7-2008

ASSESSMENT OF POTENTIAL SITE SELECTION METHODS FOR USE IN PRIORITIZING SAFETY IMPROVEMENTS ON GEORGIA ROADWAYS

Priyanka Alluri
Clemson University, palluri@clemson.edu

ASSESSMENT OF POTENTIAL SITE SELECTION METHODS FOR USE IN PRIORITIZING SAFETY IMPROVEMENTS ON GEORGIA ROADWAYS

A Thesis
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Master of Science
Civil Engineering

---

by
Priyanka Alluri
August 2008

Accepted by:
Dr. Jennifer H. Ogle, Committee Chair
Dr. Mashrur A. Chowdhury
Dr. Scott Brame

ABSTRACT

With over 40,000 people continuing to die on US roads each year, the US government has heightened the awareness of critical safety issues with the passage of SAFETEA – LU legislation in 2005. The plan requires each of the states to develop a Strategic Highway Safety Plan (SHSP) and incorporate *data-driven* approaches to prioritize and evaluate program outcomes; else federal funds will be redirected. Seeking to meet the new demands for data-driven approaches, many states are struggling to identify data collection/maintenance requirements for satisfying new approaches to highway safety analysis. Recent research has shown that selecting projects on the basis of crash frequencies and rates are misleading due to selection bias (such as greater emphasis on traffic volume and cash severity etc) and Regression-to-mean phenomena. There are several safety analysis techniques that are preferred over traditional rates and frequencies. These include level of service of safety, empirical bayes method using *SafetyAnalyst* software techniques. While all the above mentioned methods are macroscopic (giving a bigger picture of the complete road), microscopic analysis could be done using the Interactive Highway Safety Design Model (IHSDM). IHSDM is a set

of software analysis tools developed by Federal Highway Administration (FHWA) to evaluate safety on two lane rural highways.

This research aims at assessing the usability, data requirements, data availability and expertise required by different techniques that are deemed appropriate for safety analysis in Georgia. To streamline and reduce the scope of work, Cobb County was chosen as the analysis county because it had been used in a prior development effort and was expected to have the best level of completion and accuracy in the state. The procedure of using the state-of-the-art analytical tools is considered as the most comprehensive safety analysis method. Cobb County data set will be used to test the applicability of the four analysis methods: crash frequency, crash rate, critical crash rate and level of service of safety (LOSS). The results from various ranking criteria (crash frequency, crash rate, critical crash rate and LOSS) will be compared to the actual available crash data and enhanced *SafetyAnalyst* data.

SafetyAnalyst uses the Safety Performance Functions generated for northern states and it calibrated to Georgia data. SPFs applicable to Georgia data (generated from Cobb County) are compared to the non-calibrated and calibrated SPFs used in *SafetyAnalyst.* Analysis of costs and potential benefit of using various network screening methods is carried out to weigh the capabilities and limitations of various ranking methods.

DEDICATION

This thesis is dedicated to my mother, Lakshmi Alluri who inspired me to take this enlightening path. Her continuous support made this work possible.

ACKNOWLEDGMENTS

It is my pleasure to thank all the people who in one way or the other helped me in making this thesis possible. Primarily, its my privilege to thank my advisor Dr. Jennifer Harper Ogle for her expert advice and continuous encouragement. She made "crash data analysis" fun, thus, increasing my interest in the field of transportation safety.

I would like to express my sincere thanks to my committee member, Dr. Mashrur A. Chowdhury, for helping me with the benefit cost analysis. I would like to thank my committee member, Dr. Scott Brame, for getting me started on GIS and for helping me with GIS whenever I am clueless.

Its my pleasure to thank Dr. Wayne Sarasua for his friendly advice and for his invaluable help with GIS in the initial stages.

My thesis would have been incomplete without statistical analysis, and in this context, I would like to express my deepest appreciation to Dr. Julia Sharp from the Statistics Department for helping me in understanding the statistical concepts and in running SAS software. A very special thanks to Mrs. Barbara Ramirez, of the English Department, for her constructive criticism and for helping me in polishing the language in my thesis. On the same grounds, I would like to thank the students at Academic Writing Center for helping me with the grammar.

My sincere thanks go to GDOT (Georgia Department of Transportation) for funding this project. I would like to express my deepest appreciation to Mr. Norm Cressman from GDOT for his timely response to all my queries about crash data.

This project became very interesting and challenging with the use of software *SafetyAnalyst.* I would like 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 answering at least a hundred questions with a lot of patience. Their continuous and timely support made it possible to finish the data analysis in time. Also, I would like to thank Karen Richard from MRI who helped me in developing the Safety Performance Functions. Her suggestions are invaluable to this thesis. I would like to thank Angie Rios for helping me in understanding the crash database.

No project can be done in isolation. This thesis was only possible because of the help rendered by my project mates Swathi, Aimee and Chamanie. Their friendship and support are invaluable. A very special thanks to Johnatan for being my friend and for alleviating my tensions towards the end of writing this thesis. I would like to thank Jo Ann, Pranoy, Sukumar, Tabrez and Vijay for creating a friendly work environment in the department.

Sumalatha Yaski stood by my side during tough times. Its my pleasure to thank her for her friendship and support. A very special thanks to Bhanu Durga Paladugu for raising my confidence levels during difficult times. Thanks to Arpita Biswas for listening patiently to my "exciting developments in my research" everyday. I would like to thank Pavan Seemakurthy alias "Kaki" for his confidence in me, for lightening the situations when I am down and for giving me company in the department during my work. I would like to thank Sindhu Bobba for reviewing the document. I would also like to thank Anusha Pokuri for helping me with SAS.

Siva Chavali is a friend, advisor, guide, mentor, teacher and a critic to me. His continuous advice kept me on track and helped me write this thesis document. I would like to thank him for everything.

Finally, Its my pleasure to thank my best friend Aditya Pokuri for his support, guidance and inspiration. Without him, its near to impossible to pull things together to complete this thesis.

TABLE OF CONTENTS

Page

Table of Contents (Continued)

Page

LIST OF FIGURES

Figure Page

List of Figures (Continued)

Figure Page

Figure Page

LIST OF TABLES

Table Page

List of Tables (Continued)

Table Page

CHAPTER 1

INTRODUCTION

1.1 Introduction:

In 1990, the World Health Organization (WHO) identified the top 20 reasons for death; in the 9th place was road traffic crashes. On an average, 1.2 million lives are lost worldwide every year and 50 million people are injured annually. By the year 2020, it is predicted that traffic crashes will become the third cause of death from non-communicable diseases (WHO Summary report., 2007).

Traffic crashes are costing American motorists more than $160 billion each year considering property damage, travel delays, medical costs, and environmental degradation cost etc (Fox News., 2008) and nearly 117 people are dying each day on average on US roads. The severity of the situation is hence evident. Even though the statistics state that the total number of fatalities across the country has decreased from 52,627 in 1970 to 42,642 in the year 2006, road crashes are still one of the main reasons for death in the country (BTS., 2008).

Close inspection of the present transportation system in US reveals the many challenges confronting the transportation profession. Traffic management, highway safety improvement and environmental protection are some of the many major issues to be addressed. The introduction of SAFETEA – LU (*Safe Accountable Flexible Efficient Transportation Equity Act – A Legacy for Users*) in the year 2005 is a positive step in the direction to improve the country's current transportation system. The act emphasizes the following aspects: safety, equity, innovative finance, congestion relief, mobility and productivity, efficiency, environmental stewardship and environmental streamlining (Federal Highway Administration., 2008).

As the name implies, safety is the key focus of the act's overall program goals and objectives. The act requires states to develop Strategic Highway Safety Plans (SHSP) and comprehensive Highway Safety Improvement Programs (HSIP) to improve safety on highways. The states are required to submit SHSP by October 1st every year to receive safety funds. Beginning in the fiscal year 2006, HSIP authorizes federal funds to reduce traffic crashes, fatalities and serious injuries on all public roads (Federal Highway Administration., 2008). According to the Code of Federal Regulations, Title 23, Part 924, *"Each State is required to develop and implement, on a continuing basis, a Highway Safety Improvement Program (HSIP), which has* *the overall objective of reducing the number and severity of crashes and decreasing the potential for crashes on all highways."* (Epstein, et al., 2002)

According to SAFETEA-LU, all state DOTs are required to develop a Strategic Highway Safety Plan and implement Highway Safety Improvement Program emphasizing on safety improvements on highways and addressing the 4 E's (Engineering, Education, Enforcement and Emergency response) of highway safety to qualify for federal funding. It also requires the states to identify new and intense data driven approaches to crash data analysis, network screening and countermeasure selection and their evaluation.

The three main components of a Highway Safety Improvement Program that aid in achieving it's final goal are:

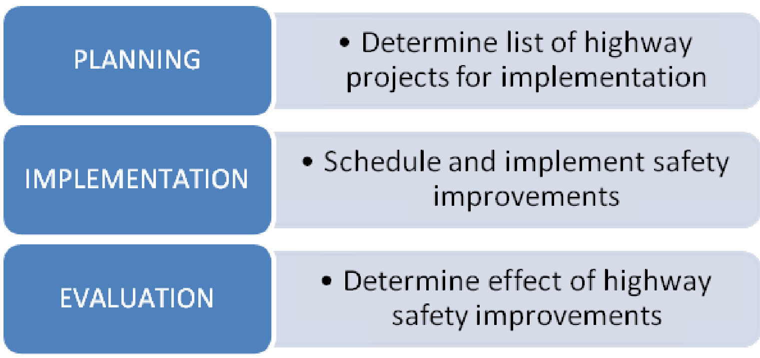


Figure 1: Components of Highway Safety Improvement Program (HSIP., 2007)

The planning phase includes collecting and maintaining data, identifying problematic locations (sites with potential for safety improvements), conducting engineering studies and establishing project priorities. The implementation phase includes scheduling projects, their design and construction and conducting operational review. The final phase, evaluation phase includes determining the effect of completed projects. SHSP must show the effectiveness of treatments through formal HSIP process. Thus, it is important to ensure proper selection of sites for countermeasure implementation (HSIP., 2007).

Newer approaches to crash data analysis and site safety improvements include the use of software like *SafetyAnalyst*, IHSDM (Interactive Highway Safety Design Module) and HSM (Highway Safety Manual). Different states have different approaches towards the highway safety problem with the bottom line of reducing the frequency and severity of crashes and improving safety. If sites are not chosen using proper methods, the effectiveness of the countermeasures will be reduced or eliminated.

For the state of Georgia, a *Strategic Highway Safety Plan* was prepared in October 2006 with a motto "Every Life counts - Strive for Zero deaths and injuries on Georgia Roads" and a goal of 1.0 fatalities per 100 million vehicle miles traveled by the year 2010 (Georgia SHSP., 2006). Comparison of the fatality trends in traffic crashes in the country and Georgia reveal the fact that since 2004, fatality rates have been above the national averages. The following figure shows the trends in Georgia and across the United States.

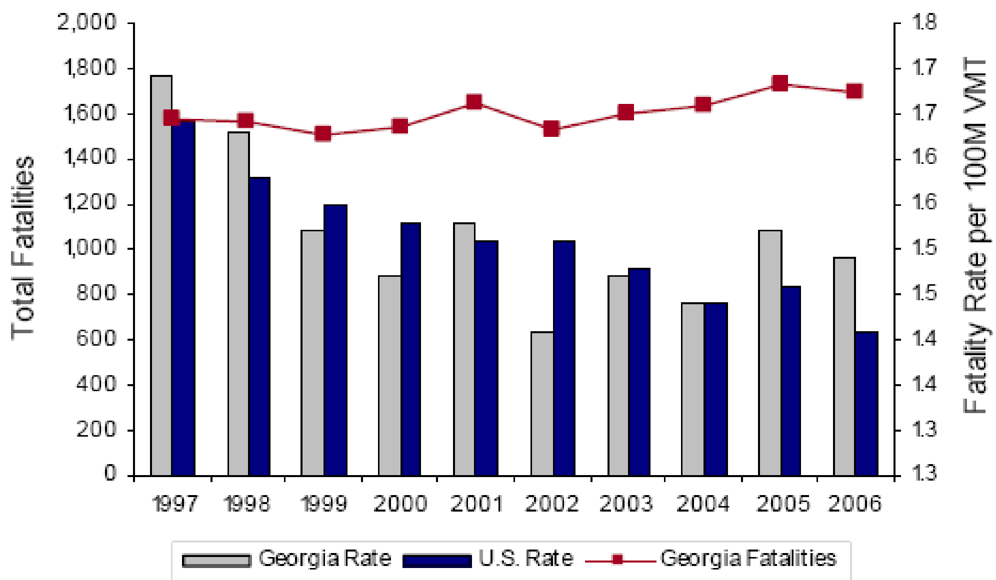


Figure 2: Traffic fatality trends in GA and US (TSP., 2008)

To continue to be eligible for safety improvement funding, Georgia (as with all states) must show continued improvement in the numbers with positive steps towards meeting their goals. To aid in this process, Georgia DOT sought the help of Clemson University to help identify appropriate data analysis techniques that will work with existing data and also to identify data needs to take advantage of new safety analysis methods.

With varying levels of available crash data, roadway characteristics, and traffic data, different states have developed different methods for analyzing crash data. Some of the most popular analytical methods include using crash frequencies and crash rates. But crash rate/crash frequency have major drawbacks like regression-to-mean effect and bias to high volume areas which can be rectified by rigorous analysis tools like the Empirical Bayes method. The notion of automation of such rigorous tools led to the creation of "*SafetyAnalyst*". *SafetyAnalyst* is a set of software tools used for highway safety management that integrates all parts of the Safety Management System. Georgia, being one among the 22 beta test states for *SafetyAnalyst* asked Clemson University to compare the traditional methods of network screening to the newer approaches.

1.2 Problem Statement:

For the state of Georgia, the total number of motor vehicle fatalities and fatal crash rates are above the national average and increasing. With limited resources, Georgia must make the best decisions about where to put its resources. For the crash data analysis and site selection, many different approaches are in practice today, some basic and some more advanced. Each approach has its own advantages and limitations. While many states are using the basic analysis methods like crash rates, crash frequencies and high proportion methods these have been shown recently to be subpar to their advanced counterparts. The Georgia Department of Transportation (GDOT) is interested in assessing new data-driven approaches for site identification and prioritization with the currently available data resources. The different approaches GDOT is interested in comparing include crash frequency, crash rate, critical crash rate, LOSS and Empirical Bayes using *SafetyAnalyst.* GDOT is concerned about the data requirements and the benefits and costs for adopting each of the above mentioned methods. In addition, there is concern that the base models (safety performance functions) included in *SafetyAnalyst* are not appropriate for Georgia because they were developed primarily for northern states.

1.3 Objectives:

Given the aforementioned needs and requirements of GDOT, the objectives of this study are:

- Review data availability, format and completeness for use in different safety data analysis methods

- Assess whether safety performance functions employed in *SafetyAnalyst* software can be properly calibrated to reflect crash distribution and conditions in Georgia
- Analyze costs and potential benefits of implementing and maintaining various methods (crash frequency, crash rate, Level Of Service of Safety and Empirical Bayes method using *SafetyAnalyst*) for selecting and prioritizing problematic crash sites by implementing these methods for Cobb County using 2004-2006 crash data.

1.4 Organization of the Thesis:

The remaining thesis describes the work completed to meet the objectives of the research. Chapter 2 provides with a brief literature review related to various network screening criteria (including crash frequencies, crash rates, generation of Safety Performance functions and *SafetyAnalyst*) and the problems, benefits and issues related to each method. Chapter 3 discusses the approach and methodology dealing with crash data analysis, generating files to be imported into *SafetyAnalyst*, generating SPFs applicable to Cobb County and benefit cost analysis of various network screening methods. Chapter 4 presents various problems and issues identified with the crash data. It also discusses the output from *SafetyAnalyst* and the identification of sites with potential for safety improvement (PSI) using various ranking criteria. Results from the comparison between the SPFs generated for Cobb County, and the non-calibrated and calibrated SPFs obtained from *SafetyAnalyst* are explained in this chapter along with the costs and potential benefits for using various network screening methods. Chapter 5 summarizes conclusions of this thesis and provides recommendations for future research.

CHAPTER 2

LITERATURE REVIEW

Transportation Safety can be defined as a transportation system lacking motor vehicle crashes and the losses resulting from property damage, injuries and fatalities (Hauer, E., 2000). Motor vehicle crashes involve a sudden collision between a vehicle and another vehicle(s) or a living or a non-living object. Individual crashes are random, unpredictable and can be difficult to evaluate. For some time, national resource constraints for safety improvements have made it impractical to implement countermeasures on all existing roadways in the country. Thus, safety analysts have, over many years, developed numerous methods for selecting intersections and road segments, referred to as sites, for further analysis and improvements. These sites should represent the shortest segments of road sections with a given set of homogeneous characteristics, at which the estimate of the expected accident frequency is largest while the coefficient of variation is smaller than a specified limiting value (Hauer, et al., 2002). The process of identifying these sites to obtain the most cost-effective solutions to safety problems is a science in itself, and also the first step in the highway safety improvement process (Hauer, et al., 2002, Hauer, et al., 2004). This process involves a multifold approach consisting of site identification, detailed engineering survey, treatments selection and prioritization. Of all the afore mentioned steps, identification of sites is the most fundamental and crucial step, since the improper identification of high priority sites result in less cost-effective solutions (Hauer, et al., 2002). Hence, site identification must be conducted with specific objectives in mind. Sites can be selected within a region, across a state, by functional classification of roads or crash types or by particular safety issues. According to Hauer (1996), the objectives of site selection should include economic efficiency, professional and institutional responsibility and fairness. These objectives help in identifying and prioritizing sites where countermeasures would prove cost effective, where engineering at the site is defective and where sites are deteriorated due to usage and where sites are unacceptably hazardous to the users.

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, et al., 2002, Hauer, et al., 2004). Several of these terms have been defined as follows:

- "Black Spot" is the general term used to recognize a hazardous location based on accident frequency and crash rates. These are the sections of roadway that are designated as being accident prone (Mandloi, et al., 2003).
- "High Crash Locations (HCLs)" are the areas that would potentially receive the largest benefit if safety funds were allocated (Pulugurtha, et al., 2003).
- "Hazardous Locations" are the sites having a potential for accident reduction based on crash frequency (Kononov, J., 2002).
- "Sites With Promise (SWiP)" are the sites in which safety can be improved cost-effectively based on Empirical Bayes methods and using Safety Performance Functions (Hauer, et al., 2004).

All the terms defined are very similar and could be used to identify problematic sites, but the underlying screening criteria are very different.

An ideal screening criterion is the one where the actual deviant sites and the sites selected for closer inspection overlap exactly (Hauer, E. and Persaud B.N., 1984). The more stringent the criteria of site selection, the more difficult it is to identify sites and the smaller are the number of inferior sites captured by the screening method. Recently, many problems have been identified with screening methods that are widely used by Departments of Transportation (DOTs) to rank problematic sites.

One of the biggest problems is with the use of just three years of crash data to identify problematic sites. At a particular site, crashes are random and it takes numerous years (for example 10 years) of crash data to identify a true average number of crashes. A mere three years of crash data is insufficient in most cases to identify problematic sites. However, using a larger number of years may have its own problems, over time, roads change, and older records may not reflect the current traffic and geometric situation (Hauer, E. and Persaud, B.N., 1984). In this situation, the data for the prior condition cannot be compared with current. The trends in the crash database reflect the changes in the factors (daily traffic, population changes) that affect the accident frequency and crash severity (vehicle fleet characteristics, speed trends). Using fewer years of crash data and a relatively fewer number of crashes for analysis results in a greater probability of error. For this reason, a practically feasible amount of crash data (5-10 years) needs to be considered for further analysis.

2.1 Basic Site Selection Criteria and Issues:

Based on the accident history and crash data, sites with potential for safety improvement can be ranked using many basic site selection criteria, such as crash frequency, crash rates, excess crash frequencies and excess crash rates, or by another criteria called "target crashes," which consider the crashes that can be affected by the proposed countermeasure (Hauer, et al., 2004). In all the screening methods, crash frequency and crash rates (or some index based thereon) are most widely used as ranking criterion (Hauer, E., 1996, Hauer, et al., 2002, Hauer, et al., 2004). Newer approaches involve more advanced statistical methods and sites are categorized as Sites With Promise (SWiP) if their long term accident record is within a multiple of a standard deviation from the normal value, which is calculated by examining similar sites within the required confidence interval (Hauer, E. and Persaud, B.N., 1984, Higle, J.L. and Witkowski, J.M., 1988). The following table briefly summarizes different site selection criteria and gives a brief description of their advantages, limitations and data requirements.

Table 1: Considerations for each selection method (HSM., 20008)

	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	Yes	Yes	No	Yes	No	No
EPDO	Yes	Yes	No	Yes	No	No
Rate	Yes	Yes	No	No	No	No
Rate Quality Control	Yes	Yes	No	No	No	No
LOSS	Yes	N/A	No	Yes	Yes	Yes
Category 2 - Screening Based on Proportions						
High Proportion of Crashes	Yes	Yes	No	Yes	No	No
Category 3 - Screening Based on Potential for Safety Improvement						
SafetyAnalyst (EB Method)	Yes	Yes	Yes	Yes	Yes	Yes

As summarized in the above table, the basic ranking criteria have many limitations. Regression-to-mean effect is not accounted for by most of the selection criteria except for *SafetyAnalyst.* Rate and rate quality method assumes a linear relationship between crashes and exposure while the relation is non-linear. This limitation is accounted for in most of the other ranking methods like crash frequency, LOSS, High proportion of crashes and *SafetyAnalyst*. Expected performance is predicted only by LOSS and *SafetyAnalyst* and both the methods need Safety Performance Functions

(SPFs). LOSS does not consider the severity of crashes while identifying problematic sites which is its major limitation. *SafetyAnalyst* uses Empirical Bayes method for predicting the expected performance which weighs the severity of a crash. Hence, *SafetyAnalyst* accounts for most of the limitations of other rankling criteria.

Ranking the problematic sites is based on an unwritten rule referred to as the Most Bang for the Buck (MBB) theory. According to Hauer (Hauer, et al., 2002), this principle emphasizes that "the money should go to where it achieves the greatest safety effect." It implies that spending money is not justified at a site where one accident can be eliminated when the same amount can eliminate several similar accidents at another site. According to this theory, network screening will tend to divert attention to sites at which the accident reduction potential is greatest. When crash frequencies are considered for site selection, accident reduction potential will be greater for sites with higher crash frequency (crashes per year). It is obvious that the crash frequencies will be comparatively higher for sites with heavier traffic such as urban roads and interstates. Thus being a biased estimate, crash frequency is not the best ranking criterion that could be used.

Further, ranking based on accident rates has its own disadvantages. "Rate measures the risk road users face while driving on specific roads" (Hauer, E., 1996). Crash rate is defined as the number of crashes per unit exposure. When proper random variables like average annual daily traffic, length of segment, lane width, shoulder width, median type etc for determining rates are not selected, crash rates appear to be misleading (Hauer, et al., 2002). Crash rates assume a linear relationship between exposure and crash frequency, but in most cases the actual relationship is non-linear (iTRANS and Human Factors North Inc., 2003). Due to this incorrect assumption, crash rates tend to identify sites that have lower traffic volumes. When traffic volumes are very low, any crash on the segment will produce a large rate. In addition, segment 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.

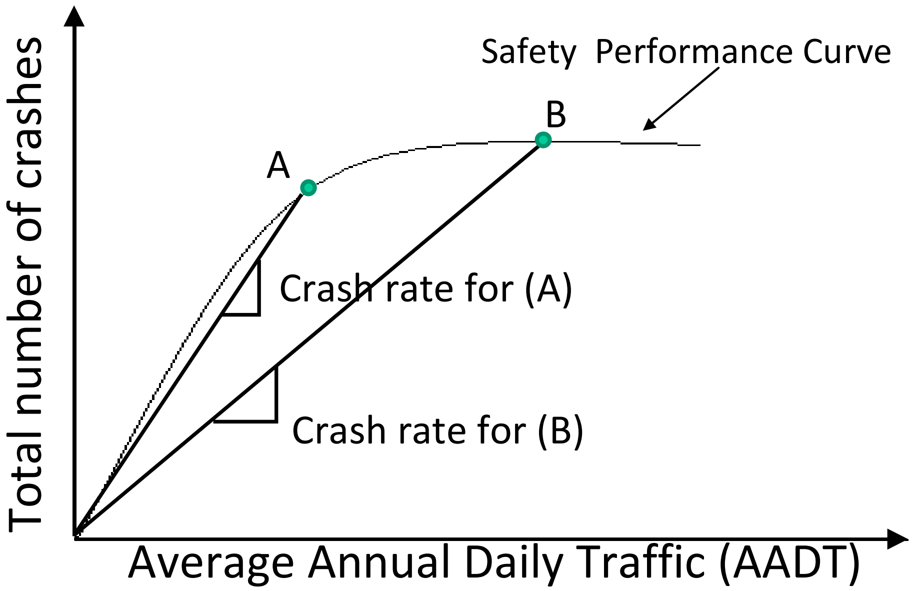


Figure 3: Rate Misleading Effect (Qin, et al., 2005)

Moreover, crash rates at different sites cannot be compared because different sites have different AADT (Average Annual Daily Traffic). To make such comparisons, accident frequencies for the same exposure need to be considered. The Rate and Number Method makes use of both of the above-mentioned approaches by comparing accident rates at sites with a predetermined minimum accident frequency (Hauer, E. and Persaud, B.N., 1984, Hauer, E., 1996). Even though this method seems to be better than considering rate alone, it is not very reliable when the minimum (or normal) accident frequency is taken into consideration as the normal accident frequency for a set of similar sites may not be normal for another set of sites

(Hauer, E. and Persaud, B.N., 1984, Hauer, E., 1996). In another screening criterion called Rate and Quality Method, the observed accident rate is compared to its critical crash rates which are specific to each site type and which depend on the degree of confidence desired for that location (Higle, J.L. and Witkowski, J.M., 1988).

Ezra Hauer, in another paper, “Identification of Sites With Promise”, mentioned that importance has to be given to the sites where severe accidents occur (Hauer, E., 1996). Analysis of crashes based on severity is deceiving since a fatal crash is given an extremely high weightage over a PDO crash (property damage only crash) that might result in false identification of SWiP. This approach resulted in the introduction of the Safety Index (Tamburri, and Smith., 1970). Safety Index requires all the crashes to be expressed as Equivalent PDO crashes (EPDOs) that could be used in ranking the SWiP based on crash severity. The reliability of this method is questionable as it is clear from research that different accident types (based on severity) have different dependencies on AADT (Hauer, et al., 2004). In the same paper, he introduced and explained the term “safety effect”, which can be estimated by the product of count of past crashes and the estimated percent reduction, severity wise. This estimation is very simple except that it has a few severe drawbacks like the exaggeration of random noise by the severity weighing of fatal accidents and the Regression-to-mean bias (Hauer, et al., 2004) that is discussed in the following paragraphs.

When the basic site selection criteria is used for network screening along with a small period of crash data (i.e. 2-3 years of crash data), a problem called "Regression-to-mean" needs to be addressed to. "Regression-to-mean bias is the phenomenon of repeated measures of data in the long run drifting towards a mean value" (iTRANS and Human Factors North Inc., 2003).

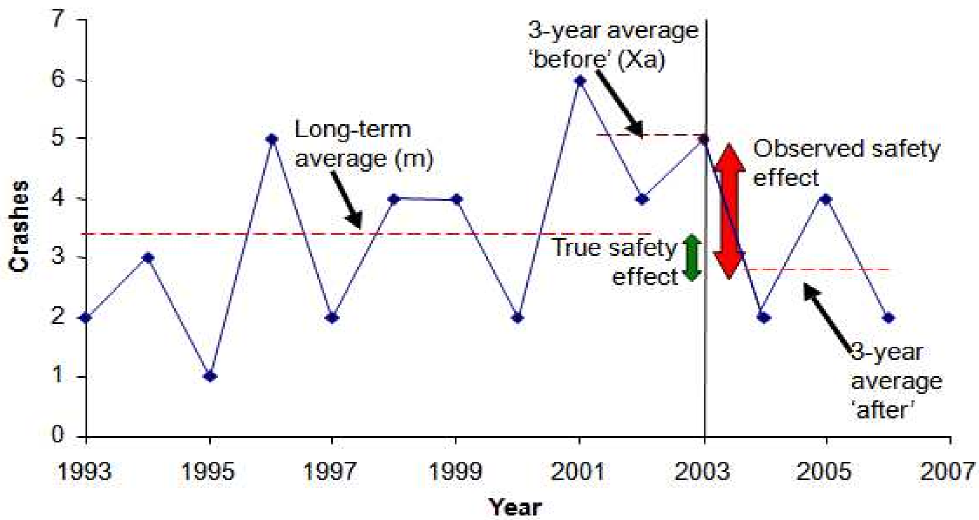


Figure 4: Regression-to-Mean effect (iTRANS and Human Factors North Inc., 2003)

The three-year average crash occurrence at a site is generally either higher or lower than the long-term average (assuming a 10-year accident history). Considering the 3-year average crash occurrence after the implementation of safety improvement, the observed safety effect (comparing the 3-year before and 3-year after periods) will be different from the true safety effect (in comparison to the long-term average crash occurrence). In the above figure, three-year observed safety effect appears larger than it really is based on 10-year data due to the random variation in year to year crash counts. Locations that have extreme variation in crash numbers are likely to have a stronger regression-to-mean effect resulting in comparatively less "practical safety effect" or vice versa depending on random increase in fluctuation or random decrease in fluctuation (Hauer, E. and Persaud, B.N., 1984). Most of the basic site selection criteria are applied to single years of crash data, and do not address this error, resulting in false identification of problematic sites. The screening methods that take into consideration the aforementioned Regression-to-mean bias will better identify Sites With Promise.

2.2 Advanced Site Selection Criteria:

Regression-to-mean effect can be corrected using advanced site selection criteria such as Empirical Bayes method of estimation (Hauer, et al., 2002). The Empirical Bayes (EB) approach is a probabilistic identification method which determines the probability that the accident rate exceeds the normal rate (Higle, J.L. and Witkowski, J.M., 1988). This method assumes that safety can only be estimated in degrees of precision which is the error measured in standard deviations (Hauer, et al., 2002). EB method is mainly based on two assumptions. First, the actual number of accidents at a site follow a Poisson distribution. The second assumption is that a site is considered to be hazardous if the probability of crash occurrence is greater than δ, that is the site's true accident rate exceeds the observed average rate across the region (Higle, J.L. and Witkowski, J.M., 1988). The traditional methods discussed earlier are also based on the assumptions that the site is deemed to be hazardous if the observed accident rate exceeds the observed average rate within an acceptable level of confidence, which is more susceptible to identifying false negatives (truly deviant sites that are not identified as SWiP) (Hauer, E. and Persaud, B.N., 1984) or if it exceeds the site's critical rate (which is a function of observed regional accident rate, traffic volumes and the desired level of confidence) (Higle, J.L. and Witkowski, J.M., 1988).

One of the drawbacks of the EB method is that it requires Safety Performance Functions. "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, et al., 2005))." (iTRANS and Human Factors North Inc., 2003). SPFs, used to identify locations with potential for accident occurrence, have no information related to the nature of the crashes. They just explain the magnitude of the problem (Kononov, J. and Janson, B.N., 2002). The nature of the problem can only be determined through direct diagnostics & pattern recognition techniques (Kononov, J. and Allery, B., 2003). The use of Safety Performance Functions is very efficient based on the fact that the relation between exposure and traffic safety (in terms of traffic crashes) is non-linear (Kononov, J. and Janson, B.N., 2002, Kononov, J., 2002, Kononov, J. and Allery, B., 2003, Qin, et al., 2005). As risk is dependent on the type of crash, different SPFs can be built for the same roadway section by disaggregating the types of crashes into four categories (single vehicle crashes, multiple vehicle crashes in the same direction, multiple vehicle crashes in opposite direction, and crashes at intersections) (Qin, et al., 2005). Research concluded that the SPFs for the above sections are also not linear (Kononov, J. and Allery, B., 2003). Further disaggregating the problem may result in other sites being selected for additional investigation.

Considering the methodological problems and resulting bias discussed earlier, EB method is reliable when limited 2-3 years accident history is available, since it increases the precision of the estimates. It uses a weight factor (which is based on logic and real data and which is a function of dispersion parameter) along with safety performance functions for predicting the expected performance at a site. This weight factor calculated based on the dispersion parameter of the SPF, addresses Regression-to-mean issues proving to be stronger (HSM., 2008).

Level Of Service of Safety (LOSS) also uses SPFs to reflect how the roadway segment is performing with regard to its expected accident frequency at a specific level of AADT. For performing these functions, the accident data is assumed to be normally distributed and a two way ANOVA test can be used to confirm this. A Poisson distribution is not suggested as the actual accident data has more widely dispersed values than its tolerable limits, it also has a limiting assumption that variance equals mean and with the accident data, variance is always greater than its mean (Hauer, E., 1996, Kononov, J. and Allery, B., 2003). According to Jake Kononov (Kononov, J. and Allery, B., 2003), LOSS uses qualitative measures that characterize safety of a roadway segment to its expected performance. An SPF that is built considering traffic accidents as random Bernoulli trials for different levels of AADT (low, medium and high) can be used to qualitatively measure the site safety from crash severity and crash frequency perspective.

While the nature of crashes needs to be considered to better understand the scenario, the above discussion helps in elucidating the magnitude of the safety problem. As mentioned earlier, such analysis of the nature can be done using direct diagnostics & pattern recognition techniques. Once sites have been selected for safety improvement, diagnostic techniques can be applied to determine appropriate countermeasures. "Detection of an accident pattern suggests a presence of an element in the roadway environment that triggered a deviation from a random statistical process in the direction of reduced safety" (Kononov, J. and Janson, B.N., 2002). Considering that the probability of success is same for all crashes and a finite number of trials, all the assumptions for Bernoulli trials are satisfied and hence the Bernoulli method can be used for calculating the probability of occurrence of an accident. (Kononov, J. and Janson, B.N., 2002, Kononov, J., 2002). The probability obtained, aided with the pattern recognition techniques, help in better analyzing the problematic sites. Even though the overall accident frequency and rate are both represented within the safety performance function, crash patterns still need to be observed. These patterns are examined visually using the crash report data sheet and sometimes specific sites are viewed on the video log. These accident patterns are considered "to provide a direct link to the development of a counter measure strategy" (Kononov, J., 2002). The limits of sections with accident patterns, if any, can be identified using the "sliding scale" technique in a Geographic Information Systems (GIS).

Geographic Information Systems, defined as a collection of hardware and software used to edit, analyze, and display geographical information stored in a spatial database, plays a vital role in transportation safety analysis (FHWA., 1999). Most of the screening methods (using accident frequency, crash rates, weight factors etc), used in combination with latest GIS tools, result in more accurate and faster identification of problematic sites. Spot or intersection analysis, strip analysis, sliding scale analysis and corridor analysis can be used for screening based on the type of analysis (FHWA., 1999).

The discussion about various screening methods, their strengths, and limitations point towards the use of Empirical Bayes approach and Safety Performance Functions as the most effective method currently available to safety analysts. The rigorous calculations involved in EB method make it tedious and automation of this process would be required for widespread adoption. Thus, twenty highway agencies along with FHWA (Federal Highway Administration) are working together in developing a software application, ***SafetyAnalyst,*** to aid the implementation and maintenance of a site safety improvement process on the basis of EB approach and use of Safety Performance Functions (Turner-Fairbank Highway Research Center., 2007).

SafetyAnalyst "provides state-of-the-art analytical 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" (Turner-Fairbank Highway Research Center., 2007). Even though the data requirements are more cumbersome compared to other conventional methods of site selection, this approach, will offset the major drawbacks like Regression-to-mean effect, over dispersion effects, non-linear relationship between crashes and exposure that were to some extent unavoidable until now (Turner-Fairbank Highway Research Center., 2007). However, *SafetyAnalyst* uses safety performance functions to identify SWiP. In the development of *SafetyAnalyst,* standard SPFs were developed from data obtained from a limited number of states (California, Minnesota, Ohio and

Washington) (Turner-Fairbank Highway Research Center., 2007). To make the SPFs applicable to specific region or states, *SafetyAnalyst* uses calibration factors to fit the areas data to the pre defined SPFs. However, *SafetyAnalyst* documentation indicates that states should consider developing their own SPFs to obtain an even better fit.

The prior discussion of various ranking criteria concludes that there is no "best" ranking criterion to adopt for all situations. Ezra Hauer et al, in his paper "How Best to Rank Sites With Promise," explains the importance of consistency in judgment while identifying the best ranking criterion suitable for a particular situation. The paper concludes that each site needs to be judged the same way with regard to the possible countermeasures and the ranking criterion (Hauer, et al., 2004).

Since no one ranking criteria is the best and each ranking criteria has its own advantages and limitations, another method categorizes sites based on two or more ranking criteria. Pair wise comparison of the results from the ranking criteria gives two sets of ranked sites. Choosing both common sites and applying a detailed engineering survey to the top ranked sites that are not common gives better SWiPs. This step is followed by estimating the anticipated costs and safety benefits at each site and calculating the benefit cost ratio. The ranking criterion that leads to the most cost effective projects is considered to be better. The larger the correlation between the rank based on screening and the rank based on cost effectiveness as established by a detailed engineering survey, the better the screening method (Hauer, et al., 2004).

The various screening methods discussed have their own data requirements that strongly influence the site selection method that is chosen for network screening. Following is the summary table of the data requirements for all the ranking criteria discussed.

Table 2: Data requirements for various site selection methods (HSM., 2008)

Method	Data and Inputs				
	Crash data by Type, Location and date	Roadway Characteristics by Location	Traffic Volume	SPF	Other
Category 1: Screening Based on Crash Counts					
Frequency	X	X			
EPDO	X	X			EPDO Weighting factors
Rate	X	X	X		
Rate Quality Control	X	X	X		
LOSS	X	X	X	X	
Category 2: Screening Based on Proportions					
High Proportion of Crashes	X	X			
Category 3: Screening Based on Potential for Safety Improvement					
SafetyAnalyst	X	X	X	X	

It is clear from the above discussion that traditional ranking criteria has limitations that need to be overcome to obtain better SWiP, while at the same time advanced ranking criteria have more intense data requirements. However, the benefits of advanced methods should outweigh the added labor and time commitments to develop and maintain the data. These developments will also likely require different levels of expertise due to the nature of the advanced statistical methods and model development. It is likely that individual states or regions will need to develop their own SPFs to achieve the greatest benefit. This is particularly true for southern states which were not included in the initial model development activities.

CHAPTER 3

METHODOLOGY

The approach towards this research is taken in stages. Figure 5 briefly mentions the various stages.

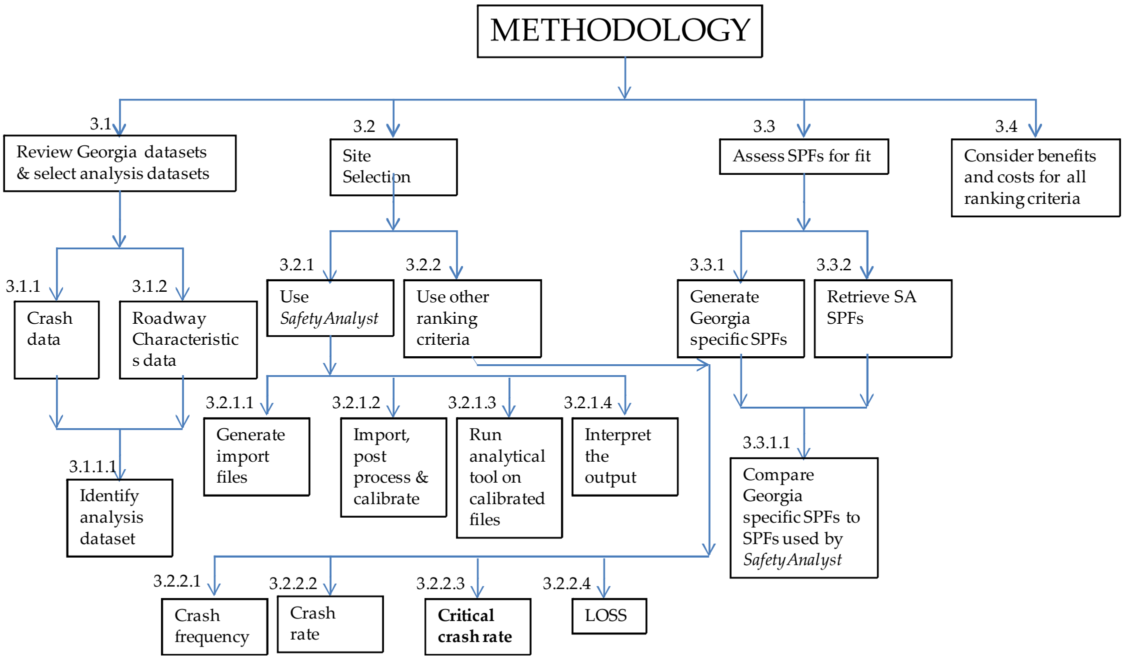


Figure 5: Various phases and steps taken towards achieving the objectives

For this project, only data from one county is considered due to time and resource constraints. Recently, GDOT has converted all Cobb County crash records to electronic format and thus they are considered to be more reliable. Moreover GDOT had an initial version of the data files for Cobb County to be imported into *SafetyAnalyst*. Hence, Cobb County is considered for further analysis.

3.1 Review Georgia datasets:

For the current study, the following datasets are reviewed and analyzed. All the datasets are obtained from Georgia Department of Transportation (GDOT).

- Crash data of Georgia for a period of three years (2004-2006)
- Roadway Characteristics data (snap shot from December 2007)
- GIS base map (snap shot from 2007)

3.1.1 Crash Data:

Georgia crash data was obtained for a period of three years (2004 - 2006). The crash database contains detailed information about the crash event, vehicles, drivers and occupants involved. A second and separate listing of crashes was also obtained which contains a spatial reference for most crashes in the state that occurred in the time period between 2004 and 2006. The crash database consisted of 1,033,517 reported crashes during the three-year period for the entire state of Georgia. Of those, 7.75 % totaling to 80,169 were reported in Cobb County. During the years 2004 through 2006, 1,032,445 crashes were spatially located for the whole state of Georgia including 80,736 crashes in Cobb County. Of the two datasets, there were some crashes in each that were not present in the other, therefore, to continue with GIS analysis, the subset of the two datasets which intersect were used. ArcGIS and Microsoft Access were used to compare the list of spatially referenced crashes with the crash database and it was found that 5% of the crashes in Georgia were not spatially located due to various reasons such as insufficient street name information. Of the 80,169 reported crashes and 80,736 spatially located crashes, only 79,726 reported crashes in Cobb County have a spatial reference attached to it. Specifically, 1,010 crashes are spatially located but not reported in the detailed crash database and 446 reported crashes were not spatially located.

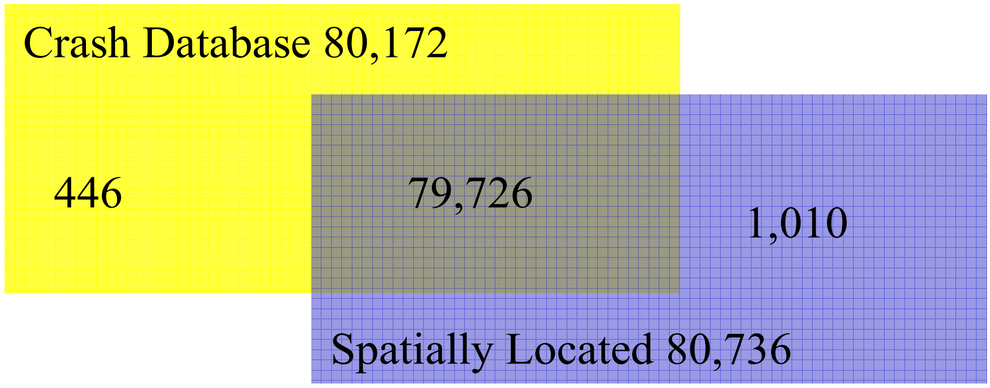


Figure 6: Summary of crashes found in crash database and also spatially located

<u>3.1.2 Roadway Characteristics and associated GIS shape files:</u>

Georgia DOT maintains a linear referencing file (LRS) for the complete state and it contains shape information for most of the roads in the state. There are 153,308 routes' records in this database. This LRS file is a shape file compatible with ArcGIS and has data stored in a dbf format to be used with other DBMS. For Cobb County, 9,109 records exist in LRS file. Each route has a unique ID or the "RCLink". The 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 detailed roadway characteristics from the roadway characteristics file (RC file).

GDOT also maintains a roadway characteristic file with detailed information about the roads such as number of lanes, type and width of shoulders, type and width of medians etc. Each route (with a unique RCLink) in LRS is divided into smaller segments consisting of similar roadway characteristics. There are 49,041 roadway segments in Cobb County which are obtained by querying the roadway characteristics table in Microsoft Access using a county code of "067". The average length of the roadway segments Cobb is 0.062 miles. Thus, one or more road characteristics changes on average every 0.062 miles. However, there are some point segments with "0" length. Each roadway segment has an RCLink, beginning milepost and an ending milepost. A unique ID is created to identify each roadway segment. The unique ID generated consisted of 15 digits. It has route type followed by the six digit route name followed by the beginning milepost (represented by four digits) and the ending milepost (represented by four digits). However, RC data is just an Access database and has no spatial reference attached to it. To obtain a spatial dimension to the RC data, a concept called "Dynamic Segmentation" is used. To carry out this, a new project in ArcGIS is created and RC_Cobb text file is imported into ArcGIS. Based on LRS data, a spatial reference is attached to this file by adding route events (by going to Tools →

Add Route Events). The segments are added along each RCLink based on start offset (beginning milepost) and end offset (ending milepost). The following is the screen shot of this step.

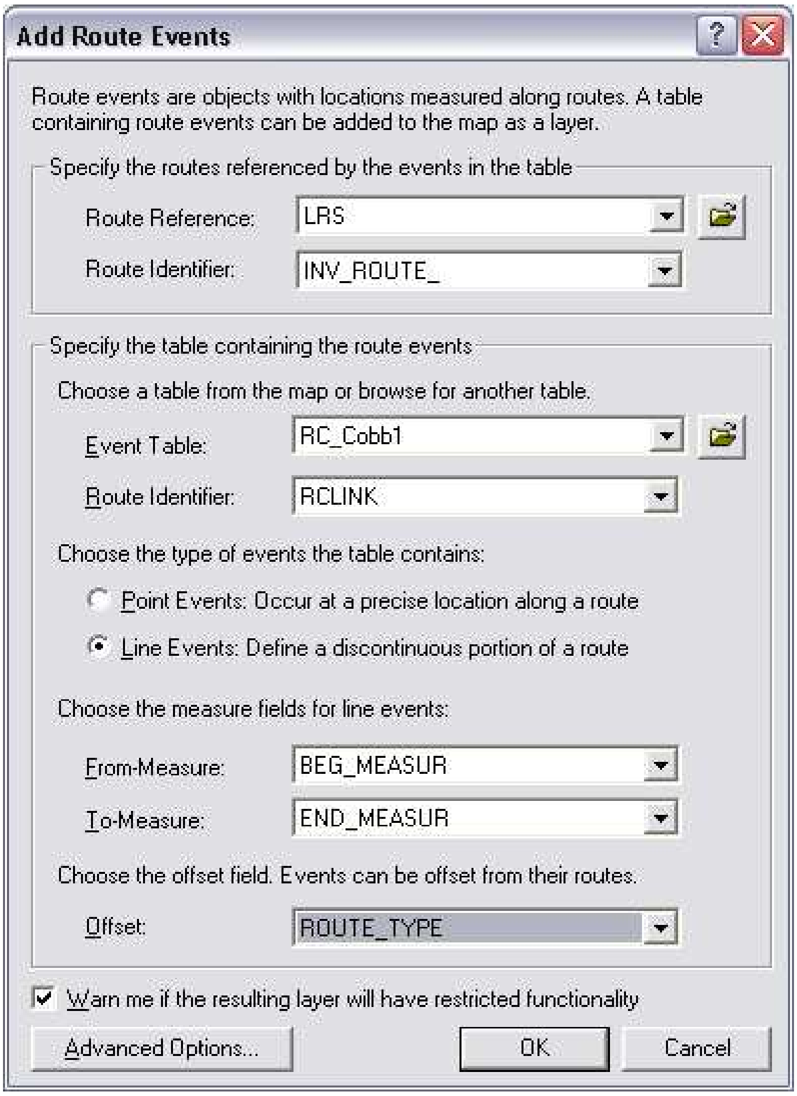


Figure 7: Add Route Events along a spatially referenced map

3.1.2.1 Identify analysis selection set:

A Microsoft Access database was created by importing the 79,726 crashes (both reported and spatially located) and the Roadway Characteristic database for the Cobb County. A crash was coded as an "intersection related crash" if it occurs within 200 ft from an intersection. All the non-intersection related crashes were identified in ArcGIS by creating a buffer of 200ft around the intersections and excluding all the crashes that fall within the buffer region. The following figure shows roadways and intersections on a typical road.

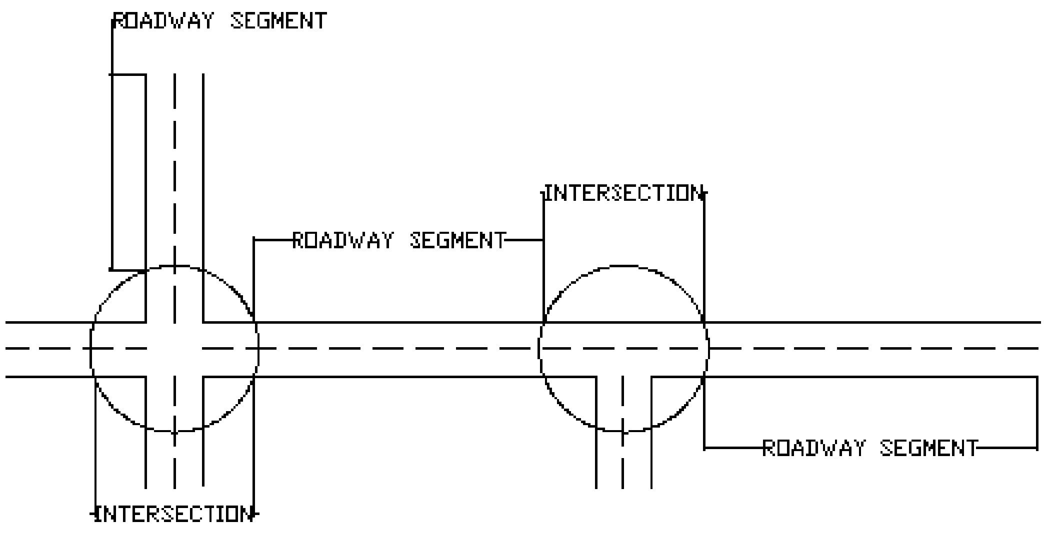


Figure 8: Roadway Segments and Intersections

32,357 spatially referenced crashes were considered to be non-intersection related for the years 2004-2006 in Cobb County. These records were imported into Microsoft Access for further analysis.

3.1.3 Discuss all selection criteria:

As mentioned in section 3.1.1, each crash had to be in crash database and spatially located to be included in the analysis. Also, to reduce the scope of the research, the analysis was limited to segment crashes only. Thus, intersection crashes were eliminated.

The other selection criteria that could be analyzed to identify sites with potential for safety improvements are intersections and ramps which are beyond the scope of this research.

3.2 Site selection:

Network screening is the process of identifying sites for further engineering study and potential countermeasure implementation. Over the past few decades, many site selection criteria are used to identify SWiP. Basic site selection methods include the use of crash frequency, crash rate, critical crash rate and high proportion of crashes. As discussed in the previous chapter, the traditional methods have many limitations like Regression-to-mean, random noise and assumption of linear relationship between crashes and exposure. These limitations are accounted for in some of the advanced ranking criteria like generation of Safety Performance Functions, use of Level Of Service of Safety, Empirical Bayes Method by using *SafetyAnalyst* software.

The following sections provide details regarding implementation requirements of various basic and advanced site selection methods. The ranking criteria assessed in this project include crash frequency, crash rate, critical crash rate, generation of Cobb County specific SPFs, LOSS, and *SafetyAnalyst*.

3.2.1 Use of *SafetyAnalyst*:

SafetyAnalyst is a set of analytical tools to aid in identifying site specific improvements to improve highway safety in a more cost effective manner. The following paragraphs discuss the process required to generate, import, post process and calibrate the files in *SafetyAnalyst* and to run the analysis.

3.2.1.1 Generate import files that are compatible with *SafetyAnalyst*:

The data requirements for *SafetyAnalyst* are comprehensive and specific. Separate files were created for the accident, roadway segment and segment traffic data and were imported into the software. SQL queries were used to pull data from GDOT crash tables (Accident, Location, Pedestrian and Occupant) and to create the import tables that have column layout and data format that is compatible. The SQL queries used are included in Appendices A and B (for accident and roadway characteristic files respectively). After running the SQL queries, data recoding is done. Following are the detailed descriptions of each import file.

a) Accident file:

Initially, a skeleton "Accident" file was generated to define the column layout and data format for the files. Next, a series of SQL queries were run in the order shown in Appendix A to populate the data file. The required fields in the *SafetyAnalyst* Accident file and the corresponding data fields in the GDOT datasets are attached in Appendix A. *SafetyAnalyst* has a very specific set of codes for each data element. Many of these variable sets had to be recoded to match the formats required by *SafetyAnalyst*. In addition, some of the required *SafetyAnalyst* elements required joining data from multiple fields and/or elements in the Georgia datasets. The data mapping guide is shown in Appendix A. The mandatory fields include accident case identifier, route type, route name, county number, accident date, accident time, relationship of accident location to junction, accident type and manner of collision, number of vehicles involved and accident severity level. The file was saved as AltAccident in csv format. This csv file contains 32,357 crashes.

b) Roadway Segment file:

Similar to Accident file, the roadway segment file started as a skeleton file structured based on *SafetyAnalyst* format. The Roadway Segment file was then generated by running a series of SQL queries on GDOT datasets to populate the skeleton file. The list of queries that were run are included in Appendix B. The fields in the Roadway Segment file to be imported into *SafetyAnalyst* and the fields in the GDOT data from where the data is taken, along with the selection criteria are attached in Appendix B. The data mapping issues are also addressed in Appendix B. The mandatory fields include agency ID, route type, route name, county number, segment length, area type, roadway class level 1, number of thru lanes in direction 1 and 2, median type level 1 and 2, shoulder type and operation type. The file was saved as AltRoadwaySegment in csv format. This csv file contains 48,565 roadway segments.

c) Segment Traffic file:

The fields in the Segment Traffic file to be imported into *SafetyAnalyst* and the fields in the GDOT data from where the data is taken, along with the selection criteria are attached in appendix C. The data mapping issues are also addressed in appendix C. The mandatory fields include agency ID (similar to the agency ID in Roadway Segment file), calendar year and the AADT for each year. The file was saved as AltSegmentTraffic in csv format. This csv file contains 242, 809 records.

Once the three files were generated, the files were opened in notepad for cleaning. In the notepad, the first row consists of the respective file name followed by many commas (","). All but one comma in the first row beside the filename was deleted and the file saved again.

The modified csv files were saved in a folder and the folder was placed in the c:drive (Note: There is a limitation in the number of characters in the file path).

3.2.1.2 Import, post process and calibrate the input files in *SafetyAnalyst*:

SafetyAnalyst version 1.4.11 was used for this project to implement the EB site selection method. Within the *SafetyAnalyst*, the Data Management tool was opened and a new dataset was created. In the import tab, the three files (AltRoadwaySegment, AltSegmentTraffic and AltAccident) were added in this specific order. (Note: Alterations in the order result in errors). The import process for Cobb County took about 12 minutes. The time required depends on the computer processor speed and the programs that were simultaneously run on the machine. At the end of the import process, *SafetyAnalyst* outputs a log of warnings and errors associated with the import process. Warnings may include zero traffic volumes for roadway segments etc. Once the import was completed without significant errors, the post process was carried out. The minimum and maximum years of the accident data to be processed needs to be given. The range of 2004 - 2006 was used for this project. For the current project, the traffic data was available for the years 2000 through 2004. If this information is left unchanged, only one year (2004) has both accident data and traffic data and *SafetyAnalyst* runs analysis for just one year. For running the analysis on all three years, maximum year of traffic data to be processed was changed to 2006. For 2005 and 2006, the software projects traffic data based on 2000-2004 trends. During post processing, homogeneous segments were created. These segments were formed by joining two or more continuous roadway segment into one depending on similar characteristics. The threshold limits for aggregating roadway segments as homogeneous segments can be input into the software. The following figure shows the screenshot of the window for editing and viewing threshold limits for homogeneous segment aggregation.

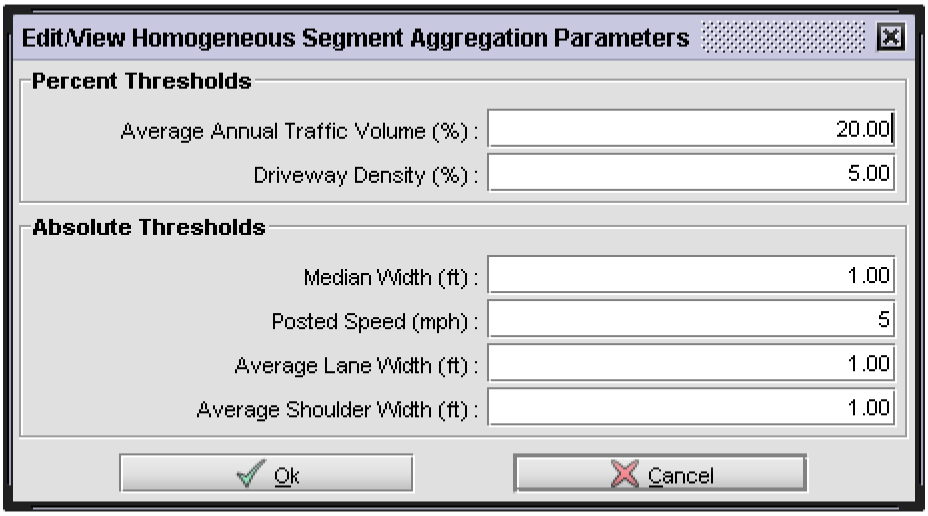


Figure 9: Screenshot of the Edit/View Homogeneous Segment Aggregation Parameters and their threshold limits

The 48,565 imported roadway segments were grouped to form 19,041 homogeneous roadway segments. For this project, post processing took about 16 minutes. Since the default SPFs used in *SafetyAnalyst* were generated from northern states' data, the SPFs need to be calibrated to the southern data to reflect the crash trends of the south. This is done in the calibration step which followed the post process step. The calibration for Cobb County data took about 2 minutes. The calibration log was saved for further reference. The calibrated data was exported to a file for import into ArcGIS to check for missing roadway segments. The exported files were automatically saved in the folder "export" in "SafetyAnalyst". *SafetyAnalyst* will not run on segments that do not have all the three components – crash, traffic volume and roadway characteristics.

The exported non-homogeneous Roadway Segment file consisted of 48,543 records. The missing 22 segments were spatially located using ArcGIS. These are found to be insignificant roads which either close as a loop or have a negligible length. The exported AltAccident file consisted of 30,023 crashes. Missing crashes were found to be on roadway segments that do not have roadway characteristics data.

3.2.1.3 Run Analytical tool on the calibrated files in *SafetyAnalyst*:

SafetyAnalyst analytical tool was used to carry out analysis on the roadway segments and accidents. This tool helps in conducting Network Screening, Diagnosis and Countermeasure Selection, Economic Appraisal and Priority Ranking and Countermeasure Evaluation. For the present project, only network screening was carried out due to the lack of sufficient data and resources for other modules. The 'Getting Started Wizard' walks users through the analytical 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 can be created and saved based on the user requirements. On the other hand, site lists can 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

Of all the above mentioned types, the *SafetyAnalyst* development team recommended "Basic Network Screening with peak searching on roadway segments" method for analysis since its results and method were verified compared to other types (Note: This research was conducted while the *SafetyAnalyst* tool was still under development). Total (Fatal, injury and PDO) crashes for all available years were considered to increase the sample size. *SafetyAnalyst* will also run for fatal and injury crashes only. Potential for safety improvement (PSI) could be calculated based on expected accident frequency or excess expected accident frequency and for this project, PSI is calculated based on expected accident frequency. Rural and urban areas are weighted equally. To exclude some of the roadway segments that have zero to minimal crashes, the crash frequency limiting values were set to 5.00 accidents/mile/year. Coefficient of variation (CV) for the roadway segments determines the number of sites to be included in the output report (the lesser the CV limit, the fewer are the sites displayed in the output report). CV limit is set to 0.50. The accident screening attribute, based on which the analysis could be done is selected and for this attribute, accident type and manner of collision was selected and all the values were selected within the attribute. Appendix D includes the screenshots of all the steps in "Analytical module" of *SafetyAnalyst*.

The network screening analysis ran for 15 minutes for this scenario. A sample of the report is attached in the appendix E.

3.2.1.4 Interpret the *SafetyAnalyst* output:

The output was saved in an excel file. *SafetyAnalyst* identified 850 roadway segments as SWiP based on total crashes. However, the software sub classifies the sites into different site subtypes. Following are the various site subtype codes for roadway segments:

Table 3: Site subtype code and description used for roadway segments

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

The various columns in the output are explained in the following table:

Table 4: 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 & the highest calculated PSI for the site</td></tr>
</table>

The observed crashes obtained from *SafetyAnalyst* and displayed in the output were normalized by mile. This is because, sites are generally less than one mile in length and normalization results in consistency.

For detailed analysis, only two site subtypes are considered due to the limited sample size. These include rural multilane divided highways (site subtype code: 103) and urban multilane undivided arterial streets (site subtype code: 152). The following table explains the logic to create site subtype codes 103 and 152 for roadway segments.

Table 5: Logic to create site subtypes 103 and 152 for roadway segments

Site Subtype ID	**Site Subtype code**	**Conversion Logic**
103	Rural multi-lane divided	Area Type = Rural Number of Through Lanes >= 4 Median Type Level 2 = Divided Two-Way Operation
152	Urban multi-lane undivided	Area Type = Urban Number of Through Lanes >= 4 Median Type Level 2 = Undivided Two-Way Operation

3.2.2: Use of other ranking criteria to identify SWiP:

A manual analysis of crash data to identify sites for study included several methods: high crash frequency, high crash rate, critical crash rate, and Level Of Service of Safety (LOSS). All manual analysis methods used three sets of data while LOSS used two sets of data. The first set (set A), consisted of all roadway segments in Cobb County. The second set (set B), includes all roadway segments that belong to site subtype 103 (Rural multilane divided highways) as defined by *SafetyAnalyst*. The third set (set C), includes all segments that belong to site subtype 152 (Urban multilane undivided arterial streets). For manual analysis, the number of crashes occurring on each homogeneous roadway segment is required. The following steps were followed to obtain the crash count on each site.

- All the roadway segment information in Cobb County exported from *SafetyAnalyst* was saved in an excel workbook.
- The excel file was imported into Microsoft Access. AADT field needs to be added to the file.
- AltSegmentTraffic file was also imported into Microsoft Access and a cross tab query was written to obtain the average AADT of each site. This query was then linked to the Roadway Segment information. However, the exported segments were homogeneous segments while AltSegementTraffic file has non homogeneous segment information. One to one linking was done between the AltSegementTraffic file and the Roadway segment table based on the first roadway segment on

homogeneous sections since the same traffic flows through all roadway segments in a homogeneous section.

- The saved query was exported into a txt file.
- The .txt file was added to GIS. It was just another table and has no spatial reference attached to it. Spatial reference was attached to it using the concept called "Dynamic Segmentation" which is explained in the earlier sections.
- Accident file was spatially joined to this layer.
- The joined shape file was exported as a dbf and later saved as an excel file
- Other workbooks were created from the excel file for datasets B and C.

Once the number of crashes occurring on each roadway segment were determined, crash frequency, crash rate, critical crash rates and LOSS can be calculated as described in the following paragraphs. Based on the number of crashes and other characteristics of roadway segment, SAS software is run and Cobb County specific SPFs are generated.

The methodology considered for identifying crashes based on crash frequency, crash rate, critical crash rate and LOSS is based on the procedures set in Highway Safety Manual Chapter 14.

3.2.2.1 High Crash Frequency:

For each set of data (A, B, and C), sites were sorted based on crash count in descending order and ranked. With this method, the site with highest crash count was ranked number 1 and the site with second highest crash count was ranked number 2 and so on.

3.2.2.2 High Crash Rate:

Total segment length for each site was calculated as the difference between the start milepost of the first segment and the end milepost of the last segment in a homogeneous segment. Exposure (EXPO) also called, million vehicle miles of travel (MVMT), was calculated using the formula,

$$EXPO = AADT * 365 * 3 * \text{Total Segment Length} / 1{,}000{,}000 \qquad \text{(Equation 1)}$$

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

The ratio between crash count and exposure was termed as "crash rate". 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.

3.2.2.3 Critical Crash Rate:

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

$$R_{Ci} = R_A + K_C * \sqrt{(R_A /(EXPO))} + (1/(2 * EXPO) \quad \text{(Equation 2)}$$

Where:

R_{Ci}: Critical crash rate for site i

R_A: Average crash rate for each reference population

K_C: 1.645 (the probability constant based on the confidence interval of 95%)

EXPO: Million vehicle miles of travel

The difference between the observed crash rate and the critical crash rate 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.

3.2.2.4 Level Of Service of Safety (LOSS):

Safety Performance Functions are required to rank sites with potential for safety improvement based on LOSS. SPFs generated for Cobb County data and the calibrated SPFs used by *SafetyAnalyst* (which are discussed in the later sections) are used to perform LOSS.

SPFs are applied to each site to obtain an estimate of the number of crashes, k, for the site under consideration. The standard deviation ($\sigma(k)$) of the above obtained estimate is calculated using the formula,

$$\sigma(k) = \sqrt{(\Phi^*(k^2))} \qquad \text{(Equation 3)}$$

Where,

$\sigma(k)$ = Standard deviation of the estimate of the expected number of crashes

Φ = dispersion parameter of the SPF used

k = the estimated number of crashes from the SPF

The observed number of crashes, K, is compared to the limits to be categorized into any one of the four categories of LOSS.

The following table describes the condition and the LOSS category along with description.

Table 6: Various LOSS, their conditions and descriptions (HSM., 2008)

LOSS	Condition	Description
I	$0<K<(k-1.5\sigma(k))$	Indicates a low potential for crash reduction
II	$(k-1.5\sigma(k))\leq K<k$	Indicates better than expected safety performance
III	$k\leq K<(k+1.5\sigma(k))$	Indicates less than expected safety performance
IV	$K\geq(k+1.5\sigma(k))$	Indicates a high potential for accident reduction

All the sites with LOSS IV are flagged and identified as SWiP. However, it is difficult to prioritize the top ranked sites without conducting a detailed engineering study.

<u>3.3 Generate Georgia specific SPFs:</u>

SafetyAnalyst identifies sites with potential for safety improvement using Empirical Bayes method. The default SPFs used by *SafetyAnalyst* are generated from northern states' data. Thus, researchers thought it important to determine if the models had an appropriate fit for Georgia data. A Safety Performance Function that fits the GDOT data needs to be generated to analyze crashes. The logic used to identify SWiP in *SafetyAnalyst* is also applied to generate SPFs to maintain consistency. Negative Binomial Regression method and not Poisson distribution is used for generating SPFs. This is mainly due to considerable difference in the mean and variance of crash data.

Number of expected crashes (crashes per mile per year) is predicted as a function of Average Annual Daily Traffic (AADT) alone. The functional form for roadway segments is found to be:

$$k = (e^{\alpha})* (ADT)^{\beta} \qquad \text{(Equation 4)}$$

Where

k – Predicted number of target crashes per mile per year

ADT – Average Annual Daily Traffic (veh/day) for roadway segments in both directions of travel.

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

$$N = (e^{\alpha})* (ADT)^{\beta} * L \qquad \text{(Equation 5)}$$

Where

N - Predicted number of target crashes per site per year

L – Length of the roadway segment in miles

To generate SPFs, all the sites in each site subtype are required. For this project, SPFs are generated for two site subtypes. They are:

103 - Rural multilane divided roads

152 - Urban multilane undivided arterial streets

Site subtype 103 had 562 homogeneous segments and site subtype 152 had 325 homogeneous segments in total. However, there were many roadway segments that are less than 0.1 miles in length. There were 315 and 185 roadway segments from site subtypes 103 and 152 respectively that were excluded due to a segment length of less than 0.1 miles. Three segments in site subtype 103 have "zero" AADT and hence, they are excluded from further analysis. Three roadway segments from site subtype 152 have extremely high AADT of about 350,000 while the AADT in this subtype range from 2031 to 50,000. Year wise and overall SPFs are generated based on the remaining segment information, AADT information and crash data. For generating SPFs, 244 roadway segments from site subtype 103, and 137 roadway segments from site subtype 152 are considered.

Statistical software, SAS, is used to generate SPFs using Negative Binomial Regression technique. Data requirements for running SAS include:

- Roadway segment ID
- Site Subtype (whether site subtype 103 or site subtype 152)
- Start Offset (starting milepost of the homogeneous segment)
- End Offset (ending milepost of the homogeneous segment)
- Segment Length (Difference between end offset and start offset)
- Log(ADT) (where ADT is the Average Annual Daily Traffic for the respective year or the average value for the three years depending on the year of analysis)
- Offset (= Log(Segment Length *Number of years of crash data available))

- Total crashes (The total number of crashes occurred on each homogenous roadway segment during a particular year or for the three years depending on the year of analysis)

SAS is used to generate SPFs specific to each year and to the complete data (for the three years 2004-2006) for the site subtypes 103 and 152.

The example of a SAS code used for this analysis is shown in Appendix F. Appendix G includes the SAS output for the two site subtypes.

3.3.1 Compare Georgia specific SPFs to SPFs used by *SafetyAnalyst*:

SafetyAnalyst uses SPFs that are generated from the northern states data calibrated to Georgia data while Cobb County specific SPFs are generated manually through negative binomial regression. Calibrated SPFs used in *SafetyAnalyst* are generated from the non calibrated SPFs by using a multiplying factor called calibration factor. The calibration factor is defined as the ratio of total number of observed crashes to the total number of expected crashes. (The number of expected crashes at each site is predicted from the SPFs). These calibrated SPFs used in *SafetyAnalyst* and Cobb County specific SPFs that were generated are used to estimate the expected number of crashes from AADT. SPFs generated for Georgia, non-calibrated and calibrated SPFs used in *SafetyAnalyst* are plotted and compared against the observed crash data. The interpretations and results are discussed in the following chapter. Freeman Tukey R^2 coefficient was used to determine the goodness of fit for the two SPFs (Fridstrom, et al, 1994). The following formulae were used for calculating Freeman Tukey R^2 coefficient (R^2_{FT}).

$$R^2_{FT} = 1-((\sum_i \hat{e}_i^2)/ (\sum_i (f_i - \bar{f})^2)) \qquad \text{(Equation 6)}$$

Where,

$$f_i = \sqrt{(y_i)} + \sqrt{(y_i+1)} \qquad \text{(Equation 7)}$$

The statistic is approximately normally distributed with mean,

$$\Phi_i = \sqrt{(4\hat{y}_i + 1)} \qquad \text{(Equation 8)}$$

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

$$\hat{e}_i = \sqrt{(y_i)} + \sqrt{(y_i+1)} - \sqrt{(4\hat{y}_i + 1)} \qquad \text{(Equation 9)}$$

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 7

$\bar{f}$ is the average of all the f_i for sites considered (Fridstrom, et al, 1994).

R^2_{FT} was calculated for both the calibrated SPFs used in *SafetyAnalyst* and for the SPFs manually generated for Cobb County for the two site subtypes (site subtype 103: Rural Multilane Divided Highways and site subtype 152: Urban Multilane Undivided Arterial). The results are explained in the next chapter.

After generating Georgia specific SPFs, *SafetyAnalyst* was run again to identify SWiP using the Georgia specific SPFs. The administration tool in *SafetyAnalyst* was used to change the default SPFs to agency specific SPFs for the two site subtypes under consideration. Once the SPF values were changed, the previously saved dataset was recalibrated in the Data Management tool and the calibration log was checked for updated SPFs. Analytical tool is run and the SWiP are obtained. The difference in the ranks are presented in the next chapter.

3.4 Consider benefits and costs for all ranking criteria:

Given roadway characteristics, AADT, and crash data, for Cobb County several different sets of sites were selected for further study using various ranking criteria and methods. Conventional ranking criteria considered in this project include crash frequency, crash rate and critical crash rate. The advanced ranking criteria include LOSS and the use of *SafetyAnalyst.*

A significant amount of time, resources and money were spent in cleaning the data and generating accident, roadway segment and segment traffic files that could be imported into *SafetyAnalyst* and for use in other methods.

Since, crash data is available only for a period of three years, and information about countermeasure selection, implementation and evaluation is unavailable, this thesis dealt only with network screening. The traditional benefit cost analysis (calculation of benefits based on the number of lives saved by implementing countermeasures on the high priority sites identified by the various network screening methods) is beyond the scope of this research and hence potential benefits are analyzed theoretically.

A lot of work that is done towards *SafetyAnalyst* overlapped with the work that needed to be done for most of the other ranking criteria. Basic ranking criteria like crash rate benefitted by creation of homogeneous segments (by joining continuous shorter segments with similar characteristics as one homogeneous segment). When calculating crash rates, short segment lengths with even just one crash generate extremely high crash rates. LOSS was carried out using both northern states' SPFs (default SPFs used by *SafetyAnalyst*) and Georgia specific SPFs (generated manually for using in *SafetyAnalyst*) saving a lot of time for this method.

To carry out any of the advanced ranking criteria, a safety specialist, a GIS professional and a statistician are required for understanding and cleaning the crash data, analyzing the problems spatially and for doing statistical tests respectively. Expertise required depends on the type of analysis. Potential benefits were analyzed in terms of data requirements, systematic procedure, ability to repeat and defend the methods and the accuracy/limitations of these methods. The resources required for various ranking criteria and potential benefits for using each method are detailed in the next chapter.

CHAPTER 4

ANALYSIS AND RESULTS

The analysis for this project was done in four phases. Following are the four phases:

1. Review Georgia datasets and identifying analysis datasets. This phase also deals with identifying potential problems and issues with the crash data and roadway characteristics data and data cleaning requirements.
2. Compare various site selection methods. The various basic site selection criteria, advanced site selection criteria like LOSS and *SafetyAnalyst* are compared. Additional problems that arose while generating, importing and post processing data into *SafetyAnalyst* are also presented in this phase.
3. Assess the fit of Safety Performance Functions to Georgia. This is carried out by developing SPFs that are applicable to Georgia and Cobb county in particular and comparing them with the default and calibrated SPFs used in *SafetyAnalyst.*
4. Considering benefits and costs for all ranking criteria

Phase 1: Review crash data and roadway characteristics data:

a) Crash data:

Two crash databases exist for each reported crash: the access database and the GIS database. Both the databases were compared to obtain a final database of crashes that were spatially located. It was found that 79,726 reported crashes were spatially located. Some of the issues that were identified include:

1. A total of **80,736** crashes were spatially located in Cobb County between years 2004 and 2006 and during the same period, **80,169** crashes were reported in Cobb County. Only **79,726** of the reported crashes were spatially located. Some crashes were identified in spatial analysis which were not in the final state crash database. Reasons for these exclusions are unknown.

2. Since the crashes were linearly referenced along routes, it is nearly impossible to cross check whether the crash is correctly located or not. The crash location completely depends on the police perception noted in the crash report form. However, researchers found that a large number of crashes can be found at 0.1 miles beyond the route start point. Thus, these

sites may produce biased results in analysis if the crashes do not actually occur at these locations.

3. The county codes used were found to be different in the accident database and in the GIS database. The accident database uses DPS (Department of Public Safety) codes for counties. The list is in alphabetic order and the Cobb County code is "033". The GIS database uses FIPS (Federal Information Processing Standard) code and the Cobb County code is "067".

4. In the accident database, for each crash, the accident mile log is noted which is later used as the basis for linear referencing in GIS. The accident mile log for 3,223 crashes is found to be "999.99". This is assumed to be a missing or unknown value since the largest route is 23.910 miles in length.

b) Roadway Characteristics data:

1. Generating a unique agency ID for each roadway characteristic record was cumbersome due to the alphanumeric nature of the route name. The unique ID generated was of the form: Route type followed by six digit "route name" followed by four digits representing the start milepost of the roadway segment followed by four digits representing the end milepost of

the roadway segment. Some of the examples of the unique ID are shown in the following table:

Table 7: Alphanumeric unique ID generated from the route name, start location and end location

unique ID	Route Type	Route Name	Start Location	End Location
10005CO01280134	1	0005CO	1.28	1.34
10005CO01340138	1	0005CO	1.34	1.38
10005CO01470152	1	0005CO	1.47	1.52
10005CO01520159	1	0005CO	1.52	1.59
10005CO01830190	1	0005CO	1.83	1.9
10005SP00830087	1	0005SP	0.83	0.87

2. All Interstates are termed as state routes due to the limitations of the coding structure.

Table 8: Table showing issues with coding structure related to route type

ROUTE TYPE	
SafetyAnalyst	GDOT
Field Name: routeType	Field Name: LOC_ROUTE_TYPE
I - Interstate US - US route SR - State route BR - Business route BL - Business loop SP - Spur route CR - County road L - Local road O - Other NA - Not applicable X – Unknown	0-Accident Not Located 1-State Route 2-County Road 3-City Street 8-Public Road 9-Collector-Distributor

3. The coding for jurisdiction is confusing. The following table shows the variations in coding structure between GDOT and *SafetyAnalyst.* Determining jurisdiction based on the route type is not a reliable way.

Table 9: The closest match to coding used in GDOT to identify jurisdiction

JURISDICTION	
SafetyAnalyst	GDOT
Field Name: jurisdiction	Field Name: ROUTE_TYPE or DESIGNATED_WAY
1 - Federal maintained 2 - State maintained 3 - County maintained 6 - Township maintained 4 - Local maintained 5 - Other maintained 99 – Unknown	1 State Route 2 County Road 3 City Street 4 Col Road 5 Unofficial Road 6 Ramp 7 Private Road 8 Public Road 9 Collector – Distributor

4. While classifying roadways, *SafetyAnalyst* needs a more detailed coding. The following table describes the coding structure in GDOT and *SafetyAnalyst.*

Table 10: Table showing the differences in coding structure for Roadway Class in GDOT and *SafetyAnalyst*

ROADWAYCLASS1	
SafetyAnalyst	GDOT
Field Name: roadwayclass1	Field Name: FUNC_CLASS
1 - Principal arterial-interstate 2 - Principal arterial-other freeway or expressway 3 - Principal arterial-other 4 - Minor arterial 5 - Major Collector 6 - Minor Collector 7 - Local 0 - Other 99 - Unknown	11-Urban-Interstate Principal Arterial 14-Urban Principal Arterial 16-Urban-Minor Arterial Street 17-Urban-Collector Street 19-Urban-Local

5. Coding for the type of median in GDOT does not match well with *SafetyAnalyst* coding and GDOT data could be more specific. To fully code this variable, other GDOT variables must be used to separate the divided and undivided roadways. In addition, HOV lanes and other specialty facilities cannot be defined using Georgia data.

Table 11: Table showing the differences in coding structure for Median Type in GDOT and *SafetyAnalyst*

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

6. GDOT has different coding for the shoulder type compared to *SafetyAnalyst*. The following table describes the differences in coding by GDOT and *SafetyAnalyst.*

Table 12: Table showing the differences in coding structure for Shoulder Type by GDOT and in *SafetyAnalyst*

SHOULDER TYPE	
SafetyAnalyst	GDOT
Field Name: shoulderType	Field Name: DIV_HWY_SHLDR_TYPE
1 - Paved 2 - Composite 3 - Gravel 4 - Turf 5 - Curb 6 - No shoulder 98 - Not applicable 99 – Unknown	G- Grass or Sod S- Gravel or Stone F- Bit. Surf. Treatment (Low) I- Bit. Conc. (High) J- Portland Cement (High) K- Curb and Gutter (Width of the gutter is not coded. Always code 00C.) N- No Identifiable Shoulder or Curb. All of roadbed used as Roadway (Soil or Gravel Road). Also if less than 1 foot paved road. D- Gutter (only) O- Bit. Conc. (High) with curb and gutter P- Bit. Surface treatment (Low) with curb and gutter C- Curb only

Geographic Information System (GIS) software is used to map the roadway characteristics file (named as RC_Cobb) to LRS file. Many errors were found during this step. The errors and constraints are discussed below:

1. It was difficult to determine whether a crash occurred at an intersection or on a road segment given the current Georgia dataset. All crashes that occurred within a distance of 200ft from an intersection were considered as “intersection related crashes”.

2. RC_Cobb has 49,041 records. The AltRoadwaySegment file (file that is imported into SafetyAnalyst) has only 48,565 records. The missing records were found to be of route type 6 which are the ramps at interchanges. There are 446 ramp segments in total. The map below shows the type of roadway segments missing.

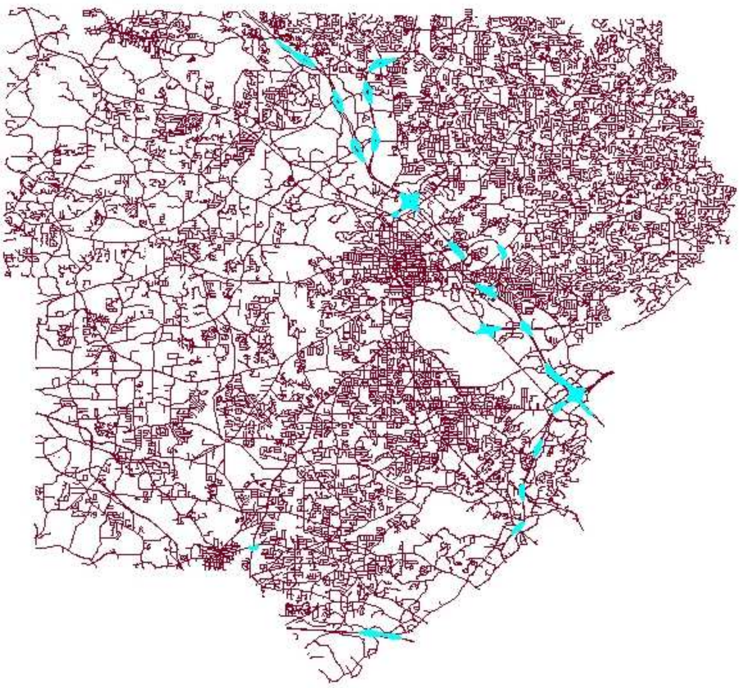


Figure 10: Cobb County with missing routes (Highlighted in blue)

3. 9,822 segments in AltRoadwaySegment file were of zero length. Some of these zero length segments were located at intersections. This problem is rectified to some extent by creating homogeneous segments while post processing. Homogeneous segments are the segments where more than

one segment with similar characteristics are combined together to form a longer segment.

4. The exported AltRoadwaySegment file consisted of 19,041 records. Twenty-two roadway segments were missing. These were found to be short loop segments. Figures 11 and 12 show the missing road segments and a detail section of one of the segments.

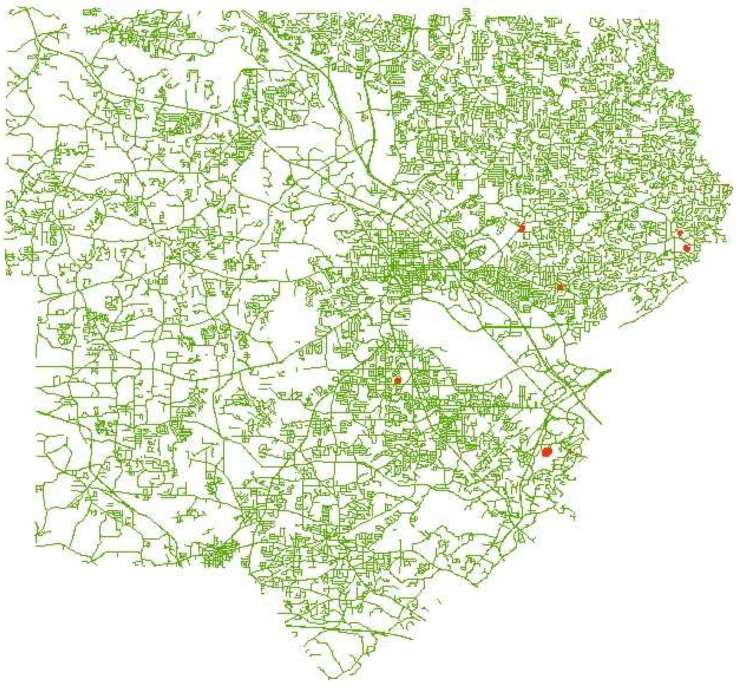


Figure 11: Cobb County with missing roadway segments after importing into *SafetyAnalyst* (highlighted in red)

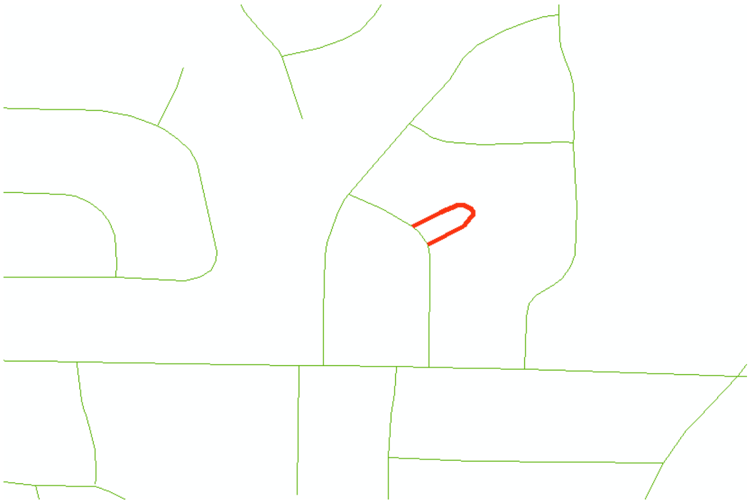


Figure 12: A detailed example of the missing Roadway segment after importing into *SafetyAnalyst*

5. When the AltAccident file from *SafetyAnalyst* was imported into GIS, dynamic segmentation should be based on the variable loc_offset (found in the AltAccident file) and not based on Acc_mile_log (found in the GIS database of the crash) due to the differences in the two columns. The following table shows several accident IDs with differing Accident Mile log and locOffset values.

Table 13: An example showing the difference between LOC_ACC_MI and locOffset

agency ID	LOC_ACC_MI (from crash database)	locOffset (from GIS database)
41220645	2.70	22.70
41470446	2.70	22.70
54580273	0.20	20.20
50030699	8.90	28.90
41120229	9.30	29.30
44270184	8.60	28.60
40740003	8.90	28.90

6. Crashes are located only on one side of the roadway on some divided roadways. This is mainly due to the missing direction coding. The screenshot of an example is shown below:

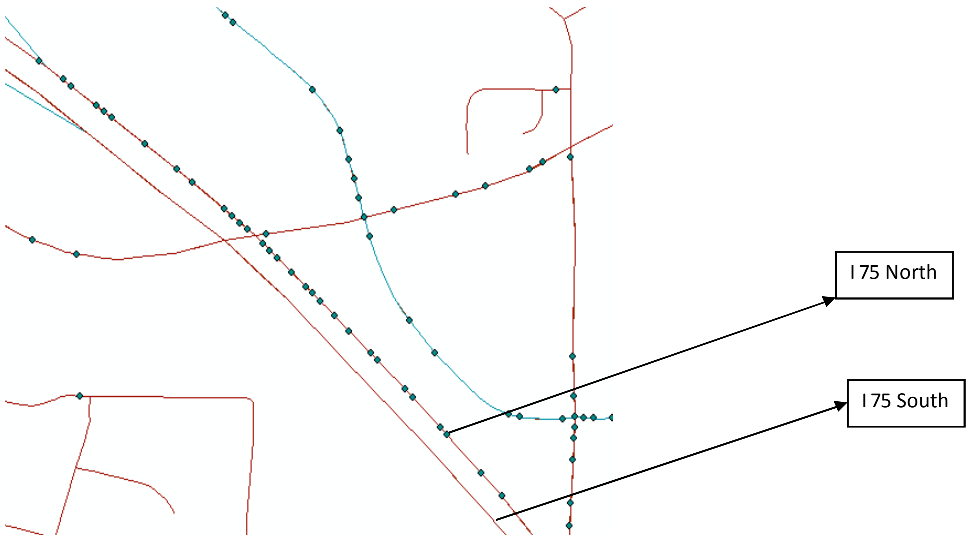


Figure 13: All the crashes are located on I 75 North and none on I 75 South

7. Coding errors were also found with the area type. Some roadways have a rural segment of 0.01 mile length in the middle of what is

otherwise coded as an urban road and vice versa. The following figure shows an example.

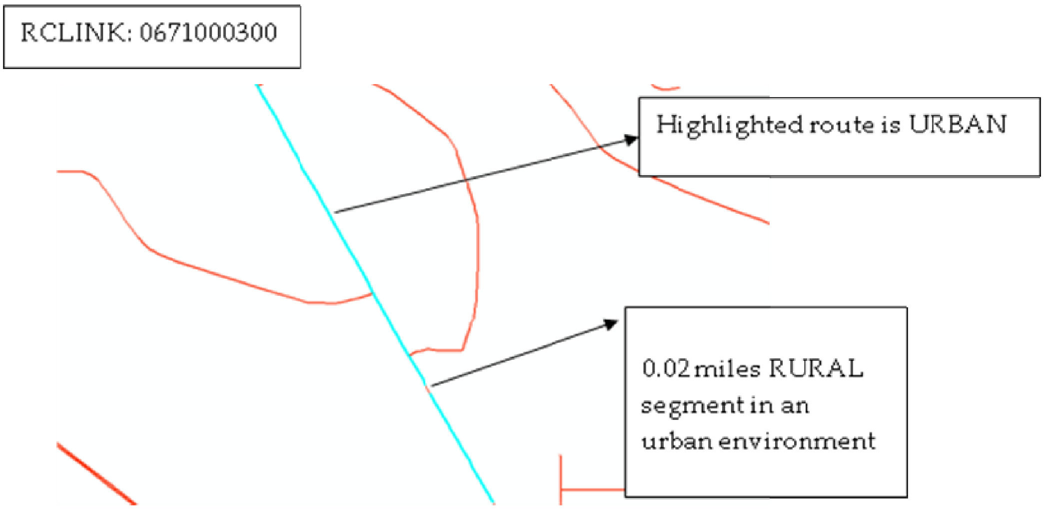


Figure 14: An example of coding error related to area type

8. Some roadway segments have missing AADTs.

9. In the Georgia roadway characteristics file, the median width, type and shoulder width can change abruptly for short segments of 0.01 miles. This caused a number of problems while generating homogeneous segments in *SafetyAnalyst*. Hence, while generating homogeneous segments, the median width and shoulder width were not considered. *SafetyAnalyst* has included a threshold level for each of these elements, whereby, a threshold of 1 ft for shoulder width would not separate two segments if their shoulder width was within 1 ft of previous. However, this function was not used for

this analysis, as a special effort would be required to determine the sensitivity levels for these attributes.

10. For predicting AADTs for the years 2005 and 2006, *SafetyAnalyst* is assuming its own growth factors since none were available from Georgia DOT. These may or may not reflect the actual trends.

<u>Phase 2: Site Selection Methods:</u>

The site selection methods used in this project include crash frequency, crash rate, critical crash rate, Level Of Service of Safety, *SafetyAnalyst* with the default SPFs and the SPFs manually generated from Cobb County data. This phase in the analysis is divided into the following sub sections.

a) Problems that arose while generating files to be imported into *SafetyAnalyst.*

    i) Accident table

    ii) Roadway Segment table

b) Generating SPFs specific to Cobb County to be imported into *SafetyAnalyst* and to perform LOSS analysis.

c) *SafetyAnalyst* with default SPFs and with Cobb County specific SPFs

d) Perform basic site selection criteria on homogeneous segments and non homogeneous segments and compare the high ranked sites in each method

The aforementioned subsections are discussed in detail in the following paragraphs:

a) Problems that arose while generating files to be imported into *SafetyAnalyst.*

    i) Accident file:

SafetyAnalyst software is run on all the non-intersection related crashes in Cobb County. It identified and ranked the top 850 sites (sites with potential for safety improvement) based on total crashes. These 850 sites belong to all site subtypes. For further analysis, two site subtypes, rural multilane divided highways (site subtype 103) and urban multilane undivided highways (site subtype 152) were considered separately since these were the only two subtypes with significant sample size.

For all the homogenous roadway segments, ranking was completed based on crash frequency, crash rate, critical crash rate and Level Of Service of Safety. Crash rates were calculated based on the exposure (in million vehicle miles travelled) of each roadway segment. For network screening based on critical crash rate, average crash rate for each reference group of population was calculated which was used to calculate the critical crash rate. Ranking was conducted based on the difference between the observed crash rate and critical rash rate. The following table describes the ranks based on the above discussed site selection criteria.

Table 14: Ranking based on different selection criteria for all site subtypes

RANKING FOR ALL SITE SUBTYPES					
SEGMENT ID	SA using default SPFs calibrated to GA	SA using GA specific SPFs	FREQ	RATE	Critical rate
104010005480549	1	1	62	65	54
104010015661567	2	2	832	719	824
104010002560257	3	3	90	97	75
223730002890291	4	4	840	551	836
100050011111112	5	5	151	15	15
10005SP00830087	6	6	76	38	33
100030003260337...100030003370342	7	7	101	177	131
228960000720074	8	8	803	182	734
220910002360237	9	10	837	45	846
104010011561161	10	11	51	157	110
217200003220323...217200003240326	11	9	605	340	551
104070003830384	12	12	237	146	121
100030002540256...100030002890295	13	13	3	101	74
104010012121213...104010012131224	14	14	37	240	155
102800012941297...102800012981301	15	15	30	34	28
104070006240625...104070006250636	16	17	112	446	287
100030001820189...100030001890192	17	16	260	321	267
100030006430651	18	18	326	307	283
100030009490950...100030009510952	19	23	127	48	43
104010003650368	20	19	102	254	179
102800006930697	21	20	139	50	45
217820005100511	22	21	284	29	29
101760000710074	23	22	829	653	832
100030018041805...100030018121815	24	25	42	66	52
10005CO00160023.10005CO01210128	25	24	4	266	166

The above table shows the top ranked sites according to five different ranking criteria. Assuming that *SafetyAnalyst* generates the list of "true deviant sites", these were compared to site ranking lists obtained using *SafetyAnalyst* with Georgia specific SPFs crash frequency, crash rate and critical crash rate. When all the roadway segments in Cobb County are considered for ranking, none of the top 10 ranked sites identified by *SafetyAnalyst* using either default SPFs or Cobb County specific SPFs are identified by any of the basic site selection methods. This demonstrates the limitations of traditional site selection methods. However, traditional ranking methods do a relatively better job of identifying problematic sites when the sites to be analyzed are regrouped into their respective site subtypes. This observation is supported with tables 17 and 18. Along with the three ranking criteria discussed above, another ranking criteria, LOSS was also considered since analysis was conducted based on particular subtypes. LOSS cannot be used on all subtypes due to the lack of the associated SPFs. Ranking based on LOSS is conducted using both calibrated default SPFs from *SafetyAnalyst* and Cobb County specific SPFs developed manually. Tables 17 and 18 show the ranks of roadway segments for the site subtypes 103 and 152 for various ranking criteria.

Table 15: Ranking based on different selection criteria for site subtype 103 (Rural multilane divided roadways)

RANKING FOR SITE SUBTYPE 103							
SEGMENT ID	SA using default SPFs calibrated to GA	SA using GA specific SPFs	FREQ	RATE	Critical rate	LOSS with GA SPF	LOSS with default SA SPF
100030001820189...100030001890192	1	1	3	4	4	4	4
10005CO00160023...10005CO01210128	2	2	2	38	27	4	4
101200001080117...101200001220123	3	3	6	6	6	4	4
100030001510154	4	4	19	3	3	3	4
228350002790281...228350003520412	5	5	9	53	37	4	4
100030021682174...100030021742180	6	6	11	5	5	4	4
100050003250333...100050005210527	7	7	1	28	21	4	4
103600003500376...103600006050649	8	8	4	70		4	4
228350001330136...228350001360141	9	9	14	11	8	3	4
100030001240132...100030001320134	10	10	12	7	7	4	4
100030001370144...100030001490151	11	11	15	12	9	3	4
100050007930796...100050008440851	12	12	10	30	22	4	4
100050008510856...100050009460950	13	13	5	45	30	4	4
100050005270531...100050006370657	14	14	7	32	24	4	4
101200006910694	15	15	45	13	11	2	2
100050009500952...100050009870989	16	16	13	46	32	3	4
100050007230727...100050007380747	17	17	16	29	25	3	4
101200001820195...101200002780293	18	19	17	94		3	4

Table 16: Ranking based on different selection for site subtype 152 (Urban multilane undivided roadways)

ID	SA using default SPFs calibrated to GA	SA using GA specific SPFs	FREQ	RATE	CR. RATE	LOSS with GA SPFs	LOSS with SA default SPFs
10005SP00830087	1	6	22	4	4	4	4
100030003260337...100030003370342	2	7	28	27	28	4	4
217200003220323...217200003240326	3	9	126	63	81	3	3
100030002540256...100030002890295	4	13	2	16	15	4	4
100030006430651	5	18	80	56	56	3	3
100030009490950...100030009510952	6	23	32	7	7	4	4
102800005960600...102800006450650	7	26	4	25	23	4	4
100030008630869	8	27	64	36	36	3	4
100030002200224...100030002430245	9	42	25	43	40	3	4
102800004810485...102800005070509	10	39	7	19	17	4	4
100050012681275...100050012891303	11	44	79	145	107	2	2
217200005300533	12	47	37	6	6	4	4
100030006510653...100030007020703	13	49	6	30	27	3	4
10120LO04520454...10120LO05420547	14	54	10	64	46	3	3
100030005920600...100030006380643	15	57	3	23	18	4	4
100030003730377...100030003780380	16	58	73	50	47	3	4
101200011721187...101200012461250	17	74	16	70	54	3	3
100060003150324...100060003440347	18	73	21	35	34	3	4
102800003220323...102800004680473	71	71	8	79	58	2	3
101200015641569...101200016211635	73	75	9	62	45	3	3
217200005380540	78	70	85	18	21	4	4
217820002530255...217820003130318	79	82	15	54	43	3	4
100050013031306...100050013181322	85	85	23	34	33	3	4
100030003000306...100030003220326	86	88	1	9	9	4	4
217200001890191...217200002350238	108	109	5	26	25	4	4

For site subtypes 103 (rural multilane divided roadways), crash frequency identified just 3 of the top 10 ranked *SafetyAnalyst* sites, whereas rates and critical rate identified only 1 out of 10 sites. The LOSS criteria based on the *SafetyAnalyst* default SPFs identified 9 out of 10 of the top ranked

SafetyAnalyst sites and LOSS criteria based on SPFs generated specifically for Cobb County identified one out of 10 sites. It would seem to make sense that Cobb County specific SPFs would perform better than default *SafetyAnalyst* SPFs, however, the limited data used to generate Cobb County specific SPFs negatively impacts the predictive capability of the SPFs. This could be improved by using more data for generating SPFs rather than using only Cobb County data.

For site subtypes 152 (rural multilane divided roadways), crash frequency identified just 3 of the top 10 ranked *SafetyAnalyst* sites, whereas rates and critical rate identified only 2 out of 10 sites. The LOSS criteria based on the *SafetyAnalyst* default SPFs identified 8 out of 10 of the top ranked *SafetyAnalyst* sites and LOSS criteria based on SPFs generated specifically for Cobb County identified six out of 10 sites. It would be unfair to predict whether Cobb County specific SPFs perform better than default *SafetyAnalyst* SPFs based on a small sample size of just one county. This prediction could be improved by using more data for generating SPFs rather than using only Cobb County data.

Phase 3: Comparison of Safety Performance Functions generated for Cobb County and the calibrated and non calibrated SPFs used in SafetyAnalyst

One of the main objectives of this research is to generate Safety Performance Functions (SPFs) that fit Georgia data and to compare them with the SPFs used by *SafetyAnalyst. SafetyAnalyst* uses SPFs that are generated from data of northern states data (California, Minnesota, Ohio and Washington) and then calibrated with Georgia data. Hence, the SPFs generated manually using Cobb County data were compared to the non-calibrated and calibrated SPFs from *SafetyAnalyst*. Due to time and resource constraints, only the three-year (2004-2006) crash and roadway inventory data from Cobb County was used for SPFs generation.

Due to the small sample size of site subtypes in Cobb County, SPFs for only two site subtypes (103 – Rural multilane divided highways and 152-Urban multilane undivided arterials) are generated. Along with the overall SPF (generated by considering three years of data), separate SPFs are generated for each year and compared to the SPFs used in *SafetyAnalyst*.

The statistical software tool, SAS (Statistical Analysis Software) is used for generating SPFs. The predicted number of crashes is considered to be a function of the traffic volume or AADT (Average Annual Daily Traffic). Because the relationship between the traffic volumes and the predicted number of crashes is typically non-linear, the independent variable is considered to be natural logarithm of AADT. The scale factor needs to be used to normalize the crash frequency to a per mile per year basis and hence an offset/ scale parameter is used. The parameter is

$$\text{Offset} = \text{Log}(3^{*}\ \text{Segment Length}) \qquad \text{(Equation 6)}$$

Where,

3 is the number of years for which crash data is available

As explained in the methodology section, for more reliable results, all the roadway segments with less than 0.1 mile length and the roadway segments with extremely high or low AADT are excluded from running the analysis because these increase potential errors. In addition to the overall 3 year PFs for site subtypes 103 and 152, SPFs are generated for every year individually. As explained in the methodology section, the form of the equation used by *SafetyAnalyst* is:

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

And the equation generated for GDOT data is of the form:

Ln(expected Number of crashes) = Intercept + Coefficient*Ln (AADT)

➔ Expected number of crashes = $e^{(\text{intercept + coefficient * Ln (AADT)})}$

➔ Expected number of crashes = $(e^{(\text{intercept})})*\text{AADT}^{(\text{coefficient})}$

The following table shows the values of intercept, coefficient, over dispersion parameter and Freeman Tukey R^2 Coefficient for Georgia specific SPFs and the calibrated SPFs used in *SafetyAnalyst.*

Table 17: Various parameters for the SPFs used for the two site subtypes

	Site Subtype	Intercept (alpha)	coefficient (Beta)	Over dispersion parameter	R^2_{FT}
GA_SPF	103	-7.0809	1.0023	3.6284	**-0.019**
SA_SPF_calibrated	103	-5.05	0.66	0.32	**0.0364**
GA_SPF	152	-3.9323	0.7409	1.8119	**0.06**
SA_SPF_calibrated	152	-10.24	1.29	0.85	**0.0874**

Freeman Tukey R^2 value is smaller for both the site subtypes. But, lower R^2 values are considered to be acceptable since the expected crashes are predicted as a function of AADT alone. It is observed from the past research that many variables like speed, weather, age of driver, etc. influence predictions of expected crashes, however, these are not considered in the model for simplicity sake and to maintain model forms accepted by SafetyAnalyst software.

For site subtype 103, R^2_{FT} value for Georgia specific SPF is 0.13 while the SPFs used by *SafetyAnalyst* has an R^2 of 0.27. This suggests that the calibrated SPFs used by *SafetyAnalyst* better fit Cobb County data. The lower fit by Georgia SPFs could be explained by the small sample size. However, for site subtype 152, as explained by the negative R^2_{FT} *SafetyAnalyst* specific SPFs do not represent the Georgia data well. Crashes on urban roads were explained well by Georgia specific SPF and this could be backed up with a positive R^2_{FT} value. The graphs in the following sheets explain how well each SPF fits the Cobb County data. The graphs also show the SPFs calibrated by *SafetyAnalyst*.

For 2006, the calibration factors calibrated by *SafetyAnalyst* and to predict yearly SPFs for site subtypes 103 and 152 are 3.597162 and 1.84415 respectively. When the *SafetyAnalyst* default SPF is plotted against the GDOT data, the default SPF falls well below the observed crashes. Hence, a calibration factor of greater than 1.00 is expected. To test to see if the data vary greatly on a yearly basis, calibration factors for the 3-year models were compared. The calibration factors for each year generated from Cobb County data and obtained from *SafetyAnalyst* are compared in the following table.

Table 18: Year wise calibration factors generated by SafetyAnalyst and manually from Cobb County data

	Site subtype 103		Site subtype 152	
Year	**SafetyAnalyst**	**Manual Calculation**	**SafetyAnalyst**	**Manual Calculation**
2004	3.629126	0.9136	1.98708	1.1036
2005	3.442261	0.8976	1.953128	0.9191
2006	3.597162	0.8104	1.844125	0.9398

The default and calibrated SPFs from *SafetyAnalyst* and Cobb County 3-year SPFs are plotted against the observed crashes. All the graphs are plotted with AADT on the X-axis and expected and observed crashes (in crashes per mile per year) on the Y-axis. Expected crashes refer to SPFs and the observed crashes refer to Cobb County site scatter points. For better visibility and consistency, the maximum value on Y-axis is kept constant at 200 crashes per mile per year and all the observed crashes beyond 200 crashes per mile per year are clipped. Rural multilane divided highways have higher AADT and the maximum AADT that is shown on graph is 400,000 vehicles/day. Urban multilane undivided arterials have a comparatively less AADT and the maximum AADT that is shown on graph is 60,000 vehicles/day. Consistency is maintained throughout the graphs with colors.

The following table describes the colors used to plot various SPFs.

Table 19: Color-coding used in the following graphs

Color	SPF
Black	Non calibrated SPF used in SafetyAnalyst
Green	SPF used in *SafetyAnalyst* that is calibrated to Georgia data and for a particular year
Blue	Non calibrated SPF manually generated for Georgia using three year crash data
Orange	SPF manually generated for Georgia using three year crash data calibrated for a particular year

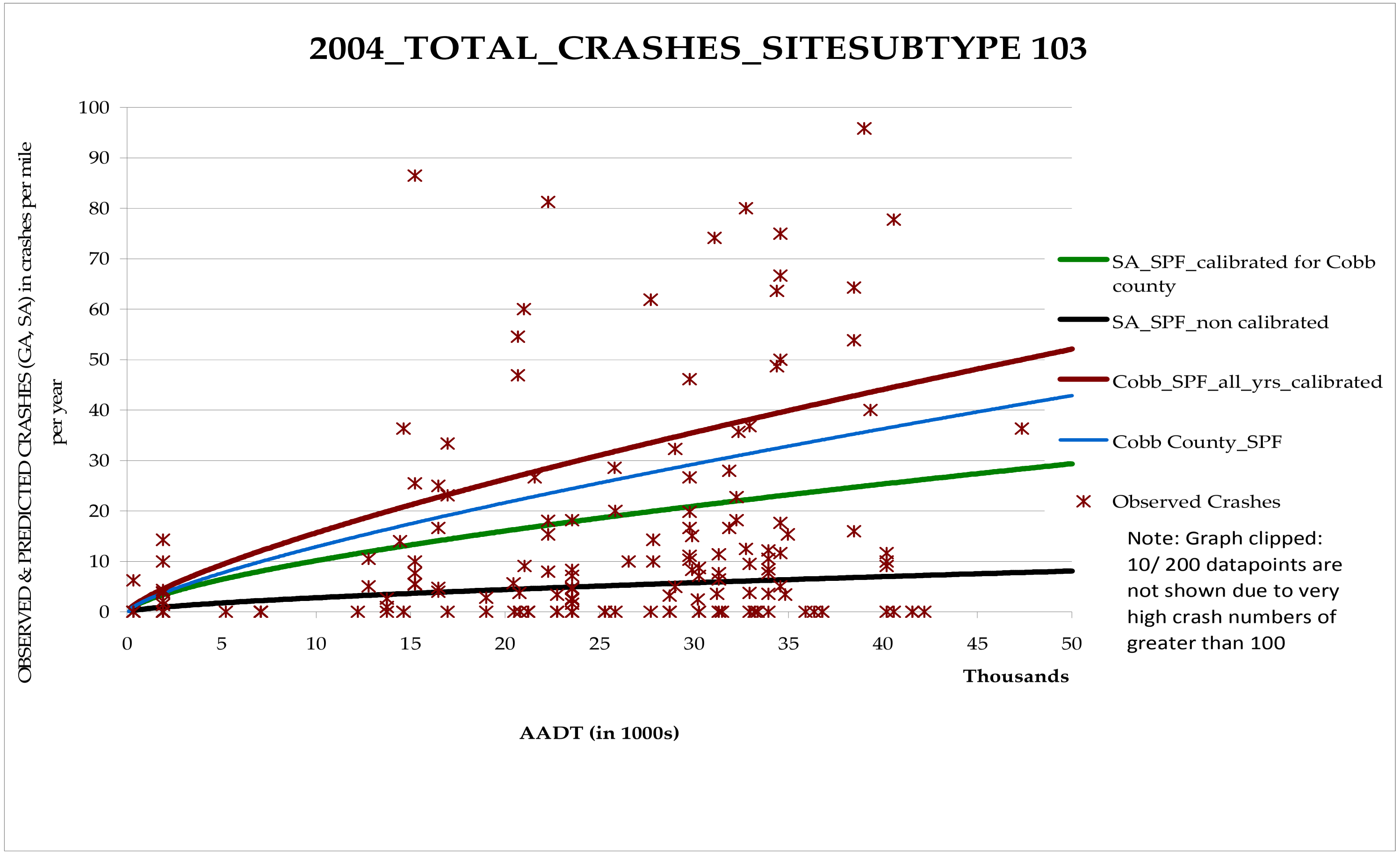


Figure 15: The calibrated and non calibrated SPFs (used by SA and generated for Georgia) for the year 2004 for site subtype 103 considering total crashes

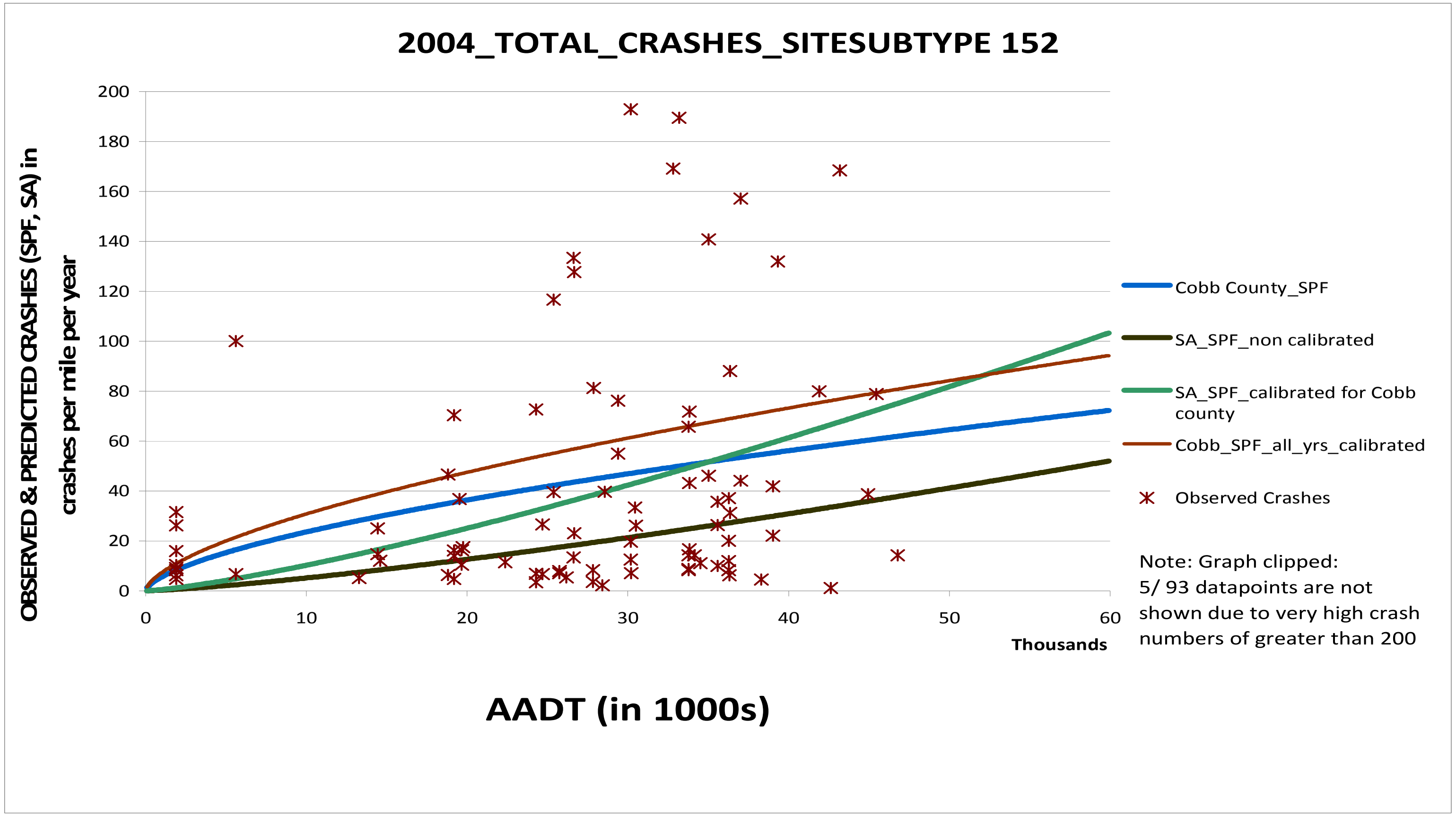


Figure 16: The calibrated and non calibrated SPFs (used by SA and generated for Georgia) for the year 2004 for site subtype 152 considering total crashes

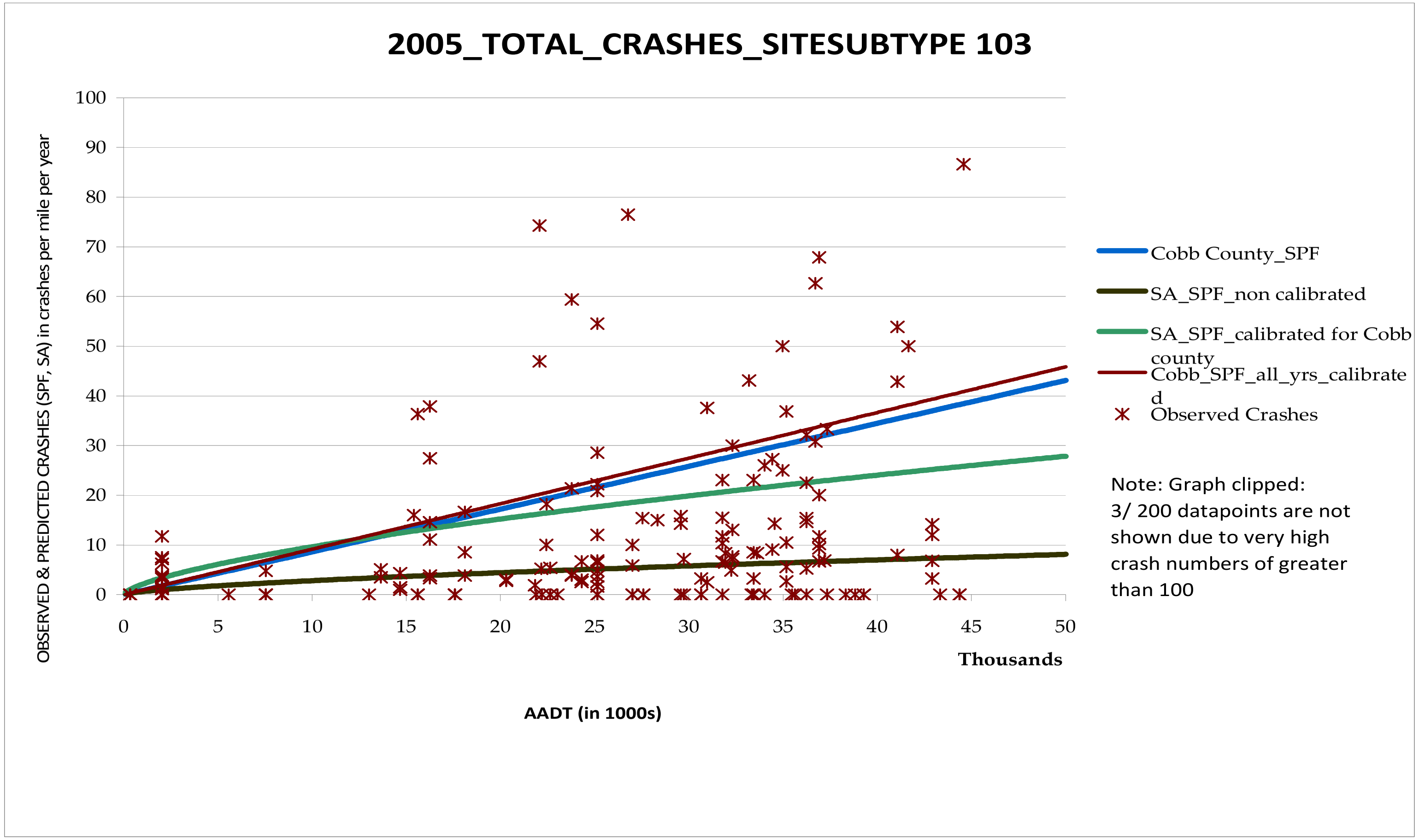


Figure 17: The calibrated and non calibrated SPFs (used by SA and generated for Georgia) for the year 2005 for site subtype 103 considering total crashes

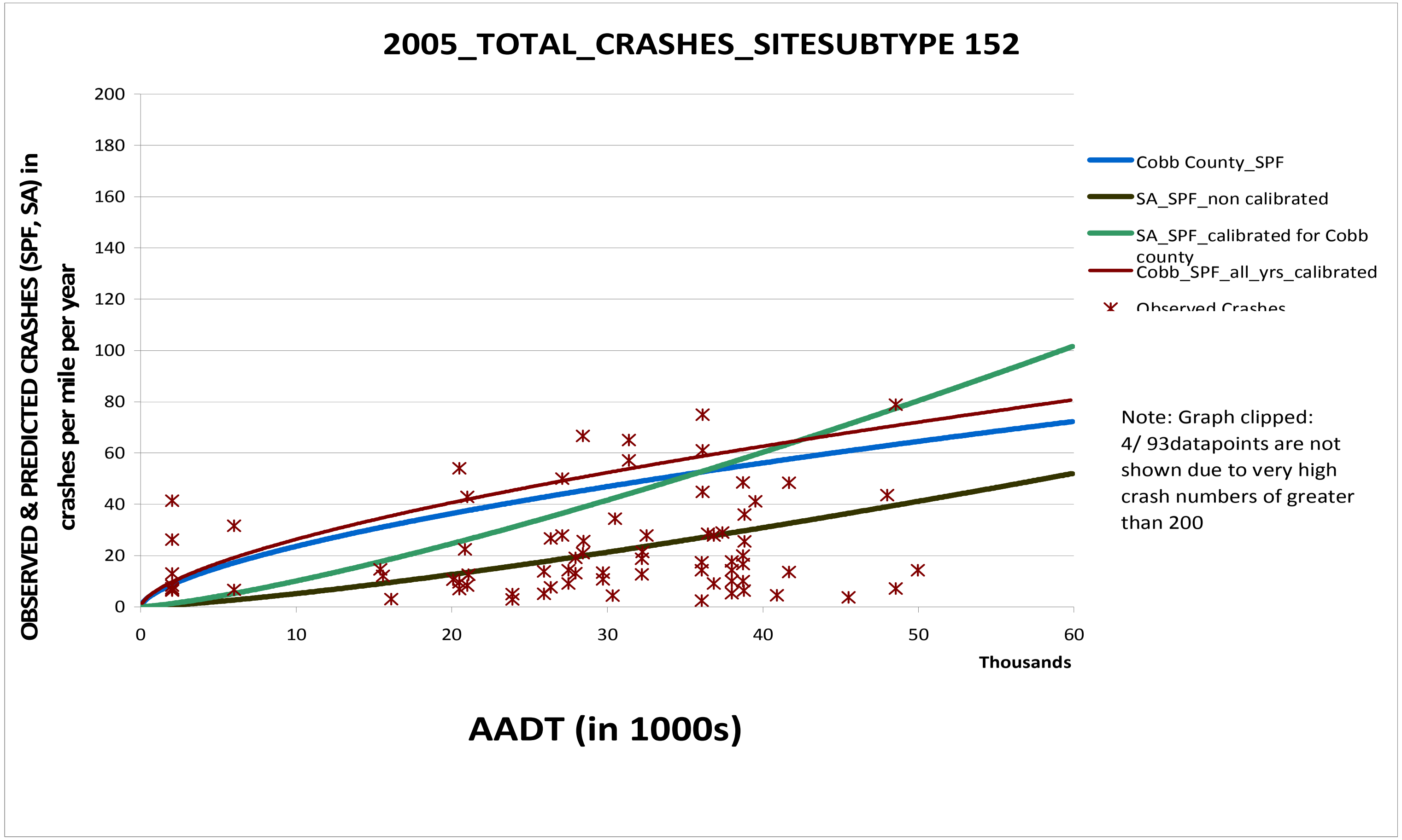


Figure 18: The calibrated and non calibrated SPFs (used by SA and generated for Georgia) for the year 2005 for site subtype 152 considering total crashes

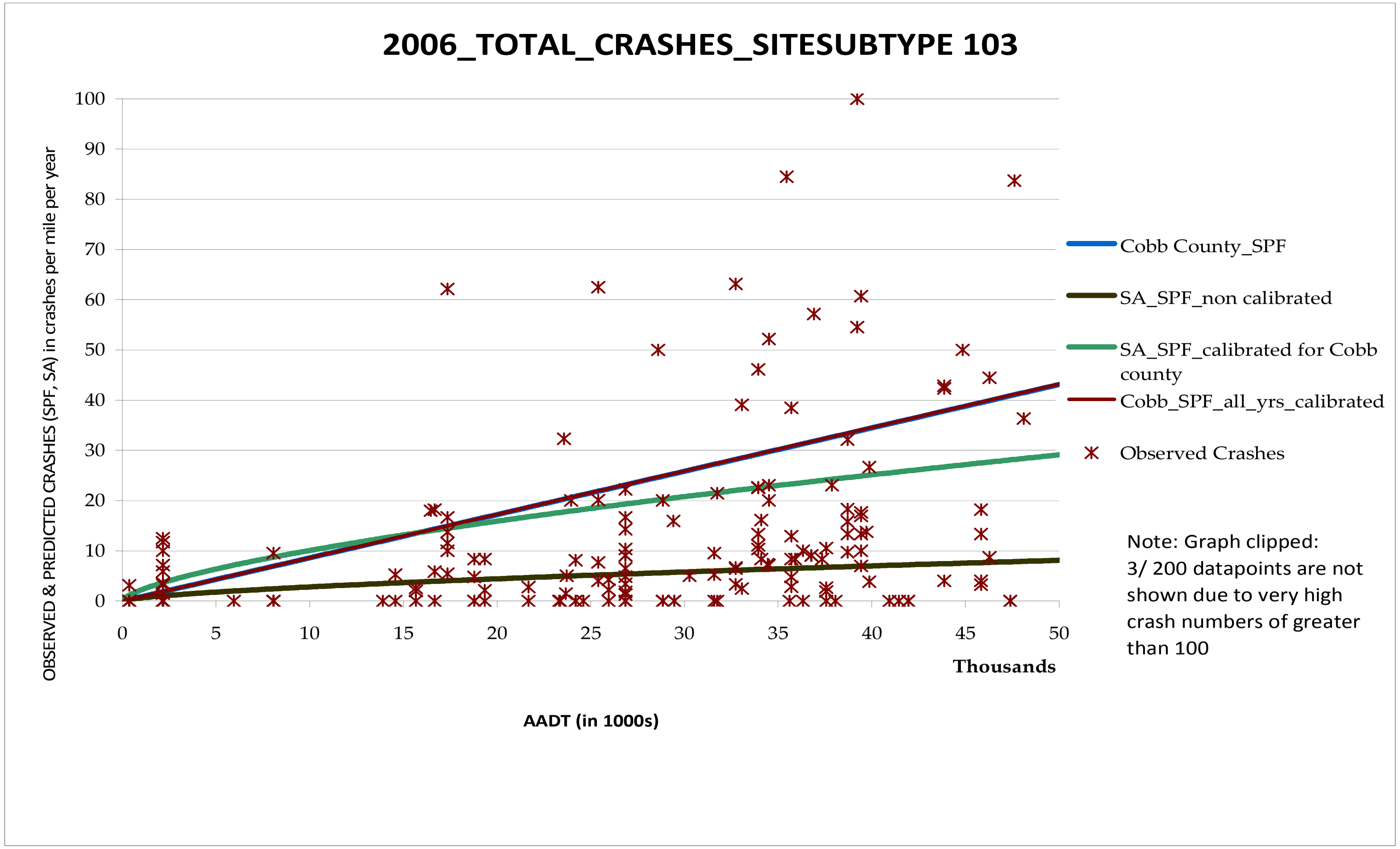


Figure 19: The calibrated and non calibrated SPFs (used by SA and generated for Georgia) for the year 2006 for site subtype 103 considering total crashes

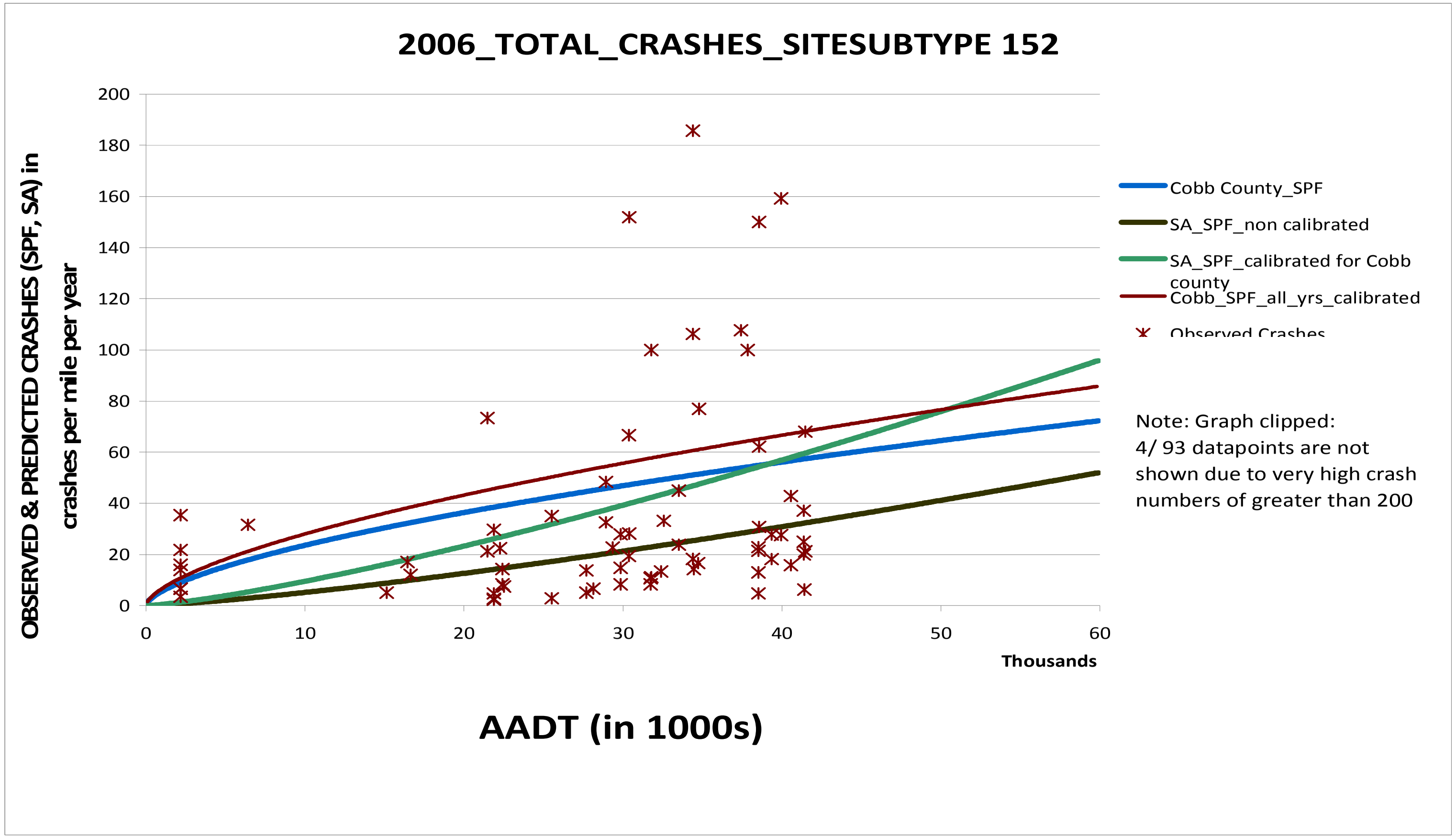


Figure 20: The calibrated and non calibrated SPFs (used by SA and generated for Georgia) for the year 2006 for site subtype 152 considering total crashes

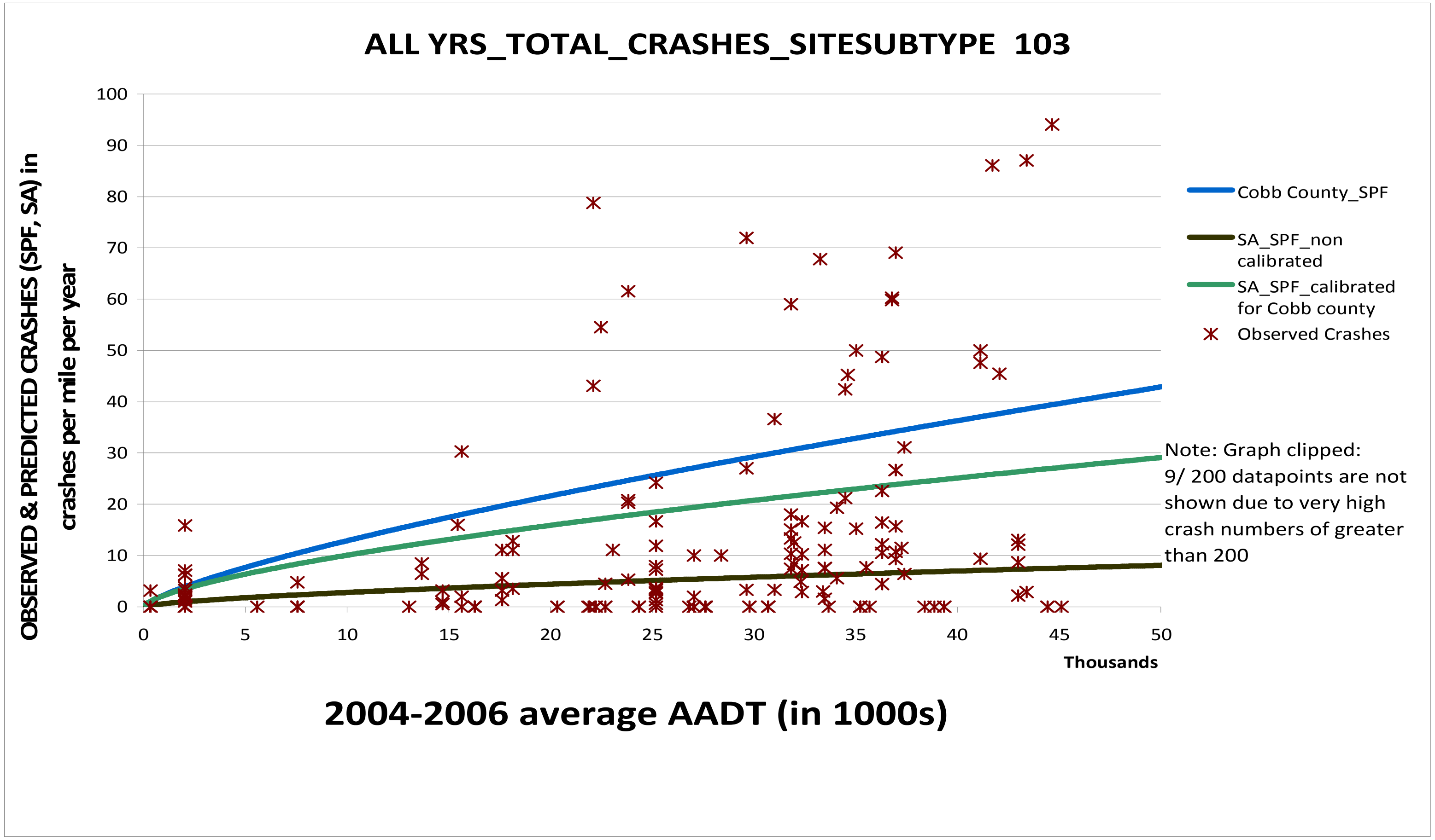


Figure 21: The calibrated and non calibrated SPFs (used by SA and generated for Georgia) for three years 04-06 for site subtype 103 considering total crashes

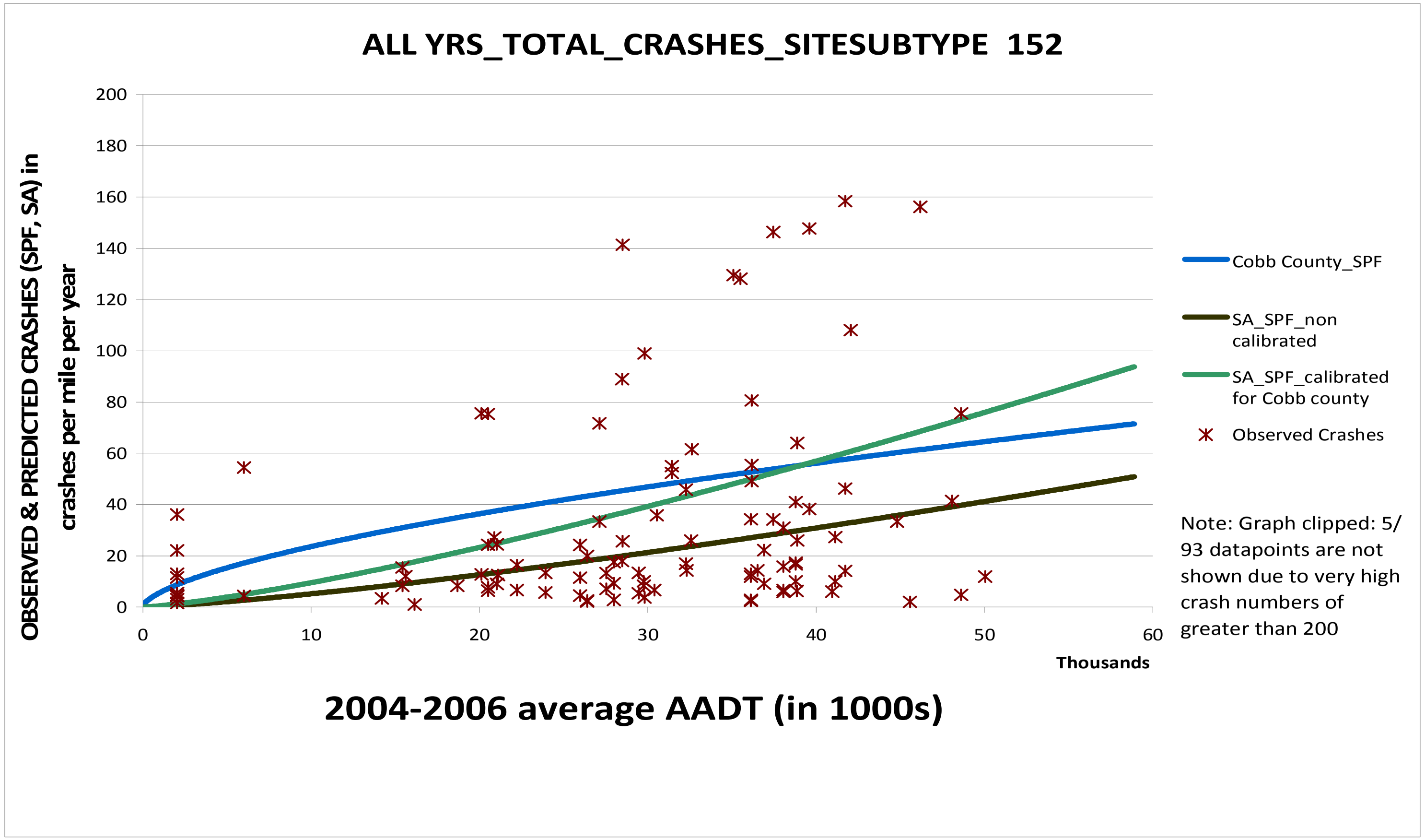


Figure 22: The calibrated and non calibrated SPFs (used by SA and generated for Georgia) for three years 04-06 for site subtype 152 considering total crashes

Figures 15, 17, 19 represent the SPFs and the observed crashes for the site subtype 103 for the three years 2004, 2005 and 2006 respectively while figure 21 represents the average crashes for three years (04-06) for site subtype 103. From the graphs, it is clear that the shapes of the SPFs used in *SafetyAnalyst* are similar to those generated for Georgia. However, S*afetyAnalyst* SPFs fit the data well compared to Cobb County specific SPFs. The fit of SPFs generated from Cobb County data could be improved by using more years of crash data and from more counties within the state.

Figures 16, 18, 20 represent the site subtype 152 for the three years 04,05 and 06 respectively while figure 22 represent the average crashes for three years (04-06) for site subtype 152. From the graphs, it is clear that the shapes of the SPFs are slightly different between those used in *SafetyAnalyst* and those generated for Cobb County, mostly due to the high crash sites in lower AADT levels. The R^2_{FT} coefficient for the *SafetyAnalyst* SPF is negative, while the SPF generated based on Cobb County data is positive, although neither has a particularly good fit. Additional data and sites would likely improve the results.

While the fit of the default *SafetyAnalyst* SPFs and Cobb County SPFs are not the same and one appears to be better than the other, no significant differences are apparent in the rankings produced by the different SPFs as shown in table 22 and table 23 for site subtypes 103 and 152 respectively.

Table 20: Ranking differences in the high ranked sites between GA specific SPFs and *SafetyAnalyst* specific SPFs for Site subtype 103

S No	ID	Site subtype	SA_SPF Rank	GA Rank
1	100030001820189...100030001890192	103	17	16
2	10005CO00160023...10005CO01210128	103	25	24
3	101200001080117...101200001220123	103	30	29
4	100030001510154	103	41	38
5	228350002790281...228350003520412	103	52	51
6	100030021682174...100030021742180	103	54	52
7	100050003250333...100050005210527	103	65	62
8	103600003500376...103600006050649	103	72	65
9	228350001330136...228350001360141	103	74	69
10	100030001240132...100030001320134	103	90	86

Table 21: Ranking differences in the high ranked sites between GA specific SPFs and *SafetyAnalyst* specific SPFs for Site subtype 152

S No	ID	Site subtype	Fed Rank	GA Rank
1	100030003260337...100030003370342	152	7	7
2	217200003220323...217200003240326	152	11	9
3	100030002540256...100030002890295	152	13	13
4	100030006430651	152	18	18
5	100030009490950...100030009510952	152	19	23
6	102800005960600...102800006450650	152	26	26
7	100030008630869	152	27	27
8	100030002200224...100030002430245	152	39	42
9	102800004810485...102800005070509	152	42	39
10	100050012681275...100050012891303	152	46	44

Phase 4: Consider benefits and costs for all ranking criteria:

The various costs required to use various ranking methods are briefly discussed in the following table:

Table 22: Resources and expertise required for various ranking methods

SITE SELECTION METHOD	**RESOURCES**				
	Time to clean data and to import data (hrs)	Expertise required	Time to run analysis (hrs)	Number of people reqd	Resources
Frequency	6	Entry level safety analyst + GIS professional	1	2	Computer, GIS
Crash rate	6	Entry level safety analyst + GIS professional	1	2	Computer, GIS
Critical Crash rate	6	Mid level safety analyst + GIS professional	2	2	Computer, GIS
LOSS	6 (assuming that SPFs exist)	Senior level safety analyst + GIS professional+ Senior level statistician	10	3	Computer, GIS, SAS
SafetyAnalyst for the first time	80	Expert + GIS professional	10	2	Computer, *SafetyAnalyst*, GIS, Access or other DBMS
SafetyAnalyst: Repetition	20	Intermediate + GIS professional	4	2	Computer, *SafetyAnalyst*, GIS, Access or other DBMS

The above table very briefly summarizes the minimum resources required for selecting sites using each ranking method. GIS is required for using every method mostly to determine the number of crashes occurring on each roadway segment. Entry level safety analyst is required for identifying SWiP based on crash frequency and crash rate. Critical crash rate requires a safety analyst with mid-level skills and a GIS professional. The Level Of Service of Safety method requires the use of SPFs, the development of which requires the expertise of a senior level statistician. A senior level safety analyst can use the LOSS methodology without statistician assuming that the SPFs for each subtype are provided. Compared to basic traditional ranking methods, LOSS and *SafetyAnalyst* require many resources and expertise to select sites. In addition to the requirements for LOSS, *SafetyAnalyst* also requires a safety analyst with proficiency in Microsoft Access. Both require a GIS specialist.

In terms of methodological limitations, *SafetyAnalyst* is assumed to be the best method for identifying SWiP because it addresses to some of the major drawbacks of traditional methods. It accounts for Regression-to-mean effect and unlike in crash rate method, linear relationship between observed number of crashes and AADT is not considered thus identifying better SWiP.

Network Screening is one of the many modules that are capable within the *SafetyAnalyst.* Diagnosis and countermeasure selection and countermeasure evaluation could also be done more systematically. The Empirical Bayes approach used in *SafetyAnalyst* is considered to be the best available method for identifying sites with greater potential for safety improvement. *SafetyAnalyst* approach is repeatable and defensible. Some of the issues dealing with small segment lengths are dealt in *SafetyAnalyst* since *SafetyAnalyst* generates homogeneous segments, thus reducing the number of shorter segments and also increasing the length of similar roadway segments. Subdivision of roadway segments based on the type of facility improves the results of the basic ranking criteria like frequency and rate.

Moreover, several types of analysis could be done with *SafetyAnalyst* very easily once the data is imported and calibrated. The process is tedious and time consuming only for the first time and its repetition doesn't require the same amount of work.

CHAPTER 5

CONCLUSIONS AND FUTURE RECOMMENDATIONS

5.1 Conclusions:

From reviewing the literature and the past work that is carried out in the area of network screening and site selection, it is clear that the conventional methods of selecting "sites with potential for safety improvement" has their own drawbacks and limitations. However, most of the DOTs use conventional methods like crash frequency and crash rate to identify SWiP resulting in improper site selection and lesser safety effect for the money spent. This research project reinforces the fact that advanced site selection methods like the use of Empirical Bayes approach, generation of Safety Performance Functions and the use of software like *SafetyAnalyst* addresses most of the limitations of traditional methods. *SafetyAnalyst* is state-of-the-art analytical tool to identify and rank SWiP, prioritize safety improvements, suggest countermeasures and evaluate countermeasures.

Cobb County is considered for analysis for this project. Most of the conventional methods and advanced site selection methods are compared to obtain the top priority sites for safety improvement. *SafetyAnalyst* uses rigorous calculations, Empirical Bayes approach and SPFs to predict the expected number of crashes in the future and to rank sites based on PSI (Potential for Safety Improvement). Assuming that the SWiP identified by *SafetyAnalyst* are the sites with greatest potential for safety improvement, These ranks are compared to the ranks obtained by frequency, rate and critical crash rate and it is found that only 50% of the top ranked crashes in *SafetyAnalyst* are identified in all the other conventional ranking criteria.

It is seen in the results that conventional ranking criteria used on a particular reference group of roadway segments yield more reliable results compared to the ranking on all site subtypes. However, serious drawbacks like Regression-to-mean and shorter segment length exists resulting in increasing the unreliability of traditional ranking methods. Use of advanced ranking criteria helps in identifying sites with greater "potential for safety improvement". Of many advanced selection criteria, *SafetyAnalyst* is a state-of-the-art analytical tool that could be used to identify and rank SWiP. This software uses SPFs generated using northern state data for the years 1997-2002. These SPFs are calibrated to the data used (for Cobb County data in this project). However, most of the factors like traffic trends, accident patterns, climate, population, geography etc change considerably among different regions. Hence, same SPFs (either calibrated or non calibrated) might not represent the "same" relationship between AADT and predicted crashes. Therefore, SPFs for each state, need to be developed and used in *SafetyAnalyst* to better identify and rank problematic sites. This observation is backed up in this research project where Cobb County data is used to compare the basic and advanced site selection criteria. The non calibrated and calibrated SPFs used in *SafetyAnalyst* are compared to the Cobb County specific SPFs generated and found that the SPFs differ considerably reinforcing the idea of generating SPFs from Georgia data to be used in *SafetyAnalyst.*

For the objectives set forth in this research project, the following conclusions are drawn:

a) Review data availability, format and completeness for use in different safety data analysis methods

- GDOT has sufficient data to conduct the basic ranking criteria. But, for advanced ranking criteria, SPFs are required along with the classification of roadway into subtypes. However, these are unavailable for Georgia.
- For LOSS, all the sites need to be divided into site subtypes and SPFs generated. This requires a lot of time and data resources.
- For *SafetyAnalyst*, the data requirements are intense. Georgia has most of the data. However, the data needs to be recoded to the format required to be imported into *SafetyAnalyst*.

b) **Assess whether safety performance functions employed in *SafetyAnalyst* software can be properly calibrated to reflect crash distribution and conditions in Georgia**

- The default SPFs used in *SafetyAnalyst* are generated from northern states data and they don't seem to fit well with GA data. This can be explained by larger yearly calibration factors for GA data. This reinforces the need for Georgia specific SPFs.
- The SPFs manually generated from Cobb County data do not fit well enough compared to the calibrated default SPFs. This is mainly because of lesser data. Conclusions cannot be drawn about the fit of SPFs on complete state just from using one county data. When the complete state's data is used in *SafetyAnalyst,* the SPF's fit might be improved.

c) **Analyze costs and potential benefits of implementing and maintaining various methods (crash frequency, crash rate, Level Of Service of Safety and Empirical Bayes method using *SafetyAnalyst*) for selecting and prioritizing problematic crash sites by implementing these methods for Cobb County using 2004-2006 crash data.**

- The basic site selection methods are easier to implement compared to the advanced methods. Entry level to mid level safety analysts are sufficient to conduct the basic selection methods. However, they do not account to some of the major drawbacks like Regression-to-mean, shorter segment length, higher AADTs and random fluctuation in crash counts over time.

- LOSS, an advanced selection criteria requires a senior level safety analyst and statistician for developing SPFs and for categorizing sites into subtypes. LOSS accounts for some of the aforementioned drawbacks, but it does not account for the severity of crashes.
- *SafetyAnalyst,* the most advanced selection criteria requires comparatively more time and resources for initial setup. A senior level safety analyst is required to generate, import, post process and calibrate files required for safety analysis. Once, this is done, the process is easily repeatable compared to other methods.
- Several types of analyses could be done easily in *SafetyAnalyst* to compare different results and to prioritize sites based on the user requirements.
- Identification of sites is just the first step. Countermeasure selection and evaluation is only possible with *SafetyAnalyst*.
- The roadway segments in Georgia are divided into small segments and in Cobb County, the average segment length is 0.062miles. Such smaller segments drastically increase rates resulting in biased results. This is accounted for in *SafetyAnalyst* since it creates homogeneous

segments based on the threshold set by the users. The following table briefly mentions the potential benefits of using various ranking criteria.

5.2 Future Recommendations:

Despite of the initial and operational costs for using *SafetyAnalyst* for network screening, it could be concluded that it better identifies and ranks sites with potential for safety improvements since it uses the most advanced and data driven Empirical Bayes approach which accounts for most of the drawbacks of basic screening methods. However, the default SPFs used in *SafetyAnalyst* were developed from northern states (California, Minnesota, Ohio and Washington). It is evident that the traffic trends, crash patterns, geography, etc are completely different in the south when compared to the north. Hence, the SPFs developed from the northern states might not exactly fit the southern crash data. *SafetyAnalyst* uses a calibration factor to fit the default SPFs to Georgia data. However, a calibration factor of about 1.00 might represent a good fit which is not the case. Higher calibration factors and graphs of the default and calibrated SPFs plotted against the observed crashes along with the R square values reinforce the fact that the calibrated SPFs do not fit the Georgia data well. In this context, SPFs were manually generated for Cobb County and were compared to the default and calibrated SPFs used in *SafetyAnalyst.* Even these SPFs do not represent the data well.

This is evidently seen from the graphs and R square values. However, conclusions cannot be drawn just based on the results obtained from this research since only one county data is used for SPF generation.

The future research for the present study might include the use of data from the whole state for generating SPFs manually and for checking the fit of the default and calibrated SPFs used in *SafetyAnalyst*. When the complete state is considered for analysis, the calibrated SPFs might fit the data well discouraging the idea of generating Georgia specific SPFs.

In this research, SPFs were generated manually considering the form of default SPFs used in *SafetyAnalyst* as a basis. This might not be the best way to develop SPFs for a southern state like Georgia since we are confining the dependant and independent variables and also the relation between them. The future research might include a study on the relationship between the dependant and independent variables.

In this research, two site subtypes (rural multilane divided roadways and urban multilane undivided roadways) are considered for generating SPFs. Sites with low AADT and high crashes are fewer in number, but, their influence is enormous and to some extent define the shapes of SPFs. Hence, sensitivity analysis might be of help to determine the effect of these “outliers” on the calibration factors and SPF development.

APPENDICES

APPENDIX A: ALTACCIDENT FILE

SQL QUERIES, DATA MAPPING AND DATA RECODING

#	Name	SQL	Description	Note
01	Create Accidents Table	CREATE TABLE ACCIDENT_SA (ACC_CASE Text,RTE_TYPE Text,RTE_NAME Text,CNTY_NUM Text, ACC_LOC Text,LOC_IDSYS Text,DIST_NUM Text,CITY_NUM Text, ACC_DATE Text,ACC_TIME Text,REL_JUNC Text,DRVWY_IND Text, LGT_COND Text,WTR_COND Text,SURF_COND Text,ACC_TYPE Text, CIRCUM_ENV Text,CIRCUM_ROAD Text,SCHLBUS_REL Text,WRKZN_REL Text, NUM_VEH Text,ACC_SEV1 Text,ACC_SEV2 Text,ALC_DRUG Text, TOW_IND Text,RUNOFF_IND Text,PED_IND Text,BIKE_IND Text, DIVHWY_FLAG Text,RTENUM_DISP Text,MLPOST_DISP Text,VEHTURN_MVT Text, DAY_WK Text,RDSEG_NUM Text,INTR_NUM Text,RAMP_NUM Text, INIT_DIR1 Text,VEH_MAN1 Text,VEH_CONF1 Text,FIRST_EVNT1 Text, DRVR_AGE1 Text,INIT_DIR2 Text,VEH_MAN2 Text,VEH_CONF2 Text, FIRST_EVNT2 Text,DRVR_AGE2 Text);	Creates the table ACCIDENT_SA	
02	Insert Inital	INSERT INTO ACCIDENT_SA (ACC_CASE, CITY_NUM, ACC_DATE, ACC_TIME, LGT_COND, WTR_COND, SURF_COND, ACC_TYPE, CIRCUM_ROAD, WRKZN_REL, NUM_VEH, RUNOFF_IND, PED_IND, DAY_WK) SELECT ACC_ID, ACC_ICO_TYPE, ACC_JULDT, ACC_ATIME, ACC_LITE_TYPE, ACC_WEAT_TYPE, ACC_SURF_TYPE, ACC_HE1_TYPE, ACC_RDD_TYPE, ACC_RDD_TYPE, ACC_TNV, ACC_LOI_TYPE, ACC_HE1_TYPE, ACC_DAYOFWEEK_TYPE FROM ACCIDENT_TBL;	Inserts all the one-to-one matches from ACCIDENT_TBL	
03	Update RTE_TYPE	UPDATE ACCIDENT_SA INNER JOIN LOCATION_TBL ON ACCIDENT_SA.ACC_CASE=LOCATION_TBL.LOC_ACC_ID SET ACCIDENT_SA.RTE_TYPE = LOCATION_TBL.LOC_ROUTE_TYPE;	Updates route type based on a match with the accident ID in the LOCATION_TBL	

#	Name	SQL	Description	Note
04	Update RTE_NAME	UPDATE ACCIDENT_SA INNER JOIN LOCATION_TBL ON ACCIDENT_SA.ACC_CASE=LOCATION_TBL.LOC_ACC_ID SET ACCIDENT_SA.RTE_NAME = LOCATION_TBL.LOC_ROUTE_IDENTIFIER+LOCATION_TBL.LOC_ROUTE_SUFFIX;	Updates route name based on a match with the accident ID in the LOCATION_TBL	
05	Update ACC_LOC	UPDATE ACCIDENT_SA INNER JOIN LOCATION_TBL ON ACCIDENT_SA.ACC_CASE=LOCATION_TBL.LOC_ACC_ID SET ACCIDENT_SA.ACC_LOC = LOCATION_TBL.LOC_ACC_MILELOG;	Updates accident location based on a match with the accident ID in LOCATION_TBL	
06	Update DIST_NUM	UPDATE ACCIDENT_SA SET ACCIDENT_SA.DIST_NUM = '7';	Sets district number to "7"	This will need to be adjusted for other counties' data sets.
07_0	Update REL_JUNC - Inter related	UPDATE ACCIDENT_SA INNER JOIN LOCATION_TBL ON ACCIDENT_SA.ACC_CASE=LOCATION_TBL.LOC_ACC_ID SET ACCIDENT_SA.REL_JUNC = LOCATION_TBL.LOC_INTERROUTE_IDENTIFIER;	Updates relationship to junction to LOC_INTERROUTE_IDENTIFIER based on a match with accident ID in LOCATION_TBL	

#	Name	SQL	Description	Note
07_1	Update REL_JUNC - At Inter	UPDATE ACCIDENT_SA INNER JOIN LOCATION_TBL ON ACCIDENT_SA.ACC_CASE=LOCATION_TBL.LOC_ACC_ID SET ACCIDENT_SA.REL_JUNC = LOCATION_TBL.LOC_SIGNAL_TYPE;	Updates relationship to junction to LOC_SIGNAL_TYPE based on a match with accident ID in LOCATION_TBL	
07_2	Update REL_JUNC - At Ramp	UPDATE ACCIDENT_SA INNER JOIN RAMP_TBL ON ACCIDENT_SA.ACC_CASE=RAMP_TBL.RMP_ACC_ID SET ACCIDENT_SA.REL_JUNC = 'Ramp';	Updates relationship to junction to "Ramp" based on a match with accident ID in RAMP_TBL	
07_3	Update REL_JUNC - At RRX	UPDATE ACCIDENT_SA INNER JOIN RRX_TBL ON ACCIDENT_SA.ACC_CASE=RRX_TBL.RRX_ACC_ID SET ACCIDENT_SA.REL_JUNC = 'Rrx';	Updates relationship to junction to "Rrx" based on a match with accident ID in RRX_TBL	
08	Update ACC_TYPE	UPDATE ACCIDENT_SA INNER JOIN ACCIDENT_TBL ON ACCIDENT_SA.ACC_CASE=ACCIDENT_TBL.ACC_ID SET ACCIDENT_SA.ACC_TYPE = ACCIDENT_TBL.ACC_MNRC_TYPE;	Updates accident type to ACC_MNRC_TYPE based on a match with accident ID in ACCIDENT_TBL	

#	Name	SQL	Description	Note
09_0	Update CIRCUM_ENV - Factor4	UPDATE ACCIDENT_SA INNER JOIN VEHICLE_TBL ON ACCIDENT_SA.ACC_CASE=VEHICLE_TBL.VEH_ACC_ID SET ACCIDENT_SA.CIRCUM_ENV = VEHICLE_TBL.VEH_CONF4_TYPE;	Updates circum env to VEH_CONF4_TYPE based on a match with accident id in VEHICLE_TBL	
09_1	Update CIRCUM_ENV - Factor3	UPDATE ACCIDENT_SA INNER JOIN VEHICLE_TBL ON ACCIDENT_SA.ACC_CASE=VEHICLE_TBL.VEH_ACC_ID SET ACCIDENT_SA.CIRCUM_ENV = VEHICLE_TBL.VEH_CONF3_TYPE;	Updates circum env to VEH_CONF3_TYPE based on a match with accident id in VEHICLE_TBL	
09_2	Update CIRCUM_ENV - Factor2	UPDATE ACCIDENT_SA INNER JOIN VEHICLE_TBL ON ACCIDENT_SA.ACC_CASE=VEHICLE_TBL.VEH_ACC_ID SET ACCIDENT_SA.CIRCUM_ENV = VEHICLE_TBL.VEH_CONF2_TYPE;	Updates circum env to VEH_CONF2_TYPE based on a match with accident id in VEHICLE_TBL	
09_3	Update CIRCUM_ENV - Factor1	UPDATE ACCIDENT_SA INNER JOIN VEHICLE_TBL ON ACCIDENT_SA.ACC_CASE=VEHICLE_TBL.VEH_ACC_ID SET ACCIDENT_SA.CIRCUM_ENV = VEHICLE_TBL.VEH_CONF1_TYPE;	Updates circum env to VEH_CONF1_TYPE based on a match with accident id in VEHICLE_TBL	
09_4	Update CIRCUM_ENV - Vision	UPDATE ACCIDENT_SA INNER JOIN VEHICLE_TBL ON ACCIDENT_SA.ACC_CASE=VEHICLE_TBL.VEH_ACC_ID SET ACCIDENT_SA.CIRCUM_ENV = VEHICLE_TBL.VEH_VOBS_TYPE;	Updates circum env to VEH_VOBS_TYPE based on a match with accident id in VEHICLE_TBL	

#	Name	SQL	Description	Note
10	Update SCHLBUS_REL	UPDATE ACCIDENT_SA INNER JOIN VEHICLE_TBL ON ACCIDENT_SA.ACC_CASE=VEHICLE_TBL.VEH_ACC_ID SET ACCIDENT_SA.SCHLBUS_REL = VEHICLE_TBL.VEH_CLASS_TYPE;	Updates schoolbus related to VEH_CLASS_TYPE based on a match with accident ID in VEHICLE_TBL	
11_0	Update ACC_SEV1 - Pedestrian	UPDATE ACCIDENT_SA INNER JOIN PEDESTRIAN_TBL ON ACCIDENT_SA.ACC_CASE=PEDESTRIAN_TBL. PED_ACC_ID SET ACCIDENT_SA.ACC_SEV1 = PEDESTRIAN_TBL.PED_INJC_TYPE;	Updates accident severity1 to PED_INJC_TYPE based on a match with accident ID in PEDESTRIAN_TBL	
11_1	Update ACC_SEV1 - Passenger	UPDATE ACCIDENT_SA INNER JOIN PASSENGER_TBL ON ACCIDENT_SA.ACC_CASE=PASSENGER_TBL.OCC_ACC_ID SET ACCIDENT_SA.ACC_SEV1 = PASSENGER_TBL.OCC_INJC_TYPE;	Updates accident severity1 to OCC_INJC_TYPE based on a match with accident ID in PASSENGER_TBL	
11_2	Update ACC_SEV1 - OccDrvier	UPDATE ACCIDENT_SA INNER JOIN OCCDRIVER_TBL ON ACCIDENT_SA.ACC_CASE=OCCDRIVER_TBL.OCC_ACC_ID SET ACCIDENT_SA.ACC_SEV1 = OCCDRIVER_TBL. OCC_INJC_TYPE;	Updates accident severity1 to OCC_INJC_TYPE based on a match with accident ID in OCCDRIVER_TBL	
12_0	Update ALC_DRUG - Pedestrian	UPDATE ACCIDENT_SA INNER JOIN PEDESTRIAN_TBL ON ACCIDENT_SA.ACC_CASE=PEDESTRIAN_TBL.PED_ACC_I D SET ACCIDENT_SA.ALC_DRUG = PEDESTRIAN_TBL.PED_DRVCND_TYPE;	Updates alcohol drug involvement to PED_DRVCND_TYPE based on a match with accident ID in PEDESTRIAN_TBL	

#	Name	SQL	Description	Note
12_1	Update ALC_DRUG - OccDriver	UPDATE ACCIDENT_SA INNER JOIN OCCDRIVER_TBL ON ACCIDENT_SA.ACC_CASE=OCCDRIVER_TBL.OCC_ACC_ID SET ACCIDENT_SA.ALC_DRUG = OCCDRIVER_TBL.OCC_DRVCND_TYPE;	Updates alcohol drug involvement to OCC_DRVCND_TYPE based on a match with accident ID in OCCDRIVER_TBL	
13	Update BIKE_IND	UPDATE ACCIDENT_SA INNER JOIN VEHICLE_TBL ON ACCIDENT_SA.ACC_CASE=VEHICLE_TBL.VEH_ACC_ID SET ACCIDENT_SA.BIKE_IND = VEHICLE_TBL.VEH_TYPE_TYPE;	Updates bike indicator to VEH_TYPE_TYPE based on a match with accident ID in VEHCILE_TBL	
14	Update RDSEG_NUM	UPDATE ACCIDENT_SA INNER JOIN LOCATION_TBL ON ACCIDENT_SA.ACC_CASE=LOCATION_TBL.LOC_ACC_ID SET ACCIDENT_SA.RDSEG_NUM = LOCATION_TBL.LOC_RCLINK_IDENTIFIER;	Updates roadseg number to LOC_RCLINK_IDENTIFIER based on a match with accident ID in LOCATION_TBL	
15	Update INTR_NUM	UPDATE ACCIDENT_SA INNER JOIN LOCATION_TBL ON ACCIDENT_SA.ACC_CASE=LOCATION_TBL.LOC_ACC_ID SET ACCIDENT_SA.INTR_NUM = LOCATION_TBL.LOC_INTERROUTE_TYPE+LOCATION_TBL.LOC_INTERROUTE_IDENTIFIER+LOCATION_TBL.LOC_INTERROUTE_SUFFIX+LOCATION_TBL.LOC_SIGNAL_TYPE;	Updates intersection number to a concatenation based on a match with accident ID in LOCATION_TBL	

#	Name	SQL	Description	Note
16	Update RAMP_NUM	UPDATE ACCIDENT_SA INNER JOIN RAMP_TBL ON ACCIDENT_SA.ACC_CASE=RAMP_TBL.RMP_ACC_ID SET ACCIDENT_SA.RAMP_NUM = RAMP_TBL.RMP_INTERCHANGE_ADDIDENTIFIER+RAMP_TBL.RMP_INTERCHANGE_IDENTIFIER+RAMP_TBL.RMP_QUADRANT_IDENTIFIER+RAMP_TBL.RMP_RAMP_IDENTIFIER+RAMP_TBL.RMP_RAMPSECTION_IDENTIFIER;	Updates ramp number to a concatenation based on a match with accident ID in RAMP_TBL	
17_0	Update INIT_DIR1	UPDATE ACCIDENT_SA INNER JOIN VEHICLE_TBL ON ACCIDENT_SA.ACC_CASE=VEHICLE_TBL.VEH_ACC_ID SET ACCIDENT_SA.INIT_DIR1 = VEHICLE_TBL.VEH_DIRT_TYPE WHERE VEHICLE_TBL.VEH_NO='01';	Updates initial direction1 to VEH_DIRT_TYPE based on a match with accident ID in VEHICLE_TBL and VEH_NO = "01"	
17_1	Update INIT_DIR2	UPDATE ACCIDENT_SA INNER JOIN VEHICLE_TBL ON ACCIDENT_SA.ACC_CASE=VEHICLE_TBL.VEH_ACC_ID SET ACCIDENT_SA.INIT_DIR2 = VEHICLE_TBL.VEH_DIRT_TYPE WHERE VEHICLE_TBL.VEH_NO='02';	Updates initial direction2 to VEH_DIRT_TYPE based on a match with accident ID in VEHICLE_TBL and VEH_NO = "02"	
18_0	Update VEH_MAN1	UPDATE ACCIDENT_SA INNER JOIN VEHICLE_TBL ON ACCIDENT_SA.ACC_CASE=VEHICLE_TBL.VEH_ACC_ID SET ACCIDENT_SA.VEH_MAN1 = VEHICLE_TBL.VEH_MANV_TYPE WHERE VEHICLE_TBL.VEH_NO='01';	Updates initial manuever1 to VEH_MANV_TYPE based on a match with accident ID in VEHICLE_TBL and VEH_NO = "01"	

#	Name	SQL	Description	Note
18_1	Update VEH_MAN2	UPDATE ACCIDENT_SA INNER JOIN VEHICLE_TBL ON ACCIDENT_SA.ACC_CASE= VEHICLE_TBL.VEH_ACC_ID SET ACCIDENT_SA.VEH_MAN2 = VEHICLE_TBL.VEH_MANV_TYPE WHERE VEHICLE_TBL.VEH_NO='02';	Updates initial manuever2 to VEH_MANV_TYPE based on a match with accident ID in VEHICLE_TBL and VEH_NO = "02"	
19_0	Update VEH_CONF1 - ComVehicle	UPDATE ACCIDENT_SA INNER JOIN COMVEHICLE_TBL ON ACCIDENT_SA.ACC_CASE=COMVEHICLE_TBL.COMV_ACC_ID SET ACCIDENT_SA.VEH_CONF1 = COMVEHICLE_TBL.COMV_CONFIG_TYPE WHERE COMVEHICLE_TBL.COMV_VEHNO='01';	Updates vehicle config1 to COMV_CONFIG_TYPE based on a match with accident ID in COMVEHICLE_TBL and COMV_VEHNO = "01"	
19_1	Update VEH_CONF1 - Vehicle	UPDATE ACCIDENT_SA INNER JOIN VEHICLE_TBL ON ACCIDENT_SA.ACC_CASE=VEHICLE_TBL.VEH_ACC_ID SET ACCIDENT_SA.VEH_CONF1 = VEHICLE_TBL.VEH_TYPE_TYPE WHERE VEHICLE_TBL.VEH_NO='01';	Updates vehicle config1 to VEH_TYPE_TYPE based on a match with accident ID in VEHICLE_TBL and VEH_NO = "01"	
19_2	Update VEH_CONF2 - ComVehicle	UPDATE ACCIDENT_SA INNER JOIN COMVEHICLE_TBL ON ACCIDENT_SA.ACC_CASE=COMVEHICLE_TBL.COMV_ACC_ID SET ACCIDENT_SA.VEH_CONF2 = COMVEHICLE_TBL.COMV_CONFIG_TYPE WHERE COMVEHICLE_TBL.COMV_VEHNO='02';	Updates vehicle config2 to COMV_CONFIG_TYPE based on a match with accident ID in COMVEHICLE_TBL and COMV_VEHNO = "02"	

#	Name	SQL	Description	Note
19_3	Update VEH_CONF2 - Vehicle	UPDATE ACCIDENT_SA INNER JOIN VEHICLE_TBL ON ACCIDENT_SA.ACC_CASE=VEHICLE_TBL.VEH_ACC_ID SET ACCIDENT_SA.VEH_CONF2 = VEHICLE_TBL.VEH_TYPE_TYPE WHERE VEHICLE_TBL.VEH_NO='02';	Updates vehicle config2 to VEH_TYPE_TYPE based on a match with accident ID in VEHICLE_TBL and VEH_NO = "02	
20_0	Update FIRST_EVNT1	UPDATE ACCIDENT_SA INNER JOIN VEHICLE_TBL ON ACCIDENT_SA.ACC_CASE=VEHICLE_TBL.VEH_ACC_ID SET ACCIDENT_SA.FIRST_EVNT1 = VEHICLE_TBL.VEH_MHE_TYPE WHERE VEHICLE_TBL.VEH_NO='01';	Updates first event1 to VEH_MHE_TYPE based on a match with accident ID in VEHICLE_TBL and VEH_NO = "01"	
20_1	Update FIRST_EVNT2	UPDATE ACCIDENT_SA INNER JOIN VEHICLE_TBL ON ACCIDENT_SA.ACC_CASE=VEHICLE_TBL.VEH_ACC_ID SET ACCIDENT_SA.FIRST_EVNT2 = VEHICLE_TBL.VEH_MHE_TYPE WHERE VEHICLE_TBL.VEH_NO='02';	Updates first event2 to VEH_MHE_TYPE based on a match with accident ID in VEHICLE_TBL and VEH_NO = "02"	
21_0	Update DRVR_AGE1	UPDATE ACCIDENT_SA INNER JOIN OCCDRIVER_TBL ON ACCIDENT_SA.ACC_CASE=OCCDRIVER_TBL.OCC_ACC_ID SET ACCIDENT_SA.DRVR_AGE1 = OCCDRIVER_TBL.OCC_DOB WHERE OCCDRIVER_TBL.OCC_VEHNO='01';	Updates driver age1 to OCC_DOB based on a match with accident ID in OCCDRIVER_TBL and OCC_VEHNO = "01"	

#	Name	SQL	Description	Note
21_1	Update DRVR_AGE2	UPDATE ACCIDENT_SA INNER JOIN OCCDRIVER_TBL ON ACCIDENT_SA.ACC_CASE=OCCDRIVER_TBL.OCC_ACC_ID SET ACCIDENT_SA.DRVR_AGE2 = OCCDRIVER_TBL.OCC_DOB WHERE OCCDRIVER_TBL.OCC_VEHNO='02';	Updates driver age2 to OCC_DOB based on a match with accident ID in OCCDRIVER_TBL and OCC_VEHNO = "02"	
22	Update CNTY_NUM	UPDATE ACCIDENT_SA INNER JOIN LOCATION_TBL ON ACCIDENT_SA.ACC_CASE=LOCATION_TBL.LOC_ACC_ID SET ACCIDENT_SA.CNTY_NUM = LOCATION_TBL.LOC_COUNTY_IDENTIFIER;	Updates county number to lOC_COUNTY_IDENTIFIER based on a match with accident ID in LOCATION_TBL	

	SafetyAnalyst		GDOT				
S No.	**SA field name**	**SA Code**	**Existing/ To be created**	**GDOT field Name**	**Data from table**	**Criteria for selection**	**GDOT Code**
1	AgencyID	unique number for each accident	Existing	Microfilm	ACCIDENT	-	-
2	LocSystem	A - Route/milepost B - Route/county/milepost C - Route/section identifier/distance D – Segment identifier/distance	New field: To be created	-	-	-	-
3	RouteType	I – Interstate US - US route SR - State route BR - Business route BL - Business loop SP - Spur route CR - County road L - Local road O - Other NA - Not applicable X – Unknown	Existing and to be changed	LOC_ROUTE_TYPE	LOCATION	-	0-Accident Not Located 1-State Route 2-County Road 3-City Street 8-Public Road 9-Collector-Distributor
4	routeName	# or name of the route	To be created	LOC_ROUTE_IDENTIFIER + LOC_ROUTE_SUFFIX	LOCATION	-	-
5	county	county number	Existing	Accident County	ACCIDENT	"067"	-
6	locOffset	- (a floating number)	Existing	LOC_ACC_MILELOG	LOCATION	-	-
7	accidentDate	YYYYMMDD	Existing	ACC_DATE	ACCIDENT	-	-
8	accidentTime	HHMM (24 hr clock)	Existing and to be changed	ACC_TIME	ACCIDENT	-	-
9	accident Severity1	K - Fatal Injury A - Incapacitating Injury B - Non-Incapacitating Injury C - Possible Injury P - Property-Damage-Only X – Unknown	Existing and to be changed	PED_INJC_TYPE, OCC_INJC_TYPE, OCC_INJC_TYPE	ACCIDENT, PEDESTRIAN, PASSENGER OCCDRIVER	-	0 - Not Injured 1 - Killed 2 - Serious 3 - Visible Injury 4 - PDO

	SafetyAnalyst		GDOT				
S No.	**SA field name**	**SA Code**	**Existing/ To be created**	**GDOT field Name**	**Data from table**	**Criteria for selection**	**GDOT Code**
10	numberOf Fatalities	total # of fatalities in each crash	Existing	ACC_TNF	ACCIDENT	-	-
11	numberOf Injuries	total # of injuries in each crash	Existing	ACC_TNI	ACCIDENT	-	-
12	junction Relationship	1 - Non-junction 2 - At intersection 3 - Intersection-related 4 - At driveway or driveway-related 5 - Entrance/exit ramp 6 - Other part of interchange 7 - Railroad/highway grade crossing 8 - Crossover related 9 - Other 99 – Unknown	To be created	LOC_INTERROUTE_IDENTIFIER, LOC_SIGNAL_TYPE	LOCATION	-	A – STOP B - Over Head Flashing Amber C-All direction Stop Sign F - Flasher other than Overhead L - Traffic Control Device with turn arrow O - Stop sign in Opposite direction of inventory P - Traffic control with Pedestrian signal R - Overhead Flashing Red Ramp - Ramp Rrx- Railroad crossing S - Traffic Control Sign W- Yield Sign in Opposite direction of inventory Y – Yield
13	driveway Indicator	1 - No - No 2 - Yes, at driveway 3 - Yes, near driveway 99 – Unknown	To be created	-	-	-	-
14	light Condition	1 - Daylight 2 - Dawn 3 - Dusk 4 - Dark-lighted 5 - Dark-not lighted 6 - Dark-unknown lighting 7 - Other 99 – Unknown	Existing and to be changed	ACC_LITE_TYPE	ACCIDENT	-	1-Daylight 2-Dusk 3-Dawn 4-Dark-Lighted 5-Dark-Not Lighted

	SafetyAnalyst		GDOT				
S No.	SA field name	SA Code	Existing/ To be created	GDOT field Name	Data from table	Criteria for selection	GDOT Code
15	weather Condition	1 - Clear 2 - Cloudy 3 - Fog, smog, smoke 4 - Rain 5 - Sleet, hail (freezing rain or drizzle) 6 - Snow 7 - Blowing snow 8 - Severe crosswinds 9 - Blowing sand, soil, dirt 10 - Other 99 – Unknown	Existing and to be changed	ACC_WEAT_TYPE	ACCIDENT	-	1- Clear 2- Cloudy 3-Rain 4-Snow 5-Sleet 6-Fog 7-Other
16	surface Condition	1 - Dry 2 - Wet 3 - Snow 4 - Slush 5 - Ice/frost 6 - Water (standing, moving) 7 - Sand 8 - Mud, dirt, gravel 9 - Oil 10 - Other 99 – Unknown	Existing and to be changed	ACC_SURF_TYPE	ACCIDENT	-	1-Dry 2-Wet 3-Snowy 4-Icy 5-Other

	SafetyAnalyst		GDOT				
S No.	SA field name	SA Code	Existing/ To be created	GDOT field Name	Data from table	Criteria for selection	GDOT Code
17	collision Type	1 - Collision with parked motor vehicle 2 - Collision with railroad train 3 - Collision with bicyclist 4 - Collision with pedestrian 5 - Collision with animal 6 - Collision with fixed object 7 - Collision with other object 8 - Other single-vehicle collision 9 - Overturn 10 - Fire or explosion 11 - Other single-vehicle non-collision 21 - Rear-end 22 - Head-on 23 - Rear-to-rear 24 - Angle 25 - Sideswipe, same direction 26 - Sideswipe, opposite direction 27 - Other multiple-vehicle collision 99 - Unknown	Existing and to be changed	ACC_MNRC_TYPE	ACCIDENT	-	1-Angle 2-Head On 3-Rear End 4-Sideswipe - Same Direction 5-Sideswipe - Opposite Direction 6-Not A Collision With A Motor Vehicle

	SafetyAnalyst		GDOT				
S No.	**SA field name**	**SA Code**	**Existing/ To be created**	**GDOT field Name**	**Data from table**	**Criteria for selection**	**GDOT Code**
18	environment Condition	1 - None 2 - Weather conditions 3 - Physical obstruction(s) 4 - Glare 5 - Animal(s) in roadway 6 - Other 99 – Unknown	Existing and to be changed	VEH_CONF1_TYPE, VEH_CONF2_TYPE, VEH_CONF3_TYPE, VEH_CONF4_TYPE	VEHICLE	-	01-No Contributing Factors 02-D.U.I 03-Following too Close 04-Failed to Yield 05-Exceeding Speed Limit 06-Disregard Stop Sign/Signal 07-Wrong Side of Road 08-Weather Conditions 09-Improper Passing 10-Driver Lost Control 11-Changed Lanes Improperly 12-Object or Animal 13-Improper Turn 14-Parked Improperly 15-Mechanical or Vehicle Failure 16-Surface Defects 17-Misjudged Clearance 18-Improper Backing 19-No Signal/Improper Signal 20-Driver Condition 21-Driverless Vehicle 22-Too Fast for Conditions 23-Improper Passing of School Bus 24-Disregard Police Officer 25-Distracted 26-Other

	SafetyAnalyst			GDOT			
S No.	**SA field name**	**SA Code**	**Existing/ To be created**	**GDOT field Name**	**Data from table**	**Criteria for selection**	**GDOT Code**
19	road Condition	1 - None 2 - Road surface condition (wet, icy, snow, slush, etc.) 3 - Debris 4 - Rut, holes, bumps 5 - Work zone (construction/maintenance/utility) 6 - Worn, travel-polished surface 7 - Obstruction in roadway 8 - Traffic control device inoperative, missing, or obscured 9 - Shoulders 10 - Non-highway work 11 - Other 99 – Unknown	Existing and to be changed	ACC_RDD_TYPE	ACCIDENT	-	1-No Defects 2-Defective Shoulders 3-Holes, Deep Ruts, Bumps 4-Loose Material on Surface 5-Water Standing 6-Road Under Construction 7-Running Water 8-Other
20	school Bus	1 - No 2 - Yes, school bus directly involved 3 - Yes, school bus indirectly involved 99 – Unknown	To be created	VEH_CLASS_TYPE	VEHICLE	VEH_CLASS_TYPE = 4	0-Unknown 1-Privately Owned 2-Police 3-Fire 4-School 5-Other Government Owned 6-Military 7-Commercial Vehicle 8-Other 9-Commercial Vehicle (No Carrier ID Available)
21	work Zone	Y - Yes N - No X – Unknown	Existing and coding error				X
22	num Vehicles	total number of vehicles involved	Existing	ACC_TNV	ACCIDENT	-	-

	SafetyAnalyst		GDOT				
S No.	SA field name	SA Code	Existing/ To be created	GDOT field Name	Data from table	Criteria for selection	GDOT Code
23	drug Involved	1 - Neither alcohol nor other drugs 2 - Yes (alcohol) 3 - Yes (drugs) 4 - Yes (alcohol and drugs) 99 – Unknown	Existing and to be changed	OCC_DRVCND_TYPE	OCCUPANT	-	1-Not Drinking 2-Not Known if U.I 3-Drinking, Not Impaired 4-U.I. Alcohol 5-U.I. Drugs 6-U.I. Alcohol and Drugs 7-Physical Impairment 8-Apparently Fell Asleep
24	v1initial Travel Direction	NB - Northbound SB - Southbound EB - Eastbound WB - Westbound NO - Not on roadway XX - Unknown	Existing and to be changed	VEH_DIRT_TYPE	VEHICLE	VEH_NUM = 1	1-North 2-South 3-East 4-West
25	v2initial Travel Direction		Existing and to be changed	VEH_DIRT_TYPE	VEHICLE	VEH_NUM = 2	
26	v1vehicle Maneuver	1 - Movements essentially straight ahead 2 - Backing 3 - Changing lanes 4 - Overtaking/passing 5 - Turning right 6 - Turning left 7 - Making U-turn 8 - Entering traffic lane 9 - Leaving traffic lane 10 - Parked 11 - Slowing 12 - Stopped in traffic 13 - Negotiating a curve 14 - Other 99 – Unknown	Existing and to be changed	VEH_MANV_TYPE	VEHICLE	VEH_NUM = 1	01-Turning Left 02-Turning Right 03-Making U-Turn 04-Stopped 05-Straight 06-Changing Lanes 07-Backing 08-Parked 09-Passing 10-Negotiating a Curve 11-Entering/Leaving Parking 12-Entering/Leaving Driveway
27	v2vehicle Maneuver		Existing and to be changed	VEH_MANV_TYPE	VEHICLE	VEH_NUM = 2	

	SafetyAnalyst			GDOT			
S No.	**SA field name**	**SA Code**	**Existing/ To be created**	**GDOT field Name**	**Data from table**	**Criteria for selection**	**GDOT Code**
28	v1vehicle Configuration	1 - Passenger car 2 - Light truck, only four tires 3 - Sport utility vehicle 4 - Motorcycle/Moped 5 - Motor home/recreational vehicle 6 - Single-unit truck - 2-axle and GVWR over 10,000 pounds 7 - Single-unit truck - 3-or-more axles 8 - Truck pulling trailer or trailers 9 - Truck tractor (bobtail) 10 - Truck tractor/semi-trailer 11 - Truck tractor/doubles 12 - Truck tractor/triples 13 - Truck over 10,000 pounds, cannot classify 14 - Bus/large van - Seats for more than 15 people, including driver 15 - Bus - Seats for 7-15 people, including driver 16 - Emergency vehicle - Fire, police, ambulance 17 - Other 99 - Unknown vehicle configuration	Existing and to be changed	VEH_TYPE_TYPE	VEHICLE	VEH_NUM = 1	01-Passenger Car 02-Pickup Truck 03-Truck Tractor (Bobtail) 04-Tractor/Trailer 05-Tractor W/Twin Trailers 06-Logging Truck 07-Logging Tractor/Trailer 08-Single Unit Truck 09-Panel Truck 10-Van 11-Utility Passenger Vehicle 12-Vehicle With Trailer 13-Bus 14-Truck Towing House Trailer 15-Ambulance 16-Motorized Recreational Vehicle 17-Motorcycle, Scooter, Minibike 18-Moped 19-Pedalcycle, Bicycle 20-Farm or Construction Equipment 21-All Terrain Vehicle 22-Other
29	v2vehicle Configuration		Existing and to be changed	VEH_TYPE_TYPE	VEHICLE	VEH_NUM = 2	

	SafetyAnalyst		GDOT				
S No.	SA field name	SA Code	Existing/ To be created	GDOT field Name	Data from table	Criteria for selection	GDOT Code
30	v1first Event	1 - Overturn/rollover - Noncollision 2 - Fire/explosion - Noncollision 3 - Immersion - Noncollision 4 - Jackknife - Noncollision 5 - Cargo/equipment loss or shift - Noncollision 6 - Fell/jumped from vehicle - Noncollision 7 - Thrown or falling object - Noncollision 8 - Other noncollision 9 - Unknown noncollision 10 - Pedestrian - Collision 11 - Bicyclist - Collision 12 - Railway vehicle - Collision 13 - Animal - Collision 14 - Motor vehicle in transport - Collision 15 - Parked motor vehicle - Collision 16 - Work zone maintenance equipment - Collision 17 - Other non-fixed object - Collision 18 - Unknown non-fixed object - Collision 19 - Impact attenuator/crash cushion - Collision Note: SA Code continued on next row	Existing and to be changed	VEH_MHE_TYPE	VEHICLE	VEH_NUM = 1	01-Overturn 02-Fire/Explosion 03-Immersion 04-Jackknife 05-Other Non-Collision 06-Pedestrian 07-Pedalcycle 08-Railway Train 09-Animal 10-Parked Motor Vehicle 11-Motor Vehicle in Motion 12-Motor Vehicle in Motion - In Other Roadway 13-Other Object (Not Fixed) 14-Deer 15-Impact Attenuator 16-Bridge Pier/Abutment 17-Bridge Parapet End Note: GDOT code continued on next row

	SafetyAnalyst		GDOT				
S No.	SA field name	SA Code	Existing/ To be created	GDOT field Name	Data from table	Criteria for selection	GDOT Code
31	v2first Event	Note: SA Code continued from previous row 20 - Bridge overhead structure - Collision 21 - Bridge pier or support - Collision 22 - Bridge rail - Collision 23 - Culvert - Collision 24 - Curb - Collision 25 - Ditch - Collision 26 - Embankment - Collision 27 - Guardrail face - Collision 28 - Guardrail end - Collision 29 - Concrete traffic barrier (Jersey barrier) - Collision 30 - Other traffic barrier - Collision 31 - Standing tree - Collision 32 - Utility pole/Light support - Collision 33 - Highway traffic sign or signpost - Collision 34 - Overhead sign or sign support - Collision 35 - Other post, pole, or support - Collision 36 - Fence - Collision 37 - Mailbox - Collision 38 - Other fixed object - Collision 39 - Unknown fixed object - Collision 41 - Other 99 – Unknown	Existing and to be changed	VEH_MHE_TYPE	VEHICLE	VEH_NUM = 2	Note: GDOT code continued form previous row 18-Bridge Rail 19-Guardrail Face 20-Guardrail End 21-Median Barrier 22-Highway Traffic Sign Post 23-Overhead Sign Support 24-Luminaire/Light Support 25-Utility Pole 26-Other Post 27-Culvert 28-Curb 29-Ditch 30-Embankment 31-Fence 32-Mailbox 33-Tree 34-Other Fixed Object
32	v1driver DOB	YYYYMMDD	Existing	OCC_DOB	OCCUPANT	VEH_NUM = 1	
33	v2driver DOB	YYYYMMDD	Existing	OCC_DOB	OCCUPANT	VEH_NUM = 2	

	AltAccident		
Sno	Field Name	Mapping required??	GDOT Code --- SA Code
1	agencyID	No	
2	locSystem	No	
3	routeType	Yes	1 ----- SR 2 ----- CR 3 ----- L 7 ----- O 8 ----- O 9 ----- O Inter—SR
4	routeName	No	
5	county	No	
6	locOffset	No	
7	accidentDate	No	
8	accidentTime	No	
9	accidentSeverity1	Yes	0 ----- O 1 ----- K 2 ----- A 3 ----- B 4 ----- C
10	numberOfFatalities	No	
11	numberOfInjuries	No	
12	junctionRelationship	Yes	A ----- 2 B ----- 2 C ----- 2 F ----- 2 L ----- 2 O ----- 2 P ----- 2 R ----- 2 Ramp ----- 5 Rrx ----- 7 S ----- 2 W ----- 1 Y ----- 1 "Blank" ----- 99

S no	Field Name	Mapping required??	GDOT Code --- SA Code
13	DrivewayIndicator	No	
14	lightCondition	Yes	1 ----- 1 2 ----- 3 3 ----- 2 4 ----- 4 5 ----- 5
15	weatherCondition	Yes	1 ----- 1 2 ----- 2 3 ----- 4 4 ----- 6 5 ----- 5 6 ----- 3 7 ----- 10
16	surfaceCondition	Yes	1 ----- 1 2 -----2 3 ----- 3 4 ----- 5 5 ----- 10 6 ----- 8 7 -----7 8 ----- 4 9 ----- 9
17	collisionType	Yes	1 ----- 24 2 ----- 22 3 ----- 23 4 ----- 25 5 ----- 26 6 ----- 8
18	environmentCondition	Yes	1 ----- 1 2 ----- 4 3 ----- 4 4 ----- 3 5 ----- 3 6 ----- 2 7 ----- 6

<table>
<tr><th>Sno</th><th>Field Name</th><th>Mapping required??</th><th>GDOT Code --- SA Code</th></tr>
<tr><td>19</td><td>roadCondition</td><td>Yes</td><td>1 ----- 1
2 ----- 9
3 ----- 4
4 ----- 6
6 ----- 5
8 ----- 11</td></tr>
<tr><td>20</td><td>schoolBus</td><td>Yes</td><td>VEH_CLASS_TYPE = 4</td></tr>
<tr><td>21</td><td>workZone</td><td>No</td><td></td></tr>
<tr><td>22</td><td>numVehicles</td><td>No</td><td></td></tr>
<tr><td>23</td><td>drugInvolved</td><td>Yes</td><td>1 ----- 1
2 ----- 99
3 ----- 2
4 ----- 2
5 ----- 3
6 ----- 4</td></tr>
<tr><td>24</td><td>v1initialTravelDirection</td><td>Yes</td><td>1 ----- NB
2 ----- SB
3 ----- EB
4 ----- WB</td></tr>
<tr><td>25</td><td>v2initialTravelDirection</td><td>Yes</td><td>1 ----- NB
2 ----- SB
3 ----- EB
4 ----- WB</td></tr>
<tr><td>26</td><td>v1vehicleManeuver</td><td>Yes</td><td rowspan="2">1 ----- 6
2 ----- 5
3 ----- 7
4 ----- 12
5 ----- 1
6 ----- 3
7 ----- 2
8 ----- 10
9 ----- 4
10 ----- 13
11 ----- 14
12 ----- 14</td></tr>
<tr><td>27</td><td>v2vehicleManeuver</td><td>Yes</td></tr>
</table>

Sno	Field Name	Mapping required??	GDOT Code --- SA Code
28	v1vehicleConfiguration	Yes	1 ----- 1 2 ----- 2 3 ----- 9 4 ----- 10 5 ----- 11 6 ----- 17 7 ----- 17 8 ----- 6 9 ----- 2 10 ----- 14
29	v2vehicleConfiguration	Yes	11 ----- 17 12 ----- 17 13 ----- 15 14 ----- 13 15 ----- 16 16 ----- 15 17 ----- 4 18 ----- 4 19 ----- 17 20 ----- 17 21 ----- 3 22 ----- 17

<table>
<tr><th>Sno</th><th>Field Name</th><th>Mapping required??</th><th>GDOT Code --- SA Code</th></tr>
<tr><td>30</td><td>v1firstEvent</td><td>Yes</td><td rowspan="2">01 ----- 1
02 ----- 2
03 ----- 3
04 ----- 4
05 ----- 8
06 ----- 10
07 ----- 11
08 ----- 12
09 ----- 13
10 ----- 15
11 ----- 14
12 ----- 14
13 ----- 17
14 ----- 13
15 ----- 19
16 ----- 21
17 ----- 21
18 ----- 22
19 ----- 27
20 ----- 28
21 ----- 29
22 ----- 33
23 ----- 34
25 ----- 32
26 ----- 35
27 ----- 23
28 ----- 24
29 ----- 25
30 ----- 26
31 ----- 36
32 ----- 37
33 ----- 34
34 ----- 38</td></tr>
<tr><td>31</td><td>v2firstEvent</td><td>Yes</td></tr>
<tr><td>32</td><td>v1driverDOB</td><td>No</td><td></td></tr>
<tr><td>33</td><td>v2driverDOB</td><td>No</td><td></td></tr>
</table>

APPENDIX B: ALTROADWAYSEGMNT FILE

SQL QUERIES, DATA MAPPING AND DATA RECODING

#	Name	SQL	Description	Note
100	Create RC_SA Table	CREATE TABLE RC_SA (SA_ID COUNTER,COUNTY Text,ROUTE_TYPE Text,ROUTE_NUM Text,BEG_MEASURE Text,END_MEASURE Text,SECTION_LENGTH Text,DESCRIPTION Text,DISTRICT Text,MAINT_AREA Text,POPULATION Text,INVENTORY_DATE Text,DESIGNATED_WAY Text,TRUCK_ROUTE Text,TRAVEL_WAY Text,RURAL_URAN Text,SPEED_LIMIT Text,FAS_NUM Text,TRUCK_ROUTE_ID Text,CONGRESS_DIST Text,STATE_ROUTE_SEQ Text,ACCESS_CONTROL Text,OPERATION Text,TOTAL_LANES Text,SPECIAL_CLASS Text,DIV_HWY_SHLDR_WIDTH_LFT Text,DIV_HWY_SHLDR_TYPE_LFT Text,DIV_HWY_SURF_WIDTH Text,DIV_HWY_SURF_TYPE Text,DIV_HWY_SHLDR_WIDTH_RT Text,DIV_HWY_SHLDR_TYPE_RT Text,DIV_HWY_MEDIAN_WIDTH Text,DIV_HWY_MEDIAN_TYPE Text,DIV_HWY_BARRIER_TYPE Text, UDIV_HWY_SHLDR_WIDTH_LFT Text,UDIV_HWY_SHLDR_TYPE_LFT Text,UDIV_HWY_SURFACE_WIDTH Text,UDIV_HWY_SURFACE_TYPE Text,UDIV_HWY_SHLDR_WIDTH_RT Text,UDIV_HWY_SHLDR_TYPE_RT Text,AUX_LANE_WIDTH_LFT Text,AUX_LANE_TYPE_LFT Text,AUX_LANE_WIDTH_RT Text,AUX_LANE_TYPE_RT Text,MAINT_YEAR Text,MAINT_TYPE Text,IMPROVE_YEAR Text,FUNC_CLASS Text,TRAFFIC_COUNT_TYPE Text,TRAFFIC_COUNT_YEAR Text,RIGHT_OF_WAY Text,RW_TYPE Text,TC_NUMBER Text,MAINTENANCE_SUR_DES Text,SIDEWALK_LEFT Text,SIDEWALK_RIGHT Text,IMPROVE_TYPE Text,TRUCK_PERCENT Text,TRUCK_PERCENT_TYPE Text,SIGNAL Text,AADT_OLD Text,INTERSECT_ROAD2 Text,S_FUNCLASS_ID Text,DUAL_MAINT_RATING		

#	Name	SQL	Description	Note
		Text,ROAD_WIDTH Text,DIVIDED Text,OPEN_TO_TRAFFIC Text,CITY_CODE Text,T_LANES_LEFT Text,T_LANES_RIGHT Text,LAND_DOMAIN Text,RCLINK Text,STEVE_MIN Text,STEVE_MAX Text);S		

#	Name	SQL	Description	Note
101	Insert RC_SA	INSERT INTO RC_SA (COUNTY, ROUTE_TYPE, ROUTE_NUM, BEG_MEASURE, END_MEASURE, SECTION_LENGTH, DESCRIPTION, DISTRICT, MAINT_AREA, POPULATION, INVENTORY_DATE, DESIGNATED_WAY, TRUCK_ROUTE, TRAVEL_WAY, RURAL_URAN, SPEED_LIMIT, FAS_NUM, TRUCK_ROUTE_ID, CONGRESS_DIST, STATE_ROUTE_SEQ, ACCESS_CONTROL, OPERATION, TOTAL_LANES, SPECIAL_CLASS, DIV_HWY_SHLDR_WIDTH_LFT, DIV_HWY_SHLDR_TYPE_LFT, DIV_HWY_SURF_WIDTH, DIV_HWY_SURF_TYPE, DIV_HWY_SHLDR_WIDTH_RT, DIV_HWY_SHLDR_TYPE_RT, DIV_HWY_MEDIAN_WIDTH, DIV_HWY_MEDIAN_TYPE, DIV_HWY_BARRIER_TYPE, UDIV_HWY_SHLDR_WIDTH_LFT, UDIV_HWY_SHLDR_TYPE_LFT, UDIV_HWY_SURFACE_WIDTH, UDIV_HWY_SURFACE_TYPE, UDIV_HWY_SHLDR_WIDTH_RT, UDIV_HWY_SHLDR_TYPE_RT, AUX_LANE_WIDTH_LFT, AUX_LANE_TYPE_LFT, AUX_LANE_WIDTH_RT, AUX_LANE_TYPE_RT, MAINT_YEAR, MAINT_TYPE, IMPROVE_YEAR, FUNC_CLASS, TRAFFIC_COUNT_TYPE, TRAFFIC_COUNT_YEAR, RIGHT_OF_WAY, RW_TYPE, TC_NUMBER, MAINTENANCE_SUR_DES, SIDEWALK_LEFT, SIDEWALK_RIGHT, IMPROVE_TYPE, TRUCK_PERCENT, TRUCK_PERCENT_TYPE, SIGNAL, AADT_OLD, HPMS_ID, PACES_RATING, AADT, INTERSECT_ROAD1, INTERSECT_ROAD2, S_FUNCLASS_ID, DUAL_MAINT_RATING,	Copy data from COBB_RC into RC_SA.	When a county dataset other than Cobb is used, COBB_RC will need to be changed in the SQL to the name of the new county table.

#	Name	SQL	Description	Note
101	Insert RC_SA	ROAD_WIDTH, DIVIDED, OPEN_TO_TRAFFIC, CITY_CODE, T_LANES_LEFT, T_LANES_RIGHT, LAND_DOMAIN, RCLINK) SELECT COUNTY, ROUTE_TYPE, ROUTE_NUM, BEG_MEASURE, END_MEASURE, SECTION_LENGTH, DESCRIPTION, DISTRICT, MAINT_AREA, POPULATION, INVENTORY_DATE, DESIGNATED_WAY, TRUCK_ROUTE, TRAVEL_WAY, RURAL_URAN, SPEED_LIMIT, FAS_NUM, TRUCK_ROUTE_ID, CONGRESS_DIST, STATE_ROUTE_SEQ, ACCESS_CONTROL, OPERATION, TOTAL_LANES, SPECIAL_CLASS, DIV_HWY_SHLDR_WIDTH_LFT, DIV_HWY_SHLDR_TYPE_LFT, DIV_HWY_SURF_WIDTH, DIV_HWY_SURF_TYPE, DIV_HWY_SHLDR_WIDTH_RT, DIV_HWY_SHLDR_TYPE_RT, DIV_HWY_MEDIAN_WIDTH, DIV_HWY_MEDIAN_TYPE, DIV_HWY_BARRIER_TYPE, UDIV_HWY_SHLDR_WIDTH_LFT, UDIV_HWY_SHLDR_TYPE_LFT, UDIV_HWY_SURFACE_WIDTH, UDIV_HWY_SURFACE_TYPE, UDIV_HWY_SHLDR_WIDTH_RT, UDIV_HWY_SHLDR_TYPE_RT, AUX_LANE_WIDTH_LFT, AUX_LANE_TYPE_LFT, AUX_LANE_WIDTH_RT, AUX_LANE_TYPE_RT, MAINT_YEAR, MAINT_TYPE, IMPROVE_YEAR, FUNC_CLASS, TRAFFIC_COUNT_TYPE, TRAFFIC_COUNT_YEAR, RIGHT_OF_WAY, RW_TYPE, TC_NUMBER, MAINTENANCE_SUR_DES, SIDEWALK_LEFT, SIDEWALK_RIGHT, IMPROVE_TYPE, TRUCK_PERCENT,	Copy data from COBB_RC into RC_SA.	When a county dataset other than Cobb is used, COBB_RC will need to be changed in the SQL to the name of the new county table.

#	Name	SQL	Description	Note
101	Insert RC_SA	TRUCK_PERCENT_TYPE, SIGNAL, AADT_OLD, HPMS_ID, PACES_RATING, AADT, INTERSECT_ROAD1, INTERSECT_ROAD2, S_FUNCLASS_ID, DUAL_MAINT_RATING, ROAD_WIDTH, DIVIDED, OPEN_TO_TRAFFIC, CITY_CODE, T_LANES_LEFT, T_LANES_RIGHT, LAND_DOMAIN, RCLINK FROM COBB_RC;	Copy data from COBB_RC into RC_SA.	When a county dataset other than Cobb is used, COBB_RC will need to be changed in the SQL to the name of the new county table.

#r	Name	SQL	Description	Note
102_0	Alter Table RC_SA	ALTER TABLE RC_SA ALTER COLUMN SA_ID Text(100);	Change the column SA_ID from type COUNTER to Text.	
102_1	Update BEG_MEASURE	UPDATE RC_SA SET RC_SA.BEG_MEASURE = '0.0' WHERE RC_SA.BEG_MEASURE='0' Or RC_SA.BEG_MEASURE='' Or IsNull(RC_SA.BEG_MEASURE);	Update BEG_MEASURE to "0.0" where it is missing or it is "0"	
102_2	Update END_MEASURE	UPDATE RC_SA SET RC_SA.END_MEASURE = '0.0' WHERE RC_SA.END_MEASURE='0' Or RC_SA.END_MEASURE='' Or IsNull(RC_SA.END_MEASURE);	Update END_MEASURE to "0.0" where it is missing or it is "0"	

#	Name	SQL	Description	Note
103	Create ROADSEG_SA Table	CREATE TABLE ROADSEG_SA (SA_ID Text,RTE_TYPE Text, RTE_NAME xt,CNTY_NUM Text, BEG_MLPOST Text,END_MLPOST Text, SEG_LEN xt,DIST_NUM Text, CITY_NUM Text,JURISDICT Text, AREA_TYPE xt,RDWY_CLASS1 Text, NUM_THRU1 Text,NUM_THRU2 Text, AUX_LANE1 xt,AUX_LANE2 Text, AVE_LANE_WIDTH Text,MED_TYPE1 Text, MED_WIDTH xt,SHLDR_TYPE_OUT1 Text, SHLDR_TYPE_IN1 Text,SHLDR_TYPE_OUT2 Text, ILDR_TYPE_IN2 Text,SHLDR_WIDTH_OUT1 Text, SHLDR_WIDTH_IN1 xt,SHLDR_WIDTH_OUT2 Text, SHLDR_WIDTH_IN2 Text,ACC_CNTRL Text, ADT_2000 Text,AADT_2001 Text, AADT_2002 Text,AADT_2003 Text, ADT_2004 Text,GRWTH_FCTR Text, PCT_HEAVY Text,POST_SPD Text, PERATION Text,INTRCHG_INFL Text, RD_SURF Text,PED_FAC Text);	Create the table ROADSEG_SA.	
104	Insert ROADSEG_SA	INSERT INTO ROADSEG_SA (SA_ID, RTE_NAME, CNTY_NUM, BEG_MLPOST, END_MLPOST, SEG_LEN, DIST_NUM, CITY_NUM, AREA_TYPE, RDWY_CLASS1, NUM_THRU1, NUM_THRU2, AUX_LANE1, AUX_LANE2, MED_WIDTH, ACC_CNTRL, PCT_HEAVY, POST_SPD, OPERATION, AADT_2003, AADT_2004) SELECT SA_ID, ROUTE_NUM, COUNTY, BEG_MEASURE, END_MEASURE, SECTION_LENGTH, DISTRICT, CITY_CODE, RURAL_URAN, FUNC_CLASS, T_LANES_RIGHT, T_LANES_LEFT, AUX_LANE_TYPE_RT, AUX_LANE_TYPE_LFT, DIV_HWY_MEDIAN_WIDTH, ACCESS_CONTROL, TRUCK_PERCENT, SPEED_LIMIT, OPERATION, AADT_OLD, AADT FROM RC_SA WHERE Not RC_SA.ROUTE_TYPE='6';	Insert all the one-to-one matches from RC_SA into ROADSEG_SA	

#	Name	SQL	Description	Note
105_0	Update RTE_TYPE - Route Num	UPDATE ROADSEG_SA INNER JOIN RC_SA ON ROADSEG_SA.SA_ID=RC_SA.SA_ID SET ROADSEG_SA.RTE_TYPE = right(RC_SA.ROUTE_NUM,2);	Update RTE_TYPE to the right 2 characters of ROUTE_NUM based on a match on SA_ID	
105_1	Update RTE_TYPE - Route Type	UPDATE ROADSEG_SA INNER JOIN RC_SA ON ROADSEG_SA.SA_ID=RC_SA.SA_ID SET ROADSEG_SA.RTE_TYPE = RC_SA.ROUTE_TYPE WHERE Not RC_SA.ROUTE_TYPE='';	Update RTE_TYPE to ROUTE_TYPE when RTE_TYPE is "" based on a match on SA_ID	
105_2	Update RTE_TYPE – Interstate	UPDATE ROADSEG_SA SET ROADSEG_SA.RTE_TYPE = "Inter" WHERE left(ROADSEG_SA.RTE_NAME,1)="4";	Update RTE_TYPE to "Inter" when the first character of RTE_NAME is "4" based on a match on SA_ID	
106_0	Update JUSRIDICT - Route Type	UPDATE ROADSEG_SA INNER JOIN RC_SA ON ROADSEG_SA.SA_ID=RC_SA.SA_ID SET ROADSEG_SA.JURISDICT = RC_SA.ROUTE_TYPE;	Update JURISDICT to ROUTE_TYPE based on a match on SA_ID	
106_1	Update JURISDICT - Designated Way	UPDATE ROADSEG_SA INNER JOIN RC_SA ON ROADSEG_SA.SA_ID=RC_SA.SA_ID SET ROADSEG_SA.JURISDICT = 'F' WHERE RC_SA.DESIGNATED_WAY='3' Or RC_SA.DESIGNATED_WAY='4' Or RC_SA.DESIGNATED_WAY='5' Or RC_SA.DESIGNATED_WAY='6' Or RC_SA.DESIGNATED_WAY='7';	Update JURISDICT to "F" when DESIGNATED_WAY is "3" or "4" or "5" or "6" or "7" based on a match on SA_ID	

#r	Name	SQL	Description	Note
107_0	Update AVE_LANE_WIDTH	UPDATE ROADSEG_SA INNER JOIN RC_SA ON ROADSEG_SA.SA_ID=RC_SA.SA_ID SET ROADSEG_SA. AVE_LANE_WIDTH = CStr(CInt(RC_SA.ROAD_WIDTH)/ CInt(RC_SA.TOTAL_LANES)) WHERE RC_SA.TOTAL_LANES <>'0' Or Not IsNull(RC_SA.TOTAL_LANES);	Update AVE_LANE WIDTH to ROAD_WIDTH / TOTAL_LANES based on a match on SA_ID	
107_1	Update AVE_LANE_WIDTH - Missing	UPDATE ROADSEG_SA SET ROADSEG_SA.AVE_LANE_WIDTH = '0' WHERE ROADSEG_ SA.AVE_LANE_WIDTH='' Or IsNull(ROADSEG_SA.AVE_LANE_WIDTH);	Update AVE_LANE_WIDTH to "0" where it is missing	
108_0	Update MED_TYPE1 - Barrier Type	UPDATE ROADSEG_SA INNER JOIN RC_SA ON ROADSEG_SA.SA_ID=RC_SA.SA_ID SET ROADSEG_SA.MED_TYPE1 = RC_SA.DIV_HWY_BARRIER_TYPE;	Update MED_TYPE_1 to DIV_HWY_BARRIER_TYPE based on a match on SA_ID	
108_1	Update MED_TYPE1 - Median Type	UPDATE ROADSEG_SA INNER JOIN RC_SA ON ROADSEG_SA.SA_ID=RC_SA.SA_ID SET ROADSEG_SA.MED _TYPE1 = 'U' WHERE RC_SA.DIV_HWY _MEDIAN_TYPE='0';	Update MED_TYPE_1 to "U" when DIV_HWY_MEDIAN_TYPE = "0" based on a match on SA_ID	
109_0	Update SHLDR_TYPE_OUT1 - Udiv	UPDATE ROADSEG_SA INNER JOIN RC_SA ON ROADSEG_SA.SA_ID=RC_SA.SA_ID SET ROADSEG_SA.SHLDR_TYPE_OUT1 = RC_SA.UDIV_HWY_SHLDR_TYPE_RT;	Update SHLDR_TYPE_OUT1 to UDIV_HWY_SHLDR_TYPE_RT based on a match on SA_ID	
109_1	Update SHLDR_TYPE_OUT1 - Div	UPDATE ROADSEG_SA INNER JOIN RC_SA ON ROADSEG_SA.SA_ID=RC_SA.SA_ID SET ROADSEG_SA.SHLDR_TYPE_OUT1 = RC_SA.DIV_HWY_SHLDR_TYPE_RT WHERE Not RC_SA.DIV_HWY_SHLDR_TYPE_RT='';	Update SHLDR_TYPE_OUT1 to DIV_HWY_SHLDR_TYPE_RT based on a match on SA_ID	

#r	Name	SQL	Description	Note
110_0	Update SHLDR_TYPE_IN1 – Udiv	UPDATE ROADSEG_SA INNER JOIN RC_SA ON ROADSEG_SA.SA_ID=RC_SA.SA_ID SET ROADSEG_SA.SHLDR_TYPE_IN1 = RC_SA.UDIV_HWY_SHLDR_TYPE_RT;	Update SHLDR_TYPE_IN1 to UDIV_SHLDR_TYPE_RT based on a match on SA_ID	
110_1	Update SHLDR_TYPE_IN1 - Div	UPDATE ROADSEG_SA INNER JOIN RC_SA ON ROADSEG_SA.SA_ID=RC_SA.SA_ID SET ROADSEG_SA.SHLDR_TYPE_IN1 = RC_SA.DIV_HWY_SHLDR_TYPE_RT WHERE Not RC_SA.DIV_HWY_SHLDR_TYPE_RT='';	Update SHLDR_TYPE_IN1 to DIV_SHLDR_TYPE_RT based on a match on SA_ID	
111_0	Update SHLDR_TYPE_OUT2 - Udiv	UPDATE ROADSEG_SA INNER JOIN RC_SA ON ROADSEG_SA.SA_ID=RC_SA.SA_ID SET ROADSEG_SA.SHLDR_TYPE_OUT2 = RC_SA.UDIV_HWY_SHLDR_TYPE_LFT;	Update SHLDR_TYPE_OUT2 to UDIV_SHLDR_TYPE_LFT based on a match on SA_ID	
111_1	Update SHLDR_TYPE_OUT2 - Div	UPDATE ROADSEG_SA INNER JOIN RC_SA ON ROADSEG_SA.SA_ID=RC_SA.SA_ID SET ROADSEG_SA.SHLDR_TYPE_OUT2 = RC_SA.DIV_HWY_SHLDR_TYPE_LFT WHERE Not RC_SA.DIV_HWY_SHLDR_TYPE_LFT='';	Update SHLDR_TYPE_OUT2 to DIV_SHLDR_TYPE_LFT based on a match on SA_ID	
112_0	Update SHLDR_TYPE_IN2 - Udiv	UPDATE ROADSEG_SA INNER JOIN RC_SA ON ROADSEG_SA.SA_ID=RC_SA.SA_ID SET ROADSEG_SA.SHLDR_TYPE_IN2 = RC_SA.UDIV_HWY_SHLDR_TYPE_LFT;	Update SHLDR_TYPE_IN2 to UDIV_HWY_SHLDR_TYPE_LFT based on a match on SA_ID	
112_1	Update SHLDR_TYPE_IN2 - Div	UPDATE ROADSEG_SA INNER JOIN RC_SA ON ROADSEG_SA.SA_ID=RC_SA.SA_ID SET ROADSEG_SA.SHLDR_TYPE_IN2 = RC_SA.DIV_HWY_SHLDR_TYPE_LFT WHERE Not RC_SA.DIV_HWY_SHLDR_TYPE_LFT='';	Update SHLDR_TYPE_IN2 to DIV_HWY_SHLDR_TYPE_LFT based on a match on SA_ID	

#	Name	SQL	Description	Note
113_0	Update SHLDR_WIDTH _OUT1 - Udiv	UPDATE ROADSEG_SA INNER JOIN RC_SA ON ROADSEG_SA.SA_ID=RC_SA.SA_ID SET ROADSEG _SA.SHLDR_WIDTH_OUT1 = RC_SA.UDIV_HWY _SHLDR_WIDTH_RT;	Update SHLDR_WIDTH _OUT1 to UDIV_HWY _SHLDR_WIDTH_RT based on a match on SA_ID	
113_1	Update SHLDR_WIDTH _OUT1 - Div	UPDATE ROADSEG_SA INNER JOIN RC_SA ON ROADSEG_SA.SA_ID=RC_SA.SA_ID SET ROADSEG_SA.SHLDR_WIDTH_OUT1 = RC_SA.DIV_HWY_SHLDR_WIDTH_RT WHERE Not RC_SA.DIV_HWY_SHLDR_WIDTH_RT='00';	Update SHLDR_WIDTH_OUT1 to DIV_HWY_SHLDR_WIDT H_RT based on a match on SA_ID	
114_0	Update SHLDR_WIDTH _IN1 - Udiv	UPDATE ROADSEG_SA INNER JOIN RC_SA ON ROADSEG_SA.SA_ID=RC_SA.SA_ID SET ROADSEG_SA.SHLDR_WIDTH_IN1 = RC_SA.UDIV_HWY_SHLDR_WIDTH_RT;	Update SHLDR_WIDTH_IN1 to UDIV_HWY_SHLDR_ WIDTH_RT based on a match on SA_ID	
114_1	Update SHLDR_WIDTH _IN1 - Div	UPDATE ROADSEG_SA INNER JOIN RC_SA ON ROADSEG_SA.SA_ID=RC_SA.SA_ID SET ROADSEG_SA.SHLDR_WIDTH_IN1 = RC_SA.DIV_HWY_SHLDR_WIDTH_RT WHERE Not RC_SA.DIV_HWY_SHLDR_ WIDTH_RT='00';	Update SHLDR_WIDTH_IN1 to DIV_HWY_SHLDR_WIDT H_RT based on a match on SA_ID	

#	Name	SQL	Description	Note
115_0	Update SHLDR_WIDTH_OUT2 - Udiv	UPDATE ROADSEG_SA INNER JOIN RC_SA ON ROADSEG_SA.SA_ID=RC_SA.SA_ID SET ROADSEG_SA.SHLDR_WIDTH_OUT2 = RC_SA.UDIV_HWY_SHLDR_WIDTH_LFT;	Update SHLDR_WIDTH_OUT2 to UDIV_HWY_SHLDR_WIDTH_LFT based on a match on SA_ID	
115_1	Update SHLDR_WIDTH_OUT2 - Div	UPDATE ROADSEG_SA INNER JOIN RC_SA ON ROADSEG_SA.SA_ID=RC_SA.SA_ID SET ROADSEG_SA.SHLDR_WIDTH_OUT2 = RC_SA.DIV_HWY_SHLDR_WIDTH_LFT WHERE Not RC_SA.DIV_HWY_SHLDR_WIDTH_LFT='00';	Update SHLDR_WIDTH_OUT2 to DIV_HWY_SHLDR_WIDTH_LFT based on a match on SA_ID	
117_0	Update INTRCHG_INFL – N	UPDATE ROADSEG_SA INNER JOIN RC_SA ON ROADSEG_SA.SA_ID=RC_SA.SA_ID SET ROADSEG_SA.INTRCHG_INFL = 'N';	Update INTRCHG_INFL to "N" based on a match on SA_ID	
117_1	Update INTRCHG_INFL – RP	UPDATE ROADSEG_SA INNER JOIN RC_SA ON ROADSEG_SA.SA_ID=RC_SA.SA_ID SET ROADSEG_SA.INTRCHG_INFL = 'Y' WHERE left(RC_SA.DESCRIPTION,2)='RP';	Update INTRCHG_INFL to "Y" when the left 2 characters of DESCRIPTION = "RP" based on a match on SA_ID	

#	Name	SQL	Description	Note
118_0	Update RD_SURF – Div	UPDATE ROADSEG_SA INNER JOIN RC_SA ON ROADSEG_SA.SA_ID=RC_SA.SA_ID SET ROADSEG_SA.RD_SURF = RC_SA.DIV_HWY_SURF_TYPE WHERE Not RC_SA.DIV_HWY_SURF_TYPE='';	Update RD_SURF to DIV_HWY_SURF_TYPE based on a match on SA_ID	
118_1	Update RD_SURF - Udiv	UPDATE ROADSEG_SA INNER JOIN RC_SA ON ROADSEG_SA.SA_ID=RC_SA.SA_ID SET ROADSEG_SA.RD_SURF = RC_SA.UDIV_HWY_SURFACE_TYPE WHERE Not RC_SA.UDIV_HWY_SURFACE_TYPE='';	Update RD_SURF to UDIV_HWY_SURF_TYPE based on a match on SA_ID	
119_0	Update PED_FAC - Left	UPDATE ROADSEG_SA INNER JOIN RC_SA ON ROADSEG_SA.SA_ID=RC_SA.SA_ID SET ROADSEG_SA.PED_FAC = RC_SA.SIDEWALK_LEFT;	Update PED_FAC to SIDEWALK_LEFT based on a match on SA_ID	
119_1	Update PED_FAC - Right	UPDATE ROADSEG_SA INNER JOIN RC_SA ON ROADSEG_SA.SA_ID=RC_SA.SA_ID SET ROADSEG_SA.PED_FAC = RC_SA.SIDEWALK_RIGHT WHERE Not RC_SA.SIDEWALK_RIGHT='';	Update PED_FAC to SIDEWALK_RIGHT based on a match on SA_ID	
120	Update AVE_LANE_WIDTH blanks	UPDATE ROADSEG_SA SET ROADSEG_SA.AVE_LANE_WIDTH = '0' WHERE ROADSEG_SA.AVE_LANE_WIDTH='' Or IsNull(ROADSEG_SA.AVE_LANE_WIDTH);	Update AVE_LANE_WIDTH to "0" where it is missing	

#	Name	SQL	Description	Note
121	Update AADT_2002	UPDATE ROADSEG_SA INNER JOIN AADT ON (ROADSEG_SA.RTE_NAME=AADT.ADT_ROUTE_IDENTIFIER) AND (ROADSEG_SA.CNTY_NUM=AADT.ADT_COUNTY_IDENTIFIER) AND (CInt(ROADSEG_SA.BEG_MLPOST)>=AADT.ADT_BEG_MILELOG) AND (CInt(ROADSEG_SA.END_MLPOST)<=AADT.ADT_END_MILELOG) SET ROADSEG_SA.AADT_2002 = AADT.ADT_ADT_COUNT WHERE AADT.ADT_YEAR='2002';	Update AADT_2002 to ADT_ADT_COUNT based on matches with AADT table	
122	Update AADT_2001	UPDATE ROADSEG_SA INNER JOIN AADT ON (CInt(ROADSEG_SA.END_MLPOST)<=AADT.ADT_END_MILELOG) AND (CInt(ROADSEG_SA.BEG_MLPOST)>=AADT.ADT_BEG_MILELOG) AND (ROADSEG_SA.CNTY_NUM=AADT.ADT_COUNTY_IDENTIFIER) AND (ROADSEG_SA.RTE_NAME=AADT.ADT_ROUTE_IDENTIFIER) SET ROADSEG_SA.AADT_2001 = AADT.ADT_ADT_COUNT WHERE AADT.ADT_YEAR='2001';	Update AADT_2001 to ADT_ADT_COUNT based on matches with AADT table	

#	Name	SQL	Description	Note
123	Update AADT_2000	UPDATE ROADSEG_SA INNER JOIN AADT ON (CInt(ROADSEG_SA.END_MLPOST)<=AADT.ADT_END_MILELOG) AND (CInt(ROADSEG_SA.BEG_MLPOST)>=AADT.ADT_BEG_MILELOG) AND (ROADSEG_SA.CNTY_NUM=AADT.ADT_COUNTY_IDENTIFIER) AND (ROADSEG_SA.RTE_NAME=AADT.ADT_ROUTE_IDENTIFIER) SET ROADSEG_SA.AADT_2000 = AADT.ADT_ADT_COUNT WHERE AADT.ADT_YEAR='2000';	Update AADT_2000 to ADT_ADT_COUNT based on matches with AADT table	
124	Update SEG_LEN	UPDATE ROADSEG_SA SET ROADSEG_SA.SEG_LEN = CStr(Round(CDbl(ROADSEG_SA.END_MLPOST)-CDbl(ROADSEG_SA.BEG_MLPOST),4));	Update SEG_LEN	

	SafetyAnalyst			GDOT			
S No.	SA field name	SA Code	Existing/ To be created	GDOT field Name	Data from table	Criteria for selection	GDOT Code
1	agencyID	unique number for each segment (18 digits)	Existing/To be created	RCLINK, BEG_MEASURE, END_MEASURE	RC_Cobb	-	-
2	locSystem	A - Route/Milepost B - Route/County/ Milepost C - Route/Section/Distance D - Section Distance	To be created	-	-	-	-
3	routeType	I - Interstate US - US route SR - State route BR - Business route BL - Business loop SP - Spur route CR - County road TR - Township road L - Local road O - Other X – Unknown	Existing and to be changed	LOC_ROUTE_TYPE	LOCATION	-	0-Accident Not Located 1-State Route 2-County Road 3-City Street 8-Public Road 9-Collector-Distributor
4	routeName	# or name of the route	To be created	LOC_ROUTE_IDENTIFIER + LOC_ROUTE_SUFFIX	LOCATION	-	-
5	county	county number	Existing	Accident County	ACCIDENT	"067"	-
6	startOffset	Start Measure of the segment	Existing	BEG_MEASURE	RC_Cobb	-	-
7	endOffset	End Measure of the segment	Existing	END_MEASURE	RC_Cobb	-	-
8	segmentLength	Length of the segment	Existing	SECTION_LENGTH	RC_Cobb	-	-
9	district	District in which the segment is located	Existing	DISTRICT	RC_Cobb	-	-
10	city	City in which the segment is located	Existing	CITY_CODE	RC_Cobb	-	-
11	jurisdiction	1 - Federal maintained 2 - State maintained 3 - County maintained 6 - Township maintained 4 - Local maintained 5 - Other maintained 99 – Unknown	To be created	ROUTE_TYPE or DESIGNATED_WAY	RC_Cobb	JURISDICTION = "F", if DESIGNATED_WAY = 3 or 4 or 5 or 6 or 7	1 State Route 2 County Road 3 City Street 4 Col Road 5 Unofficial Road 6 Ramp 7 Private Road 8 Public Road 9 Collector

	SafetyAnalyst		GDOT				
S No.	SA field name	SA Code	Existing/ To be created	GDOT field Name	Data from table	Criteria for selection	GDOT Code
12	areaType	U - Urban R - Rural X – Unknown	Existing and to be changed	RURAL_URAN	RC_Cobb	-	7 – Rural 8 – Urban
13	roadwayClass1	1 - Principal arterial-interstate 2 - Principal arterial-other freeway or expressway 3 - Principal arterial-other 4 - Minor arterial 5 - Major Collector 6 - Minor Collector 7 - Local 0 - Other 99 - Unknown	Existing and to be changed	FUNC_CLASS	RC_Cobb	-	11-Urban-Interstate Principal Arterial 14-Urban Principal Arterial 16-Urban-Minor Arterial Street 17-Urban-Collector Street 19-Urban-Local
14	d1numThruLane	Total number of thru lanes in direction 1	Existing	T_LANES_LEFT	RC_Cobb	-	-
15	d2numThruLane	Total number of thru lanes in direction 1	Existing	T_LANES_RIGHT	RC_Cobb	-	-
16	medianType1	1 - Rigid barrier system (i.e., concrete) 2 - Semi - rigid barrier system (i.e., box beam, W - beam strong post, etc.) 3 - Flexible barrier system (i.e., cable, W - beam weak post, etc.) 4 - Raised median with curb 5 - Depressed median 6 - Flush paved median [at least 4 ft in width] 7 - HOV lane(s) 8 - Railroad or rapid transit 9 - Other divided 0 - Undivided 98 - Not applicable 99 – Unknown	Existing and to be changed	MEDIAN_TYPE	RC_Cobb	-	0-Undivided Road 1-Grass 2-Soil, Stone 3-Park, Business 4-Couplet 5-Concrete 6-Other 7-Roadway Separated by Barrier Only
17	medianWidth	-	Existing	DIV_HWY_MEDIAN_WIDTH	RC_Cobb	-	-

	SafetyAnalyst			GDOT			
S No.	SA field name	SA Code	Existing/ To be created	GDOT field Name	Data from table	Criteria for selection	GDOT Code
18	d1shoulderTypeOut		Existing and to be changed	DIV_HWY_SHLDR_TYPE_RT	RC_Cobb	-	
19	d1shoulderTypeIn		Existing and to be changed	DIV_HWY_SHLDR_TYPE_RT	RC_Cobb	-	
20	d2shoulderTypeOut		Existing and to be changed	DIV_HWY_SHLDR_TYPE_LT	RC_Cobb	-	
21	d2shoulderTypeIn	1 - Paved 2 - Composite 3 - Gravel 4 - Turf 5 - Curb 6 - No shoulder 98 - Not applicable 99 – Unknown	Existing and to be changed	DIV_HWY_SHLDR_TYPE_LT	RC_Cobb	-	G- Grass or Sod S- Gravel or Stone F- Bit. Surf. Treatment (Low) I- Bit. Conc. (High) J- Portland Cement (High) K- Curb and Gutter (Width of the gutter is not coded. Always code 00C.) N- No Identifiable Shoulder or Curb. All of roadbed used as Roadway (Soil or Gravel Road). Also if less than 1 foot paved road. D- Gutter (only) O- Bit. Conc. (High) with curb and gutter P- Bit. Surface treatment (Low) with curb and gutter C- Curb only

Sno	Field Name	Mapping required??	GDOT Code --- SA Code
1	agencyID	No	
2	locSystem	No	
3	routeType	Yes	1 ----- SR 2 ----- CR 3 ----- L 7 ----- O 8 ----- O 9 ----- O Inter-- SR
4	routeName	No	
5	county	No	
6	startOffset	No	
7	endOffset	No	
8	segmentLength	No	
9	district	No	
10	city	No	
11	jurisdiction	Yes	1 ----- 2 2 ----- 3 3 ----- 4 4 ----- 4 5 ----- 98 7 ----- 5 8 ----- 1 9 ----- 1 F ----- 1
12	areaType	Yes	7 ----- R 8 ----- U
13	roadwayClass1	Yes	11 ----- 1 14 ----- 3 16 ----- 4 17 ----- 5 19 ----- 7
14	d1numThruLane	No	
15	d2numThruLane	No	
16	medianType1		
17	medianWidth	No	

<table>
<tr><th></th><th>Field Name</th><th>Mapping required??</th><th>GDOT Code --- SA Code</th></tr>
<tr><td>18</td><td>d1shoulderTypeOut</td><td>Yes</td><td rowspan="4">F ----- 1
I ----- 1
J ----- 1
D ----- 1
S ----- 3
G ----- 4
C ----- 5
O ----- 5
P ----- 5
K ----- 5
N ----- 6</td></tr>
<tr><td>19</td><td>d1shoulderTypeIn</td><td>Yes</td></tr>
<tr><td>20</td><td>d2shoulderTypeOut</td><td>Yes</td></tr>
<tr><td>21</td><td>d2shoulderTypeIn</td><td>Yes</td></tr>
<tr><td>22</td><td>d1avgShoulderWidthOut</td><td>No</td><td></td></tr>
<tr><td>23</td><td>d1avgShoulderWidthIn</td><td>No</td><td></td></tr>
<tr><td>24</td><td>d2avgShoulderWidthOut</td><td>No</td><td></td></tr>
<tr><td>25</td><td>d2avgShoulderWidthIn</td><td>No</td><td></td></tr>
<tr><td>26</td><td>accessControl</td><td>Yes</td><td>F ----- 1
P ----- 2
U ----- 3</td></tr>
<tr><td>27</td><td>growthFactor</td><td>No</td><td></td></tr>
<tr><td>28</td><td>postedSpeed</td><td>No</td><td></td></tr>
<tr><td>29</td><td>operationWay</td><td>Yes</td><td>1 ----- 1
2 ----- 2
0 ----- 99</td></tr>
<tr><td>30</td><td>interchangeInfluence</td><td>No</td><td></td></tr>
</table>

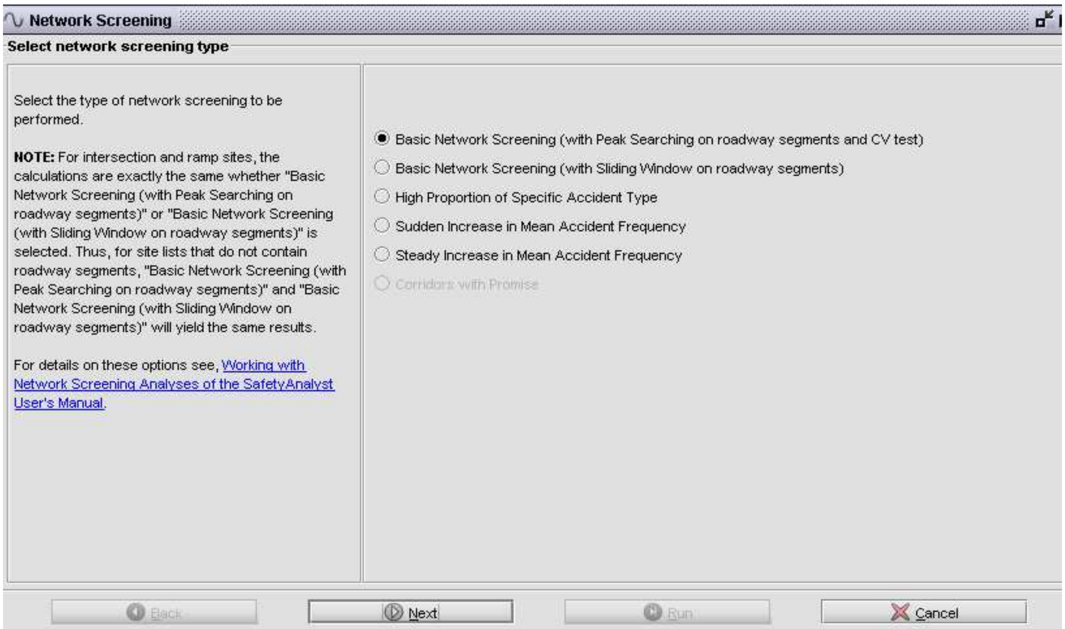
APPENDIX C: ALTSEGMNTTRAFFIC FILE

DATA MAPPING AND DATA RECODING

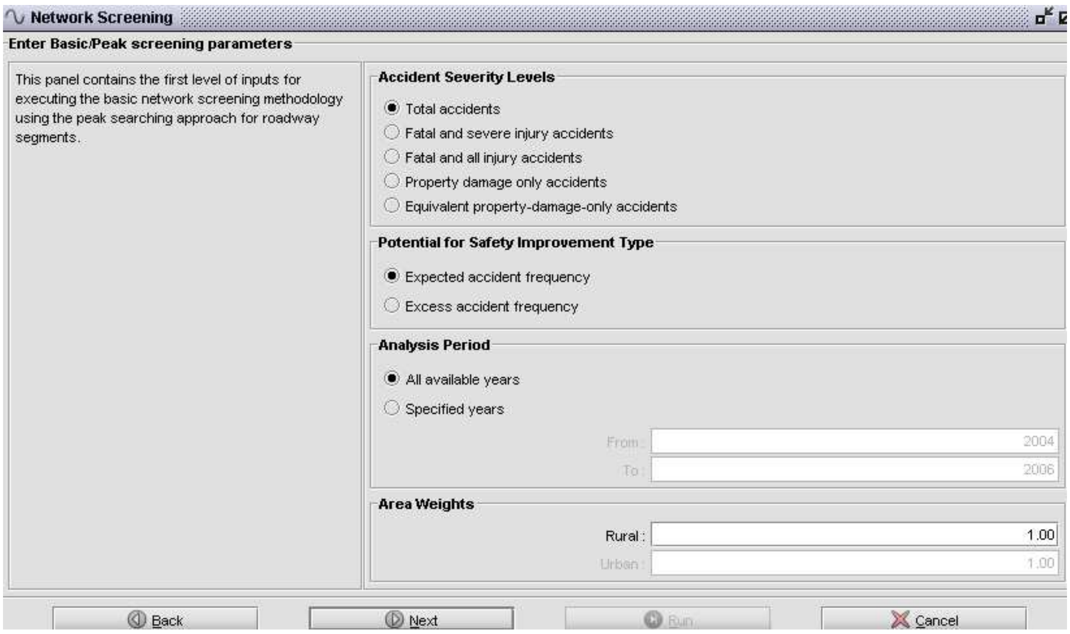
	SafetyAnalyst		GDOT				
S No.	SA field name	SA Code	Existing/ To be created	GDOT field Name	Data from table	Criteria for selection	GDOT Code
1	agencyID	unique number for each segment	To be created	RCLINK, Beg Milepost, End Milepost	RC_Cobb	-	-
2	calendarYear	Year for which traffic data is collected	Existing	ADT_YEAR	AADT	County = "067" RC_Cobb.BegMlpost>=AADT .BegMlpost AND RC_Cobb.EndMlpost>=AADT .EndMlpost	-
3	aadtVPD	AADT for the year mentioned	Existing	ADT_ADT_COUNT	AADT	-	-
4	percentHeavyVehicles	Not required for initial analysis					
5	peakHourlyVolume	Not required for initial analysis					
6	Comment	Not required for initial analysis					

	AltSegmentTraffic		
Sno	**Field Name**	**Mapping required??**	**GDOT Code --- SA Code**
1	agencyID	No	
2	calendarYear	No	
3	aadtVPD	No	
4	percentHeavyVehicles	No	
5	peakHourlyVolume	No	
6	comment	No	

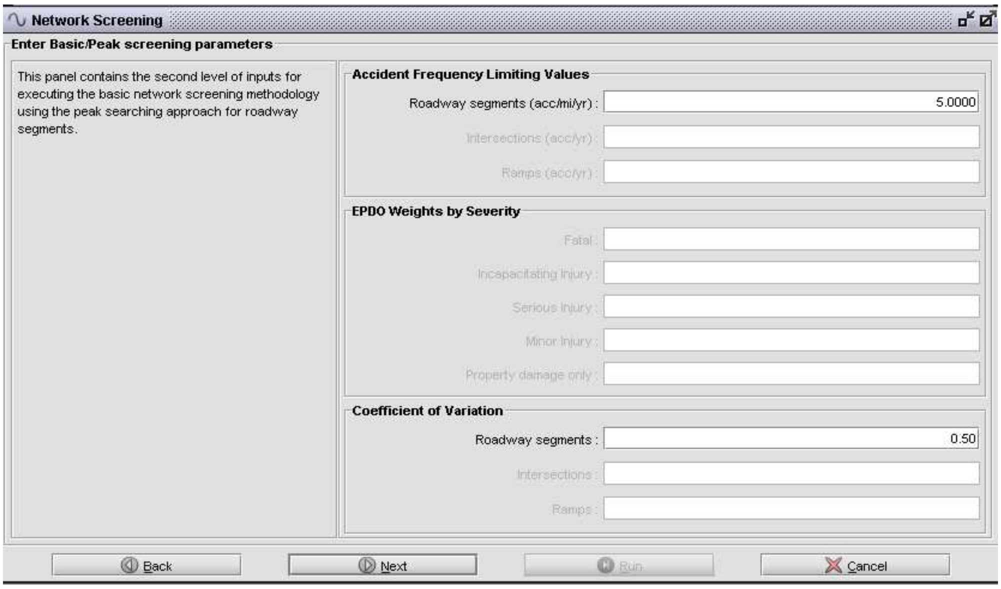
APPENDIX D: *SafetyAnalyst* ANALYTICAL TOOL: SCREENSHOT OF THE STEPS



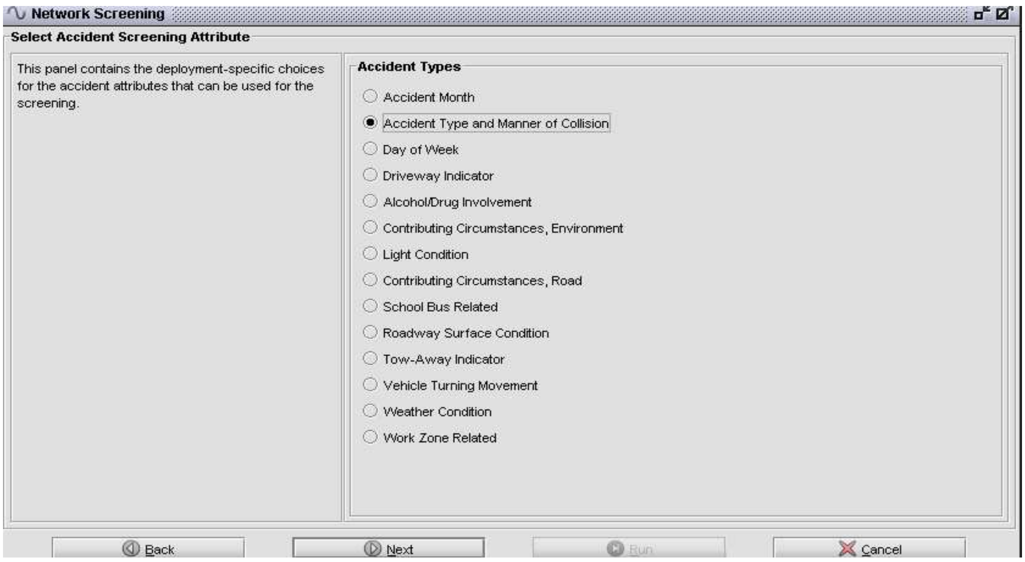
Select Network screening method



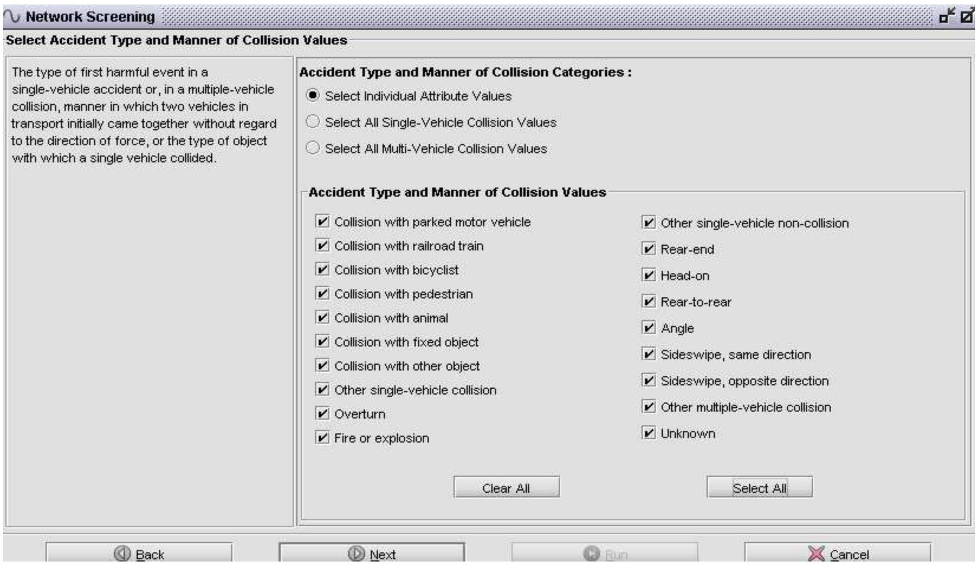
Select Accident Severity Level, PSI type, Analysis period and Area weights



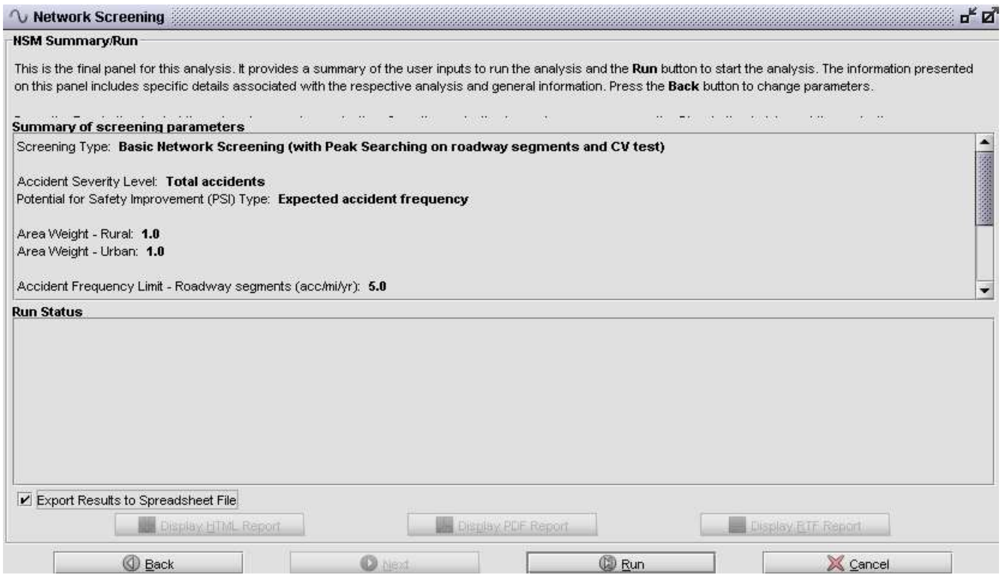
Select limiting value for accident frequency and the coefficient of variation



Select the accident type to be analyzed



Select attributes for Accident type and manner of collision



Final step in the “Network Screening” module

APPENDIX E: *SafetyAnalyst* NETWORK SCREENING SAMPLE REPORT

SafetyAnalyst

Network Screening Report

Jun 11, 2008

Disclaimer

This document is disseminated under the sponsorship of the Department of Transportation in the interest of information exchange. The United States Government assumes no liability for its content or use thereof. This document does not constitute a standard, specification, or regulation.

The United States Government does not endorse products or manufacturers. Trade and manufacturers' names may appear in this document only because they are considered essential to the objective of the document.

Limited Warranty and Limitations of Remedies

This software product is provided "as-is," without warranty of any kind-either expressed or implied (but not limited to the implied warranties of merchantability and fitness for a particular purpose). The FHWA and distributor do not warrant that the functions contained in the software will meet the end-user's requirements or that the operation of the software will be uninterrupted and error-free.

Under no circumstances will the FHWA or the distributor be liable to the end-user for any damages or claimed lost profits, lost savings, or other incidental or consequential damages rising out of the use or inability to use the software (even if these organizations have been advised of the possibility of such damages), or for any claim by any other party.

Notice

The use and testing of the SafetyAnalyst software is being done strictly on a voluntary basis. In exchange for provision of SafetyAnalyst, the user agrees that the Federal Highway Administration (FHWA), U.S. Department of Transportation and any other agency of the Federal Government shall not be responsible for any errors, damage or other liability that may result from any and all use of the software, including installation and testing of the software. The user further agrees to hold the FHWA and the Federal Government harmless from any resulting liability. The user agrees that this hold harmless provision shall flow to any person to whom or any entity to which the user provides the SafetyAnalyst software. It is the user's full responsibility to inform any person to whom or any entity to which it provides the SafetyAnalyst software of this hold harmless provision.

Caution

The Analytical Tool processing modules in this version of SafetyAnalyst have not been fully conformance tested. Results from these modules, although representative for the types of analysis performed, should not be considered usable for decision making.

1. Network Screening Report
Basic Network Screening
SafetyAnalyst: v1.4.11, packaged: Apr 18, 2008 3:25 PM on sa_dev.systems.de.ittind.com
Data set title: 0601GDOT
Data set comment: own SPFs
Data set created: Jun 1, 2008 1:29 PM
Roadway Segments: Peak Searching
Accident Severity Level: Fatal and all injury accidents
Site Types: Segments
Accident Types: Accident Type and Manner of Collision; Rear-end
Potential for Safety Improvement Using: Expected accident frequency
Analysis Period: From 2004 To 2006
CV limit (roadway segments): 0.5
Area Weights (Rural): 1.0
Area Weights (Urban): 1.0
Limiting Value (Roadway Segments): 5.0 acc/mi/yr

ID	Site Type	Site Subtype	Cnty	Rte	Site Start Loc	Site End Loc	Avg Obs Acc for Entire Site	Location with Highest Potential for Safety Improvement								Rank	Ad dtl. Wndow of Intrst
								Avg Obs Acc	Pred Acc Freq	PSI Exp Acc Freq	Var	Start Loc	End Loc	#. of Exp Fat	# of Exp Inj		
10401001 5661567	Segment	Seg/Rur; Fwy in intchng area (6+ ln)	1067	SR104 0100	15.66	15.67	333.33	333.33	3.668	229.3	5,236	15.66	15.67	0	0	1	
10401000 5480549	Segment	Seg/Urb ; Fwy in intchng area (8+ ln)	1067	SR104 0100	5.48	5.49	166.66	166.66	16.66	187.6	7,015	5.48	5.49	0	0	2	
10401000 2560257	Segment	Seg/Rur; Fwy in intchng area (6+ ln)	1067	SR104 0100	2.56	2.57	166.66	166.66	10.98	143.7	4,091	2.56	2.57	0	0	3	
10003000 2540256... 10003000 2890295	Segment	Seg/Urb ; Multilane undivided	1067	SR100 0300	2.54	2.95	22.76	76.66	4.159	76.50	253	2.64	2.74	0	0	4	2.74 - 2.84

Table 1 Site Data Summary

* - Units for Observed, Predicted and Expected Accident Frequency
- Roadway Segments (acc/mi/yr)
- Intersections (acc/yr)
- Ramps (acc/yr)

** - Units for Variance
- Roadway Segments (acc/mi**2/yr)
- Intersections (acc/yr)
- Ramps (acc/yr)

APPENDIX F: SAS CODE

```
DM
'LOG;CLEAR;OUT;CLEAR;';
OPTIONS
NODATE NONUMBER LS=90 PS=80;
DATA
alluri;
INFILE
'U:\profile.cu\My Documents\My SAS Files\0514_103_152_SAS.csv'
delimiter=
',' firstobs=2;
INPUT ID SiteSubtype $ SiteStLoc SiteEndLoc length logADT logLengthYrs
TotAcc
;
Proc print;
PROC
GENMOD; BY SiteSubtype;
MODEL TotAcc=logADT /
LINK = Log DIST = NEGBIN  OFFSET = logLengthYrs;
run;quit;
```

APPENDIX G: SAS OUTPUT

The SAS System (years 2004-2006)

------------------------------------ SiteSubtype=103 -------------------------------------

The GENMOD Procedure

Model Information

Data Set	WORK.ALLURI
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	TotAcc
Offset Variable	logLengthYrs

Number of Observations Read	200
Number of Observations Used	200

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	198	206.5289	1.0431
Scaled Deviance	198	206.5289	1.0431
Pearson Chi-Square	198	391.7024	1.9783
Scaled Pearson X2	198	391.7024	1.9783
Log Likelihood	13462.7995		

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Standard Estimate	Wald Error	95% Confidence Limits		Chi-Sqr	Pr> Chi Sq
Intercept	1	-7.0809	1.2059	-9.4445	-4.7173	34.48	<.0001
logADT	1	1.0023	0.1225	0.7621	1.2425	66.90	<.0001
Dispersion	1	3.6284	0.4048	2.8349	4.4218		

NOTE: The negative binomial dispersion parameter was estimated by maximum likelihood.

The SAS System (years 2004-1006)

----------------------------------- SiteSubtype=152 ------------------------------------

The GENMOD Procedure

Model Information

Data Set	WORK.ALLURI
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	TotAcc
Offset Variable	logLengthYrs

Number of Observations Read	136
Number of Observations Used	136

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	134	158.8497	1.1854
Scaled Deviance	134	158.8497	1.1854
Pearson Chi-Square	134	196.1335	1.4637
Scaled Pearson X2	134	196.1335	1.4637
Log Likelihood	19802.4904		

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Standard Estimate	Wald Error	95% Limits	Confidence	Chi-Sqr	Pr> Chi Sq
Intercept	1	-3.9323	1.0906	-6.0698	-0.2194	13.00	0.000 3
logADT	1	0.7409	0.1089	0.5275	0.9544	46.29	<.000 1
Dispersio n	1	1.8119	0.2194	1.3819	2.2420		

NOTE: The negative binomial dispersion parameter was estimated by maximum likelihood.

The SAS System (Year 2004)

----------------------------------- SiteSubtype=103 ------------------------------------

The GENMOD Procedure

Model Information

Data Set	WORK.ALLURI
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	TotAcc
Offset Variable	logLengthYrs

Number of Observations Read	200
Number of Observations Used	200

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	198	198.3593	1.0018
Scaled Deviance	198	198.3593	1.0018
Pearson Chi-Square	198	386.6234	1.9526
Scaled Pearson X2	198	386.6234	1.9526
Log Likelihood	3991.8579		

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Standard Estimate	Wald Error	95% Limits	Confidence	Chi-Sqr	Pr> Chi Sq
Intercept	1	-8.2320	1.3948	-10.9657	-5.4983	34.83	<.0001
logADT	1	1.1288	0.1398	0.8547	1.4029	65.15	<.0001
Dispersion	1	2.4831	0.3104	1.8747	3.0914		

NOTE: The negative binomial dispersion parameter was estimated by maximum likelihood.

The SAS System (Year 2004)

----------------------------------- SiteSubtype=152 ------------------------------------

The GENMOD Procedure

Model Information

Data Set	WORK.ALLURI
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	TotAcc
Offset Variable	logLengthYrs

Number of Observations Read	136
Number of Observations Used	136

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	134	149.2595	1.1139
Scaled Deviance	134	149.2595	1.1139
Pearson Chi-Square	134	162.4551	1.2124
Scaled Pearson X2	134	162.4551	1.2124
Log Likelihood	4664.2976		

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Standard Estimate	Wald Error	95% Limits	Confidence	Chi-Sqr	Pr> Chi Sq
Intercept	1	-3.4535	1.2027	-5.8107	-1.0963	8.25	0.0041
logADT	1	0.7047	0.1205	0.4685	0.9408	34.21	<.0001
Dispersion	1	2.0431	0.2767	1.5008	2.5855		

NOTE: The negative binomial dispersion parameter was estimated by maximum likelihood.

The SAS System (Year 2005)

----------------------------------- SiteSubtype=103 ------------------------------------

The GENMOD Procedure

Model Information

Data Set	WORK.ALLURI
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	TotAcc
Offset Variable	logLengthYrs

Number of Observations Read	200
Number of Observations Used	200

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	198	197.1298	0.9956
Scaled Deviance	198	197.1298	0.9956
Pearson Chi-Square	198	527.4591	2.6639
Scaled Pearson X2	198	527.4591	2.6639
Log Likelihood	3272.5412		

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Standard Estimate	Wald Error	95% Limits	Confidence	Chi-Sqr	Pr> Chi Sq
Intercept	1	-8.6193	1.3415	-11.2486	-5.9899	41.28	<.0001
logADT	1	1.1541	0.1340	0.8915	1.4168	74.17	<.0001
Dispersion	1	2.2198	0.2705	1.6897	2.7499		

NOTE: The negative binomial dispersion parameter was estimated by maximum likelihood.

The SAS System (Year 2005)

----------------------------------- SiteSubtype=152 ------------------------------------

The GENMOD Procedure

Model Information

Data Set	WORK.ALLURI
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	TotAcc
Offset Variable	logLengthYrs

Number of Observations Read	136
Number of Observations Used	136

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	134	149.0491	1.1123
Scaled Deviance	134	149.0491	1.1123
Pearson Chi-Square	134	194.5024	1.4515
Scaled Pearson X2	134	194.5024	1.4515
Log Likelihood	4574.5887		

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Standard Estimate	Wald Error	95% Limits	Confidence	Chi-Sqr	Pr> Chi Sq
Intercept	1	-3.4164	1.1696	-5.7087	-1.1241	8.53	0.0035
logADT	1	0.6862	0.1163	0.4582	0.9142	34.80	<.0001
Dispersion	1	1.8224	0.2524	1.3277	2.3172		

NOTE: The negative binomial dispersion parameter was estimated by maximum likelihood.

The SAS System (Year 2006)

----------------------------------- SiteSubtype=103 ------------------------------------

The GENMOD Procedure

Model Information

Data Set	WORK.ALLURI
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	TotAcc
Offset Variable	logLengthYrs

Number of Observations Read	200
Number of Observations Used	200

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	198	202.0204	1.0203
Scaled Deviance	198	202.0204	1.0203
Pearson Chi-Square	198	506.9512	2.5604
Scaled Pearson X2	198	506.9512	2.5604
Log Likelihood	3267.6993		

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Standard Estimate	Wald Error	95% Limits	Confidence	Chi-Sqr	Pr> Chi Sq
Intercept	1	-7.3085	1.2840	-9.8250	-4.7919	32.40	<.0001
logADT	1	1.0237	0.1280	0.7729	1.2745	63.98	<.0001
Dispersion	1	2.2358	0.2687	1.7091	2.7626		

NOTE: The negative binomial dispersion parameter was estimated by maximum likelihood.

The SAS System (Year 2006)

----------------------------------- SiteSubtype=152 -------------------------------------

The GENMOD Procedure

Model Information

Data Set	WORK.ALLURI
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	TotAcc
Offset Variable	logLengthYrs

Number of Observations Read	136
Number of Observations Used	136

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	134	148.1276	1.1054
Scaled Deviance	134	148.1276	1.1054
Pearson Chi-Square	134	221.8631	1.6557
Scaled Pearson X2	134	221.8631	1.6557
Log Likelihood	4482.5550		

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Standard Estimate	Wald Error	95% Limits	Confidence	Chi-Sqr	Pr> Chi Sq
Intercept	1	-4.1450	1.2452	-6.5855	-1.7046	11.08	0.0009
logADT	1	0.7573	0.1230	0.5162	0.9984	37.91	<.0001
Dispersion	1	2.0512	0.2805	1.5014	2.6009		

NOTE: The negative binomial dispersion parameter was estimated by maximum likelihood.

REFERENCES

Al-Ghamdi, A.S. (2000). Probability approach for ranking high accident locations. *Urban Transport IV; urban transport and the environment for the 21st century, 6,* 523-531.

Bureau of Transportation Statistics. (2008, May 21). *Transportation fatalities by mode*. Retrieved May 25, 2008 from US Department of Transportation, Research and Innovative Technology Administration Web site:

Cafiso, S., Cava, G.L., & Montella, A. (2007). Safety index for evaluation of two- lane rural highways . *Transportation Research Record: Journal of the Transportation Research Board, 2019,*136-145.

Czerniak, R.J. (2002). NCHRP synthesis of highway practice 310: collecting processing and integrating GPS data into GIS. *Transportation Research Record: Journal of the Transportation Research Board, 301,* 1-39.

Ei-Basyouny, K., & Sayed, T. (2006). Comparison of two negative binomial regression techniques in developing accident prediction models. *Transportation Research Record: Journal of the Transportation Research Board, 1950,* 9-16.

Epstein, K., Corino, G., & Neumann, D. (2002). National review of the Highway Safety Improvement Program . *Public Roads, 65*(5), .

Federal Highway Administration (June 6, 2008). *SAFETEA-LU*. Retrieved July 11, 2008 from , Web site: http://safety.fhwa.dot.gov/safetealu/

Fox news. (2008, March 5). *Study: Traffic crashes cost Americans $164 billion a year*. Retrieved May 25, 2008 from Fox News.com, Web site: http://www.foxnews.com/story/0,2933,335150,00.html

Fridstrom, L., Ifver, J., Ingerbrigtsen, S., Kulmala, R., & Thomson, L.K. (1995). Measuring the contribution of randomness, exposure, weather, and daylight to the variation in road accident counts. *Accident Analysis and Prevention, 27*(1), 1-20.

Georgia Department of Transportation. (2006). *Georgia Strategic Highway Safety Plan.* Atlanta, GA.

Hauer, E. (1996). Identification of sites with promise. *Transportation Research Record: Journal of the Transportation Research Board, 1542,* 54-60.

Hauer, E., & Persaud, B.N. (1984). Problem of identifying hazardous locations using accident data. *Transportation Research Record: Journal of the Transportation Research Board, 975,* 36-43.

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,* 48-54.

Hauer, E., Harwood, D.W., Council, F.M., & Griffith, M.S. (2002). Estimating safety by the empirical bayes method. *Transportation Research Record: Journal of the Transportation Research Board, 1784,* 126-131.

Hauer, E., Kononov, J., Allery, B.K., & Griffith, M.S. (2002). Screeningthe road network for sites with promise. *Transportation ResearchRecord: Journal of the Transportation Research Board, 1784,* 27-32.

Hermans, E., Bossche, F.V., & Wets, G. (2007, January). *Impact of methodological choices on road safety ranking*. Paper presented at the Transportation Research Board Annual Meeting 2007 , Washington, DC.

Higle, J.L., & Witkowski, J.M. (1988). Bayesian identification of hazardous locations. *Transportation Research Record: Journal of the Transportation Research Board, 1185,* 24-36. http://209.85.215.104/search?q=cache:oazHdoruMxYJ:www.bts.gov/publications/national_transportation_statistics/html/table_02_01.html+Annual+Traffic+Fatality+Totals+Nationally+52,627+42,642&hl=en&ct=clnk&cd=1&gl=us http://www.who.int/violence_injury_prevention/publications/road_traffic/world_report/chapter1.pdf

iTRANS Consulting Ltd., & Human Factors North INC (2003). *New approaches to highway safety analysis*. : US Department of Transportation Federal Highway Administration.

Kononov, J. (2002). Identifying locations with potential for accident reductions: Use of direct diagnostics and pattern recognition methodologies. *Transportation Research Record: Journal of the Transportation Research Board, 1784,* 153-158.

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., & Janson, B.N. (2002). Diagnostic Methodology for the detection of safety problems at intersections. *Transportation Research Record: Journal of the Transportation Research Board, 1784,* 51-56.

Krull, K. A., Hummer, J.E., Crouch, J.A., & Chatterjee, A. (2000, July). *A realistic test of Hauer's promising site identification method*. Accidents on secondary highways and countermeasures, Knoxville.

Lei, Y., Fengxiang, Q., & Linhua, L. (2007). Bayesian-based approach for updating characteristics of crash data: Frequency, duration, and lanes blocked. *Transportation Research Record: Journal of the Transportation Research Board, , .*

Mandloi, D., & Gupta, R. (2003, January). *Evaluation of accident black spots on roads using geographical information systems*. Paper presented at the Map India 2003 conference, New Delhi, India.

Persaud, B., Cook, W., & Kazakov, A. (1997, September). *Demonstration of new approaches for identifying hazardous locations and prioritizing safety improvement*. Paper presented at the Traffic Safety on Two Continents, Lisbon, Portugal.

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), 345-351.

Sando, T., Mussa, R., Wu, H., Sobanjo, J., & Spainhour, L. (2004). *A cost effective GIS safety analysis tool for improving highway safety*. Paper presented at the Twenty-Fourth Annual ESRI User Conference, San Diego, CA.

Tamburri, T.N., & Smith, R.N. (1970). The safety index: A method of evaluating and rating safety benefits. *Highway Research Record, 332*, 28-43.

Turner Fairbank Highway Research Center (n.d.). *Highway Safety Information System: GIS safety tools*. Retrieved April 12, 2008from Federal Highway Administration, Web site: http://www.hsisinfo.org/hsis.cfm?num=8&page=1

Turner Fairbank Highway Research Center (n.d.). *SafetyAnalyst*. Retrieved June 20, 2008 from Federal Highway Administration, Web site: http://www.safetyanalyst.org/mri_brochure.pdf

US Department of Transportation Federal Highway Administration. (n.d.). *Motor vehicle crashes are a leading cause of death in the United States*. Retrieved April 11, 2008 from The Transportation Safety Planning, Web site: ftp://ftp.camsys.com/temp/outgoing/Safety_Fact_Sheets/Georgia.pdf

World Health Organization. (2004). *World report on road traffic injury prevention*. Retrieved April 12, 2008 from World Health Organization, Violence and Injury Prevention and Disability Website: http://www.who.int/violence_injury_prevention/publications/road_traffic/world_report/summary_en_rev.pdf

(2008). *Highway safety manual data needs guide* (1st edition). : National Cooperative Highway Research Program.