. I don't know
   d. Other (please specify) ___________________________

Page 4: General questions about data: traffic data

1. How many years of traffic data (adt) do you maintain in your traffic database?
   a. Not sure
   b. 1 year
   c. 2 years
   d. 3 years
   e. 4 years
   f. 5 years
   g. 6 years
   h. 7 years
   i. 8 years
   j. 9 years
   k. 10 years
   l. > 10 years
   m. Other (please specify) ___________________________

2. Do you have a comprehensive traffic database for the entire state?
   a. Yes
   b. No
   c. I don't know
   d. Other (please specify) ___________________________

3. For the types of roadways below, approximately what percent of roadway miles do you have ACTUAL ADT count data for (not estimated)

Interstates	< 25%	25%-50%	50%-75%	>75%
State Routes	< 25%	25%-50%	50%-75%	>75%
Secondary Routes	< 25%	25%-50%	50%-75%	>75%
County Routes	< 25%	25%-50%	50%-75%	>75%

City Routes	< 25%	25%-50%	50%-75%	>75%
Other	< 25%	25%-50%	50%-75%	>75%
Low Volume Routes	< 25%	25%-50%	50%-75%	>75%

Comment: ____________________________

4. Do you estimate ADT data for roads which are not actually counted?
   a. Yes
   b. No
   c. May be
   d. I don't know
   e. Comment ____________________________
5. Do you have a documented method for estimating ADT?
   a. Yes
   b. No
   c. May be

   If yes, would you be able to share it with us?
6. For each type of roadway below, approximately what percent of your system miles are covered by <u>actual + estimated</u> ADT values:

Interstates	< 25%	25%-50%	50%-75%	>75%
State Routes	< 25%	25%-50%	50%-75%	>75%
Secondary Routes	< 25%	25%-50%	50%-75%	>75%
County Routes	< 25%	25%-50%	50%-75%	>75%
City Routes	< 25%	25%-50%	50%-75%	>75%
Other	< 25%	25%-50%	50%-75%	>75%
Low Volume Routes	< 25%	25%-50%	50%-75%	>75%

Comment: ____________________________

Page5: General questions about safety data analyses

1. For your primary safety analysis, do you use contract services or perform analysis in house. If you use outside services, please elaborate
   a. We perform all of our safety analysis
   b. We do some of our own analysis and contract the remainder
   c. We outsource
   d. Not sure
   e. Other (please specify) ____________________________
2. Is your safety analysis centralized or decentralized or both (i.e. if sites and improvements are selected and funded at state level, it would be considered centralized vs district or other sub sections being able to select their own sites/ treatments) Choose all that apply
   a. Centralized
   b. Decentralized
   c. I don't know

Comment: ______________________________

3. How are the safety funds distributed in your state (centralized or decentralized or both) Choose all that apply
   a. Centralized
   b. Decentralized
   c. I don't know
   d. Comment: ______________________________
4. Before running analysis, do you classify the roadway sections? (choose all that apply)
   a. We run analysis on the complete state data as a whole
   b. We broadly classify the sections by a couple of variables (i.e. by area type and functional classification)
   c. We specifically classify segments using multiple variables (i.e. using area type, # of lanes, functional classification, median divide etc)
   d. Not sure
   e. Comment ______________________________
5. Before running analysis, do you classify the intersections? (choose all that apply)
   a. We run analysis on the complete state data as a whole
   b. We broadly classify the intersections by a couple of variables (i.e. by area type and functional classification)
   c. We specifically classify segments using multiple variables (i.e. using area type, # of approach lanes, functional classification, traffic control etc)
   d. Not sure
   e. Comment ______________________________
6. Have you changed your tools/ measures/ methods that you are using for safety analysis within the last 2-5 years?
   a. We have completely switched methods in the last 2-5 years
   b. We are using a combination of traditional methods as well as new methods adopted within the last 2-5 years
   c. We are using methods that were used 5 or more years ago
   d. Other: ______________________________
   e. Comment: ______________________________
7. If you have completely switched methods in the past 2-5 years, what tool(s)/ measure(s) did you use prior to your current tool(s)/ measure(s)? (choose all that apply)
   a. Crash Frequency: Sites are ranked based on the number of crashes that have occurred at the site
   b. Equivalent Property Damage Only (EPDO): Each crash is multiplied by a weight based on the crash severity (injury, fatality, or PDO). Weights relative to property damage only crashes are developed for injury and fatality crashes and applied to all severe crashes at a site. The weighted

sum of crashes determines the rank of the sites. The weights are usually the relative monetary value of crashes by severity

c. Relative Severity Index: Monetary values are assigned to each crash based on the crash type (e.g. rear-end, angle, sideswipe crashes). For each site, the monetary values for each crash type are multiplied by the number of crashes of that specific type that occurred at the site. After each crash has been multiplied by the appropriate monetary value based on the type of crash it is, the resulting monetary values are summed and used to rank the site.
d. Crash Rate: Combines crash frequency and vehicles exposed (e.g., total number of entering vehicles for intersections or million vehicle-miles traveled for sections). Sites are ranked depending on their calculated crash rate.
e. Rate Quality Control: Compares the observed crash rate at each site with a calculated critical crash rate unique to each site. Sites that have observed crash rates greater than their critical crash rate are identified for further analysis. The critical crash rate is calculated based on the average rate for sites with similar characteristics, the traffic volume at the site, and a statistical constant that represents the desired confidence level for estimating the critical crash rate.
f. Level of Service Of Safety – LOSS: This method uses SPFs to identify high crash sites. The LOSS for a site depends on the degree to which crash frequency and severity deviate from the mean for sites estimated by a SPF. Sites can be assigned to one of four levels, ranging from LOSS I to LOSS IV. LOSS I indicates a low potential for crash reduction and LOSS IV indicates a high potential for crash reduction.
g. High Proportion of Specific Crash Types: Ranks sites according to their probability of having a specific crash type in a proportion that is higher than a threshold value. The threshold value is either calculated from historic data or user specified depending on the data available to the analyst.
h. Rank Based on Expected Crashes: Uses empirical Bayes (EB) methodology to predict the expected number of crashes per year. Sites are ranked from highest to lowest expected number of crashes per year. This procedure essentially estimates a weighted average of a SPF prediction for similar sites and the crash history of the specific site.
i. Rank Based on Excess Expected Crashes: This method also uses EB methodology to predict an expected number of crashes per year at a particular site. The expected crash frequency is then compared to a crash frequency prediction from a SPF.

- j. CARE: (Critical Analysis Reporting Environment) is a data analysis software package designed for problem identification and countermeasure development purposes.
- k. SafetyAnalyst software: Provide state-of-the-art analytical tools for use in the decision-making process to identify and manage a systemwide program of site-specific improvements to enhance highway safety by cost-effective means.
- l. Highway Safety Manual: Provides tools to conduct quantitative safety analyses, allowing for safety to be quantitatively evaluated alongside other transportation performance measures
- m. Not sure
- n. Other (Please describe): ___________________________

8. What tools or measures do you currently use for selecting sites for safety improvements (choose all that apply)
   - a. Crash Frequency: Sites are ranked based on the number of crashes that have occurred at the site
   - b. Equivalent Property Damage Only (EPDO): Each crash is multiplied by a weight based on the crash severity (injury, fatality, or PDO). Weights relative to property damage only crashes are developed for injury and fatality crashes and applied to all severe crashes at a site. The weighted sum of crashes determines the rank of the sites. The weights are usually the relative monetary value of crashes by severity
   - c. Relative Severity Index: Monetary values are assigned to each crash based on the crash type (e.g. rear-end, angle, sideswipe crashes). For each site, the monetary values for each crash type are multiplied by the number of crashes of that specific type that occurred at the site. After each crash has been multiplied by the appropriate monetary value based on the type of crash it is, the resulting monetary values are summed and used to rank the site.
   - d. Crash Rate: Combines crash frequency and vehicles exposed (e.g., total number of entering vehicles for intersections or million vehicle-miles traveled for sections). Sites are ranked depending on their calculated crash rate.
   - e. Rate Quality Control: Compares the observed crash rate at each site with a calculated critical crash rate unique to each site. Sites that have observed crash rates greater than their critical crash rate are identified for further analysis. The critical crash rate is calculated based on the average rate for sites with similar characteristics, the traffic volume at the site, and a statistical constant that represents the desired confidence level for estimating the critical crash rate.
   - f. Level of Service Of Safety – LOSS: This method uses SPFs to identify high crash sites. The LOSS for a site depends on the degree to which

crash frequency and severity deviate from the mean for sites estimated by a SPF. Sites can be assigned to one of four levels, ranging from LOSS I to LOSS IV. LOSS I indicates a low potential for crash reduction and LOSS IV indicates a high potential for crash reduction.

g. High Proportion of Specific Crash Types: Ranks sites according to their probability of having a specific crash type in a proportion that is higher than a threshold value. The threshold value is either calculated from historic data or user specified depending on the data available to the analyst.
h. Rank Based on Expected Crashes: Uses empirical Bayes (EB) methodology to predict the expected number of crashes per year. Sites are ranked from highest to lowest expected number of crashes per year. This procedure essentially estimates a weighted average of a SPF prediction for similar sites and the crash history of the specific site.
i. Rank Based on Excess Expected Crashes: This method also uses EB methodology to predict an expected number of crashes per year at a particular site. The expected crash frequency is then compared to a crash frequency prediction from a SPF.
j. CARE: (Critical Analysis Reporting Environment) is a data analysis software package designed for problem identification and countermeasure development purposes.
k. SafetyAnalyst software: Provide state-of-the-art analytical tools for use in the decision-making process to identify and manage a systemwide program of site-specific improvements to enhance highway safety by cost-effective means.
l. Highway Safety Manual: Provides tools to conduct quantitative safety analyses, allowing for safety to be quantitatively evaluated alongside other transportation performance measures
m. Not sure
n. Other (Please describe): ____________________________

9. Do you have plans to use any of the new highway safety analysis tools (IHSDM, SafetyAnalyst, HSM)?
   a. Yes
   b. No
   c. I don’t know
   d. Comment: ____________________________
10. Are you currently working with any of the new highway safety analysis tools (IHSDM, SafetyAnalyst, HSM)?
   a. Yes
   b. No

Page6: General questions about SafetyAnalyst

Survey on road safety analysis

1. How long have you been working with SafetyAnalyst?
   a. < 6 months
   b. 6 months – 1 year
   c. 1-2 years
   d. > 2 years
2. What modules are you currently using in SafetyAnalyst? (choose all that apply)
   a. Network Screening
   b. Diagnosis and Countermeasure Selection
   c. Economic Appraisal and Priority Ranking
   d. Countermeasure Evaluation
3. What data have you successfully imported and post processed into SafetyAnalyst? (choose all that apply)
   a. Crashes
   b. Traffic
   c. Segments
   d. Intersections
   e. Ramps
4. Are you using the reports from SafetyAnalyst to select and/or prioritize safety improvements in your state? (Choose all that apply)
   a. Select sites
   b. Prioritize sites
   c. Not yet
   d. Comment: ____________________________
5. How easy is it for you to interpret the SafetyAnalyst reports?
   a. Extremely difficult
   b. Difficult
   c. Normal
   d. Easy
   e. Extremely easy

Page7: Questions about Safety Performance Functions

1. Are you using Safety Performance Functions (SPFs) provided within SafetyAnalyst?
   a. Yes
   b. For some site subtypes
   c. No, we have our own SPFs for our state
   d. Not sure
2. If yes for all or some subtypes, do you think that your state's data is well represented by the default SPFs from SA?
   a. Yes

   a. No
   b. For some site subtypes
   c. Other ______________________________

3. If you are currently using the default SPFs available with SafetyAnalyst, are you planning on developing your state specific SPFs?
   a. Yes
   b. No
   c. For some site subtypes
   d. Other ______________________________
4. Have you analyzed the fit of state specific SPFs (if you have state specific SPFs) versus default national SPFs?
   a. Yes
   b. No
   c. For some site subtypes
   d. Other ______________________________
5. How well did the calibrated SPFs from SafetyAnalyst match the ones developed specifically for your state?
   a. Not at all
   b. Matches to a little extent
   c. Matches to some extent
   d. Matches to a considerable extent
   e. Matches very well
   f. Other ______________________________
6. Would you be able to share the info about the SPFs of your state?
   a. Yes
   b. No
   c. May be
   d. Other ______________________________
7. How difficult was it to generate import files for SafetyAnalyst?
   a. Extremely difficult
   b. Difficult
   c. Normal
   d. Easy
   e. Extremely easy

Page8: Questions about SafetyAnalyst implementation

1. Approximately, how many man hours did it take for your agency to import data into SafetyAnalyst (ex: 6 man-months)?

   ____

2. How long do you think it will take for your agency to repeat the process of working with SA when updates are needed to base maps, roadway characteristics, or traffic data?
   ____
3. In general, what is the expertise of the people working with SafetyAnalyst and on this project? (Choose all that apply)
   a. High school
   b. Bachelor in Engineering/ Math/ Statistics/ IT
   c. Masters in Engineering/ Math/ Statistics/ IT
   d. PhD in Engineering/ Math/ Statistics
   e. Other (please specify) ______________________________
4. Did you receive many errors and warnings when you tried to import files into SafetyAnalyst for the first time?
   a. Yes
   b. No
   c. Some
   d. I don't know
   e. Other ______________________________
5. What is the nature of the errors and warnings? Were you able to fix them? Did you ignore them? Do you still have existing errors and warnings?
   ____
6. Do you allow SafetyAnalyst to generate homogeneous segments with your data?
   a. Yes
   b. No
   c. Sometimes
   d. I don't know
   e. Comment ______________________________
7. Have you bought or do you intend to buy the license for SafetyAnalyst from AASHTO and continue to use it for your safety analysis?
   a. Yes
   b. No
   c. I don't know
   d. Comment ______________________________
8. Would you recommend other states to work with SafetyAnalyst?
   a. Yes
   b. No
   c. I don't know
   d. Comment ______________________________
9. What do you think are the top 5 most difficult hurdles you faced in the whole process that deals with SafetyAnalyst in the order as "ONE" being the toughest hurdle?

____
____
____
____
____

10. What 3 tips would you have for other states planning to implement SafetyAnalyst?

    ____
    ____
    ____

11. Would you be willing to share examples of how you are using SafetyAnalyst within your state?
    a. Yes
    b. No
    c. May be
    d. I don't know
    e. Other _______________________________________
12. Have you received any funding from the federal government to implement new data needs, for example, collection and processing of data, to support the implementation of SafetyAnalyst?
    a. Yes
    b. No
    c. May be
    d. I don't know
    e. Other _______________________________________
13. Are you using SafetyAnalyst to generate priorities for SHSP?
    a. Yes
    b. No
    c. Some priorities
    d. May be
    e. I don't know
    f. Other _______________________________________

Page7: Questions about Highway Safety Manual Implementation (If the states are currently using/ planning to use HSM)

1. How many copies of manual have you received?

   ____
2. Who received these manuals?

   ____
3. Do you have an implementation plan for HSM deployment?

   ____

4. Do you have a specific person responsible for HSM implementation?

   ____

5. What time frame are you looking at for complete deployment?

   ____

6. Will you make a complete conversion to HSM or supplement current practices?

   ____

Page8: Thank you

Thank you for taking the survey

APPENDIX I: SEVEN ROADSIDE HAZARD RATING LEVELS

Clear zone greater than or equal to 30 ft sideslope flatter than 1V:4H, recoverable.

Figure 62: Typical roadway with RHR of 1

Clear zone between 20 and 25 ft; sideslope about 1V:4H, recoverable.

Figure 63: Typical roadway with RHR of 2

Clear zone about 10 ft; sideslope about 1V:3H, marginally recoverable.

Figure 64 : Typical roadway with RHR of 3

Clear zone between 5 and 10 ft; sideslope about 1V:3H or 1V:4H, marginally forgiving, increased chance of reportable roadside crash.

Figure 65: Typical roadway with RHR of 4

Clear zone between 5 and 10 ft; sideslope about 1V:3H, virtually non-recoverable.

Figure 66: Typical roadway with RHR of 5

Clear zone less than or equal to 5 ft; sideslope about 1V:2H, non-recoverable.

Figure 67: Typical roadway with RHR of 6

Clear zone less than or equal to 5 ft; sideslope about 1V:2H or steeper, non-recoverable with high likelihood of severe injuries from roadside crash.

Figure 68: Typical roadway with RHR of 7