CIVIL ENGINEERING STUDIES
Illinois Center for Transportation Series No. 10-066
UILU-ENG-2010-2006
ISSN: 0197-9191

DEVELOPMENT AND APPLICATION OF SAFETY PERFORMANCE FUNCTIONS FOR ILLINOIS

Prepared By
Robert A. Tegge
Jang-Hyeon Jo
Yanfeng Ouyang
University of Illinois at Urbana-Champaign

Research Report ICT-10-066

A report of the findings of
ICT-R27-20
Development of Safety Performance Functions for Illinois

Illinois Center for Transportation

March 2010

Technical Report Documentation Page

1. Report No. FHWA-ICT-10-066	**2. Government Accession No.**	**3. Recipient's Catalog No.**
4. Title and Subtitle Development and Application of Safety Performance Functions for Illinois		**5. Report Date** March 2010
		6. Performing Organization Code
7. Author(s) Robert A. Tegge, Jang-Hyeon Jo, Yanfeng Ouyang		**8. Performing Organization Report No.** ICT-10-066 UILU-ENG-2010-2006
9. Performing Organization Name and Address Illinois Center for Transportation Department of Civil and Environmental Engineering University of Illinois at Urbana-Champaign 205 N. Mathews Ave., MC-250 Urbana, IL 61801		**10. Work Unit (TRAIS)**
		11. Contract or Grant No. ICT R27-20
		13. Type of Report and Period Covered
12. Sponsoring Agency Name and Address Illinois Department of Transportation Bureau of Materials and Physical Research 126 East Ash Street Springfield, IL 62704		**14. Sponsoring Agency Code**
15. Supplementary Notes		

16. Abstract

This report explains the procedure involved in developing Safety Performance Functions (SPF) for the road segments and intersections in the state of Illinois, under the jurisdiction of the Illinois Department of Transportation. SPFs predict the relationship among traffic, geometric conditions of the road and crash density, crash severity, and crash type. The SPFs are developed in such a way that they can be easily incorporated into the SafetyAnalyst tool that is being developed by FHWA to analyze and improve the safety of road elements. The SPFs are used to calculate a given site's Potential for Safety Improvement (PSI) and thus help in identifying the locations that have the highest potential for improvement.

A literature review was conducted as a part of the study to identify the methodology that would be needed. The literature review included studies on identifying the statistical techniques best suited for the requirement and identifying road element variables that have to be considered while developing SPFs. This report also includes the background and rationale behind the use of techniques such as Empirical Bayesian method, Sliding Window technique, and other such procedures that have been used in the analysis. As part of the study, road segments and intersections were classified into peer groups such that members of a peer group would have homogenous characteristics. Network Screening was conducted for all state-maintained (marked and unmarked) routes to identify high-crash locations, which directly supports the development of the 2008 Illinois Five Percent Report to FHWA.

This project also develops a VBA (Visual Basic for Applications) software tool that can be directly used by Illinois Department of Transportation (IDOT) officials to update SPFs and PSI screening in the future. This tool incorporates the statistical and computation models in an easy-to-use Excel spreadsheet environment. The software also automates the decision support process for identifying high crash sites in the Illinois roadway network. The system requirements and the procedure involved in using the software are explained in this report.

17. Key Words	**18. Distribution Statement** No restrictions. This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161.

19. Security Classif. (of this report) Unclassified	**20. Security Classif. (of this page)** Unclassified	**21. No. of Pages**	**22. Price**

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

ACKNOWLEDGMENTS

This publication is based on the results of research project ICT-R27-20, *Developing Safety Performance Functions for Illinois*. This project was conducted in cooperation with the Illinois Center for Transportation, the Illinois Department of Transportation, and the U.S. Department of Transportation, Federal Highway Administration. The research team thanks the Technical Review Panel (TRP) for its guidance, direction, and support. The members of TRP, and others who provided valuable support, include (alphabetically):

Jim Allen
Jessica Baker
Donald Berndt
Patty Broers
Jim Conlon
Mike Gillette
Sharon Haasis
Michael Kane
Jon-Paul Kohler
Kimberly Kolody
Leigh Ann Lare
Heidi Liske
Karen Magee
Melanie McDaniel
Lori Midden
Dan Mlacnik
Roseanne Nance
Mehdi Nassirpour
Mary Ann Paulis
David Piper (TRP Chair)
Dan Pratte
Rob Robinson
Mike Staggs
Priscilla Tobias
Peggy Westerfield

DISCLAIMER

The contents of this report reflect the view of the authors, who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the Illinois Center for Transportation, the Illinois Department of Transportation, or the Federal Highway Administration. This report does not constitute a standard, specification, or regulation.

EXECUTIVE SUMMARY

This report explains the procedure involved in developing Safety Performance Functions for the road segments and intersections in the state of Illinois, under the jurisdiction of the Illinois Department of Transportation. Safety Performance Functions (SPF) predict the relationship among traffic, geometric conditions of the road and crash density, crash severity, and crash type. The SPFs are developed in such a way that they can be easily incorporated into the *SafetyAnalyst* tool that is being developed by FHWA to analyze and improve the safety of road elements. The SPFs are used to calculate a given site's Potential for Safety Improvement (PSI) and thus help in identifying the locations that have the highest potential for improvement.

A literature review was conducted as a part of the study to identify the methodology that would be needed. The literature review included studies on identifying the statistical techniques best suited for the requirement and identifying road element variables that have to be considered while developing SPFs. This report also includes the background and rationale behind the use of techniques such as Empirical Bayesian method, Sliding Window technique and other such procedures that have been used in the analysis. As part of the study, road segments and intersections were classified into peer groups such that members of a peer group would have homogenous characteristics. Network Screening was conducted for all state-maintained (marked and unmarked) routes to identify high-crash locations, which directly supports the development of the 2008 Illinois Five Percent Report to FHWA.

This project also includes the development of a VBA (Visual Basic for Applications) software tool that can be directly used by Illinois Department of Transportation (IDOT) officials to update SPFs and PSI screening in the future. This tool incorporates the statistical and computation models in an easy-to-use Excel spreadsheet environment. The software also automates the decision support process for identifying high crash sites in the Illinois roadway network. The system requirements and the procedure involved in using the software are explained in this report.

TABLE OF CONTENTS

CHAPTER 1 INTRODUCTION

Whenever a person is driving on a roadway, the risk of being involved in a collision exists. The risks can be associated with driver reactions, weather conditions, roadway geometries, bad luck, or a combination of several factors. Even though all roadways have some inherent level of risk, some roadway sites (e.g., segments and intersections) are considered more hazardous than others, but what is the best way to identify "unsafe" sites? In the past, agencies would measure the rate of crashes to traffic volumes, use civilian input, or measure the absolute number of crashes at a location to declare whether a location was a safety concern. However, these techniques tend to be subjective, shortsighted, and/or an outdated view on transportation safety.

Transportation safety has become an intensively researched topic with the goal of better understanding why crashes occur. If they know why crashes occur, agencies will be able to identify safety improvement projects more efficiently and effectively. Recent studies have shown the relationship between crash frequency and traffic volume is nonlinear and depends on several variables. Traffic can behave extremely different depending on functional class, area type, median width, etc. Treating all situations the same does not reflect the impact a roadway's geometries and surroundings have on safety. The following study will provide an objective and unbiased method of measuring safety for Illinois roadways by implementing current safety research.

1.1 MOTIVATION FOR STUDY

According to statistics generated by the National Highway Traffic Safety Administration (NHTSA) in their Fatality Analysis Reporting System, the United States had 38,253 fatal crashes in 2004 nationwide. Of these crashes, 1,175 occurred in Illinois (NHTSA, 2005). The Illinois Department of Transportation (IDOT) is aware of this statistic and established a goal of reducing traffic-related fatalities and severe injuries. To achieve this goal, they need to better understand how traffic volumes (and other risk exposure variables) affect crash frequency (number of crashes per unit of time per unit roadway length). This relationship can be predicted with statistical models referred to as Safety Performance Functions (SPFs). The models will provide realistic and accurate predictions of crash frequencies as a function of traffic volume and roadway geometries over a roadway network (segments or intersections). By knowing what the expected crash frequency is for a location, comparisons can be made with the observed crash frequency to determine relative safety. When the observed number of crashes is greater than the predicted, IDOT can classify that a location is less safe. On the contrary, if the observed number of crashes is less than the predicted, then a site can be classified as being safer. The comparisons allow IDOT to better understand which locations, along its network, are a potential safety concern, and therefore spend its time, energy, and money accordingly. Analyzing sites through this methodology, as opposed to previous techniques, will produce consistent, unbiased, and objective results. In addition, IDOT can use the findings of these models in their comprehensive highway safety plan (CHSP) that focuses on the 4E's (Engineering, Enforcement, Education, and Emergency Medical Services).

1.2 SCOPE OF STUDY

In conjunction with IDOT, the University of Illinois at Urbana-Champaign (UIUC) is developing Illinois specific statistical models for various roadway and severity types in order to determine the relationship between crash frequency and roadway characteristics. Due to the difficulties in performing controlled experiments, observational studies have been adapted as the most practical method to explore the relationships between crashes and the causes of them. Controlled laboratory experiments would produce quicker and more accurate results because of the ability to change a single variable, but these methods are typically not available in the field of safety analysis due to variation in conditions and safety concerns (Hauer, 1997). Observational studies, on the other hand, rely on field data that is statistically analyzed to discover any trends that would show correlation between variables.

Past studies indicate that crash reports provide sufficient observational data, therefore experimental trials do not need to be run. The reports often indicate variables such as pavement conditions, driver characteristics, and location of crashes. Other datasets can provide information on the roadway characteristics such as average annual daily traffic (AADT), grade, and median type. With the information, SPFs can be developed to provide a statistical relationship between the expected number of crashes per year and roadway characteristics. Using the SPFs, an analysis of the Illinois roadway network can be performed to identify locations that have a high potential for improving safety.

1.3 OBJECTIVES

The first objective of the study is to develop Illinois-specific SPFs that will be used for network screening. The developed SPFs will predict crash frequency based solely on traffic volumes for various types of roadways (peer groups). Using the predicted number of crashes from the SPFs and the number of observed crashes, the Illinois network can be analyzed to determine which sites (segments and intersections) have the highest potential for safety improvement (PSI). Site's PSIs can be ranked in numerous ways to provide flexibility in analyzing safety. The analysis and techniques developed in the study will allow IDOT to better assess and improve the safety of its network through an objective method.

One of the limitations of SPFs using only traffic volumes is their inability to identify the reason for a crash. The models tend to be reactionary by using only traffic volumes as a variable. In order to determine the reason accidents occurred and provide a more proactive model, a multivariate analysis is necessary to determine how a variety of variables can contribute to crashes. These variables could include everything from lighting of the roadway, weather conditions, roadway geometrics, pavement conditions, and countless other variables. Knowing which roadway characteristics lead to an increase or decrease in crash frequency would allow an agency to better engineer roadways to help mitigate safety concerns.

The next objective of the study is to incorporate the developed SPFs into the R27-18 Illinois Center for Transportation (ICT) Project, "Crash Data Analysis and Engineering Solutions for Local Agencies," as a method to screen Illinois sites for PSI. This has been completed. In addition, the SPFs will also be input into a larger safety management program sponsored by the Federal Highway Administration (FHWA) entitled *SafetyAnalyst*. This program is a nationwide safety analysis software that allows agencies to better manage the safety of their roadways. *SafetyAnalyst* will be discussed further in the literature review, since it provides much of the inspiration and knowledge base for the development of Illinois-specific SPFs.

The final objective of this study is to automate the SPF development process and PSI calculation procedure by developing an easy-to-use computer program. This will allow IDOT officials to update the computation as data become available in the future.

This report is organized as follows. Following this introduction, the report provides a literature review, which consists of a detailed description of the development, application, and implementation of SPFs, EB Methodology, and PSI. The review was conducted based on published articles, reports, and books written on the topic of SPFs and statistical modeling. Following the literature review is a summary of the data processing. This section summarizes the data received in year 2007 (crash data from 2001 to 2005) and the methodology used to organize the datasets into a single dataset for analyzing. After the data processing section, the Illinois-specific SPFs for each peer group and severity type will be presented. The fourth section illustrates the site screening procedure and identifies sites with the highest PSIs for each severity type. A second round of site screening was conducted as new 2002-2006 crash data became available in early 2008, and the result was used to support the development of the FHWA 5-percent report. The sixth section shows the multivariate process and explains the significant variables leading to crash occurrence. The seventh section explains the development of an Excel spreadsheet software that can be used to update the SPFs and PSI site screenings automatically. The final section summarizes the study, reiterates the limitations, and suggests potential future studies.

CHAPTER 2 LITERATURE REVIEW

According to statistics generated by the Fatality Analysis Reporting System (FARS), 38,253 fatal crashes occurred in the United States in 2004. Of these fatalities, 1,175 occurred in Illinois (NHTSA, 2005). Transportation safety has become an intensively researched topic, while the goal is to better understand why crashes occur and to identify safety improvement projects to make roadways safer. However, controlled studies of crashes are difficult to undertake due to the extreme amounts of variation involved. Roadway, weather, and driver factors can all lead to a crash, and often the causality involves a combination of several different factors. Granted, randomized laboratory experiments would produce quicker and more accurate results by the ability to change a single variable, but these methods are typically not available in the field of safety analysis (Hauer, 1997). Due to the difficulties of performing controlled experiments, observational studies have been adapted as the most practical method to explore the relationships between crashes and the causes of them. Observational studies rely on field data that is statistically analyzed to discover any trends that would show correlation between variables. Crash reports provide sufficient observational data so that experimental trials do not need to be run. The reports often indicate variables such as pavement conditions, driver characteristics, and location of crashes. Other datasets can provide information on the roadway characteristics such as AADT, grade, median type, etc. With all of the information, Safety Performance Functions (SPF) can provide a statistical relationship between the expected number of crashes per year and roadway characteristics. The subsequent sections will provide a review of existing literature on the development and implementation of SPF. The sections will outline the different types of SPFs and the commonly used variables, the statistical modeling techniques utilized in developing SPFs, and the role of Empirical Bayesian Relationships and Potential for Safety Improvements in the network screening process.

2.1 DEVELOPMENT OF SAFETY PERFORMANCE FUNCTIONS

Crashes are inherent as vehicles travel on a roadway network (segments and intersections). SPFs help determine what the expected number of crashes should be for various settings and design variables. Since roadway segments and intersections have distinctly different characteristics, different sets of SPF are required to model the crash frequencies. For example, a roadway segment depends on characteristics such as AADT, lane width, median type, and shoulders, where as an intersection depends on characteristics such as control type, turning movement, and AADT for both the major and minor legs. The following sections will describe these two distinct elements of the roadway network, the variables used in developing statistical models, and a summary of other studies on SPFs.

2.1.1 APPLICATION TO ROADWAY SEGMENTS

Segments make up the majority of the roadway network. Roadways segments travel through a variety of terrains, area types, and experience numerous geometric changes as they transport people and goods from one location to another. The following sections describe the different techniques used to subdivide roadways into smaller segments, the different levels of SPF, variables used for roadway segments, and the SPF developed in other studies.

2.1.1.1 Segment Length Selection

Several ways exist to analyze roadways by just looking at the roadway through different analysis “windows.” For starters, the roadway can be treated as a single segment for its entire length. The problem with treating roadways as a large entity is the inability to spot locations where safety is a concern. Since the safety will be averaged over the entire extent of the roadway, areas of peak crash frequencies will be smoothed from areas that are not so high. Another problem is that roadway lengths vary tremendously. When peaks occur on longer roadways, they are reduced by averages more than a shorter roadway where a peak crash location will not be reduced as much. This means that during site selection, shorter roadways will have a higher representation than longer roadways even though the safety concerns could be the same (Hauer, 2002b).

The second method for analyzing segments is to break a roadway into fixed lengths (*SafetyAnalyst*, Module 1). Breaking a roadway into fixed segments allows comparable units to be used on roadways that have different lengths. With small enough subdivisions, typically 0.1 miles, identifying locations of safety concerns can be easier because they will not be averaged over a larger distance. The main concern in using a fixed scale is that it is difficult to locate a peak safety concern unless the peak is located completely within the analysis segment. If the peak is on the edge of two segments, the peaks will be distributed over two windows, and therefore the segments could be determined safer for the segments than what actually exists (Hauer, 2002b).

The final technique in analyzing a roadway segment is to use a sliding window. A sliding window takes a set length of roadway and continuously moves the window for the length of the roadway. The purpose is that areas of peak safety concern can be accurately identified since all points along the roadway can be compared to their surroundings more effectively. In other words, each segment will overlap the previous and next segment (*SafetyAnalyst*, Module 1). The problem that arises with the sliding window is that it may require the use of several SPF for a given segment. For segments of fixed length, the lengths and segments can be selected so that the roadway has homogenous features in the area of study. However, the sliding window does not allow for this since it is a continuous procedure, so combining SPFs is necessary to represent the expected number of crashes.

2.1.1.2 Types of Analysis

Two types of SPFs are used for representing crash frequencies as a function of given variables. The first type of SPF is a Level I, or descriptive analysis model, which determine crash frequencies based only on traffic volumes (AADT). They typically take on the following functional form:

$$\mu_i = (Segment\ Length)_i \cdot \alpha_0 \cdot (AADT_i)^{\alpha_1}$$

where μ_i *is* the predicted crash frequency per year and α_0 and α_1 are regression parameters. From past studies, AADT has the largest impact on crash frequencies. Currently, the FHWA program *SafetyAnalyst*, only uses Level I SPF for analyzing safety (*SafetyAnalyst*, Module 1).

Level II SPF, classified as multivariate models or causality models, incorporates a variety of variables other than just traffic volumes. In Level II SPF, variables such as weather conditions, roadway geometries, traffic data, and human factors to calculate the crash frequencies (*SafetyAnalyst*, Module 1). These models have the following functional form for a given segment:

$$\mu_i = (Segment\ Length)_i \cdot \alpha_0 \cdot (AADT_i)^{\alpha_1} \cdot X_{2i}^{\alpha_2} \cdots X_{qi}^{\alpha_q} \quad \text{(Equation 1)}$$

Or

$$\mu_i = (Segment\ Length)_i \cdot \alpha_0 \cdot (AADT_i)^{\alpha_1} \cdot \exp(\alpha_2 X_{2i} + \alpha_3 X_{3i} + \cdots + \alpha_q X_{qi}) \quad \text{(Equation 2)}$$

where μ_i is the predicted crash frequency per year, $\alpha_{0,}\ \alpha_{1,}\ \ldots,\ \alpha_n$ are regression parameters, and X_{1i}, X_{2i}, …., X_{qi} are the variables of interest (Hauer, 2007). The additional variables provide information for making future improvements to a given roadway. By realizing the impacts that changing a given geometry will have allows engineers to make better safety judgments. The information provided by the additional variables also allows for benefits such as education and enforcement. By recognizing situations, habits, and other conditions that pose safety concerns, law enforcement and drivers can adapt driving practices to increase their potential for safe commutes.

Even though both of these levels of SPFs involve different variables, they still take on the same statistical model. Since crash is a rare occurrence on roadway segments and demonstrates significant overdispersion, utilization of negative binomial distribution accurately models the crash frequencies (Persaud, 2001). However, during the modeling process, the overdispersion factor, similar to the crash frequency, is determined per a given length. As the segment length increases, so does the overdispersion factor.

2.1.1.3 Variables for Roadway Segments

Variables that depict roadway characteristics, driver behavior, climatic conditions, and traffic data can be included in the development of SPF. These variables are broken into two classifications for modeling purposes: quantitative and categorical. Quantitative variables are discrete values that represent a condition, characteristic, or parameter. In past studies, the quantitative variables used for developing SPF for roadway segments include the following:

- Access point density
- Average Annual Daily Traffic (AADT)
- Grade
- Lane width
- Median width
- Number of lanes
- Peak hour volume/design hour volume
- Percent heavy vehicles
- Radius of Curvature
- Segment length
- Shoulder width
- Speed (85th percentile or posted speed)

When modeling, these values can be entered directly into the SPF for determining the crash frequency (*SafetyAnalyst*, Module 1). As Equation (1) and (2) suggest, the logarithm of AADT is often used in the model specifications.

The other type of variables used in developing SPF is categorical variables. These variables are non-numerical and are ways to describe a given situation. Categorical variables used by others in SPF are as follows:

- Functionality classification of the roadway
- Auxiliary lanes (TWLTL/passing lane/climbing lane/other auxiliary lane)
- Terrain
- Median type
- Area type (rural/urban)
- Shoulder type

During modeling, these variables are represented as binary dummies for all of the different possible scenarios (Bauer and Harwood, 1999).

Another technique to reduce the number of categorical and quantitative variables is to design more SPF for various scenarios (Persaud, 2001). This technique presorts the crash data into given classes. SPF are then developed for these classes in an effort to increase the precision of the SPF while reducing the number of inputs. Some of the common classes of SPF are shown below:

- Rural two-lane highway segments
- Rural multilane undivided highway segments
- Rural multilane divided highway segments
- Rural freeway segments – 4 lanes
- Rural freeway segments – 6+ lanes
- Rural freeway segments within an interchange area – 4 lanes
- Rural freeway segments within an interchange area – 6+ lanes
- Urban two-lane arterial segments
- Urban multilane undivided highway segments
- Urban multilane divided highway segments
- Urban multilane divided highway segments
- Urban freeway segments – 4 lanes
- Urban freeway segments – 6 lanes
- Urban freeway segments – 8+ lanes
- Urban freeway segments within an interchange area – 4 lanes
- Urban freeway segments within an interchange area – 6 lanes
- Urban freeway segments within an interchange area – 8+ lanes

As can be seen, the classes utilize several variable categories that would no longer need to be a part of the statistical model (SafetyAnalyst, Module 1).

2.1.1.4 Safety Performance Functions Developed by Others

There have been a significant number of studies performed in developing SPF for various agencies across the United States and Canada. In the year 2001, the Federal Highway Administration (FHWA) commissioned the development of software tools for safety management known as the *SafetyAnalyst*. Even though studies occurred before 2002, the development of this software spurned several explorations into the most appropriate method of analyzing safety and developing statistical models. Currently, the *SafetyAnalyst* program supports only Level I SPF, but studies have looked into multivariate models to determine the impact other variable have on the crash frequency for a given type of roadway. Attached in the Appendix A is a more comprehensive summary of specific SPFs for rural roadway segments, taken from the progress report for ICT Project R27-18 “Crash Data Analysis and Engineering Solutions for Local Agencies.” As can be seen from the summarized studies, a majority of SPFs developed occurred before 2002 and classified as Type II SPFs. However, the *SafetyAnalyst* development team developed a Level I SPF relating crash frequencies to AADT in 2000.

2.1.2 Application to Intersections

Intersections occur when two or more roads meet at a common point. Due to this, crashes tend to be more common due to the conflicting traffic movements. As such, SPFs have been developed to accurately model the crash frequencies for various types of intersections based on unique roadway characteristics. The following sections describe difficulties in defining an intersection-related crash, levels of SPF, variables used for intersections, and the SPF developed in other studies.

2.1.2.1 Defining an Intersection

As a whole, intersections are similar to roadway segments, except for one distinct difference (SafetyAnalyst, Module 1). Intersections are discrete locations, so crash frequencies are not determined per unit length. No longer do the subdivisions of a segment have significance, but instead the difficulties are determining if a crash was a direct result of an intersection. In current practice, most crash reports will identify a crash being "at an intersection", "intersection related, but not at an intersection", or "not intersection related." However, this does not define which takes place at the intersection and which does not. The wording can be interpreted in a variety of ways. Previous research has defined intersection crashes are all at-intersection crashes and all intersection-related crashes that occur within 250 feet of the intersection (Harwood and Bauer, 2000). Another current practice includes marking all at-intersection and intersection-related crashes to the milepost of the intersection. Finally, other states using a link-node system where intersection related crashes are given one number for a node and roadway section crash is represented by two numbers for the two nodes that act as end-points of the segment. Ultimately, when forming SPF for intersection crashes, a system needs to be adopted and maintained for consistency.

2.1.2.2 Types of Analysis

Similar to roadway segments, there are two types of SPFs that are used in representing the crash frequencies as a function of variables. Level I SPFs are models based on only on traffic volumes. However, as opposed to roadway segments, intersections have two traffic volumes to consider. Both the AADT from the minor and major roads are part of the SPF. The functional form of a Level I SPF is as follows:

$$\mu_i = \alpha_0 \cdot (AADT_{Major,i})^{\alpha_1} \cdot (AADT_{Minor,i})^{\alpha_2}$$

where μ_i represents the predicted crash frequency per year at a given intersection and α_0, α_1 and α_2 are the regression parameters. These AADT encompass both the through and turning movements on the given roadway. Typically, the AADT from the major road has the largest impact and therefore has a larger coefficient in the model (Development of SPF for *SafetyAnalyst*). Level I SPF allow for simple comparisons of intersections safety based on a single parameter, traffic volume.

Similar to roadway segments, Level II SPFs incorporate variables other than just traffic volumes. In Level II SPF, variables such as weather conditions, roadway geometries, traffic data, and human factors to calculate the crash frequencies (SafetyAnalyst, Module 1). The variables are based on characteristics of both the minor and major roadways. For example, the type of median or shoulder width is important on both the major and minor roadways. Typically, a Level II SPF for intersections takes on the following functional form:

$$\mu_i = \alpha_0 \cdot (AADT_{Major,i})^{\alpha_1} \cdot (AADT_{Minor,i})^{\alpha_2} \cdot \exp(\alpha_3 X_{3i} + \alpha_4 X_{4i} + \cdots + \alpha_q X_{qi})$$

where μ_i is the predicted crash frequency per year at a given intersection, α_3, α_4, …, α_q are additional regression parameters, and X_{3i}, X_{4i}, …., X_{qi} are the variables of interest (Harwood and Bauer, 2000). The additional variables provide information for making future improvements to a given roadway. By realizing the impacts of changing a given geometry will have a certain safety impact allows engineers to have better safety judgment. Also, the additional variables allow for benefits such as education and enforcement.

Even though both of these Levels of SPFs involve different variables, they still use the same statistical model. Again, utilization of negative binomial distribution accurately models the crash frequencies. However, the overdispersion factor is treated differently for intersections. Whereas the overdispersion parameter varies for segments based on segment lengths, intersections are discrete entities where the overdispersion parameter does not vary (*SafetyAnalyst*, Module 1).

2.1.2.3 Variables for Intersections

Variables that depict roadway characteristics, driver behavior, climatic conditions, and traffic data are often included in the development of SPF. Similar to the roadway segments, the variables are broken into either quantitative or categorical variables. Quantitative variables are discrete values that represent a condition, characteristic, or parameter. In past studies, the quantitative variables used for developing SPFs for intersections include the following:

- Grade of Major Road
- Grade of Minor Road
- Lane width
- Median width on major road
- Median width on minor road
- Number of legs
- Number of through lanes on major road
- Number of left-turn lanes on major road
- Number of right-turn lanes on major road
- Number of through lanes on minor road
- Number of left-turn lanes on minor road
- Number of right-turn lanes on minor road
- Peak hour volume/design hour volume on major road
- Peak hour volume/design hour volume on minor road
- Shoulder width
- Traffic volume (AADT) on major road
- Traffic volume (AADT) on minor road
- Turning volumes

When modeling, these values can be entered directly into the SPF for determining the crash frequency.

The other type of variables used in developing SPFs is categorical variables. These variables are non-numerical and are ways to describe a given situation. Categorical variables used by others in SPFs are as follows:

- Access control
- Area type (rural/urban)
- Channelization
- Functional classification of roadway
- Lighting
- Median type on major road
- Median type on minor road
- Terrain types
- Traffic control type

During modeling, as previously described, categorical variables are often represented by 1 or 0 depending on whether the situation exists or not.

To reduce the number of categorical and quantitative variables, SPFs are designed for various scenarios (Persaud, 2001). The main categorical variables that can be incorporated into various scenarios are area type, number of legs, and intersection control type. The main classes of SPF are shown below:

- Rural four-legged minor stop control
- Rural four-legged all-way stop control
- Rural three-legged minor stop control
- Rural four-legged signalized
- Rural three-legged signalized
- Urban four-legged minor stop control
- Urban three-legged stop control
- Urban four-legged signalized
- Urban three-legged signalized
- Urban four-legged all-way stop control

As can be seen, the classes utilize several variable categories that would no longer need to be a part of the statistical model.

2.1.2.4 Safety Performance Functions Developed by Others

Similar to roadway segments, a significant number of studies have developed SPFs for various classes of intersections, with a large number of these coming after the announcement of *SafetyAnalyst* in 2001. Appendices B and C include a comprehensive summary of specific SPFs for various intersection classifications.

The tables are broken into two parts according to different modeling techniques. The first table focuses on SPFs with the assumption of a Poisson distribution. Vogt and Bared performed these studies in 1998 for Minnesota and Washington. The studies use a multivariable analysis and the variables include AADT, grade, roadside hazards, and speed limit. The second table assumes a negative binomial distribution. As shown, there were three main studies performed on various types of intersections in rural areas. The first study was a multivariate analysis of crashes in California and Michigan performed by Vogt. Harwood conducted the two remaining studies. One was a multivariable analysis in 1998 and the other was a comprehensive Level I SPF study of various states in 2002. For the Harwood 2002 study, a base SPF was found, and state specific correction factors were determined.

2.2 STATISTICAL MODELING

SPFs are based on statistical models of relating crash frequencies to roadway and driver characteristics. The crash frequencies are compared using regression analysis to determine which variables produce a significant cause-and-effect relationship. Whether a variable is significant is based on a user-defined parameter 'level-of-significance', α. The level-of-significance measures the maximum probability that the statistic would be observed. For most statistical models, α-values range from 0.01 to 0.10 (Hauer, 1996b). A smaller α indicates that it is more difficult to declare a variable significant. Since crashes are a serious subject, a larger value of α is used in order to include more variables into the model. A level of significance of 0.10 typically is used for the development of SPFs.

In the past, several statistical methods have been used in developing SPF; however, some represent the data more accurately. Currently, the two most widely used statistical models are lognormal regression and loglinear regression models (Harwood and Bauer, 2000).

2.2.1 Lognormal Regression Models

Lognormal regression models are used when the distribution of data is skewed. The lognormal distribution assumes that the natural log of the crash frequency has a normal distributed with mean μ and variance σ^2. This model proves to be especially effective when the data is inherently non-negative and the mean is relatively large. This type of distribution is common with intersections that have high volumes, such as at a signalized control.

The model for the i^{th} roadway segment/intersection with q parameters, X_{i1}, X_{i2}, …., X_{iq}, takes on the following form:

$$\log(\mu_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_q X_{iq}$$

where $\beta_0, \beta_1, \cdots, \beta_q$ are constant coefficients that need to be estimated. The above expression can also be written in the form

$$\mu_i = \exp(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_q X_{iq})$$

The main assumption is that the logarithm of the number of crashes is normally distributed. Typically, the linear regression coefficients are estimated using ordinary least-squared method (Harwood and Bauer, 2000).

2.2.2 Loglinear Regression Models

A loglinear model is a specific type of generalized linear model that the conditional relationship between two or more variables is analyzed by taking the logarithm of the dependent variable (Jeansonne, Loglinear Models). The two main types of loglinear models are the Poisson model and negative binomial model. Both of these models are described in more detail below.

2.2.2.1 Poisson Models

The Poisson distribution is a discrete distribution that expresses the probability of a certain number of events occurring in a given amount of time. These events occur with a known probability and are independent of the previous event (Hogg and Tanis, 2001). The Poisson is a limiting case of the binomial distribution. The probability of y_i events occurring in a given time interval is expressed as follows:

$$P(y_i) = \frac{\exp(-\mu_i)(\mu_i)^{y_i}}{y_i!}$$

where μ_i is the expected number of occurrences for a given interval (Vogt and Bared, 1998).

The linear model for the i^{th} roadway segment/intersection with q characteristics, X_{i1}, X_{i2}, …., X_{iq}, and regression coefficients β_0, β_1, β_2,…, β_q, takes on the form as follows:

$$\log(\mu_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_q X_{iq}$$

As opposed to the lognormal models, the model assumes that the crashes follow a Poisson distribution.

2.2.2.2 Negative Binomial Models

Similar to Poisson distributions and models, negative binomial distributions describe the occurrence of random and rare events. However, unlike the Poisson distribution where it is assumed that the mean is equal to variance, the negative binomial distribution compensates for situations where the variation is larger than the mean, or overdispersed. The variance in a negative binomial distribution can be expressed as $\mu_i + k(\mu_i)^2$ where k is the overdispersion parameter. Essentially, the negative binomial allows for variation caused by variables that are not included in the model. From past studies, crashes have been modeled better using negative binomial distribution due to variables that are not accounted for in the model.

Due to the overdispersion, negative binomial models utilize the following distribution function (Vogt and Bared, 1998):

$$P(y_i) = \frac{\Gamma\left(y_i + \frac{1}{k}\right)}{y_i!\,\Gamma\left(\frac{1}{k}\right)} \left(\frac{k\mu_i}{1+k\mu_i}\right)^{y_i} \left(\frac{1}{1+k\mu_i}\right)^{\frac{1}{k}}$$

As the overdispersion parameter approaches zero, meaning there is less variation, the distribution approaches a Poisson model.

The development of a linear model for the *i*th roadway segment/intersection with *q*th parameters, X_{i1}, X_{i2}, …., X_{iq}, and regression coefficients β_0, β_1, β_2,…, β_q, still takes on the following form:

$$\log(\mu_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_q X_{iq}$$

However, the linear model now assumes that the number of crashes follows a negative binomial distribution with parameters *α* and *k* (Harwood and Bauer, 2000).

2.3 SITE SCREENING PROCEDURE

Safety Performance Functions predict the number of crashes on a segment or at an intersection. However, a more involved analysis must be conducted to determine the actual safety of a specific location. To accomplish this, a two-step procedure that utilizes both the predicted crashes found by the SPFs and the actual number of crashes observed at a specified location was used. The first step of the procedure combines these values using an Empirical Bayesian relationship to find the estimated performance of the site, and the second step compares the estimated performance to the predicted performance by calculating the site's Potential for Safety Improvements.

2.3.1 Empirical Bayesian Relationship

Safety Performance Functions are only part of the overall safety evaluation process. The crashes that actually observed on a given segment or at an intersection also need to be accounted for while determining safety. However, the number of observed crashes can often be a misleading statistic due to the regression-to-the-mean phenomenon (Hauer, 2002). A location may have randomly high crash frequencies for one period, and if given the opportunity, the number of crashes would decrease in the following period without any safety improvements being made. Due to this, this location may be declared a site with safety concerns when in reality it was just experiencing statistical randomness. Conversely, a hazardous site may experience a period of randomly low crash frequencies and may be overlooked as a location in need of safety improvements. Regression-to-the-mean has a larger impact when a smaller sample sets are used.

The Empirical Bayesian method increases the precision of safety estimation by correcting for the regression-to-the-mean bias (Hauer, 2002). By calculating a weighted combination of the predicted with the observed number of crashes, the Empirical Bayesian method is able to provide an estimated number of crashes for a specific roadway segment or intersection.

The implementation of the Empirical Bayesian Method is connected with the results from the statistical modeling performed during the development of SPF. Using the overdispersion parameter found during modeling, a weight can be determined as follows:

$$w = \frac{1}{1 + k\sum_{n=1}^{N} P_n}$$

where k is the overdispersion parameter and P_n is the predicted number of crashes for a given roadway in year n. From the Empirical Bayesian procedure, the weight factor is then applied to the predicted and observed number of crashes to determine the estimated number of crashes as follows:

$$E_n = wP_1 + \frac{(1-w)\sum_{n=1}^{N} O_n}{\sum_{n=1}^{N} C_n}$$

where w is the aforementioned weight factor, P_1 is the predicted crashes for year 1, O_n is the observed number of crashes for a given year, and C_n is a yearly correction factor (*SafetyAnalyst*, Module 1s). Typically, this formula can be redefined as:

$$m = w \cdot P + (1-w)\frac{F}{n}$$

where m is the estimated number of crashes, w is the determined weight factor, P is the number of predicted crashes, F is the total number of crashes observed in n years. Based on this analysis, the longer the observations are made, the smaller the weight factor, which makes the estimated number of crashes weighted more towards the observed number. The result is consistent with the purpose of using the Empirical Bayesian procedure to increase precision by correcting for regression-to-the-mean; i.e. as the period of observation increases, the regression-to-the-mean phenomenon is not as severe.

2.3.2 Calculations of the Potential for Safety Improvement

The final step in screening sites is to determine a site's Potential for Safety Improvements (PSI). There are two methods of calculating a sites PSI: based on a site's expected crash frequency or based on a site's expected *excess* crash frequency (*SafetyAnalyst*, Module 1). The method is based on the site's expected crash frequency is merely the total predicted value determined after the Empirical Bayesian procedure. This frequency can easily be compared to other sites to determine where a greater number of crashes are more likely to occur. However, one of the limitations of this approach is that as AADT increases, so does the expected and predicted number of crashes, so comparisons made between sites are not as valid. Sites with higher traffic volumes will typically have higher predicted crashes.

The second method of analyzing a site's excess expected frequency is based on the difference between the predicted crashes frequency determined by Empirical Bayesian estimation and the predicted crash frequencies, as shown in Figure 2.1. The advantage to this technique is that the predicted crash frequencies are taken into account. When comparisons are made between different sites with varying AADT, they are made based on the how much the estimated crash frequencies surpass the expected prediction. Therefore, the natural growth in crash frequencies caused by increasing AADT does not contribute to the ranking as significantly.

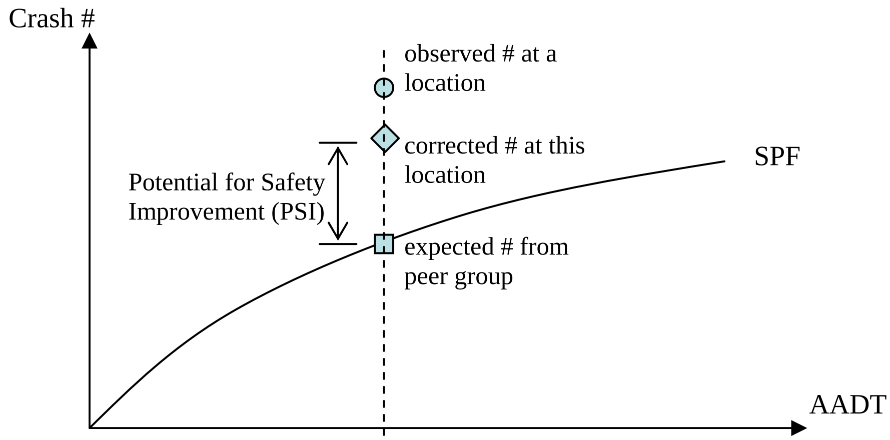


Figure 2.1 Empirical Bayesian method for PSI computation.

2.4 SUMMARY

In practice, one of the most common methods of defining the safety of a roadway segment, intersection, or ramp was to determine the crash rate for a given location. After several statistical studies, it was found that crashes do not exhibit a linear relationship. In reality, variables such as traffic volume and area type have non-linear impacts on crash frequencies. These relationships led to the development of SPFs as a way to assess the safety of a given site. SPFs are statistical models that relate crash frequencies to given variables. They are solved using general linear models and assuming a given distribution such as Poisson, negative binomial, or lognormal. Most SPFs use negative binomial distribution because the overdispersion parameters corrects for missing variables within the model. Using an Empirical Bayesian approach based on the overdispersion factor, SPFs are combined with existing crash data to determine sites that demonstrate a Potential for Safety Improvement (PSI).

Even though there has been a significant amount of research conducted on SPFs, more still needs to be done. As previously mentioned, the FHWA is developing *SafetyAnalyst* that uses SPF during the site-screening step of the safety analysis. In order to better identify the safety of sites, a greater number of SPFs need to be developed. From the existing literature, each state has a unique set of SPFs. By continuing the development of SPF for additional states, the site screening can more precisely rank sites with PSI because the SPFs are more situational specific.

CHAPTER 3 2001-2005 DATA SUMMARIES AND PROCESSING

The main components necessary for developing SPFs are data on crashes and roadway characteristics. The crash data provides the number of crashes and severity while the roadway data provides important information such as traffic volumes, number of lanes, functional classifications, and area type. With the aid of IDOT, the investigators have been able to obtain, process, and clean relevant data to form a dataset usable in developing SPFs. The following sections present a detailed description of the available datasets and the methodology used to combine the datasets to develop SPFs.

3.1 SUMMARY OF RECEIVED DATA

IDOT provided four datasets for the development of SPFs. The datasets include information on the roadways, the crashes, location of intersections, and translation tables. A brief description of each data set and its role in the developing SPFs is provided below.

3.1.1 Roadway Data

IDOT provided the roadway dataset Hwy04_sw.shp, which is in GIS format. The dataset includes 113 fields of various roadway and traffic characteristics. The fields that are the most instrumental to the development of SPFs are:

- Average Annual Daily Traffic (AADT)
- Beginning Station
- County
- Ending Station
- Functional Class
- Inventory (Key Route)
- Median Type
- Number of Lanes
- Segment Length
- Township
- Urban Code

Subsequent sections will describe the precise use for each of these fields. Even though only 11 of the 113 fields were used for SPFs, the remaining variables do provide great information that can be included in the development of multi-variate causality models. The remaining variables summarize a variety of roadway characteristics such as the shoulder information, lane widths, and roadway distress information. A complete list of the variables is in Appendix D.

3.1.2 Crash Data

Similar to roadway information, IDOT provided the 2001-2005 crash datasets saf03_2005.shp, saf03_2004.shp, saf03_2003.shp, saf03_2002.shp, and saf03_2001.shp. Similar to the roadway file, these files are also in GIS formatting. The crash datasets include 57 fields with information about the cause, location, and time of the crash. The most relevant information to the development of SPFs from these datasets is:

- Case-ID
- County

- Crash-Type
- Mile
- Route Name
- TS Route

Subsequent sections of this report will describe the precise use for each of these fields. The remaining variables may have some use in a multi-variate setting (such as lighting, weather, etc.), but others merely describe the crash and do not have much foreseeable use. A complete list of the variables is in Appendix D.

3.1.3 Intersection Data

The intersection datasets include information of where intersections are located on IL Routes, Interstate, and U.S. Routes. The files do not have information about intersections along TS Routes and local roads. IDOT provided two intersection datasets. The first one, Reference.shp, contains 48,153 point locations on State Routes where there is a reference point described by the IRIS mainframe system as of the year-end 2004 file. The second file, Traffic_Control.shp, contains 39,290 point locations on State Routes where a traffic control code exists that indicates some type of an intersection as of the year-end 2004 file. The discrepancies occur when the main route intersects a small local road where there is insufficient information available. Staff resources are limiting the effort to create a complete and comprehensive dataset. The most relevant information in the datasets for the development of SPFs is:

- County
- Inventory of Major Road
- Minor Road Name
- Minor Route Reference Name
- Station
- Township
- Traffic Control

The subsequent sections will describe the precise use for each of these fields.

3.1.4 Translation Data

When trying to merge the datasets together, a problem exists in the way datasets classify the route. The crash datasets utilizes the TS Route Number and the roadway datasets use Inventory Numbers. IDOT provided translation tables that converted TS Route Number to Inventory Numbers based on the beginning and ending station of the route. In other words, the tables provide the beginning and ending station of all the TS routes and provide the beginning and ending station of the corresponding Inventory Number. IDOT supplied six translation tables because the roadway network changes on a yearly basis with the realignment of roads, addition/subtraction of jurisdiction of roads, and other factors. The subsequent sections will describe the translation tables and their role in developing SPFs.

3.2 DATA PROCESSING AND ORGANIZING

After receiving the datasets from IDOT, organizing and processing the datasets is necessary to form a comprehensive dataset that is useable to develop SPFs. The following sections provide a detailed methodology to merge the data, organize the data into intersection/segment related crashes, and separate the datasets into different peer groups.

3.2.1 Merging of Datasets

As illustrated, the Crash and Roadway datasets represent roads with different systems. The crash data uses the TS Route Number, which provides a numeric representation of the route, while the roadway dataset utilizes Inventory Number, a string of numbers that represent of the route name and other roadway characteristics through a sequence of numbers. In addition to different numbering systems, the two datasets use different methods for marking position along the given route. The crash dataset uses mileposts, which act as an absolute numbering system over the whole state. For example, if an Interstate stretches the entire length of the state, the milepost would be continuous for the entire duration of the Interstate. In a contrary technique, the GIS roadway dataset utilizes stationing, which provides a localized representation of position.

An extreme amount of effort is required to ensure that the two datasets are compatible. Ultimately, a combination of two techniques is used to create a common field in order to merge the two datasets. The first technique utilizes the translation tables provided by IDOT. Since the roadway network changes each year, a translation table is available for each year. SAS merges the translation tables with the corresponding crash data from the respective year to transform the TS Route and Milepost into Key Route Number and Station, respectively. The second technique involves a self-made conversion table developed to create a common field. The conversion table recognizes the Route Name, county, and milepost in the crash datasets and assigns the crash with the appropriate Inventory Number and Station based on the appropriate conversion. These two techniques allow for the recognition and assignment of over 99.6% of the crashes. After the creation of a common field, SAS merges a total crash dataset (all five years merged) with the roadway dataset through a series of SQL commands based on the station and Inventory of the crash and the Beginning Station, Ending Station, and Inventory of the roadway. Figure 3.1 shows a sample of the merging technique.

Crash Dataset

Crash Number	TS Route	Inventory	Milepost	Station	Severity
1	12	1	3.5	12.5	K
2	12	1	9.3	18.3	A
3	13	2	12.8	21.2	A
:	:	:	:	:	:
:	:	:	:	:	:

Roadway Dataset

Segment Number	Inventory	Beginning Station	Ending Station	AADT
1	1	18.1	18.9	16000
2	2	20.8	21.4	17500
3	2	21.4	21.9	17700
:	:	:	:	:
:	:	:	:	:

Figure 3.1. Example of merging datasets.

After merging the two datasets, crash characteristics and roadway characteristics are used to define each crash.

3.2.2 Separation of Intersections

A key aspect of developing SPFs is to distinguish the difference between intersection related crashes and segment related crashes. Per the standard set forth by *SafetyAnalyst*, an intersection related crash is defined as a crash that occurs within 250 feet of an intersection. In order to determine whether a crash occurs within 250 feet of an intersection, the crash dataset is merged with the intersection dataset.

Prior to merging the crash dataset with the intersection dataset, some manipulation and organization is necessary to complete the intersection dataset. The first step is to combine the two distinct intersection datasets into one usable dataset. By merging the two datasets into one complete data set, all of the intersections and traffic control device, if available, are available for analysis. For the intersections where traffic control information is not available, then the traffic control is set as "undetermined." The second challenge to overcome is that the dataset references the minor route by road name, marked route name, or reference name, but not Inventory Number. In addition, the intersection dataset does not contain the station along the minor route, so it is impossible to determine site-specific roadway information for the minor route. The solution is to use localized averages for the minor routes. Averages are calculated based on the roadway dataset for each road name, marked route name, and reference name after being segregated into counties and townships. This allows for a minor route AADT to be used in the development of intersection SPFs. However, not all of the minor route reference names are available in the roadway dataset, especially for smaller local roads. For these circumstances, IDOT provided a table of average AADTs for rural and urban local roads in each county. The final challenge is that the intersection datasets include ramp locations and some overpass intersections (grade separations) along interstates. Per the request of IDOT, these intersections are eliminated from the dataset, so all crashes that occur within these vicinities become a part of the corresponding segment.

Based on the aforementioned circumstances, the intersections are merged with the minor route. Similar to the SQL techniques utilized in merging crashes with roadways, intersections are merged with major route roadway based on Inventory Number and Station. Careful considerations and adjustments are necessary when an intersection falls between two segments. Once the intersections have both major and minor roadway characteristics, the new intersection dataset is merged with the translated crash dataset based on Inventory Number and whether the station falls within 250 feet of the intersection. In situations where a crash occurred within 250 feet of two intersections, the crash is assigned to the closer of the two intersections. By comparing the original crash dataset with the crashes that are classified as intersection related, the remaining are segment related crashes.

After the crashes are separated into intersection related crashes and segment related crashes, the number of crashes that occur on each specific segment and intersection is determined. Utilizing SAS, the number of crashes for each severity type is calculated that occurred on given segments. The summation of crashes are then merged back with the roadway datasets creating an additional field on the roadway dataset that expresses the number of crashes. Figure 3.2 shows a simplified example on how the files would appear.

Segment Number	Inventory	Beginning Station	Ending Station	AADT	Number of Crashes
1	1	18.1	18.9	16000	1
2	2	20.8	21.4	17500	2
3	2	21.4	21.9	17700	0
:	:	:	:	:	:
:	:	:	:	:	:

Figure 3.2. Sample merged dataset.

After determining the number of crashes for each severity, the only remaining step in organizing the data is to divide the segments and intersections into particular peer groups.

3.2.3 Formation of Peer Groups

As mentioned in Section 2.1.1.3 and Section 2.1.2.3, SPFs are divided into peer groups that allow similar segments and intersections to be modeled together. By doing this, some categorical variables, such as functional class and area type, are accounted for in the model without having them represented by variables. In total, there are 17 different peer groups for segments and 10 different peer groups for intersection suggested by *SafetyAnalyst*. However, the roadway dataset did not include enough information to form all the different peer groups. As a result, the analysis only includes the following twelve peer groups for segments:

- Rural Two-Lane Highway
- Rural Multilane Undivided Highway
- Rural Multilane Divided Highway
- Rural Freeway, 4 Lanes
- Rural Freeway, 6+ Lanes
- Urban Two-Lane Highway
- Urban One-Way Arterial
- Urban Multilane Undivided Highway
- Urban Multilane Divided Highway
- Urban Freeway, 4 Lanes
- Urban Freeway, 6 Lanes
- Urban Freeway, 8+ Lanes

The main limitation is the inability to identify interchange areas. No field in the available IDOT databases provides that information, and *SafetyAnalyst* does not provide guidance on what defines an interchange area.

Similar to segments, the development of intersection SPFs is only able to include the following eight peer groups:

- Rural Minor Leg Stop Control
- Rural All-Way Stop Control
- Rural Signalized Intersection
- Rural Undetermined
- Urban Minor Leg Stop Control
- Urban All-Way Stop Control
- Urban Signalized Intersection
- Urban Undetermined

The limitation for intersections, as previously mentioned, is the lack of information on the minor routes; the inability to recognize the number of legs at intersections limited the number peer groups. However, two additional peer groups entitled "Rural Undetermined" and "Fatal Undetermined" is formed as a method to group all of the unknown intersections together.

The methodology for how roadway characteristics are divided into the given peer groups are presented below:

Area Type – The Urban field of the roadway dataset dictates whether a segment or intersection is in an urban or rural area

Number of Lanes – The lanes field in the roadway dataset states the number lanes of a given segment

Undivided Highway – The median type field dictates what type of median is used. If no median is present, then the segment is classified being an undivided highway

Divided Highway – The median type field dictates what type of median is used. If any type of median is present, then the segment is classified being a divided highway
Functional Class – The functional class field states whether a roadway segment is an interstate, arterial, collector, or local road
Traffic Control Type – The traffic control field in the intersection file indicates whether an intersection is two-way stop control, all-way stop control, signalized, or undetermined

After the categorization of segments into the specific peer groups occurred, modeling commenced.

CHAPTER 4 SAFETY PERFORMANCE FUNCTIONS FOR ILLINOIS

The Illinois-specific SPFs were developed based on a negative binomial regression (except for fatal intersections which used Poisson) using SAS GENMOD software. As stated in Section 2.2, the negative binomial regression is an appropriate regression method for data that has low occurrence frequency and where the variance exceeds the mean. Due to the nature of the data, the SPFs were developed for a five-year period. This means, that the results of the regression produce the expected number of crashes per five years. The reason for using five years is that only one-year worth of roadway data was available. If we were able to match each year with more time consistent data, then regression on a per year basis would have been possible. So essentially, the total number of crashes for five years was expressed as one observation as opposed to five observations.

4.1 SEGMENT SAFETY PERFORMANCE FUNCTIONS

Safety Performance Functions for segments include three aspects to predict the expected number of crashes: segment length, traffic volume, and regression parameters. Due to past research in the area of statistical modeling crashes, the fundamental form of SPFs is

$$\mu_i = (SL)_i \cdot e^a \cdot (AADT_i)^b$$

where μ_i is the expected number of crashes for given segment *i*, SL_i is the segment length in miles of segment *i*, $AADT_i$ is the Average Annual Daily Traffic of segment *i*, and *a* and *b* are the regression coefficients. Provided in the following sections are a summary of the Illinois Segment SPFs for Fatal crashes, Type-A injury crashes, Type-B injury crashes, and Fatal and Injury crashes.

4.1.1 Fatal Crashes

Using SAS, estimates of the regression parameters were found for the various Fatal SPF peer groups. Shown below in Table 4.1.1 are the regression coefficients, dispersion parameter, number of segments, and total miles within each peer group. Also, a summary of these parameters and additional regression related information can be found in Appendix G.

Table 4.1.1. Fatal SPF Parameters and Characteristics for Segments

Peer Group	Regression Coefficients		Dispersion Parameter Per Mile	Number of Segments	Total Miles
	Intercept (a)	LogAADT (b)			
Rural Two-Lane Highway	-7.249	0.521	77.886	25,885	7,968.06
Rural Multilane Undivided Highway	-3.925	0.068	25.776	217	33.20
Rural Multilane Divided Highway	-9.693	0.725	136.217	2,022	307.82
Rural Freeway, 4 Lanes	-10.575	0.881	20.675	3,116	1,337.30
Rural Freeway, 6+ Lanes	-33.403	2.924	87.322	138	25.28
Urban Two-Lane Highway	-6.899	0.423	146.043	10,091	1,366.28
Urban One-Way Arterial	-82.156	8.386	278.130	1,263	110.05
Urban Multilane Undivided Highway	-7.707	0.475	195.963	4,285	529.53
Urban Multilane Divided Highway	-8.865	0.606	271.873	9,118	1,030.25
Urban Freeway, 4 Lanes	-16.256	1.371	20.282	2,215	524.92
Urban Freeway, 6 Lanes	-6.927	0.499	23.361	1,453	310.61
Urban Freeway, 8+ Lanes	-15.855	1.247	21.943	437	63.51

Appendix E presents a graphical representation of these models.

4.1.2 Type-A Injury Crashes

Using SAS, estimates of the regression parameters were found for the various Type-A SPF peer groups. Shown below in Table 4.1.2 are the regression coefficients, dispersion parameter, number of segments, and total miles within each peer group. Also, a summary of these parameters and additional regression related information can be found in Appendix H.

Table 4.1.2. Type-A SPF Parameters and Characteristics for Segments

Peer Group	Regression Coefficients		Dispersion Parameter Per Mile	Number of Segments	Total Miles
	Intercept (a)	LogAADT (b)			
Rural Two-Lane Highway	-5.194	0.472	26.569	25,885	7,968.06
Rural Multilane Undivided Highway	-16.855	1.630	17.061	217	33.20
Rural Multilane Divided Highway	-6.327	0.583	42.128	2,022	307.82
Rural Freeway, 4 Lanes	-9.339	0.952	6.764	3,116	1,337.30
Rural Freeway, 6+ Lanes	-3.983	0.358	15.007	138	25.28
Urban Two-Lane Highway	-5.979	0.571	42.354	10,091	1,366.28
Urban One-Way Arterial	-3.398	0.302	154.099	1,263	110.05
Urban Multilane Undivided Highway	-7.786	0.728	41.662	4,285	529.53
Urban Multilane Divided Highway	-6.667	0.659	36.957	9,118	1,030.25
Urban Freeway, 4 Lanes	-10.045	1.013	7.582	2,215	524.92
Urban Freeway, 6 Lanes	-7.910	0.815	4.868	1,453	310.61
Urban Freeway, 8+ Lanes	-12.906	1.226	4.548	437	63.51

Appendix E presents a graphical representation of these models.

4.1.3 Type-B Injury Crashes

Using SAS, estimates of the regression parameters were found for the various Type-B SPF peer groups. Shown below in Table 4.1.3 are the regression coefficients, dispersion parameter, number of segments, and total miles within each peer group. Also, a summary of these parameters and additional regression related information can be found in Appendix I.

Table 4.1.3. Type-B SPF Parameters and Characteristics for Segments

Peer Group	Regression Coefficients		Dispersion Parameter Per Mile	Number of Segments	Total Miles
	Intercept (a)	LogAADT (b)			
Rural Two-Lane Highway	-5.039	0.523	19.055	25,885	7,968.06
Rural Multilane Undivided Highway	-17.955	2.066	20.564	217	33.20
Rural Multilane Divided Highway	-7.293	0.795	25.334	2,022	307.82
Rural Freeway, 4 Lanes	-4.897	0.548	4.638	3,116	1,337.30
Rural Freeway, 6+ Lanes	-4.274	0.452	11.678	138	25.28
Urban Two-Lane Highway	-6.354	0.698	26.203	10,091	1,366.28
Urban One-Way Arterial	-5.563	0.551	88.212	1,263	110.05
Urban Multilane Undivided Highway	-6.551	0.714	26.398	4,285	529.53
Urban Multilane Divided Highway	-7.062	0.780	23.635	9,118	1,030.25
Urban Freeway, 4 Lanes	-9.628	1.052	5.277	2,215	524.92
Urban Freeway, 6 Lanes	-11.567	1.235	3.353	1,453	310.61
Urban Freeway, 8+ Lanes	-12.060	1.301	2.211	437	63.51

Appendix E presents a graphical representation of these models.

4.1.4 Fatal and Injury Crashes

Using SAS, estimates of the regression parameters were found for the various Fatal and Injury SPF peer groups. Shown below in Table 4.1.4 are the regression coefficients, dispersion parameter, number of segments, and total miles within each peer group. Also, a summary of these parameters and additional regression related information can be found in Appendix J.

Table 4.1.4. Fatal and Injury SPF Parameters and Characteristics for Segments

Peer Group	Regression Coefficients		Dispersion Parameter Per Mile	Number of Segments	Total Miles
	Intercept (a)	LogAADT (b)			
Rural Two-Lane Highway	-4.435	0.525	14.386	25,885	7,968.06
Rural Multilane Undivided Highway	-3.005	0.259	17.016	217	33.20
Rural Multilane Divided Highway	-7.767	0.923	18.273	2,022	307.82
Rural Freeway, 4 Lanes	-6.687	0.804	3.934	3,116	1,337.30
Rural Freeway, 6+ Lanes	-7.406	0.872	9.671	138	25.28
Urban Two-Lane Highway	-3.557	0.462	22.543	10,091	1,366.28
Urban One-Way Arterial	-5.495	0.689	78.242	1,263	110.05
Urban Multilane Undivided Highway	-4.876	0.647	25.556	4,285	529.53
Urban Multilane Divided Highway	-6.206	0.761	20.435	9,118	1,030.25
Urban Freeway, 4 Lanes	-10.369	1.201	4.594	2,215	524.92
Urban Freeway, 6 Lanes	-13.022	1.425	3.032	1,453	310.61
Urban Freeway, 8+ Lanes	-10.508	1.228	2.586	437	63.51

Appendix E presents a graphical representation of these models.

4.2 INTERSECTION SAFETY PERFORMANCE FUNCTIONS

Safety Performance Functions for intersection vary from segments because of the inclusion of a minor route AADT and the fact that intersections are discrete locations that do not have any length scaling necessary. Therefore, the three aspects necessary for an intersection SPF to predict the expected number of crashes are major route traffic volume, minor route traffic volume, and regression parameters. Due to past research in the area of statistical modeling crashes, the fundamental form of SPFs is

$$\mu_i = e^a \cdot (AADT_{Major,i})^b \cdot (AADT_{Minor,i})^c$$

where μ_i is the expected number of crashes for given intersection *i*, $AADT_{Major,i}$ is the Average Annual Daily Traffic of the major route at intersection *i*, $AADT_{Minor,i}$ is the Average Annual Daily

Traffic of the minor route at intersection *i*, and *a*, *b*, and *c* are the regression coefficients. It should be noted that while performing a regression analysis, some of the minor route AADTs were found to be non-significant with an alpha level of significance of 0.10. In these situations, the variable was eliminated from the analysis and the regression rerun to produce a new set of regression parameters. Provided in the following sections are a summary of the Illinois Intersections SPFs for Fatal crashes, Type-A injury crashes, Type-B injury crashes, and Fatal and Injury crashes.

4.2.1 Fatal Crashes

Using SAS, estimates of the regression parameters were found for the various Fatal SPF peer groups. Shown below in Table 4.2.1 are the regression coefficients, dispersion parameter, number of segments, and total miles within each peer group. Also, a summary of these parameters and additional regression related information can be found in Appendix G.

Table 4.2.1. Fatal SPF Parameters and Characteristics for Intersections

Peer Group	Regression Coefficients			Dispersion Parameter	Number of Intersections
	Intercept (a)	$LogAADT_{Major}$ (b)	$LogAADT_{Minor}$ (c)		
Rural Minor Leg Stop Control	-7.738	0.215	0.355	1.000	14,933
Rural All-Way Stop Control	-25.464	2.520	0.000	1.000	351
Rural Signalized Intersection	-16.691	1.501	0.000	1.000	199
Rural Undetermined	-7.288	0.240	0.187	1.000	5,579
Urban Minor Leg Stop Control	-9.329	0.386	0.305	1.000	12,121
Urban All-Way Stop Control	3.518	-0.839	0.000	1.000	132
Urban Signalized Intersection	-13.380	0.890	0.213	1.000	4,311
Urban Undetermined	-7.838	0.429	0.000	1.000	4,250

Appendix F presents a graphical representation of these models.

4.2.2 Type-A Injury Crashes

Using SAS, estimates of the regression parameters were found for the various Fatal SPF peer groups. Shown below in Table 4.2.2 are the regression coefficients, dispersion parameter, number of segments, and total miles within each peer group. Also, a summary of these parameters and additional regression related information can be found in Appendix H.

Table 4.2.2. Type-A SPF Parameters and Characteristics for Intersections

Peer Group	Regression Coefficients			Dispersion Parameter	Number of Intersections
	Intercept (a)	$LogAADT_{Major}$ (b)	$LogAADT_{Minor}$ (c)		
Rural Minor Leg Stop Control	-8.574	0.601	0.293	2.139	14,933
Rural All-Way Stop Control	-6.095	0.544	0.000	2.153	351
Rural Signalized Intersection	-11.243	1.190	0.000	1.203	199
Rural Undetermined	-7.132	0.565	0.067	2.483	5,579
Urban Minor Leg Stop Control	-7.795	0.556	0.214	0.979	12,121
Urban All-Way Stop Control	-7.825	0.772	0.000	0.337	132
Urban Signalized Intersection	-9.384	0.765	0.259	0.695	4,311
Urban Undetermined	-6.456	0.519	0.000	1.475	4,250

Appendix E presents a graphical representation of these models.

4.2.3 Type-B Injury Crashes

Using SAS, estimates of the regression parameters were found for the various Fatal SPF peer groups. Shown below in Table 4.2.3 are the regression coefficients, dispersion parameter, number of segments, and total miles within each peer group. Also, a summary of these parameters and additional regression related information can be found in Appendix I.

Table 4.2.3. Type-B SPF Parameters and Characteristics for Intersections

Peer Group	Regression Coefficients			Dispersion Parameter	Number of Intersections
	Intercept (a)	$\text{LogAADT}_{\text{Major}}$ (b)	$\text{LogAADT}_{\text{Minor}}$ (c)		
Rural Minor Leg Stop Control	-9.220	0.764	0.265	1.338	14,933
Rural All-Way Stop Control	-5.927	0.456	0.177	1.907	351
Rural Signalized Intersection	-14.389	1.482	0.170	1.103	199
Rural Undetermined	-7.547	0.690	0.050	1.859	5,579
Urban Minor Leg Stop Control	-8.306	0.670	0.266	0.936	12,121
Urban All-Way Stop Control	-7.982	0.893	0.000	0.989	132
Urban Signalized Intersection	-8.661	0.801	0.254	0.649	4,311
Urban Undetermined	-6.775	0.649	0.000	1.415	4,250

Appendix E presents a graphical representation of these models.

4.2.4 Fatal and Injury Crashes

Using SAS, estimates of the regression parameters were found for the various Fatal SPF peer groups. Shown below in Table 4.2.4 are the regression coefficients, dispersion parameter, number of segments, and total miles within each peer group. Also, a summary of these parameters and additional regression related information can be found in Appendix J.

Table 4.2.4. Fatal and Injury SPF Parameters and Characteristics for Intersections

Peer Group	Regression Coefficients			Dispersion Parameter	Number of Intersections
	Intercept (a)	$\text{LogAADT}_{\text{Major}}$ (b)	$\text{LogAADT}_{\text{Minor}}$ (c)		
Rural Minor Leg Stop Control	-8.005	0.674	0.272	1.429	14,933
Rural All-Way Stop Control	-5.907	0.507	0.171	1.638	351
Rural Signalized Intersection	-13.502	1.443	0.151	1.103	199
Rural Undetermined	-6.638	0.631	0.065	1.990	5,579
Urban Minor Leg Stop Control	-7.580	0.639	0.254	0.905	12,121
Urban All-Way Stop Control	-6.596	0.778	0.000	1.126	132
Urban Signalized Intersection	-8.248	0.793	0.252	0.664	4,311
Urban Undetermined	-5.958	0.602	0.000	1.285	4,250

Appendix E presents a graphical representation of these models.

CHAPTER 5 SITE SCREENING WITH 2001-2005 DATA

Utilizing the EB procedure, we screened the Illinois roadway network to determine sites with a potential for safety improvements (PSI). Similar to the SPFs, the PSI represents the safety of segments and intersections over a five-year period. Two separate analysis techniques were used in determining the PSI. The first looked at segments and intersections as individual elements, and the second approach utilized a sliding window.

5.1 SITE SPECIFIC

Performing a site-specific analysis of the roadway networks PSI treats each individual segment and intersection as a separate entity. In other words, each segment is compared with every other segment as opposed to grouping adjacent segments together to perform a corridor study. Intersections were analyzed in a similar manner. The segment and intersection PSIs were calculated for five different situations; Fatal crashes only, Type-A Injury crashes only, Type-B Injury crashes only, Fatal and Injury crashes together (based on SPF developed for a pool of all crashes), and a Weighted average of PSIs (weights of 25 for Fatal PSIs, 5 for Type-A PSIs, and 1 for Type-B PSIs). The technical review panel decided on the weighting scale of 25/5/1 in the August 2007 meeting because the scale reflects the relative significance of each severity. The scale changed from the original of 3,760,000/188,000/48,200 to avoid being too heavily dependent on fatal crashes. The following sections provide the analysis of segment and intersection PSIs for the different severities.

5.1.1 Fatal Site Screening of Segments

A Fatal PSI was calculated for all roadway segments. Table 5.1.1 shows the top ten segments with the highest Fatal PSI.

Table 5.1.1. Fatal PSI for Segments

Rank	Inventory	Route Name	Beginning Station	Ending Station	AADT	Peer Group	Fatal PSI Per Mile
1	016 10094 000000	I 090	30.41	30.62	296,400	Urban Freeway, 8+ Lanes	6.13
2	016 10057 000000	I 057	17.33	17.78	145,900	Urban Freeway, 6 Lanes	5.42
3	016 10290 000000	I 290	15.68	15.81	207,100	Urban Freeway, 8+ Lanes	5.17
4	016 10094 000000	I 094	33.86	34.52	234,900	Urban Freeway, 8+ Lanes	4.66
5	101 20517 000000	US020B	3.99	4.14	34,300	Urban Multilane Divided Highway	4.32
6	056 20336 000000	IL031	16.95	17.28	30,000	Urban Multilane Divided Highway	4.09
7	016 10094 000000	I 094	34.81	35.33	248,200	Urban Freeway, 8+ Lanes	4.08
8	099 10080 000000	I 080	11.61	12.18	94,000	Urban Freeway, 4 Lanes	3.71
9	016 10094 000000	I 094	43.51	44.1	153,000	Urban Freeway, 6 Lanes	3.55
10	016 20397 000000	IL083	3.32	3.69	36,600	Urban Multilane Undivided Highway	3.42

5.1.2 Type-A Injury Site Screening of Segments

A Type-A PSI was calculated for all roadway segments. Table 5.1.2 shows the top ten segments with the highest Type-A PSI.

Table 5.1.2. Type-A PSI for Segments

Rank	Inventory	Route Name	Beginning Station	Ending Station	AADT	Peer Group	Type-A PSI Per Mile
1	016 10094 000000	I 090	19.65	19.77	270,800	Urban Freeway, 8+ Lanes	55.59
2	082 10070 000000	I 055	0	0.34	121,800	Urban Freeway, 8+ Lanes	39.77
3	016 10094 000000	I 090	27.75	28.06	186,600	Urban Freeway, 8+ Lanes	36.15
4	016 10094 000000	I 090	31.13	31.93	293,600	Urban Freeway, 8+ Lanes	30.38
5	016 10094 000000	I 090	20.12	20.28	264,200	Urban Freeway, 8+ Lanes	28.93
6	016 10094 000000	I 090	30.72	31.02	297,400	Urban Freeway, 8+ Lanes	28.47
7	016 10057 000000	I 057	17.33	17.78	145,900	Urban Freeway, 6 Lanes	27.19
8	022 10290 000000	I 290	3.69	3.99	144,300	Urban Freeway, 6 Lanes	26.99
9	016 10290 000000	I 290	15.84	15.97	207,100	Urban Freeway, 8+ Lanes	26.53
10	101 20517 000000	US020B	4.87	4.99	32,800	Urban Multilane Divided Highway	25.53

5.1.3 Type B-Injury Site Screening of Segments

A Type-B PSI was calculated for all roadway segments. Table 5.1.3 shows the top ten segments with the highest Type-B PSI.

Table 5.1.3. Type-B PSI for Segments

Rank	Inventory	Route Name	Beginning Station	Ending Station	AADT	Peer Group	Type-B PSI Per Mile
1	016 10094 000000	I 090	19.65	19.77	270,800	Urban Freeway, 8+ Lanes	289.61
2	016 20341 000000	US041	38.2	38.27	132,700	Urban Multilane Divided Highway	172.73
3	016 10094 000000	I 090	27.75	28.06	186,600	Urban Freeway, 8+ Lanes	139.33
4	016 10094 000000	I 090	28.81	28.86	228,000	Urban Freeway, 8+ Lanes	130.33
5	016 10094 000000	I 090	27.05	27.15	286,600	Urban Freeway, 8+ Lanes	127.80
6	016 10290 000000	I 290	20.1	20.2	200,600	Urban Freeway, 8+ Lanes	112.82
7	016 10094 000000	I 090	23.42	23.53	291,300	Urban Freeway, 8+ Lanes	101.73
8	082 10070 000000	I 055	0	0.34	121,800	Urban Freeway, 8+ Lanes	100.76
9	016 10290 000000	I 290	15.84	15.97	207,100	Urban Freeway, 8+ Lanes	97.56
10	016 10290 000000	I 290	9.21	9.36	186,600	Urban Freeway, 6 Lanes	94.99

5.1.4 Fatal and Injury Site Screening of Segments

A Fatal and Injury PSI was calculated for all roadway segments. Table 5.1.4 shows the top ten segments with the highest Fatal and Injury PSI.

Table 5.1.4. Fatal and Injury PSI for Segments

Rank	Inventory	Route Name	Beginning Station	Ending Station	AADT	Peer Group	FI PSI Per Mile
1	016 10094 000000	I 090	19.65	19.77	270,800	Urban Freeway, 8+ Lanes	424.49
2	016 10094 000000	I 090	28.81	28.86	228,000	Urban Freeway, 8+ Lanes	263.62
3	016 20341 000000	US041	38.2	38.27	132,700	Urban Multilane Divided Highway	236.01
4	016 10094 000000	I 090	33.41	33.43	279,300	Urban Freeway, 8+ Lanes	220.15
5	016 10094 000000	I 090	27.75	28.06	186,600	Urban Freeway, 8+ Lanes	165.40
6	057 20704 000000	I 055B	6.22	6.32	40,700	Urban Multilane Divided Highway	155.82
7	016 10290 000000	I 290	15.84	15.97	207,100	Urban Freeway, 8+ Lanes	147.92
8	082 10070 000000	I 055	0	0.34	121,800	Urban Freeway, 8+ Lanes	144.76
9	016 10290 000000	I 290	20.1	20.2	200,600	Urban Freeway, 8+ Lanes	140.16
10	016 20345 000000	US020	1.43	1.62	37,700	Urban Multilane Undivided Highway	130.63

5.1.5 Weighted Site Screening of Segments

A Weighted PSI was calculated for all roadway segments. The weights were 25 for the Fatal PSI, 5 for the Type-A PSI, and 1 for the Type-B PSI. Table 5.1.5 shows the top ten segments with the highest Weighted PSI.

Table 5.1.5. Weighted PSI for Segments

Rank	Inventory	Route Name	Beginning Station	Ending Station	AADT	Peer Group	Total PSI Per Mile
1	016 10094 000000	I 090	19.65	19.77	270,800	Urban Freeway, 8+ Lanes	604.78
2	016 10094 000000	I 090	27.75	28.06	186,600	Urban Freeway, 8+ Lanes	354.79
3	016 10057 000000	I 057	17.33	17.78	145,900	Urban Freeway, 6 Lanes	350.27
4	082 10070 000000	I 055	0.00	0.34	121,800	Urban Freeway, 8+ Lanes	296.57
5	016 20341 000000	US 041	38.2	38.27	132,700	Urban Multilane Undivided Highway	285.97
6	016 10094 000000	I 094	33.86	34.52	234,900	Urban Freeway, 8+ Lanes	263.21
7	016 10094 000000	I 090	26.8	27.01	233,300	Urban Freeway, 8+ Lanes	252.12
8	016 10094 000000	I 090	31.13	31.93	293,600	Urban Freeway, 8+ Lanes	250.56
9	016 10290 000000	I 290	15.84	15.97	207,100	Urban Freeway, 8+ Lanes	227.84
10	016 10094 000000	I 090	23.42	23.53	291,300	Urban Freeway, 8+ Lanes	220.24

5.1.6 Fatal Site Screening of Intersections

A Fatal PSI was calculated for all intersections. Table 5.1.6 shows the top ten intersections with the highest Fatal PSI.

Table 5.1.6. Fatal PSI for Intersections

Rank	Major Route Inventory	Major Route Name	Station	Minor Route Name	Major AADT	Minor AADT	Peer Group	Fatal PSI
1	016 20345 000000	IL 019	21.9	KIMBALL AV	37,800	13,095	Urban Signalized Intersection	0.35
2	100 20331 000000	IL 013	3.07	OR 100	25,800	240	Rural Signalized Intersection	0.34
3	049 20346 000000	US 041	13.18	BUCKLEY RD	54,900	20,744	Urban Signalized Intersection	0.31
4	016 20350 000000	IL 050	19.3	47TH ST	58,400	15,350	Urban Signalized Intersection	0.31
5	016 20307 000000	IL 064	12	WESTERN AV	24,800	30,804	Urban Signalized Intersection	0.30
6	049 20342 000000	IL 120	3.69	HUNT CLUB RD	27,800	17,550	Urban Signalized Intersection	0.29
7	016 20351 000000	US 006	6.01	HARLEM AV	41,300	39,563	Urban Signalized Intersection	0.29
8	016 20350 000000	IL 050	21.85	MARQUETTE RD	52,400	14,307	Urban Signalized Intersection	0.29
9	016 20350 000000	IL 050	24.55	88TH ST	51,300	1,100	Urban Signalized Intersection	0.28
10	016 20350 000000	IL 050	18.57	I-55 WB TO IL-50	45,000	17,200	Urban Signalized Intersection	0.27

5.1.7 Type-A Injury Site Screening of Intersections

A Type-A PSI was calculated for all intersections. Table 5.1.7 shows the top ten intersections with the highest Type-A PSI.

Table 5.1.7. Type-A PSI for Intersections

Rank	Major Route Inventory	Major Route Name	Station	Minor Route Name	Major AADT	Minor AADT	Peer Group	Type-A PSI
1	016 20341 000000	IL 072	15.18	BUSSE RD	37,800	25,644	Urban Signalized Intersection	11.02
2	022 20344 000000	IL 083	15.39	63RD	56,000	13,800	Urban Signalized Intersection	10.27
3	016 20351 000000	US 006	11.19	KEDZIE AV	35,700	20,772	Urban Signalized Intersection	8.15
4	016 20341 000000	IL 072	5.7	MOON LAKE BLVD	26,200	1,100	Urban Signalized Intersection	7.75
5	016 20334 000000	US 012	1.33	DUNDEE RD	36,300	126,726	Urban Signalized Intersection	7.57
6	016 20344 000000	IL 083	14.05	CHICAGO-JOLIET RD	14,600	15,427	Urban Signalized Intersection	7.53
7	016 20029 000000	US 012	10.79	HALSTED ST	35,900	25,180	Urban Signalized Intersection	7.44
8	016 20350 000000	IL050	30.36	135TH ST	46,400	14,425	Urban Signalized Intersection	7.41
9	016 20350 000000	IL050	38.59	VOLLMER RD	16,500	19,240	Urban Signalized Intersection	7.36
10	016 20559 000000	IL 058	5.88	ROSELLE RD	41,700	32,833	Urban Signalized Intersection	7.28

5.1.8 Type B-Injury Site Screening of Intersections

A Type-B PSI was calculated for all intersections. Table 5.1.8 shows the top ten intersections with the highest Type-B PSI.

Table 5.1.8. Type-B PSI for Intersections

Rank	Major Route Inventory	Major Route Name	Station	Minor Route Name	Major AADT	Minor AADT	Peer Group	Type-B PSI
1	016 20348 000000	IL 043	24.55	47TH ST	38,200	1,100	Urban Signalized Intersection	25.88
2	016 20341 000000	IL 072	15.18	BUSSE RD	37,800	25,644	Urban Signalized Intersection	23.96
3	016 20348 000000	IL 043	36.9	143RD ST	42,900	275	Urban Signalized Intersection	23.50
4	016 20029 000000	US 012	13.79	STONY ISLAND AV	33,000	50,998	Urban Signalized Intersection	22.65
5	099 20607 000000	US 052	6.92	LARKIN AVE	29,900	19,869	Urban Signalized Intersection	21.73
6	016 20348 000000	IL 043	3.8	WILLOW RD	27,200	31,809	Urban Signalized Intersection	21.57
7	016 20350 000000	IL 050	13.29	CHICAGO AV	34,200	15,580	Urban Signalized Intersection	20.71
8	016 20341 000000	US 041	39.79	MONROE ST	139,100	15,900	Urban Undetermined	19.88
9	045 20365 000000	IL 056	3.25	FARNSWORTH	17,400	29,900	Urban Signalized Intersection	19.11
10	016 93730 000000	IL 001	18.11	111TH ST	31,300	13,100	Urban Signalized Intersection	17.78

5.1.9 Fatal and Injury Site Screening of Intersections

A Fatal and Injury PSI was calculated for all intersections. Table 5.1.9 shows the top ten intersections with the highest Fatal and Injury PSI.

Table 5.1.9. Fatal and Injury PSI for Intersections

Rank	Major Route Inventory	Major Route Name	Station	Minor Route Name	Major AADT	Minor AADT	Peer Group	FI PSI
1	016 20341 000000	IL 072	15.18	BUSSE RD	37,800	25,644	Urban Signalized Intersection	39.53
2	016 20348 000000	IL 043	24.55	47TH ST	38,200	1,100	Urban Signalized Intersection	36.95
3	016 20341 000000	US 041	39.79	MONROE ST	139,100	15,900	Urban Undetermined	31.25
4	016 20344 000000	IL 083	14.05	CHICAGO-JOLIET RD	14,600	15,427	Urban Signalized Intersection	30.36
5	016 20348 000000	IL 043	36.9	143RD ST	42,900	275	Urban Signalized Intersection	30.01
6	016 20350 000000	IL 050	13.29	CHICAGO AV	34,200	15,580	Urban Signalized Intersection	28.64
7	099 20607 000000	US 052	6.92	LARKIN AVE	29,900	19,869	Urban Signalized Intersection	27.18
8	016 20029 000000	US 012	13.79	STONY ISLAND AV	33,000	50,998	Urban Signalized Intersection	27.05
9	016 20348 000000	IL 043	3.8	WILLOW RD	27,200	31,809	Urban Signalized Intersection	26.00
10	016 20350 000000	IL 050	38.59	VOLLMER RD	16,500	19,240	Urban Signalized Intersection	23.89

5.1.10 Weighted Site Screening of Intersections

A Weighted PSI was calculated for all intersections. The weights were 25 for the Fatal PSI, 5 for the Type-A PSI, and 1 for the Type-B PSI. Table 5.1.10 shows the top ten intersections with the highest Weighted PSI.

Table 5.1.10. Weighted PSI for Intersections

Rank	Major Route Inventory	Major Route Name	Station	Minor Route Name	Major AADT	Minor AADT	Peer Group	Total PSI
1	016 20341 000000	IL 072	15.18	BUSSE RD	37,800	25,644	Urban Signalized Intersection	81.94
2	016 20348 000000	IL 043	24.55	47TH ST	38,200	1,100	Urban Signalized Intersection	56.05
3	022 20344 000000	IL 083	15.39	63RD	56,000	13,800	Urban Signalized Intersection	55.57
4	016 20344 000000	IL 083	14.05	CHICAGO-JOLIET RD	14,600	15,427	Urban Signalized Intersection	54.99
5	016 20341 000000	US 041	39.79	MONROE ST	139,100	15,900	Urban Undetermined	53.21
6	016 20351 000000	US 006	11.19	KEDZIE AV	35,700	20,772	Urban Signalized Intersection	50.16
7	016 20350 000000	IL 050	38.59	VOLLMER RD	16,500	19,240	Urban Signalized Intersection	48.92
8	016 20350 000000	IL 050	13.29	CHICAGO AV	34,200	15,580	Urban Signalized Intersection	47.35
9	016 20341 000000	IL 072	5.7	MOON LAKE BLVD	26,200	1,100	Urban Signalized Intersection	46.66
10	016 20334 000000	US 012	1.33	DUNDEE RD	36,300	126,726	Urban Signalized Intersection	45.69

5.2 SLIDING WINDOW APPROACH

A sliding window analysis was performed to assess the PSI of corridors in Illinois. As described in a previous section, a sliding window defines a set length as the window and continuously moves the window over the length of the roadway. The purpose is that areas of peak safety concern can be identified since all points along the roadway can be compared to its surroundings more effectively. In other words, each segment will overlap with the previous and next segment. Per the request of IDOT, rural and urban segments used different windows. The urban segments used a 0.25 mile window while rural segments used 1 mile segments. For the analysis, a window started at the beginning of every segment and the characteristics of the next segment (where the ending station of the previous segment equaled the beginning station of the next segment) were added if the total length was less still less than the predefined window length. The process was iterated until all segments met the windowed length. The following sections provide the site rankings of the sliding window for the Fatal PSI, Type-A PSI, Type-B PSI, Fatal and Injury PSI, and a Weighted PSI.

5.2.1 Fatal Site Screening of Urban Sliding Window

A Fatal PSI was calculated for all urban roadway segments utilizing the sliding window technique. Table 5.2.1 shows the top ten segments with the highest Fatal PSI using the sliding window.

Table 5.2.1. Fatal PSI for Urban Sliding Window

Rank	Inventory	Route Name	Beginning Station	Ending Station	Segment Length	Fatal PSI Per Mile
1	016 10057 000000	I 057	17.33	17.78	0.45	5.42
2	016 10057 000000	I 057	17.31	17.78	0.47	5.19
3	016 10094 000000	I 090	33.86	34.52	0.66	4.66
4	016 10094 000000	I 090	33.85	34.52	0.67	4.59
5	016 10094 000000	I 090	30.41	30.71	0.3	4.26
6	056 20336 000000	US 336	16.95	17.28	0.33	4.09
7	016 10094 000000	I 090	34.81	35.33	0.52	4.08
8	016 10094 000000	I 090	34.8	35.33	0.53	4.01
9	016 10094 000000	I 090	33.75	34.52	0.77	3.98
10	099 10080 000000	I 080	11.61	12.18	0.57	3.71

5.2.2 Type-A Injury Site Screening of Urban Sliding Window

A Type-A PSI was calculated for all urban roadway segments utilizing the sliding window technique. Table 5.2.2 shows the top ten segments with the highest Type-A PSI using the sliding window.

Table 5.2.2. Type-A PSI for Urban Sliding Window

Rank	Inventory	Route Name	Beginning Station	Ending Station	Segment Length	Type-A PSI Per Mile
1	082 10070 000000	I 070	0	0.34	0.34	39.77
2	016 10094 000000	I 090	27.75	28.06	0.31	36.15
3	016 10094 000000	I 090	31.13	31.93	0.8	30.38
4	016 10094 000000	I 090	31.12	31.93	0.81	30.00
5	016 10094 000000	I 090	31.09	31.93	0.84	28.97
6	016 10094 000000	I 090	30.72	31.02	0.3	28.47
7	016 10094 000000	I 090	31.07	31.93	0.86	28.28
8	016 10094 000000	I 090	31.05	31.93	0.88	27.69
9	016 10094 000000	I 090	30.71	31.02	0.31	27.55
10	016 10094 000000	I 090	19.65	19.94	0.29	27.45

5.2.3 Type B-Injury Site Screening of Urban Sliding Window

A Type-B PSI was calculated for all urban roadway segments utilizing the sliding window technique. Table 5.2.3 shows the top ten segments with the highest Type-B PSI using the sliding window.

Table 5.2.3. Type-B PSI for Urban Sliding Window

Rank	Inventory	Route Name	Beginning Station	Ending Station	Segment Length	Type-B PSI Per Mile
1	016 10094 000000	I 090	27.75	28.06	0.31	139.33
2	016 10094 000000	I 090	19.65	19.94	0.29	132.90
3	016 10094 000000	I 090	19.51	19.77	0.26	130.47
4	016 10094 000000	I 090	19.62	19.94	0.32	123.14
5	016 10094 000000	I 090	27.65	28.06	0.41	102.07
6	082 10070 000000	I 070	0	0.34	0.34	100.76
7	016 10055 000000	I 055	14.07	14.4	0.33	79.41
8	016 10057 000000	I 057	17.33	17.78	0.45	78.84
9	016 10057 000000	I 057	17.31	17.78	0.47	75.69
10	016 20341 000000	US 041	38.2	38.5	0.3	75.15

5.2.4 Fatal and Injury Site Screening of Urban Sliding Window

A Fatal and Injury PSI was calculated for all urban roadway segments utilizing the sliding window technique. Table 5.2.4 shows the top ten segments with the highest Fatal and Injury PSI using the sliding window.

Table 5.2.4. Fatal and Injury PSI for Urban Sliding Window

Rank	Inventory	Route Name	Beginning Station	Ending Station	Segment Length	FI PSI Per Mile
1	016 10094 000000	I 090	19.51	19.77	0.26	179.87
2	016 10094 000000	I 090	19.65	19.94	0.29	178.48
3	016 10094 000000	I 090	19.62	19.94	0.32	169.23
4	016 10094 000000	I 090	27.75	28.06	0.31	165.40
5	082 10070 000000	I 070	0.00	0.34	0.34	144.76
6	016 10094 000000	I 090	27.65	28.06	0.41	111.76
7	016 10057 000000	I 057	17.33	17.78	0.45	99.59
8	016 20341 000000	US 041	38.02	38.27	0.25	99.41
9	016 20341 000000	US 041	38.20	38.50	0.30	96.84
10	016 20341 000000	US 041	38.00	38.27	0.27	96.57

5.2.5 Weighted Site Screening of Urban Sliding Window

A Weighed PSI was calculated for all urban roadway segments utilizing the sliding window technique. Table 5.2.5 shows the top ten segments with the highest Weighted PSI using the sliding window.

Table 5.2.5. Weighted Total PSI for Urban Sliding Window

Rank	Inventory	Route Name	Beginning Station	Ending Station	Segment Length	Total PSI Per Mile
1	016 10094 000000	I 090	27.75	28.06	0.31	354.79
2	016 10057 000000	I 057	17.33	17.78	0.45	350.27
3	016 10057 000000	I 057	17.31	17.78	0.47	335.56
4	082 10070 000000	I 070	0	0.34	0.34	296.57
5	016 10094 000000	I 090	19.51	19.77	0.26	293.95
6	016 10094 000000	I 090	19.65	19.94	0.29	282.77
7	016 10094 000000	I 090	33.86	34.52	0.66	263.21
8	016 10094 000000	I 090	19.62	19.94	0.32	262.90
9	016 10094 000000	I 090	27.65	28.06	0.41	261.17
10	016 10094 000000	I 090	33.85	34.52	0.67	259.32

5.2.6 Fatal Site Screening of Rural Sliding Window

A Fatal PSI was calculated for all rural roadway segments utilizing the sliding window technique. Table 5.2.6 shows the top ten segments with the highest Fatal PSI using the sliding window.

Table 5.2.6. Fatal PSI for Rural Sliding Window

Rank	Inventory	Route Name	Beginning Station	Ending Station	Segment Length	Fatal PSI Per Mile
1	026 10070 000000	I 070	12.15	13.64	1.49	1.63
2	089 20301 000000	US 020	9.70	10.70	1.00	1.60
3	089 20301 000000	US 020	9.78	10.78	1.00	1.60
4	089 20301 000000	US 020	9.83	10.83	1.00	1.60
5	082 2060 0000000	IL 159	14.89	15.98	1.09	1.55
6	042 30747 000000	IL 003	4.23	5.29	1.06	1.55
7	056 20305 000000	US 014	6.30	7.32	1.02	1.50
8	011 20075 000000	IL 029	5.88	7.03	1.15	1.50
9	089 20301 000000	US 020	9.71	10.78	1.07	1.49
10	089 20301 000000	US 020	9.63	10.70	1.07	1.49

5.2.7 Type-A Injury Site Screening of Rural Sliding Window

A Type-A PSI was calculated for all rural roadway segments utilizing the sliding window technique. Table 5.2.7 shows the top ten segments with the highest Type-A PSI using the sliding window.

Table 5.2.7. Type-A PSI for Rural Sliding Window

Rank	Inventory	Route Name	Beginning Station	Ending Station	Segment Length	Type-A PSI Per Mile
1	056 20525 000000	US 020	7.58	8.60	1.02	7.15
2	100 20331 000000	IL 013	0.64	1.76	1.12	6.80
3	100 20331 000000	IL 013	0.53	1.76	1.23	6.48
4	100 20331 000000	IL 013	1.26	2.32	1.06	6.30
5	100 20331 000000	IL 013	0.52	1.76	1.24	6.28
6	100 20331 000000	IL 013	1.25	2.32	1.07	6.28
7	100 20331 000000	IL 013	0.50	1.76	1.26	6.23
8	058 20322 000000	IL 013	21.76	22.76	1.00	6.11
9	058 20322 000000	IL 013	21.66	22.66	1.00	6.11
10	058 20322 000000	US 051	21.55	22.66	1.11	6.06

5.2.8 Type B-Injury Site Screening of Rural Sliding Window

A Type-B PSI was calculated for all rural roadway segments utilizing the sliding window technique. Table 5.2.8 shows the top ten segments with the highest Type-B PSI using the sliding window.

Table 5.2.8. Type-B PSI for Rural Sliding Window

Rank	Inventory	Route Name	Beginning Station	Ending Station	Segment Length	Type-B PSI Per Mile
1	038 10057 000000	I 057	20.98	21.98	1.00	10.03
2	038 10057 000000	I 057	20.96	21.96	1.00	10.03
3	038 10057 000000	I 057	20.92	21.96	1.04	9.64
4	100 20331 000000	IL 013	0.23	1.25	1.02	9.41
5	100 20331 000000	IL 013	0.17	1.25	1.08	9.41
6	100 20331 000000	IL 013	0.12	1.25	1.13	9.39
7	100 20331 000000	IL 013	0.11	1.25	1.14	9.37
8	100 20331 000000	IL 013	0.02	1.25	1.23	9.12
9	100 20331 000000	IL 013	0	1.25	1.25	9.08
10	055 20310 000000	US 067	6.03	7.95	1.92	8.97

5.2.9 Fatal and Injury Site Screening of Rural Sliding Window

A Fatal and Injury PSI was calculated for all rural roadway segments utilizing the sliding window technique. Table 5.2.9 shows the top ten segments with the highest Fatal and Injury PSI using the sliding window.

Table 5.2.9. Fatal and Injury PSI for Rural Sliding Window

Rank	Inventory	Route Name	Beginning Station	Ending Station	Segment Length	FI PSI Per Mile
1	056 20525 000000	US 020	7.58	8.6	1.02	15.12
2	038 10057 000000	I 057	20.98	21.98	1	14.80
3	038 10057 000000	I 057	20.96	21.96	1	14.80
4	038 10057 000000	I 057	20.92	21.96	1.04	14.23
5	056 20525 000000	US 020	7.84	9.09	1.25	14.03
6	056 20525 000000	US 020	7.83	9.09	1.26	13.92
7	071 20553 000000	IL 072	9.85	10.91	1.06	13.83
8	058 20322 000000	US 051	21.55	22.66	1.11	13.69
9	100 20331 000000	IL 013	0.12	1.25	1.13	13.59
10	100 20331 000000	IL 013	0.02	1.25	1.23	13.58

5.2.10 Weighted Site Screening of Rural Sliding Window

A Weighted was calculated for all rural roadway segments in Illinois utilizing the sliding window technique. Table 5.2.10 shows the top ten segments with the highest Weighted PSI using the sliding window.

Table 5.2.10. Weighted PSI for Rural Sliding Window

Rank	Inventory	Route Name	Beginning Station	Ending Station	Segment Length	Total PSI Per Mile
1	042 30747 000000	IL 109	4.23	5.29	1.06	58.93
2	056 20305 000000	US 014	6.3	7.32	1.02	44.36
3	053 10055 000000	I 055	22.45	23.72	1.27	44.31
4	056 20305 000000	US 014	6.43	7.46	1.03	44.12
5	082 20600 000000	IL 159	14.89	15.98	1.09	43.65
6	038 10057 000000	I 057	20.98	21.98	1	43.31
7	038 10057 000000	I 057	20.96	21.96	1	43.31
8	056 20525 000000	US 020	7.58	8.6	1.02	41.92
9	008 20308 000000	IL 084	4.75	5.77	1.02	41.67
10	038 10057 000000	I 057	20.92	21.96	1.04	41.64

5.3 NETWORK SCREENING WITH 2002-2006 DATA

The SPF equations and network screening products are based on crash data from 2001 through 2005. This information was made available to the districts as supplemental information going into the FY-2008 Highway Safety Improvement Program (HSIP.). In early 2008, a new network screening procedure was conducted using 2002 through 2006 crash data and revised PSI weighting factors. This second screening was completed by May 2008 so that the screening result could be used to support the calendar 2008 “Five Percent Report” for the Highway Safety Improvement Program. This work is now being made available to districts to help select HSIP projects for the FY-2009 program.

This round of screening was based on the SPF results developed in Section 4 (based on 2001-2005 data), PSI computation is updated for all Illinois state marked routes (for segments, intersections, and sliding windows). During this process, the 2002-2006 crash data and unmarked roadway data are prepared into formats that are suitable for SPF development for IL state routes. The University of Illinois research teach worked with CH2M HILL to identify all crash cases that occur at selected 5% locations in Illinois. The results are summarized in the 2008 “Five Percent Report,” prepared by CH2M HILL.

In the screening, a new set of weights across different crash severity types is used. FHWA stipulates that while selecting road segments and intersections for safety improvement, priority has to be given to road elements that cause severe crashes which include fatal crashes and Type A injury crashes. In our PSI update, weighting coefficients of 25, 10, and 1 are used for a fatal crash, Type A injury crash and Type B injury crash, respectively.

Only limited information was available for local roads (county, municipal, and township roads), especially regarding the traffic and geometric conditions. Such lack of information hindered the attempt to spatially match crashes and roadway sites for further analysis. Hence, Safety Performance Functions were not developed and network screening was not conducted for Illinois local roads.

For detailed description of the new screening results, please refer to the 2008 Illinois "Five-Percent Report," submitted to FHWA, at the following URL.

http://safety.fhwa.dot.gov/hsip/fivepercent/2008/08il.htm

CHAPTER 6 MULTIVARIATE ANALYSIS OF ILLINOIS ROADWAYS

Since IDOT has set a goal of reducing traffic related fatalities and severe injuries, they need to understand not only the relationship between traffic volumes and severe crash frequency, but also the contribution of roadway characteristics to those crashes. Such additional explanatory variables, as described in detail in the following subsections, include information on roadway design features (e.g., functional class, lane width, shoulder and median type) and traffic operational features (e.g., speed limit, traffic control). These variables can be incorporated into statistical multivariate regression models to obtain more comprehensive crash predictions. Such findings can serve as a critical component for safety considerations in highway planning and design and can further be utilized in the CHSP that focuses on the 4E's (Engineering, Enforcement, Education, and Emergency Medical Services). The following sections will describe the multivariate analysis used and provide an interpretation of the results.

6.1 CALIBRATED MULTIVARIATE ANALYSIS

Similar to the development of the SPFs, the multivariate model was developed for a five-year period. This means the predictions from the regression model produces the expected number of crashes (K, A, or B crashes only) per five years. As previously stated, the total number of crashes for five years was expressed as one observation as opposed to five observations. This was due to having only one-year worth of roadway data to match with the number of crashes

For the regression models, only roadway data from the IRIS was used. Other factors, such as weather, driver information, traffic characteristics, etc. could have been included if aggregate information was available. The crash datasets provided information on driver characteristics, weather, and vehicle type. However, the aggregation of this information and application to the segments as a roadway characteristic would produce an extreme bias in the analysis. In order to include these human factors, data would need to be available on all drivers who travel along each segment as opposed to just the ones who are involved in crashes. In terms of weather and traffic conditions, data was not readily available to include these variables.

The multivariate analysis utilizes a negative binomial regression within the SAS GENMOD software. The negative binomial regression is an appropriate regression method for data that has low occurrence frequency and where the variance exceeds the mean. Multivariate models were developed to determine which variables produced a significant cause-and-effect relationship. Since crashes are a serious subject, a larger value of α was used in order to include more variables into the model. For developing the multivariate model, an α level-of-significance of 0.10 was used to retain variables in the model. The multiplicative form shown in Equation 6.1 represents the multivariate model, where $\beta_0, \beta_1, \ldots, \beta_n$ are the parameter estimates for *n-1* variables, X_{ij} represents the j^{th} segment of the i^{th} predictor variable, and $\ln(\mu_j)$ is the natural log of the expected crash frequency per five years per mile of the j^{th} segment.

$$\ln(\mu_j) = \beta_0 + \sum_{i=1}^{n} \beta_i X_{ij} \qquad (6.1)$$

Table 6.1 shows the estimates of the significant parameters, the corresponding standard errors, and *p-value* (significance).

Table 6.1. Parameter Estimation for Crash Frequency on Roadway Segments

Variable	Parameter Estimate	Standard Error	p-Value
Intercept	-0.2054	0.1777	0.2476
Average Annual Daily Traffic per 1,000 Vehicles (Log)	0.8338	0.2060	< 0.0001
Access Control (1 = Uncontrolled, 0 = Otherwise)	0.4133	0.0902	< 0.0001
Access Control (1 = Partially Controlled, 0 = Otherwise)	0.4880	0.0910	< 0.0001
Conditional Rating System	0.0428	0.0098	< 0.0001
Functional Class (1 = Interstate, 0 = Otherwise)	-0.5992	0.1111	< 0.0001
Functional Class (1 = Urban Freeway and Expressway, 0 = Otherwise)	-0.8733	0.1503	< 0.0001
Functional Class (1 = Other Principal Arterial, 0 = Otherwise)	-0.1472	0.0388	0.0001
Functional Class (1 = Rural Minor Arterial, 0 = Otherwise)	-0.0756	0.0424	0.0748
Inside Type 1 Shoulder (1 = Surface Treated, 0 = Otherwise)	0.6844	0.3422	0.0455
Inside Type 1 Shoulder (1 = Bituminous, 0 = Otherwise)	0.2060	0.0744	0.0057
Inside Type 2 Shoulder (1 = Sod, 0 = Otherwise)	-0.4884	0.2058	0.0176
Inside Type 2 Shoulder (1 = Aggregate, 0 = Otherwise)	-0.2872	0.0724	< 0.0001
Outside Type 1 Shoulder (1 = Earth, 0 = Otherwise)	-1.1984	0.6639	0.071
Outside Type 1 Shoulder (1 = Bituminous, 0 = Otherwise)	-0.2109	0.0369	< 0.0001
Outside Type 1 Shoulder (1 = "V" Gutter, 0 = Otherwise)	-0.3085	0.0899	0.0006
Outside Type 2 Shoulder (1 = Curb and Gutter, 0 = Otherwise)	0.2846	0.1015	0.0051
Outside Type 2 Shoulder (1 = Earth, 0 = Otherwise)	-0.9785	0.3100	0.0016
Outside Type 2 Shoulder (1 = Sod, 0 = Otherwise)	0.3158	0.0705	< 0.0001
Outside Type 2 Shoulder (1 = Aggregate, 0 = Otherwise)	0.2537	0.0652	< 0.0001
Outside Type 2 Shoulder (1 = Concrete-Tied, 0 = Otherwise)	1.2291	0.7424	0.0978
Outside Type 2 Shoulder (1 = "V" Gutter, 0 = Otherwise)	-0.2787	0.1389	0.0448
Inside Type 1 Shoulder Width	-0.0162	0.0097	0.0941
Outside Type 1 Shoulder Width	-0.0174	0.0041	< 0.0001
Outside Type 2 Shoulder Width	-0.0408	0.0128	0.0015
International Roughness Index	0.0020	0.0003	< 0.0001
Number of Lanes	0.0721	0.0168	< 0.0001
Lane Width	0.0161	0.0078	0.0386
Median Type (1 = M-2.12 Traversable, 0 = Otherwise)	0.2989	0.1198	0.0126
Median Type (1 = Unprotected, 0 = Otherwise)	-0.2879	0.0565	< 0.0001
Median Type (1 = Curbed, 0 = Otherwise)	0.1854	0.0452	< 0.0001
Median Type (1 = Painted, 0 = Otherwise)	0.1026	0.0433	0.018
One or Two Way (1 = One-Way, 0 = Two-Way)	0.3499	0.0827	< 0.0001
Rut Depth	-0.1259	0.0126	< 0.0001
Speed Limit	-0.0026	0.0007	0.0004
Surface Type (1 = Portland Cement Concrete, 0 = Asphalt Concrete)	-0.2371	0.0408	< 0.0001
Area Type (1 = Rural, 0 = Urban)	-0.5259	0.0364	< 0.0001
Number of Intersections	-0.0397	0.0100	< 0.0001
Average Annual Daily Traffic per 1,000 Vehicles of Minor Legs	0.0441	0.0040	< 0.0001
Dispersion per Mile	7.5692	0.0586	
Log-Likelihood	3162797.1200		

6.2 EFFECT OF PREDICTOR VARIABLES ON CRASH FREQUENCY

Even though the parameter estimates from the model can indicate how a given variable will influence the crash frequency on a given segment, extreme caution must be taken in interpreting the actual estimate by itself. The variables in the roadway dataset were correlated amongst themselves. When predictor variables are correlated, the standard error of the regression coefficients increases, meaning the estimates are not as precise. However, the standard errors of the predictions do not change, which means that the predictions are still accurate to a level of significance. With that stated, the variables included in the model are detailed below with an explanation of their influence on the prediction of crash frequency.

Average Annual Daily Traffic – AADT has a strong effect on the expected crash frequency for a given segment of roadway. The parameter estimate is positive, which is expected since the number of crashes would logically increase as the number of vehicles traveling on a segment increases. In addition, the AADT has the largest impact on crash frequency. Of all of the variables, AADT was the only one modeled as a logarithm, ln(AADT). By being modeled in this manner, the AADT accounts for the majority of the prediction while the other variables refined the prediction further. Other models were created with other variables playing a more significant role, like AADT, but the Log-Likelihood statistic was larger than the one from this model, indicating that the other models did not represent the data as well.

Access Control – The IRIS manual defines three types of access control along a segment: uncontrolled, partial control, and full control. Since access control is a categorical variable, the fully controlled segment (a segment that restricts access points) was the base condition. The regression analysis shows that both the uncontrolled and partially controlled segment conditions are significant with positive parameter estimates. This means that as a segment has more access points, the expected crash frequency increases. The positive coefficients are expected for access control. By having no control or partial control, a segment has additional access points that allow vehicles to enter and exit the segment. Access points create conflict points for vehicles to crash.

Conditional Rating System – The conditional rating system is a discrete, quantitative variable between one and nine that indicates the condition of the lanes carrying traffic. The rating system considers nine as new pavement and one as pavement in a critical condition. Based on the analysis, conditional rating system is statistically significant with a positive parameter estimate. This result indicates that newer pavements are expected to have a higher crash frequency than older pavements if all other variables are held constant. Intuition would suggest that older, deteriorating pavement would cause more crashes. However, studies show drivers are aware of poor infrastructure and adjust accordingly (Shafizadeh & Mannering, 2006), which is consistent with the findings of the study.

Functional Class – The IRIS manual defines functional class on the service provided by a highway, typically based on the traffic volumes and access control of a given route. IDOT recognizes nine different functional classes: interstate, urban freeway, other principal arterial, rural minor arterial, rural major collector, rural minor collector, rural local road, urban minor arterial, urban collector, and urban local road. For the purpose of analysis, the rural local streets were used as the base condition that the other indicator variables are referenced to. Of the functional classes, interstate, urban freeway and expressway, other principal arterial, and rural minor arterial are statistically significant with negative parameter estimates, which are expected. As the functional class of roadway gets busier, access control becomes stricter, so there are less conflict points that can cause crashes to occur.

Inside Type 1 Shoulder – The inside type 1 shoulder is the shoulder type that occurs from the edge of pavement to the change in slope on the median shoulder, as shown in the example in Figure 6.1.

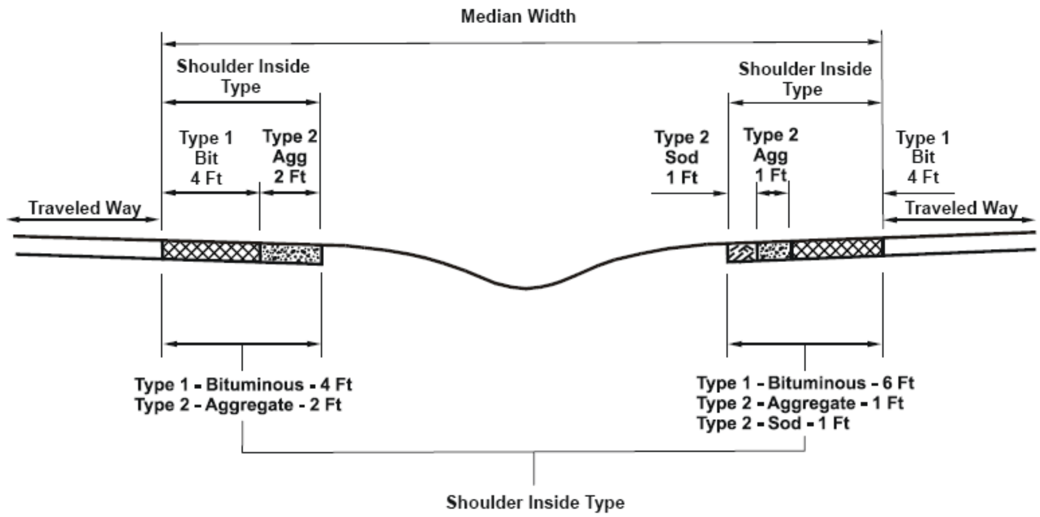


Figure 6.1. Example of Inside Shoulder Type (IDOT, 2001)

IDOT defines ten different shoulder types: no shoulder, earth, sod, aggregate, surface treated, bituminous, concrete-untied, concrete-tied, curb and gutter, and "V" gutter. For the multivariate analysis, the no shoulder variable was defined as the base case that all of the variables would be derived from.[1] Of the different shoulder types, only the surface treated and bituminous are statistically significant, with both having a positive parameter estimate. This means that if bituminous treated shoulder types are used, the expected crash frequency increases when compared to having no shoulder.

Inside Type 2 Shoulder – The inside type 2 shoulder is the shoulder that occurs from the edge of the type 1 shoulder to the end of the shoulder on the median shoulder, as shown in Figure 6.1. There are ten different shoulder types: no shoulder, earth, sod, aggregate, surface treated, bituminous, concrete-untied, concrete-tied, curb and gutter, and "V" gutter. For the multivariate analysis, the no shoulder variable was defined as the base case that all of the variables were derived from. Of the different types, the sod and aggregate variables are statistically significant with both having a negative parameter estimate. This means that if one of these shoulder types is used, the expected crash frequency decreases when compared to having no shoulder if all other variables are constant.

Outside Type 1 Shoulder – The IRIS manual defines the outside type 1 shoulder as the shoulder type that occurs from the edge of pavement to the change in slope on the non-median shoulder, as shown in the example in Figure 6.2.

lv

[1] For inside type-1 shoulder, "no shoulder" feature is really indicating those roads that do not have a median shoulder (undivided highways, or highways with a flush median). So technically this variable, and also the "Inside Type-2 Shoulder" variable, are highly correlated with the median variables. The statistical results are likely resulted from two interrelated comparisons: median vs no median, and no shoulder v.s. other shoulder types. Hence, in practice the statistical interpretation shall be used with caution.

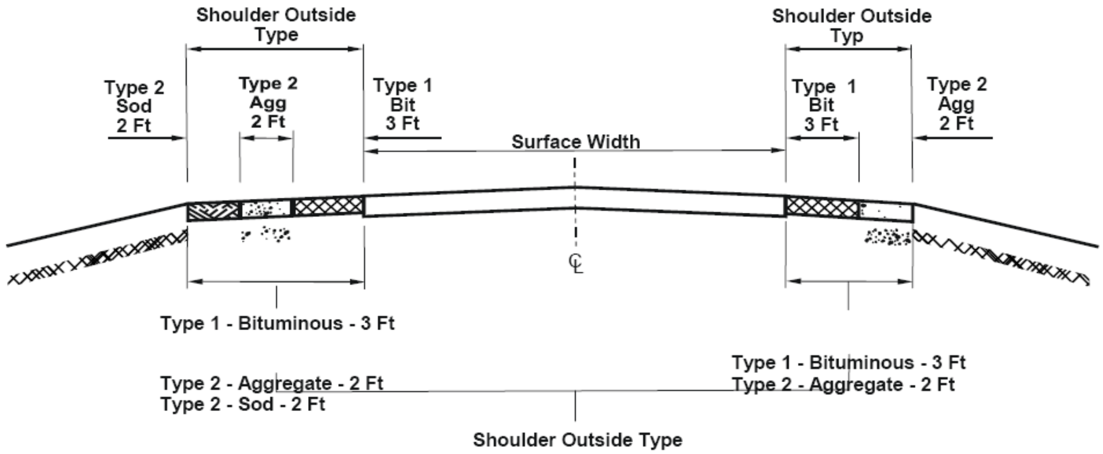


Figure 6.2. Example of outside shoulder type (IDOT, 2001).

IDOT defines ten different shoulder types: no shoulder, earth, sod, aggregate, surface treated, bituminous, concrete-untied, concrete-tied, curb and gutter, and "V" gutter. For the multivariate analysis, the no shoulder variable was the base case.[2] Of the different shoulder types, the earth, bituminous, and "V" gutter are statistically significant, with all of them having a negative parameter estimate. This means that if these shoulder types are used, the expected crash frequency decreases when compared to having no shoulder.

Outside Type 2 Shoulder – The IRIS manual defines the outside type 2 shoulder as the shoulder type that occurs from the edge of type 1 shoulder to the end of the shoulder on the non-median shoulder, as shown in the example in Figure 6.2. There are ten different shoulder types: no shoulder, earth, sod, aggregate, surface treated, bituminous, concrete-untied, concrete-tied, curb and gutter, and "V" gutter. For the multivariate analysis, the no shoulder variable was the base case. Of the different types, curb and gutter, sod, aggregate, and concrete-tied are statistically significant with a positive parameter estimates and the earth and "V" gutter are statistically significant with negative parameter estimates.

Shoulder Width – The type 1 shoulder width is the distance from the edge of pavement to where the shoulder changes slope and the type 2 shoulder width is the width from the edge of type 1 to where the shoulder ends, as shown in Figure 6.1. The variables are quantitative, continuous variables that range from 0 to 23 feet. From the multivariate analysis, the inside type 1, outside type 1, and outside type 2 shoulder widths are statistically significant with negative parameter estimates. These estimates infer that as the shoulder increases, the expected crash frequency decreases if all other variables remain constant.

International Roughness Index – International Roughness Index (IRI) is a continuous, quantitative variable that measures the roughness of a road experiences for a given mile. Agencies measure the IRI of a roadway by mounting a laser to a specialized van. As the laser moves, it records the total vertical movement over the stretch of roadway. An increase in the IRI represents a rougher road. A perfectly smooth roadway segment would have a theoretical IRI of zero. The multivariate analysis determined that the IRI is statistically significant with a positive parameter estimate. This means that as the IRI increases, the expected crash frequency increases while all other variables remain constant.

Number of Lanes – The IRIS manual defines the number of lanes as the prevailing number of through-traffic lanes in both directions during peak hour operation. The parameter

lvi

[2] Virtually all roads have some outside shoulder (unlike inside shoulder). The statistical result on outside shoulders does not heavily correlated with other variables.

estimate is statistically significant with a positive value. This result indicates that as the number of lanes increases, the segment is expected to have a higher crash frequency on average, which is an expected result. We shall note that the number of lanes is generally positively correlated with the AADT of a given segment.

Lane Widths – Lane width is the prevailing lane width for through-traffic lanes. The parameter has a positive parameter estimate that indicates that as the lane width increases, the expected crash frequency increases. This estimate has some support. Studies indicate that roadways can be narrowed in an effort to improve driver's awareness by creating less comfortable situations (Ewing, 1999). Similarly, by improving the driver's awareness, narrowing the lane widths will reduce the crash frequency. Other studies show and Accident Modification Factor of less than 1 for increasing lane widths, at least up to common values (i.e., 12 feet). (Hauer, E., 2000a; 2000b)

Median Type – Medians are the barrier separating opposing directions of traffic. They are necessary to prevent vehicles from crossing over into opposing traffic. Illinois currently uses eight median options: no median, unprotected, curbed, positive barrier, rumble strip, painted, high-tension cables, and M-2.12 traversable median. The multivariate analysis found four types to be statistically significant when compared against the base case of no median. The M-2.12 traversable, curbed, and painted medians had positive parameter estimates meaning that they would increase the expected crash frequency when compared to segments with no median. The unprotected median had a negative regression parameter indicating a reduction in crash frequency compared to segments with no median. The other median types are statistically insignificant, or there was no difference between a segment having the median and having no median. Another interpretation is that urban roads with medians have more crashes than undivided roads (mostly rural) and divided highways with depressed medians (unprotected and includes rural Interstate) have fewer crashes than other rural undivided highways.

One or Two Way – The IRIS manual defines the one or two way categorical variable as whether a segment operates in one or two direction during peak hours of operations. In other words, does the roadway segment allow one or two way traffic? The parameter estimate is positive, which indicates that a roadway that operates as a one-way facility during peak hours has a higher crash frequency than one that operates as a two-lane facility if all other variables are constant. A possible explanation of this result is that one-way roads are often located in urban settings where traffic volume is high and driving environment is complex. Such unobserved factors may contribute to more traffic crashes.

Rut Depth – The IRIS manual defines the rut depth as the average depth of wear occurring in the wheel pathway along a highway section carrying traffic in the route direction-of-inventory. Based on the multivariate analysis, rut depth is statistically significant with a negative parameter estimate. This result implies that as the rut depth increases, the expected crash frequency decreases if all other variables are held constant. Rut depth depends highly on traffic volumes, heavy vehicle traffic, functional class, etc. All of these factors are present on interstates and larger facilities. These roadway classifications typically have fewer crashes since they are limited in access control.

Speed Limit – The speed limit is a quantifiable variable that measures the legal rate for vehicles to travel. Typically, higher speeds are associated with functional classes that allow for greater traffic volumes. For example, interstates have the highest speeds and traffic volumes. From the multivariate analysis, speed is statistically significant with a negative parameter estimate, which indicates as the speed limit increases, the expected frequency of crashes decreases.

Surface Type – The type of pavement is defined as the material used on the surface of the roadway segment. For this study, the several types of surface treatments were summarized into either Portland cement concrete (PCC) or asphalt concrete (AC). For the multivariate analysis, AC was the base condition. From the model, the use of PCC surface treatment is statistically significant with a negative parameter estimate. The result implies that the use of PCC reduces the expected crash frequency along a given segment for Illinois roadways when compared to AC.

Area Type – The IRIS defines an urban area as a location with a population greater than 5,000. According to the database, there are 84 urban areas located in Illinois. Using an urban setting as a base condition, the multivariate regression model shows that the area type is a highly significant variable with a negative parameter estimate. Being negative indicates that an urban setting has a higher expected crash frequency than a rural setting, which is expected. Urban areas typically have larger facilities, higher AADT, more conflict points, and therefore more crashes.

Number of Intersections – The number of intersections is not a field in the IRIS, but it was calculated based on the Illinois Intersection file, which identifies the location of intersections along state routes. The number of intersections is the number of roads that intersect a given segment. This varies from access control because it does not take into account secondary access points such as driveways and provides a quantitative instead of a categorical variable. From the multivariate analysis, the number of intersections is significant with a negative parameter estimate. The result is not expected. The intuitive expectation is that as the number of intersection increases, the number of crashes would increase. A possible explanation is that a majority of segments do not have intersections, as defined in the roadway file. Therefore, more crashes do occur on no intersection segments, so the parameter estimate is consistent in that regard.

Average Annual Daily Traffic on Minor Roads – The AADT of the minor routes averages the AADTs of the minor legs that intersect a given segment. The parameter estimate is positive, which is expected since the number of crashes would logically increase as the number of vehicles intersecting a given segment increases. The result is consistent with Level I SPFs that use minor route AADT as an entry variable.

6.3 REMARKS

To determine the reason accidents occur and provide a more proactive model, a multivariate analysis is necessary to show how certain variables can contribute to crashes. As shown from the multivariate analysis, a large number of variables have an impact on the frequency of crashes on Illinois roadways. Some of these variables can arguably have a larger impact; however, all of them play a statistically significant role (based on a 0.10 level of significance) in crash frequencies. IDOT can use the findings of the multivariate model as part of their comprehensive highway safety plan in an effort to reduce fatalities and severe injuries resulting from traffic crashes.

Developing a multivariate analysis for Illinois roadways is limited to the information provided in the datasets. The study utilized a crash and roadway datasets. However, the crash dataset did not provide any variables due to lack of aggregate information that would have biased the results of the study. Despite not using the crash datasets variables, sufficient information existed to create models to help determine why crashes occur. Even with all of the variables included in the datasets, several factors such as lighting, weather conditions, demographic information, and additional roadway geometrics were not included in the study that could have a significant impact on crash frequency. More information and studies would need to be conducted to determine the impact of such variables.

CHAPER 7 EXCEL VBA SOFTWARE FOR SPF ESTIMATION AND PSI CALCULATION

The SPF development and network screening is needed on a continuing basis to support the HSIP program and to help calibrate the *SafetyAnalyst* software as it is released and implemented. Hence, an automated system is needed for IDOT to update SPF and PSI. This chapter briefly describes the development process of an Excel VBA (Visual Basic for Applications) software tool for SPF estimation and PSI calculation. The software includes data matching algorithms and statistical models in the Excel spreadsheet environment, and it automates the decision support process for identifying high crash sites in the Illinois roadway network.

7.1. SOFTWARE SETUP AND SYSTEM REQUIREMENTS

The computer program software was developed and tested with Microsoft Office Excel 2007 and Microsoft Visual Basic 6.5 (embedded in Office Excel 2007), on the Microsoft Windows XP Professional operating system. It is also expected to run on any Microsoft Windows operating system with Office Excel 2007 or (future) higher versions. At this point, the software may not work on any Macintosh OS version with Office 2008 because Visual Basic support is not available in Microsoft Office for Macintosh 2008. We have not tested the software on Microsoft Excel 2003 or earlier versions.

All Excel files (input file and user interface file) should be put into the same folder (anywhere in the computer). Open "User_Interface.xlsm" to run the software. Macros must be enabled in Excel. For Excel 2007, if the security level of Excel is "Medium," Excel will show a security warning when opening the file. Click the "Enable Macros" button to enable macros. If the security level is "High," user needs to change the security level to "Medium" or "Low." Open Excel 2007, select "Tools" menu, select "Macro," and then "Security." In the "Security" dialog box, select "Medium" or "Low," and click "OK."

The software uses the Excel Solver Add-in, which should be included in the Microsoft Office Package. If this add-in has not been installed, the user will receive an error message "Compile Error: Can't find project or library" when running the software. Please contact your computer administrator to have the Solver Add-in installed. If you plan to install it yourself, open Excel, select "Tools" menu, "Add-ins", and see if the "Solver Add-in" option is checked. If not, please check it and proceed to install the add-in. You will need the original Microsoft Office Package.

Sometimes, the user may still receive an error message "Compile Error: Can't find project or library" when running the software. This is a common issue with Excel VBA. If the Solver Add-in has been installed, then this message is a false alarm. To solve this problem, follow these steps:

1. Make sure Excel is open.
2. Open the Visual Basic window by pressing Alt+F11, or select "Tools" menu, "Macro", "Visual Basic Editor". The Visual Basic window may have been opened automatically when the user receives the error message.
3. In the "Visual Basic" window, select "Tools", "References". If "References" option is grey (disabled), select "Run", "Reset" first to stop the program from running.
4. In the "References" dialog box, uncheck all checkboxes marked with "MISSING: <referencename>." Press "OK" button to close the dialog box.
5. Save and run the program again.

More information about this Excel problem can be found in the Microsoft Help and Support Webpage: http://support.microsoft.com/kb/283806/en-us.

7.2. INPUT DATA

This Excel VBA software requires one input file (roadway data and matched crashes) in the Microsoft Excel spreadsheet format. Figure 7.1 illustrates the format of the input file.

	A	B	C	D	E	F	G	H	I	J	K
1	OBJECTID_1	OBJECTID	INVENTORY	BEG_STA	END_STA	AADT	AADT_YR	ACC_CNTL	ANN_BK_NBR	BLT	CH
2	1	37730	016 10055 000000	0	0.17	171700	2005	0		4	0001A
3	2	37729	016 10055 000000	0.17	0.2	171700	2005	0		4	0001A
4	3	41518	016 10055 000000	0.2	0.24	177300	2005	0		5	00000
5	4	53750	016 10055 000000	0.24	0.35	181200	2005	0		5	00000
6	5	76494	016 10055 000000	0.35	0.52	158600	2005	0		5	00000
7	6	76096	016 10055 000000	0.52	0.57	158600	2005	0		5	00000
8	7	41586	016 10055 000000	0.57	0.75	146500	2005	0		5	00000
9	8	41585	016 10055 000000	0.75	0.76	160600	2005	0		5	00000
10	9	41585	016 10055 000000	0.76	0.77	160600	2005	0		5	00000

Figure 7.1. Sample input file.

The majority of the data fields are from the IRIS database. Spatial locations shall be represented by inventory numbers and stations. The required fields of roadway site information include inventory number, beginning station, ending station, AADT, AADT year, road name, segment length, roadway functional class, county name, township/municipality name, and peer group information. Some of this information is needed for SPF development and PSI computation. Other information from the IRIS database (e.g., that on traffic and geometry) can also be included, but the Excel VBA software uses only those necessary fields.

The peer groups shall be explicitly defined in the input data, based on the following rules:

[Roadway Segments]

- Rural Two-Lane Highway (Peer Group 1)
- Rural Multilane Undivided Highway (Peer Group 2)
- Rural Multilane Divided Highway (Peer Group 3)
- Rural Freeway, 4 Lanes (Peer Group 4)
- Rural Freeway, 6+ Lanes (Peer Group 5)
- Urban Two-Lane Highway (Peer Group 6)
- Urban One-Way Arterial (Peer Group 7)
- Urban Multilane Undivided Highway (Peer Group 8)
- Urban Multilane Divided Highway (Peer Group 9)
- Urban Freeway, 4 Lanes (Peer Group 10)
- Urban Freeway, 6 Lanes (Peer Group 11)
- Urban Freeway, 8+ Lanes (Peer Group 12)

[Intersections]

- Rural Minor Leg Stop Control (Peer Group 1)
- Rural All-Way Stop Control (Peer Group 2)
- Rural Signalized Intersection (Peer Group 3)
- Rural Undetermined (Peer Group 4)
- Urban Minor Leg Stop Control (Peer Group 5)
- Urban All-Way Stop Control (Peer Group 6)
- Urban Signalized Intersection (Peer Group 7)
- Urban Undetermined (Peer Group 8)

Such definitions are consistent with those used in the previous chapters. The pseudo-codes for peer group definitions are listed in Appendix L.

The input data should also include the number of crashes (by crash severity type) occurred on each roadway site, as shown in Figure 7.2. The last three columns in the input data record the counts of crashes by severity type (K, A, and B).

DF	DG	DH	DI	DJ	DK	DL	DM	DN	DO
TWP_TEXT	DIVIDER	DIR_TRAVEL	Urban_Rura	RoadType	PeerGroup	Divided	K	A	B
Champaign_City	N	T	Urban	Two_Lane	1	No	5	1	4
Champaign_City	N	T	Urban	Two_Lane	1	No	4	4	5
Champaign_City	N	T	Rural	Two_Lane	1	No	4	3	3
Champaign_City	N	T	Rural	Two_Lane	1	No	6	9	14
Champaign_City	N	T	Rural	Two_Lane	1	No	9	9	8
Champaign_City	N	T	Rural	Two_Lane	1	No	5	8	11
Champaign_City	N	T	Rural	Two_Lane	1	No	7	6	9
Champaign_City	N	T	Rural	Two_Lane	1	No	2	4	3
Champaign_City	N	T	Rural	Two_Lane	2	No	6	7	12
Champaign_City	N	T	Rural	Two_Lane	2	No	4	10	6

Figure 7.2. Crash information in the input data.

The input data file described above will be prepared by IDOT's Data Mart system. Appendix K provide guidelines on how to develop the input data from the current IDOT IRIS and crash information databases.

7.3. USER INTERFACE, FUNCTIONALITY AND ALGORITHM

When the user opens the Microsoft Excel file named "User_Interface," the menu is similar to the one shown in Figure 7.3 will be displayed. Instructions for software implementation are given on the top. The user needs to open the input data file and type down its filename in the designated cell. In addition, the relative weighting scales of different crash severities (i.e., K, A, and B) need to be specified for weighted PSI calculation. If the user chooses to conduct sliding window analysis, window lengths for urban and rural areas should also be defined. For example, Figure 7.3 shows that the input data should come from an Excel file named "roadway.xls," and that one fatal crash (K) is equivalent to 25 type-B injury crashes, while one type-A injury crash is equivalent to 5 type-B injury crashes. By default, urban and rural window lengths are specified to be 0.25 mile and 1 mile, respectively.

The Excel VBA software has three main functionalities: SPF estimation, PSI calculation, and sliding window analysis. The three buttons for these functionalities should be pushed sequentially, and the results will be stored in separate output spreadsheets.

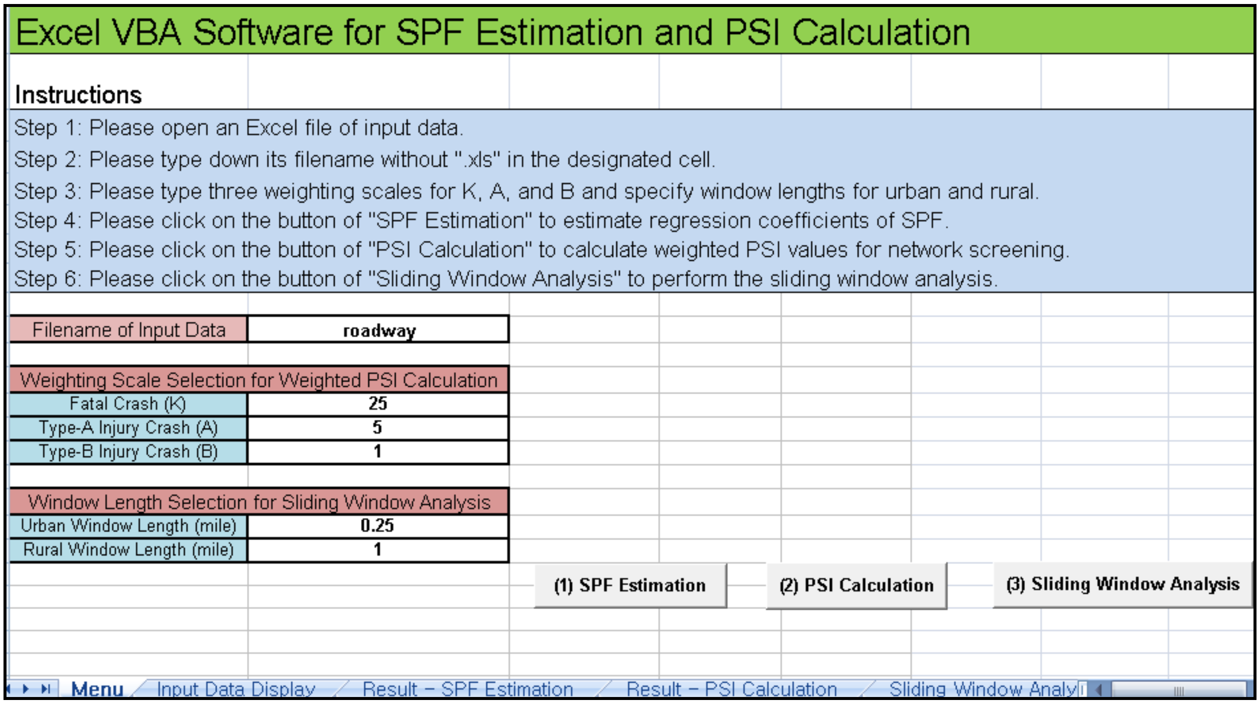


Figure 7.3. Screenshot of the user interface.

7.3.1. SPF Estimation

Once the input data is ready, Illinois-specific SPFs can be estimated by clicking on the "SPF Estimation" button. Herein, the Excel VBA software estimates the SPFs by peer group and crash severity type (K, A, and B). The SPFs for segments include three aspects to estimate the predicted number of crashes: segment length, traffic volume, and regression coefficients. The fundamental form of the segment SPF is:

$$\mu_i = (SL)_i \cdot e^a \cdot (AADT_i)^b$$

where μ_i is the predicted number of crashes for a given segment *i*, SL_i is the length in miles of segment *i*, $AADT_i$ is the Average Annual Daily Traffic of segment *i*, and *a* and *b* are regression coefficients.

For the SPF development, the negative binomial regression model is used with the following distribution function:

$$P(y_i) = \frac{\Gamma\left(y_i + \frac{1}{k}\right)}{y_i!\,\Gamma\left(\frac{1}{k}\right)} \left(\frac{k\mu_i}{1+k\mu_i}\right)^{y_i} \left(\frac{1}{1+k\mu_i}\right)^{\frac{1}{k}}$$

where *k* is the overdispersion parameter and y_i is the number of observed crashes on segment *i*, which is same as the matched crash numbers in the input data.

The software estimates the two regression coefficients (*a* and *b*) and the overdispersion parameter (*k*) using the Maximum Likelihood Estimation (MLE) method. The VBA software incorporates the *Excel Solver* to execute the MLE method. The following steps explain the process of MLE method for the SPF development.

- Step 1 - Initialization: Set the two regression coefficients and the overdispersion parameter to be zero (*a* = *b* = *k* = 0).
- Step 2 - Calculation of μ_i: With the initial coefficients, calculate the predicted number of crashes (μ_i) by road segment and crash severity.
- Step 3 - ln $P(y_i)$ Calculation: Using the μ_i from the step 2 and the initial *k* value, calculate the ln $P(y_i)$ values by road segment and crash severity.
- Step 4 - Summation of ln $P(y_i)$: Sum up all road segments' ln $P(y_i)$ values by crash severity.
- Step 5 - Maximization of Summation of ln $P(y_i)$: Through iterations (changing *a*, *b*, and *k* values), the *Excel Solver* will find the coefficient values that maximize the summation of ln $P(y_i)$ for each crash severity.

Once the summation of ln $P(y_i)$ reaches the maximum value, the values of *a*, *b*, and *k* are recorded in a separate output spreadsheet, as shown in Figure 7.4.

Reset

Peer Group - Severity K	Regression Coefficients		Dispersion Parameter Per Mile
	Intercept(a)	LogAADT(b)	
PeerGroup 1 - Rural Two-Lane Highway	4.187	0.001	0.097
PeerGroup 2 - Rural Multilane Undivided Highway	-15.479	1.658	0.522
PeerGroup 3 - Rural Multilane Divided Highway	4.596	0.000	0.940
PeerGroup 4 - Rural Freeway, 4 Lanes	-41.941	3.907	1.460
PeerGroup 5 - Rural Freeway, 6+ Lanes	2.933	0.000	0.683
PeerGroup 6 - Urban Two-Lane Highway	-14.651	1.470	0.151
PeerGroup 7 - Urban One-way Arterial	-113.734	9.889	1.904
PeerGroup 8 - Urban Multilane Undivided Highway	3.840	0.000	1.137
PeerGroup 9 - Urban Multilane Divided Highway	3.474	0.000	1.198
PeerGroup 10 - Urban Freeway, 4 Lanes	1.973	0.000	0.068
PeerGroup 11 - Urban Freeway, 6 Lanes	2.944	0.000	0.710
PeerGroup 12 - Urban Freeway, 8+ Lanes	4.308	0.029	1.243

Figure 7.4. Sample results of SPF estimation.

Since SPFs are developed for each crash severity, there are four SPF outcome tables:

- SPF of Crash Severity K
- SPF of Crash Severity A
- SPF of Crash Severity B
- SPF of Crash Severity K+A+B

7.3.2. PSI Calculation

When the user clicks on the "PSI Calculation" button, the VBA software calculates weighted PSI values based on the estimated SPFs. For roadway segments, the steps are as follows.

▪ Step 1 - Calculation of μ_i: Using the estimated regression coefficients by peer group, the predicted number of crashes (μ_i) for K, A, and B are respectively calculated by road segment.

▪ Step 2 - Calculation of Weight Factor (w_i): With the μ_i from the step 1 and the estimated k, calculate the weight factor for K, A, and B by road segment using the following equation:

$$w_i = \frac{1}{1 + (k \cdot segment\_length)(\mu_i)}$$

▪ Step 3 - Calculation of EB Approach Estimate (m_i): Using the μ_i, w_i, and observed crashes (matched crashes, y_i), calculate the estimated number of crashes (m_i) for K, A, and B by road segment using the following equation:

$$m_i = w_i \cdot \mu_i + (1 - w_i) y_i$$

▪ Step 4 - Calculation of PSI: With the calculated μ_i and w_i, calculate the PSI for K, A, and B by road segment using the following equation:

$$\text{PSI} = \frac{m_i - \mu_i}{segment\_length}$$

▪ Step 5 - Weighting PSI Values: Using the calculated PSI for K, A, and B, weight the PSI values as follows:

Weighted PSI = weight scale (K) $\cdot$ PSI (K) + weight scale (A) $\cdot$ PSI (A) + weight scale (B) $\cdot$ PSI (B)

The weighted PSI values will be listed in a separate outcome spreadsheet similar to the one shown in Figure 7.5.

ROAD_NAME	SEG_LENGTH	PeerGroup	K	A	B	Weighted PSI
STEVENSON EXPWY	0.01	4	6	6	10	8.352
STEVENSON EXPWY	0.02	3	5	5	4	5.423
STEVENSON EXPWY	0.06	2	7	7	9	5.231
STEVENSON EXPWY	0.01	12	5	1	3	5.033
STEVENSON EXPWY	0.02	3	4	5	6	3.625
STEVENSON EXPWY	0.01	7	3	3	2	3.180
STEVENSON EXPWY	0.01	12	3	4	1	2.495
STEVENSON EXPWY	0.02	7	3	2	2	2.238
STEVENSON EXPWY	0.01	2	6	7	12	2.178
STEVENSON EXPWY	0.01	3	3	4	6	1.855
STEVENSON EXPWY	0.01	3	3	3	5	1.855
STEVENSON EXPWY	0.01	7	2	2	3	1.628

Figure 7.5. Sample results of weighted PSI calculation.

Optionally, if the user clicks on the “Network Screening: Listing road segments in a descending order of weighted PSI” button in the outcome spreadsheet, the sites are sorted in descending order of weighted PSI values. The top entries in the list have high potential for safety improvement.

7.3.3. Sliding Window Analysis

Instead of computing site-specific weighted PSI, the sliding window analysis uses a window of user-specified length and moves it along the roadway. The window always starts at the beginning of a segment and the next segment is added to the window if the total length is still less than the predefined window length. The process is iterated until the total length of included segments is no shorter than the specified window length. In order for the sliding window approach to work well, the stations of adjacent roadways segments should be continuous, as shown in Figure 7.6. Once the segments in a window are determined, the weighted PSI value for this entire window is calculated in a similar way as the calculation of segment-specific PSI values (as described in the previous section).

C	D	E
INVENTORY	BEG_STA	END_STA
016 10055 000000	0	0.17
016 10055 000000	0.17	0.2
016 10055 000000	0.2	0.24
016 10055 000000	0.24	0.35
016 10055 000000	0.35	0.52
016 10055 000000	0.52	0.57
016 10055 000000	0.57	0.75
016 10055 000000	0.75	0.76
016 10055 000000	0.76	0.77
016 10055 000000	0.77	0.94
016 10055 000000	0.94	1.03
016 10055 000000	1.03	1.21
016 10055 000000	1.21	1.26
016 10055 000000	1.26	1.38
016 10055 000000	1.38	1.5

Figure 7.6. Example of continuous roadway segments.

By clicking on the "Sliding Window Analysis" button, the Excel software calculates the weighted PSI values by window and displays results in the separate spreadsheet as shown in Figure 7.7.

INVENTORY	BEG_STA	END_STA	SEG_LENGTH	K	A	B	Weighted PSI
016 10055 000000	0	1.03	1.03	57	65	81	5356.558
016 10055 000000	0.17	1.21	1.04	61	70	82	4288.276
016 10055 000000	0.2	1.21	1.01	57	66	77	4160.227
016 10055 000000	0.24	1.26	1.02	56	70	84	3780.991
016 10055 000000	0.35	1.38	1.03	53	67	76	3317.596
016 10055 000000	0.77	2.08	1.31	52	59	67	2632.471
016 10055 000000	16.16	16.45	0.29	9	10	8	2558.717
016 10055 000000	0.57	1.68	1.11	55	65	74	2526.405
016 10055 000000	1.26	2.32	1.06	37	39	46	2503.380
016 10055 000000	0.52	1.56	1.04	54	67	79	2451.000
016 10055 000000	1.21	2.24	1.03	37	42	52	2190.372
016 10055 000000	1.38	2.55	1.17	37	35	44	2106.067
016 10055 000000	16.24	16.66	0.42	13	15	11	2024.820
016 10055 000000	2.08	3.36	1.28	34	37	45	1916.240
016 10055 000000	1.56	2.58	1.02	35	35	43	1914.343
016 10055 000000	1.76	2.82	1.06	29	31	42	1863.339
016 10055 000000	2.59	3.77	1.18	19	21	22	1790.694

Figure 7.7. Sample results of sliding window analysis.

CHAPTER 8 SUMMARY AND RECOMMENDATIONS

The Illinois Department of Transportation has set a goal of reducing traffic-related fatalities and severe injuries. To accomplish this goal, IDOT must better understand the relationships between traffic volumes, other risk exposure variables, crash density, and crash severity. To meet this need, the UIUC has developed SPFs for various roadway and severity types. The SPFs will provide a foundation to screen the IDOT roadway network in order to identify sites with high-risk of fatal, A-injury, and B-injury crashes to occur. In addition to the site-screening capabilities of SPFs, a multivariate analysis was performed to determine the statistical significant of given roadway criteria on crash density. The following sections will summarize the research, the limitations of the research, and suggestions for how the project can continue in the future.

8.1 SUMMARY OF RESEARCH

The development of SPFs is critical in the site screening procedure. SPFs are statistical models that predict the expected number of crashes per year based on given roadway characteristics. The study developed SPFs for 12 segment peer groups and eight intersection peer groups for fatal, A-injury, and B-injury; and for combined fatal and injury crashes. The different peer groups and severity types allow segments to be grouped in order to create more accurate prediction models for various types of crashes. Agencies can use the SPFs to help identify objectively areas of safety concerns by comparing the predicted and estimated crash frequency. The system allows for unbiased analysis of the roadway network since personal perception, public scrutiny, and experiences do not influence the analysis.

After developing Illinois-specific SPFs, the study analyzed the Illinois state maintained roads to determine which locations had the largest safety concerns. The analysis can rank sites by fatal PSI, type-A injury PSI, type-B injury PSI, fatal and injury PSI, weighted PSI, or any other criteria. The various rankings provided IDOT with an extreme amount of flexibility in identifying locations with a safety concern and developing safety projects in the future to help mitigate more severe locations. An example of the flexibility is ranking PSIs by districts or counties so local agencies can identify the more relevant locations.

To determine the reason accidents occur and provide a more proactive model, a multivariate analysis is necessary to show how certain variables can contribute to crashes. These variables could include everything from lighting of the roadway, weather conditions, roadway geometrics, pavement conditions, and countless other variables. As shown from the multivariate analysis, 37 variables have an impact on the frequency of crashes on Illinois roadways. Some of these variables have a larger impact, but all of them are statistically significant (based on a 0.10 level of significance) in crash occurrences. IDOT can use the multivariate model for their CHSP that focuses to reduce fatalities and severe injuries resulting from traffic crashes. By knowing which roadway characteristics increase the risk of fatal or injury crashes, safety criteria can be established to help engineers design safer roadways, similar to site distance standards. Knowing that a given driver behavior leads to fatal and severe injury crashes, enforcement can be trained to identify these in an effort to reduce crashes. Educating drivers that certain weather conditions increase their risk to injuries is another use of developing a multivariate analysis. Finally, an increase in fatal and severe accidents requires the use and deployment of EMS vehicles. By recognizing conditions that cause these types of crashes, it can reduce the number of crashes and therefore deployments, and provide safer routes for EMS vehicles to travel on if they recognize some severe conditions.

The SPFs developed in this project have been used in ICT R27-18 Project, “Crash Data Analysis and Engineering Solutions for Local Agencies,” as a method to screen sites for potential for safety improvements. The R27-18 research project develops an Internet-based GIS program, similar to SafetyAnalyst, that will allow local agencies to analyze their networks to determine which sites have safety concerns. The program will have the ability to analyze the site on a crash-by-crash basis to determine the most common type of crash, and provide countermeasures to help mitigate the safety concerns. In the future, the Illinois specific SPFs will have the ability to be inputted into a larger safety management program, SafetyAnalyst. By being a part of SafetyAnalyst, IDOT, and other Illinois agencies, can use the Illinois-specific SPFs to analyze the roadways. Illinois-specific SPFs will provide a more accurate representation of the roadways as opposed to using SPFs based on other states roadways.

8.2 LIMITATIONS OF WORK

Throughout the study, challenges were encountered that often slowed progress or caused the scope of the project to be adjusted in order to be successful. Despite the challenges that were overcome, several persisted that produced limitations within the study and affected the study's extent and accuracy. The current remaining limitations are the lack of information that could have been included in the analysis, differences between the roadway and crash datasets, and the extent of roadways analyzed.

Developing statistical models to analyze the Illinois roadways is limited to information provided by the datasets. The crash dataset provides information on the driver behavior and circumstances surrounding the crash while the roadway dataset supplies the necessary information regarding roadway characteristics. Between these datasets, there is sufficient information to create models to help shed some light on why crashes occur. However, the datasets limited the extent of the study by not including variables that could have a large impact on crashes. Information from the crash dataset could not be formed into variables due to concerns of biasing the results. As a result, the models only include variables from the roadway dataset, even though it is missing several variables that could have a large impact on crash frequency. These variables could include everything from lighting of the roadway, weather conditions, demographic information, additional roadway geometrics, and countless other variables.

The method in which the datasets are maintained also produced a source of error that could have limited the accuracy of the models. Separate divisions within IDOT manage and maintain the crash database and road geometry datasets. The two datasets have different techniques for identifying segments and locations on the segments. To remedy this situation, a large amount of effort was necessary to create conversion tables between the crash and roadway datasets in order to successfully merge the two datasets and create models. Another source of error was that the crash locations were merged to a single roadway dataset. The Illinois roadway network changes every year due to the creation of additional intersections, gaining jurisdiction over segments, losing jurisdiction over segments, realignment projects, and other capital improvements. By using a single roadway file, the current roadway characteristics assigned to a crash may not accurately reflect the actual roadway at the time of crash. For example, a crash can be considered intersection related, but at the time of the crash, the intersection did not exist. This inconsistency may create errors in the analysis and provide less than precise SPFs and PSI calculations.

The last limitation of the study is the inability to analyze the complete roadway network. The scope of the analysis consisted of marked state routes. These include interstates, U.S. routes, and IL routes. The local roads are not included in the study due to lack of information regarding the roadway characteristics on the routes. Local agencies maintain local roads and often do not have updated information or information at all, in some cases. Local routes also lack crash information. The crash dataset is for state-maintained routes only, so crash reports are not available for local roads.

8.3 FUTURE DEVELOPMENTS

After completing the study and reflecting on the scope of the project, several additional aspects should be investigated to provide easy implementation and increase the accuracy of the models. Future developments of the project would include developing an automated system for easy analysis of the Illinois roadway network, providing a crash correction methodology to compensate for errors in the datasets, and to implement the SPFs in both the state's safety programs and *SafetyAnalyst*.

Because IDOT initiated the project in an effort to analyze the Illinois roadway network, software should be developed to provide easy and accurate ways to rank sites based on given criteria. An ideal program would allow a user to input a list of locations with the number of observed crashes over a given period and roadway characteristics at the locations. The program would then be able to calculate automatically the PSI for any specified road locations and severities of interest by selecting the appropriate SPF, using the EB methodology to determine the estimated number of crashes, and comparing the two values. The output of the program would be a list of locations with descending PSI values for further investigation or any other ranking scheme desired by the user.

In order to develop accurate and meaningful statistical models for safety analysis, a process to improve the accuracy of the crash locations is necessary. In an effort to adjust for this deficiency, a future development would be to determine probabilistically the most-likely location for actual crashes to occur. Inconsistencies arise due to IDOT utilizing separate divisions to develop and maintain the crash and roadway datasets, which causes problems when merging them together. In addition, crash information is recorded initially by the police, which introduces a human element to the error. Officers can range in experience on the job or experience in the area of a crash. In other words, a crash may be incorrectly assigned to a segment adjacent or near the actual segment. When regression occurs, the models do not accurately portray the situation and may declare an insignificant variable to be significant or incorrectly assign a parameter estimate with a larger/lesser coefficient. A possible method to determine the optimal crash location is by declaring that the likelihood function of observing a set of crash locations must be maximized in an effort to estimate the most-probable locations that a crash occurred. Establishing an optimal network design problem can accomplish this. Based on the optimal network, statistical models can be constructed based on the estimated actual crash locations.

REFERENCES

Carlin, B. P., and T. A. Louis, *Bayes and Empirical Bayes Methods for Data Analysis*. New York: Chapman and Hall, 1996.

CH2M HILL, Five Percent Report to Highway Safety Improvement Program (FHWA), IDOT, 2008.

Development of SPF for *SafetyAnalyst*, Interim Tools, Technical Memorandum. FHWA, 2004.

Hauer, E., D. W. Harwood, F. M. Council, and M. S. Griffith, *The Empirical Bayes Method for Estimating Safety: A Tutorial*. Transportation Research Record 1784, pp. 126-131. National Academies Press, Washington, D.C.. 2002.

Hauer, E., *Observational Before-After Studies in Road Safety*, Pergamon Press, 1997.

Hauer, E., "Shoulder Width, Shoulder Paving and Safety," 2000a. <www.roadsafetyresearch.com> Accessed December 2008

Hauer, E., The Median and Safety: Review of Literature for the Interactive Highway Safety Design Model. 2000b. <www.roadsafetyresearch.com>. Accessed December 2008.

Hauer, E., J. Kononov, B. K. Allery, and M. S. Griffith, *Screening the Road Network for Sites with Promise*. Transportation Research Record 1784, pp 27-31 National Academies Press, Washington, D.C.., 2002.

Hauer, E., *Statistical Test of the Difference Between Expected Crash Frequencies*. Transportation Research Record 1542, 24-29, 1996b.

Hauer, E., “Statistical Safety Modeling” *RoadSafetyResearch.com* 21 Jan. 2007 < http://roadsafetyresearch.com/>

Harwood, D. W. and K. M. Bauer, *Statistical Models of At-Grade Intersections Crashes – Addendum,* FHWA-RD-99-094, Federal Highway Administration, March 2000.

Hogg, R. and E. Tanis. *Probability and Statistical Inference, Sixth Edition*. New Jersey: Prentice Hall, 2001.

Jeansonne, A., “Loglinear Models” *Log Linear Models.pdf* 25 Jan. 2007 < http://online.sfsu.edu/~efc/classes/biol710/loglinear/Log%20Linear%20Models.pdf>

Kononov, J., and B. Janson, *Diagnostic Methodology for the Detection of Safety Problems at Intersections*. Transportation Research Record 1784, pp 51-56 National Academies Press, Washington D.C., 2002.

NHTSA. *Fatal Crash Reporting System 2004.* National Highway Traffic Safety Administration, U.S. Department of Transportation, Washington D.C., 2005.

Persaud, B. N., *Statistical Methods in Highway Safety Analysis*, National Cooperative Highway

Research Program Synthesis 295, Transportation Research Board, 2001.

Qin, X., and J. N. Ivan, *Hierarchical Bayesian Estimation of Safety Performance Functions for Two-Lane Highways Using Markov Chain Monte Carlo Modeling*. Journal of Transportation Engineering 5 (May 1, 2005), 345-351.

SafetyAnalyst: Software Tools for Safety Management of Specific Highway Sites, Task K, SafetyAnalyst for Module 1 – Network Screening, FHWA, 2002.

Vogt, A., and J. Bared, *Crash Models for Two-Lane Rural Roads: Segments and Intersections*, FHWA-RD-98-133.

APPENDICES

APPENDIX A SAFETY PERFORMANCE FUNCTIONS (SPFS) FOR RURAL ROADWAY SEGMENTS

(Source: Progress Report, ICT Project R27-18, "Crash Data Analysis and Engineering Solutions for Local Agencies")

Study	Variable	2-Lane/Multilane
	Dependent Variable: ln(Total crashes/(mile-year))	
Zegeer et al. [1987]	Intercept	ln(0.0015)
	ln(Average daily traffic, veh/day)	0.9711
	Lane width, ft	ln(0.8897)
	Averaged paved shoulder width, ft	ln(0.9403)
	Average unpaved shoulder width, ft	ln(0.9602)
	Median or roadside hazard rating (1 to 7)	ln(1.2)
	Dependent Variable: ln(Related crashes/(mile-year))	
	Intercept	-6.2659
	ln(Average daily traffic, veh/day)	0.8824
	Lane width, ft	ln(0.8786)
	Averaged paved shoulder width, ft	ln(0.9192)
	Average unpaved shoulder width, ft	ln(0.9316)
	Median or roadside hazard rating (1 to 7)	ln(1.2365)
	Dummy variable, 1 if terrain is flat, 0 otherwise	ln(0.8822)
	Dummy variable, 1 if terrain is mountainous, 0 otherwise	ln(1.3221)
Persaud [1992]	Intercept	-4.2934
	ln(Average daily traffic, veh/day)	0.618
	(Model for divided multilane rural roads)	
	Intercept	-9.3145
	ln(Average daily traffic, veh/day)	1.129
	(Model for undivided multilane rural roads)	
Forkenbrock et al. [1994]	Intercept	-8.5753
	ln(Average daily traffic, 1,000 veh/day)	ln(1,214)
	Right shoulder width, ft	ln(0.974)
	Pavement present serviceability rating	ln(0.972)
	Degree of sharpest curve	ln(1.068)
	Percent of steepest grade	ln(1.051)
	Dummy variable, 1 if with passing restrictions, 0 otherwise	ln(1.179)
	Dummy variable, 1 if 4 lanes, 0 otherwise	ln(0.933)
Tarko et al. [1999]	Intercept	-0.4634
	ln(Average daily traffic, veh/day)	1.156
	Median width, ft	-0.175
	Dummy variable, 2 for full access control, 1 if partial access control, 0 otherwise	-0.995
	(Model for 2-lane rural roads)	
	Intercept	-1.0354
	ln(Average daily traffic, veh/day)	1.019
	Lane width, ft	-0.453
	Pavement present serviceability index	-0.027
	Dummy variable, 1 if PCC pavements, 0 otherwise	-0.973
	(Model for multilane rural roads)	
Harwood et al., [2000]	Intercept	-8.402
	ln(Average daily traffic,)	1

APPENDIX B SPFS FOR RURAL INTERSECTIONS – POISSON DISTRIBUTION

(Source: Progress Report, ICT Project R27-18, "Crash Data Analysis and Engineering Solutions for Local Agencies")

Study	Variable	3-Leg Stop Controlled	4-Leg Stop Controlled
	Dependent Variable	ln(Vehicle crashes/year)	
Minnesota [Vogt and Bared, 1998]	Intercept	-12.5714	-10.5546
	ln(Major road average daily traffic)	0.8524	0.6517
	ln(Minor road average daily traffic)	0.4466	0.6089
	Horizontal curvature index	0.0473	0.0334
	Vertical curve grade index	0.3313	0.3805
	Major road speed limit	0.0190	0.0166
	Major road roadside hazardous rating index	0.1788	-0.0425
	Major road number of driveways	-0.0441	0.1165
	Major road right-turn channelization		
	- No free right turn	0	0
	- Provision for free right turn	0.2684	-0.0803
	Intersection angle	0.0060	-0.0044
Washington [Vogt and Bared, 1998]	Intercept	-10.4414	-10.7648
	ln(Major road average daily traffic)	0.6569	0.3710
	ln(Minor road average daily traffic)	0.5219	0.7934
	Horizontal curvature index	-0.0018	-0.4329
	Vertical curve grade index	-0.2430	-0.0064
	Major road speed limit	0.0062	0.0630
	Major road roadside hazardous rating index	0.0995	-0.2050
	Major road number of driveways	-0.0342	0.0546
	Major road right-turn channelization		
	- No free right turn	0	0
	- Provision for free right turn	0.1472	-0.7261
	Intersection angle	-0.0073	0.0309

APPENDIX C SPFS FOR RURAL INTERSECTIONS – NEGATIVE BINOMIAL DISTRIBUTION

(Source: Progress Report, ICT Project R27-18, "Crash Data Analysis and Engineering Solutions for Local Agencies")

Study	Variable	3-Leg Stop Controlled	4-Leg Stop Controlled
	Dependent Variable	ln(Vehicle crashes/year)	
IHSDM Model [Bauer and Harwood, 1999]	Intercept	-9.178	-10.025
	ln(Major road average daily traffic)	0.830	0.758
	ln(Minor road average daily traffic)	0.383	0.532
	Major road number of lanes		
	- 3 or less		0.321
	- 4 or more		0
	Major road design speed		0.009
	Major road access control		
	- None	0.225	0.200
	- Partial	0	0
	Major road functional class		
	- Principal arterial	0	0
	- Minor arterial	0.145	0.181
	- Major collector	0.211	0.173
	Terrain		
	- Level	-0.045	0.053
	- Rolling	0	0
	- Mountainous	0.095	-0.159
	Major road outside shoulder width	-0.017	
	Major road left-turn channelization	0.213	
	- No left turn	0	
	- Painted left-turn lane	0.124	
	- Curbed left turn lane		
	Major road right-turn channelization		
	- No free right turn		0.157
	- Provision for free right turn		0
	Lighting		
	- Yes		0
	- No		0.122
Iowa [Harwood et al., 2002]	Intercept	-12.153	-8.136
	ln(Major road average daily traffic)	1	0.298
	ln(Minor road average daily traffic)	0.633	0.856
	State coefficient, dummy variable	-2.232	0
Illinois [Harwood et al., 2002]	Intercept	-12.153	-8.136
	ln(Major road average daily traffic)	1	0.298
	ln(Minor road average daily traffic)	0.633	0.856
	State coefficient, dummy variable	-1.145	0

Safety Performance Functions for Rural Intersections – Negative Binomial Distribution (Con'd)
(Source: Progress Report, ICT Project R27-18, "Crash Data Analysis and Engineering Solutions for Local Agencies")

Study	Variable	3-Leg Stop Controlled	4-Leg Stop Controlled	4-Leg Signalized
	Dependent Variable	ln(Vehicle crashes/year)		
Louisana Nebrask Virginia [Harwood et al., 2002]	Intercept	-12.153	-8.136	
	ln(Major road average daily traffic)	1	0.298	
	ln(Minor road average daily traffic)	0.633	0.856	
	State coefficient, dummy variable	0	0	
North Carolina [Harwood et al., 2002]	Intercept	-12.153	-8.136	
	ln(Major road average daily traffic)	1	0.298	
	ln(Minor road average daily traffic)	0.633	0.856	
	State coefficient, dummy variable	-0.406	0	
Oregon [Harwood et al., 2002]	Intercept	-12.153	-8.136	
	ln(Major road average daily traffic)	1	0.298	
	ln(Minor road average daily traffic)	0.633	0.856	
	State coefficient, dummy variable	-1.242	0	
Minnesota [Harwood et al., 2002]	Intercept	-11.28	-9.34	
	ln(Major road average daily traffic)	0.79	0.60	
	ln(Minor road average daily traffic)	0.49	0.61	
	Major road roadside hazard rating	0.19		
	Number of driveways on major road legs		0.13	
	Intersection angle		-0.0054	
	Major road right-turn channelization			
	- No free right turn	0		
	- Provision for free right turn	0.28		
California Michigan [Vogt, 1999]	Intercept	-12.2196	-9.46311	-6.954
	ln(Major road average daily traffic)	1.1479	0.8503	0.620
	ln(Minor road average daily traffic)	0.2624	0.3294	0.395
	Major road median width	-0.0546		
	Number of driveways on major road legs	0.0391		0.041
	Major road left-turn channelization			
	- No free left turn		0	0
	- Provision for free left turn		-0.4841	-0.675
	Average of vertical grades on both roads			0.130
	Major road peak truck percentage			0.0315
	Major road peak left-turn percentage		0.1100	
	Minor road peak left-turn percentage			-0.0142

APPENDIX D DATASET FIELDS

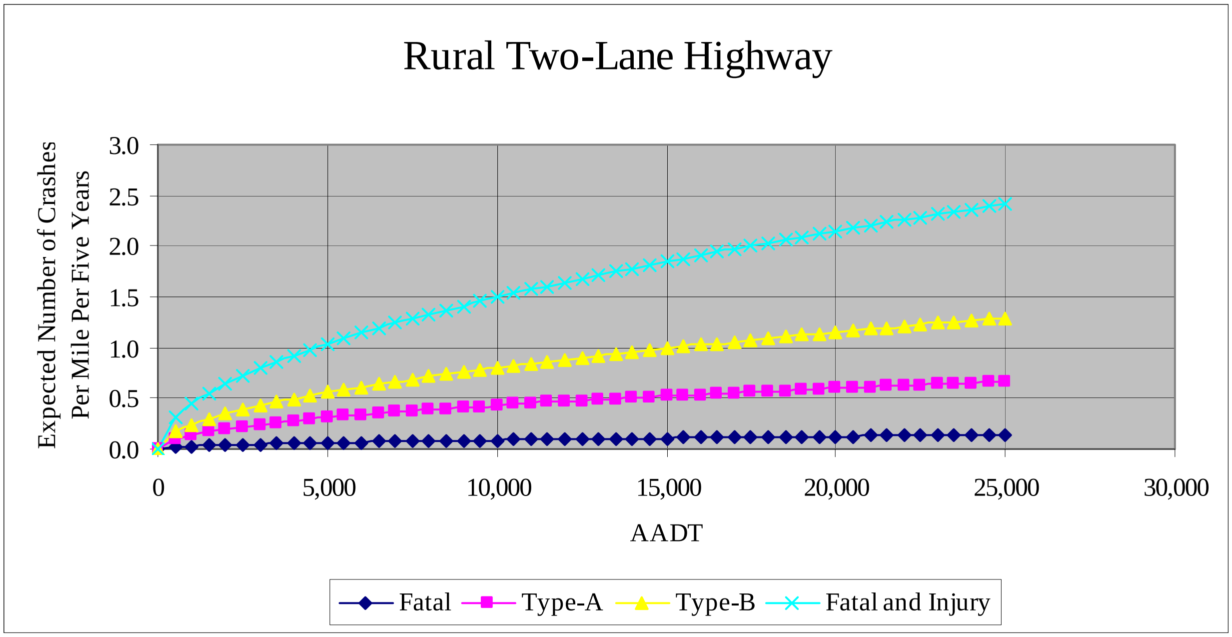
Crash Dataset Fields

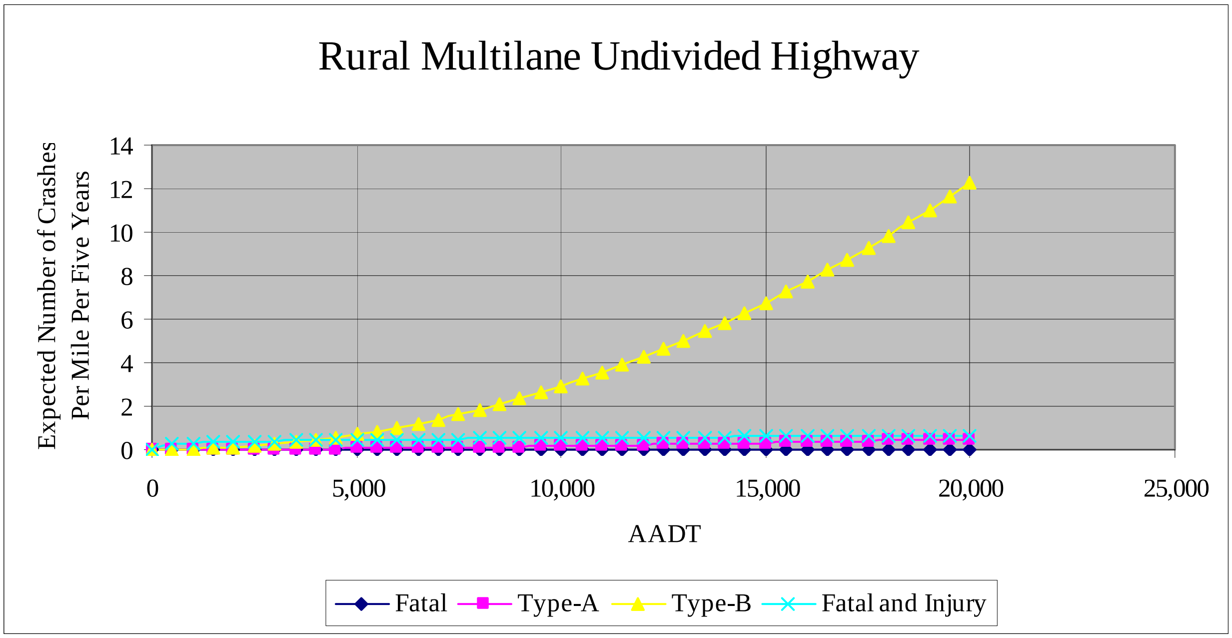
1	Field	30	CITY
2	FID	31	MODIFIED
3	Shape	32	DYNSEG_RTE
4	AREA	33	DYNSEG_MI
5	PERIMETER	34	DRIVER_1
6	TEMP2001_	35	VEH1_TYPE
7	TEMP2001_I	36	VEH1_SPECL
8	ROUTE	37	VEH1_DIR
9	CROSS_ROAD	38	VEH1_MANUV
10	CASE_ID	39	VEH1_EVNT1
11	YEAR	40	VEH1_LOC1
12	MONTH	41	VEH1_EVNT2
13	DAY	42	VEH1_LOC2
14	HOUR	43	VEH1_EVNT3
15	DAY_O_WEEK	44	VEH1_LOC3
16	NUM_VEH	45	DRIVER_2
17	INJURIES	46	VEH2_TYPE
18	FATALITIES	47	VEH2_SPECL
19	COLL_TYPE	48	VEH2_DIR
20	WEATHER	49	VEH2_MANUV
21	LIGHTING	50	VEH2_EVNT1
22	SURF_COND	51	VEH2_LOC1
23	RD_DEFECT	52	VEH2_EVNT2
24	RD_FEATURE	53	VEH2_LOC2
25	TRAF_CNTRL	54	VEH2_EVNT3
26	COUNTY	55	VEH2_LOC3
27	TOWNSHIP	56	EDIT_IND
28	TS_ROUTE	57	REC_TYPE
29	MILE		

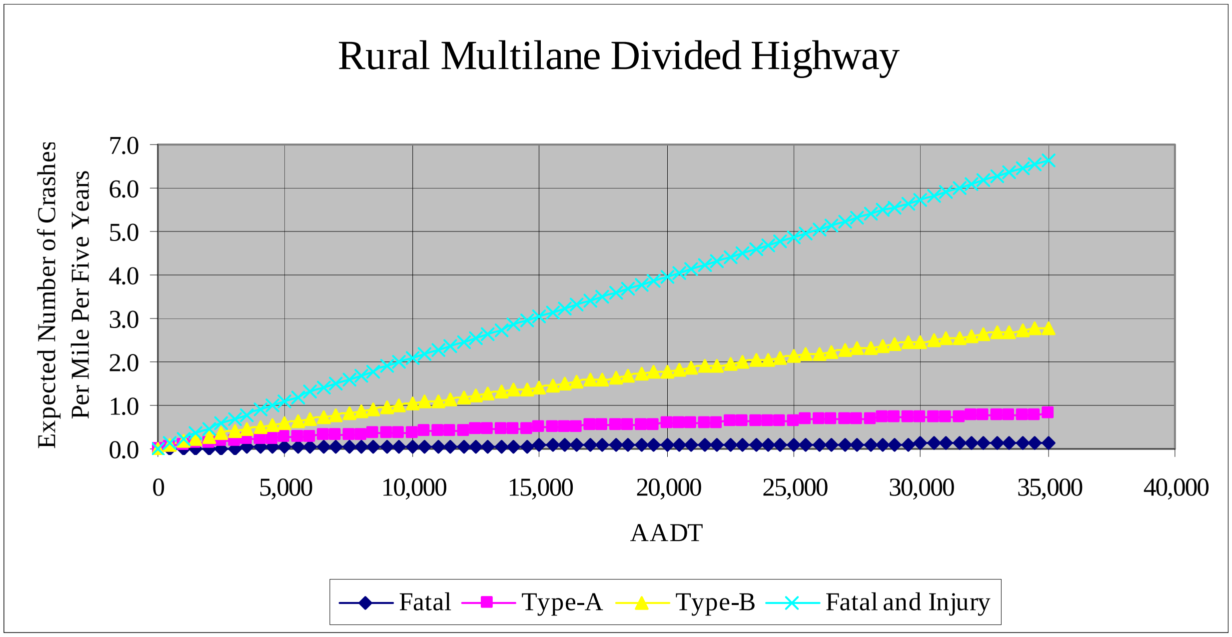
Roadway Dataset Fields

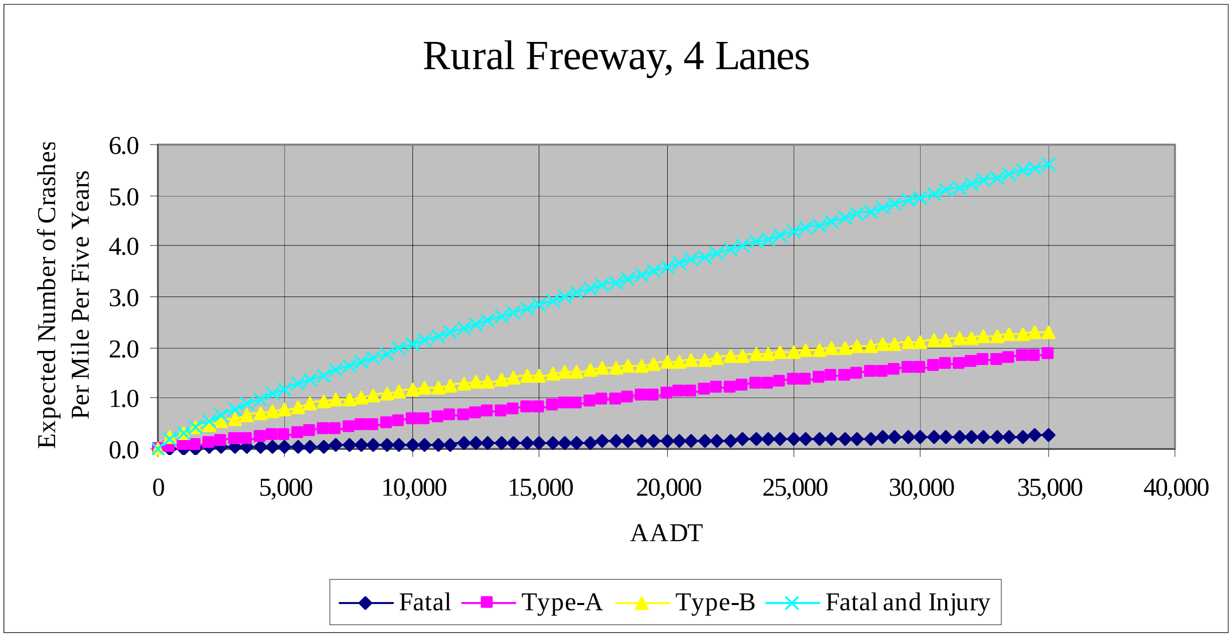
1	Field	58	LN_SPC_NBR
2	FID	59	LN_SPC_WTH
3	Shape	60	LN_WTH
4	FNODE_	61	LNS
5	TNODE_	62	MAINT_TYPE
6	LPOLY_	63	MARKED_RT
7	RPOLY_	64	MARKED_RT2
8	LENGTH	65	MARKED_RT3
9	HWY01_S_	66	MARKED_RT4
10	HWY01_S_ID	67	MED_TYP
11	INVENTORY	68	MED_WTH
12	BEG_STA	69	MNT_1
13	END_STA	70	MNT_2
14	AADT	71	MNT_DIST
15	AADT_YR	72	MNT_SECT
16	ACC_CNTL	73	MRK_RT_TY2
17	ANN_BK_NBR	74	MRK_RT_TY3
18	BLT	75	MRK_RT_TY4
19	CH	76	MRK_RT_TYP
20	CO_ADJ	77	MU_VOL
21	CONG	78	MUNI
22	CONG_ADJ	79	MUNI_ADJ
23	CRS_LOW	80	NHS
24	CRS_OPP	81	NON_ATTAIN
25	CRS_WITH	82	O_SHD1_TYP
26	CRS_YR	83	O_SHD1_WTH
27	DIST	84	O_SHD2_TYP
28	DTRESS_OPP	85	O_SHD2_WTH
29	DTRESS_WTH	86	OP_1_2_WAY
30	FAULT_LOW	87	PL_AGY
31	FAULT_OPP	88	PL_AGY_ADJ
32	FAULT_WITH	89	PRK_LT
33	FC	90	PRK_RT
34	GAP	91	REP
35	HCV	92	REP_ADJ
36	HCV_MU_YR	93	ROAD_NAME
37	HPMS_SECT	94	ROW
38	I_SHD1_TYP	95	ROW_AVL
39	I_SHD1_WTH	96	RUT_LOW
40	I_SHD2_TYP	97	RUT_OPP
41	I_SHD2_WTH	98	RUT_WITH
42	INV_CO	99	SEG_LENGTH
43	IRI_LOW	100	SP_LIM
44	IRI_OPP	101	SPEC_SYS
45	IRI_WITH	102	SU_VOL
46	JUR_1	103	SURF_TYP
47	JUR_2	104	SURF_WTH
48	JUR_TYPE	105	SURF_YR
49	K2	106	SURFACE1
50	KEY_RT_APN	107	TOLL
51	KEY_RT_APP	108	TRK_RT
52	KEY_RT_NBR	109	TWP
53	KEY_RT_SEG	110	TWP_ADJ
54	KEY_RT_SUF	111	URB_LOC
55	KEY_RT_TYP	112	URBAN
56	LABEL_1	113	labelr
57	LN_SPC		

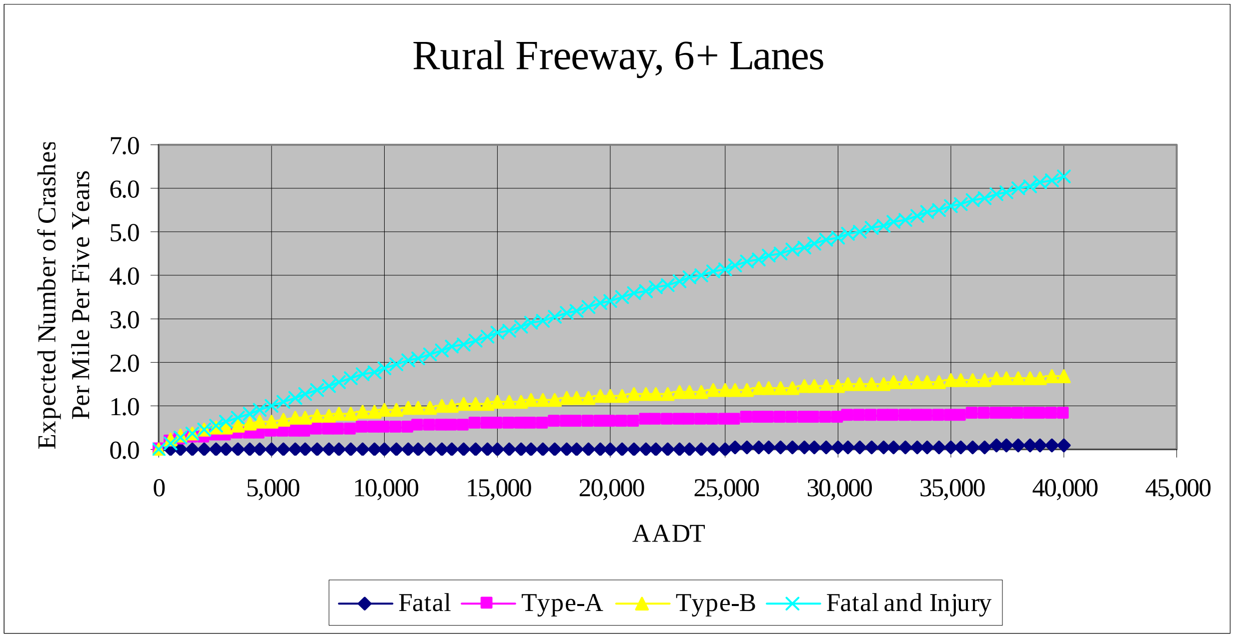
APPENDIX E GRAPHICAL REPRESENTATIONS OF SEGMENT SPFS

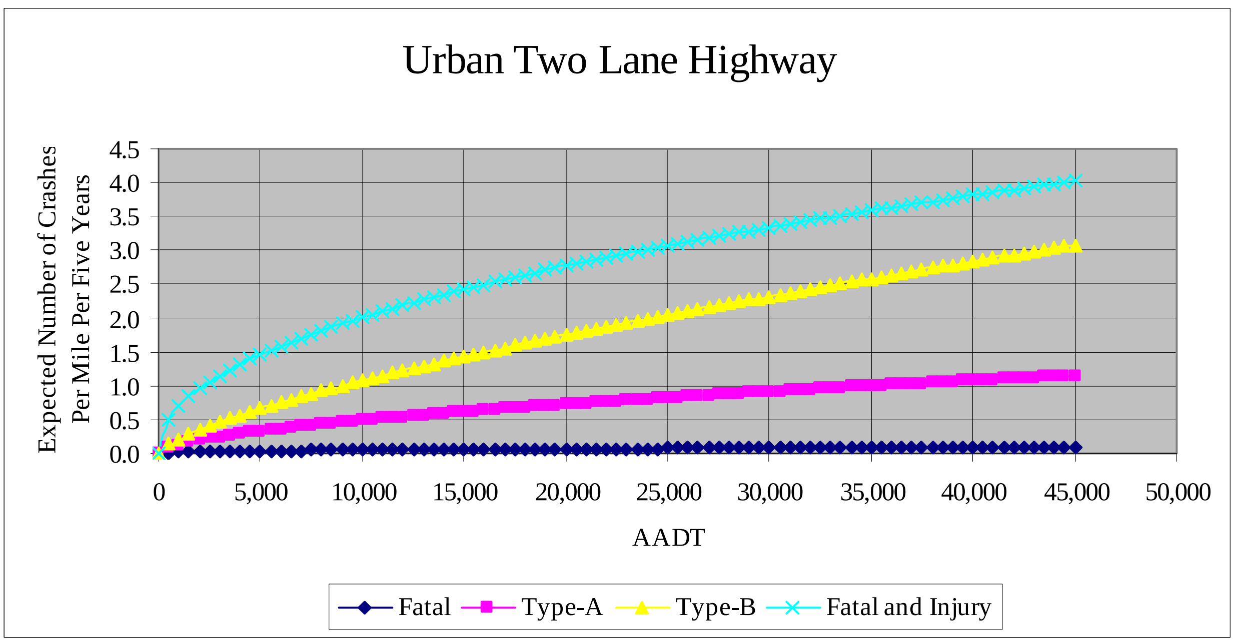


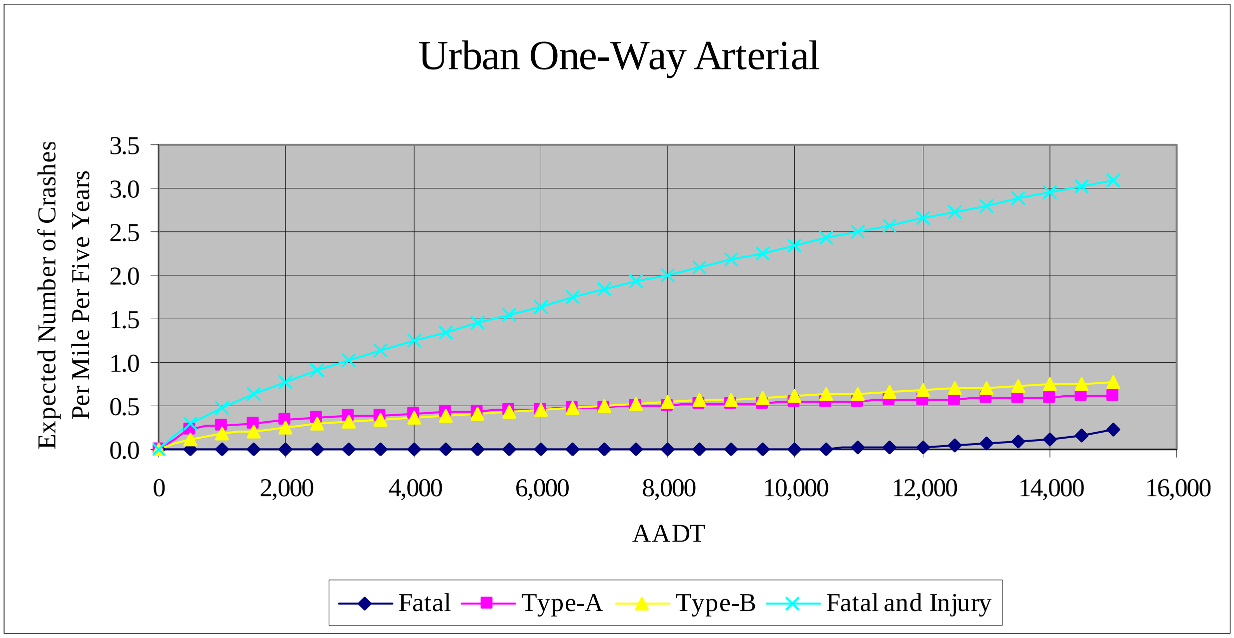


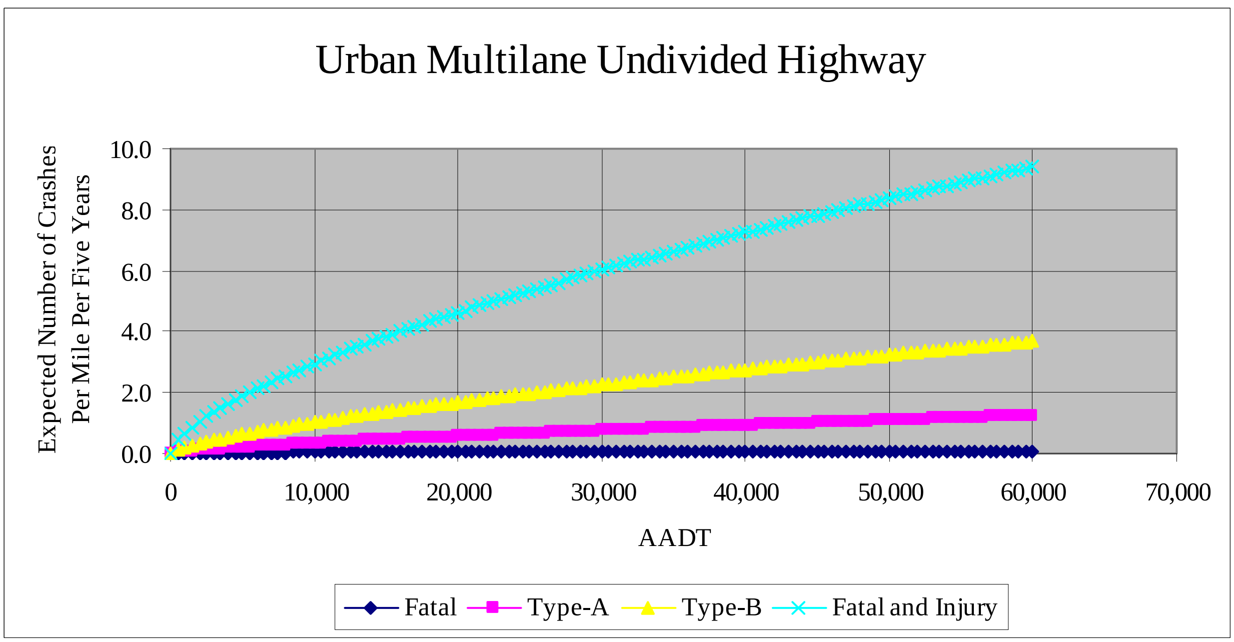

Rural Multilane Divided Highway
Expected Number of Crashes Per Mile Per Five Years
7.0
6.0
5.0
4.0
3.0
2.0
1.0
0.0
0
5,000
10,000
15,000
20,000
25,000
30,000
35,000
40,000
AADT
Fatal
Type-A
Type-B
Fatal and Injury

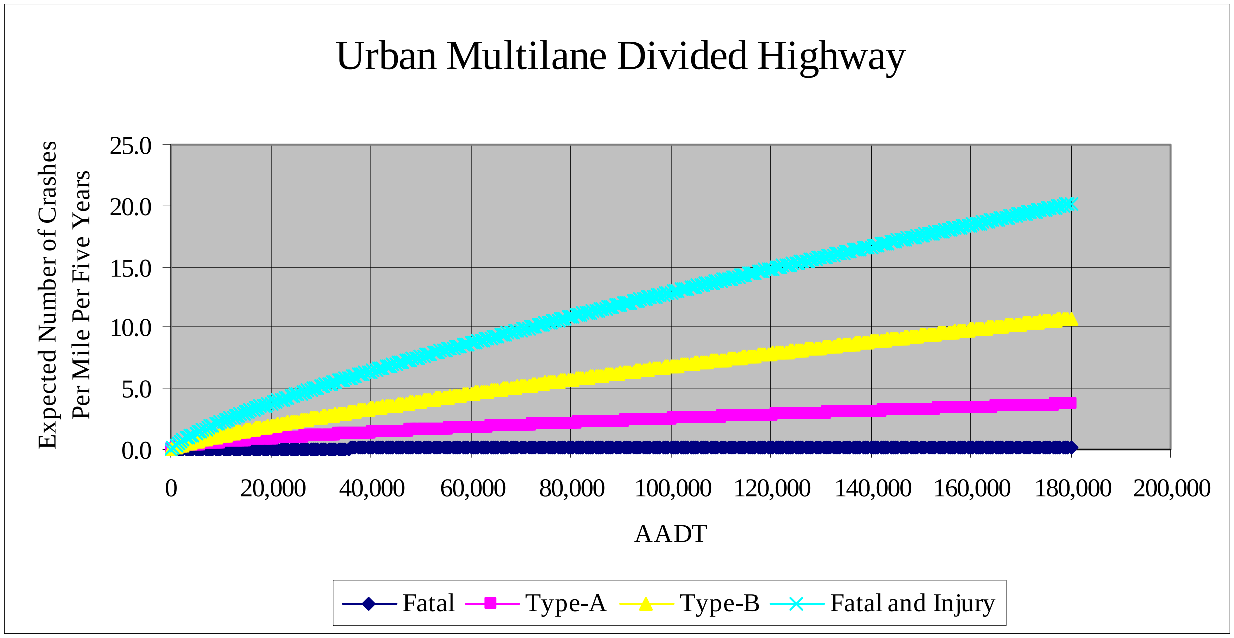

Rural Freeway, 4 Lanes
Expected Number of Crashes Per Mile Per Five Years
6.0
5.0
4.0
3.0
2.0
1.0
0.0
0
5,000
10,000
15,000
20,000
25,000
30,000
35,000
40,000
AADT
Fatal
Type-A
Type-B
Fatal and Injury

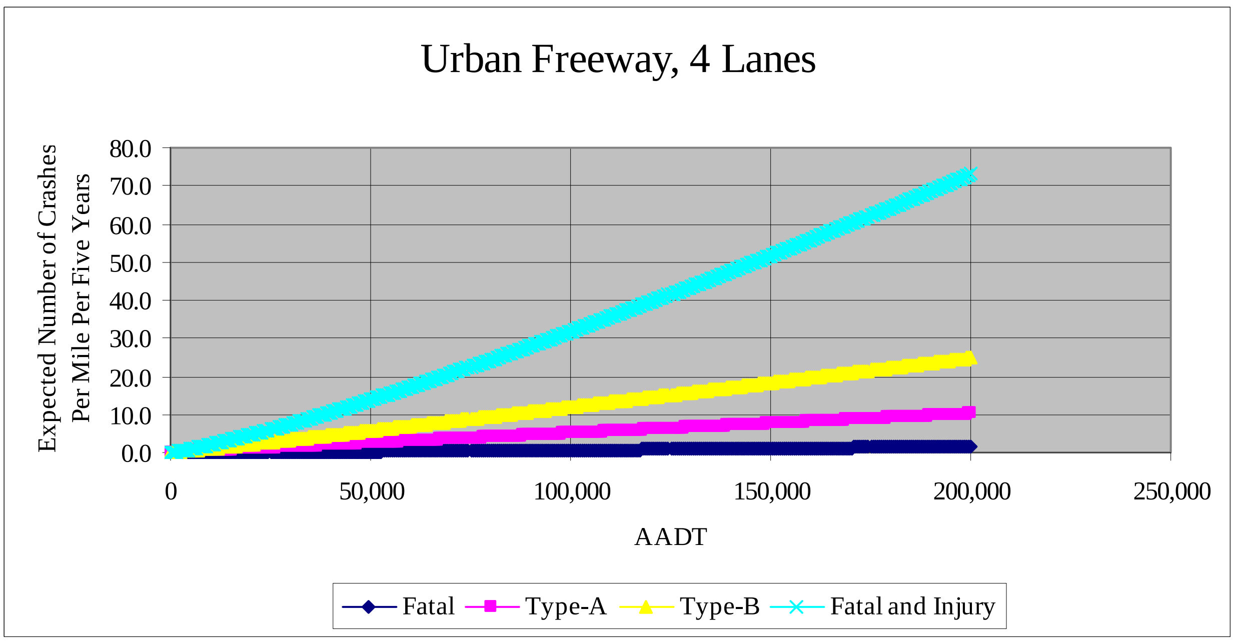

Rural Freeway, 6+ Lanes
Expected Number of Crashes Per Mile Per Five Years
7.0
6.0
5.0
4.0
3.0
2.0
1.0
0.0
0
5,000
10,000
15,000
20,000
25,000
30,000
35,000
40,000
45,000
AADT
Fatal
Type-A
Type-B
Fatal and Injury

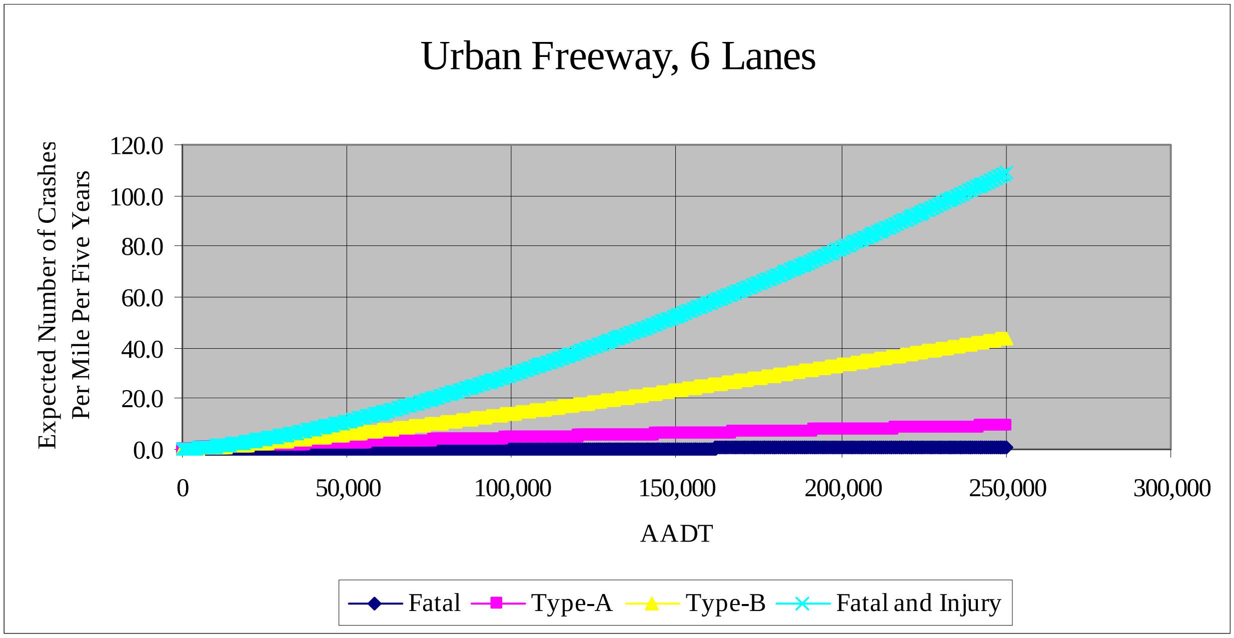

Urban Two Lane Highway
Expected Number of Crashes Per Mile Per Five Years
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0
5,000
10,000
15,000
20,000
25,000
30,000
35,000
40,000
45,000
50,000
AADT
Fatal
Type-A
Type-B
Fatal and Injury

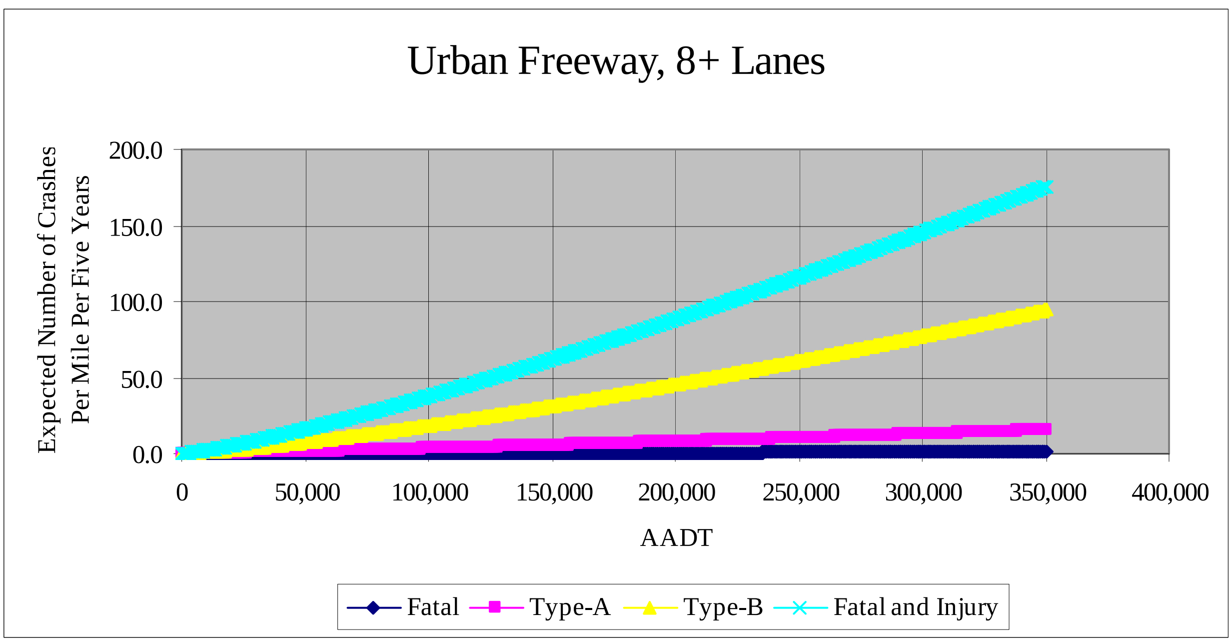

Urban One-Way Arterial
Expected Number of Crashes Per Mile Per Five Years
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0
2,000
4,000
6,000
8,000
10,000
12,000
14,000
16,000
AADT
Fatal
Type-A
Type-B
Fatal and Injury


Urban Multilane Undivided Highway
Expected Number of Crashes Per Mile Per Five Years
10.0
8.0
6.0
4.0
2.0
0.0
0
10,000
20,000
30,000
40,000
50,000
60,000
70,000
AADT
Fatal
Type-A
Type-B
Fatal and Injury


Urban Multilane Divided Highway
Expected Number of Crashes Per Mile Per Five Years
25.0
20.0
15.0
10.0
5.0
0.0
0
20,000
40,000
60,000
80,000
100,000
120,000
140,000
160,000
180,000
200,000
AADT
Fatal
Type-A
Type-B
Fatal and Injury

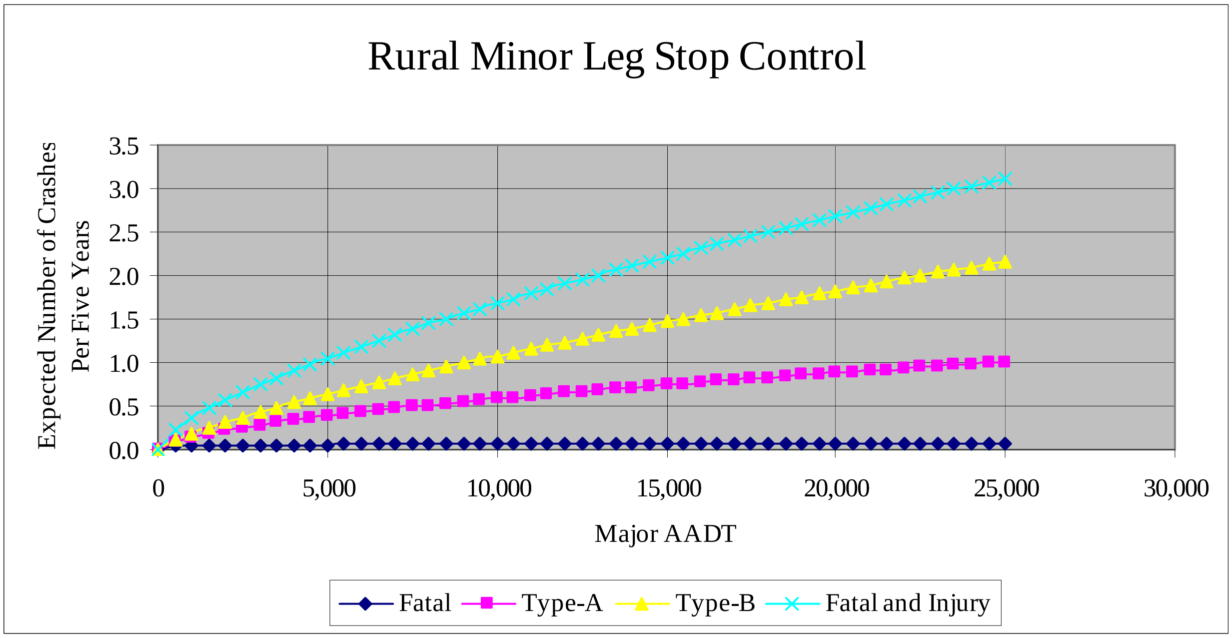

Urban Freeway, 4 Lanes
Expected Number of Crashes Per Mile Per Five Years
80.0
70.0
60.0
50.0
40.0
30.0
20.0
10.0
0.0
0
50,000
100,000
150,000
200,000
250,000
AADT
Fatal
Type-A
Type-B
Fatal and Injury

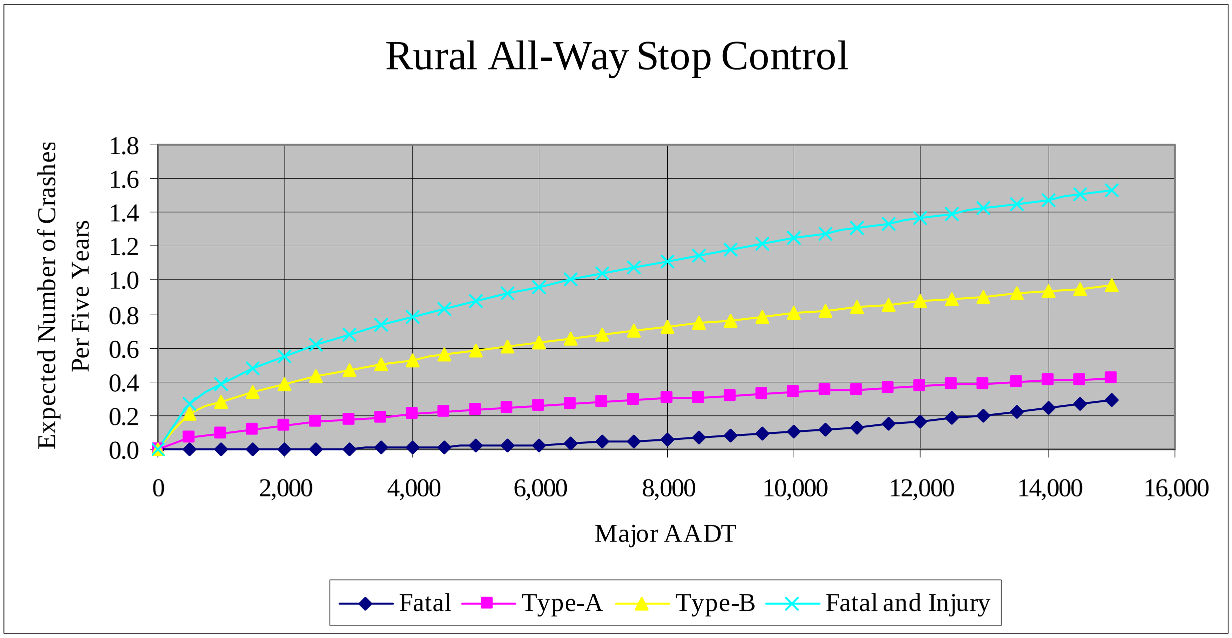

Urban Freeway, 6 Lanes
Expected Number of Crashes Per Mile Per Five Years
120.0
100.0
80.0
60.0
40.0
20.0
0.0
0
50,000
100,000
150,000
200,000
250,000
300,000
AADT
Fatal
Type-A
Type-B
Fatal and Injury

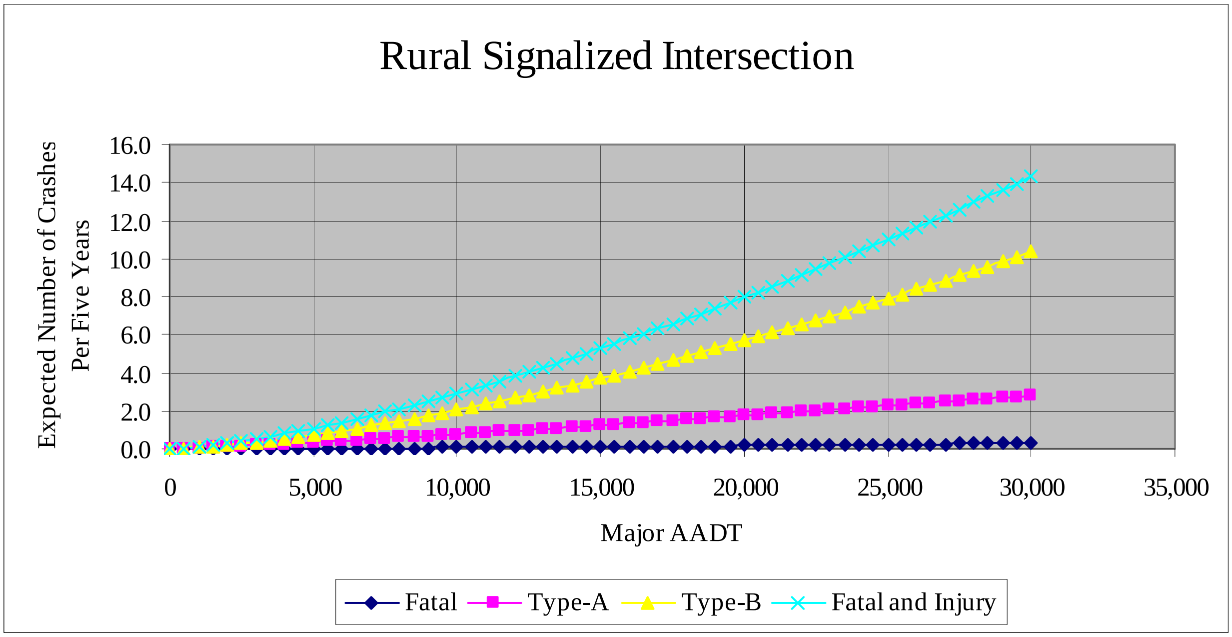

Urban Freeway, 8+ Lanes
Expected Number of Crashes Per Mile Per Five Years
200.0
150.0
100.0
50.0
0.0
0
50,000
100,000
150,000
200,000
250,000
300,000
350,000
400,000
AADT
Fatal
Type-A
Type-B
Fatal and Injury

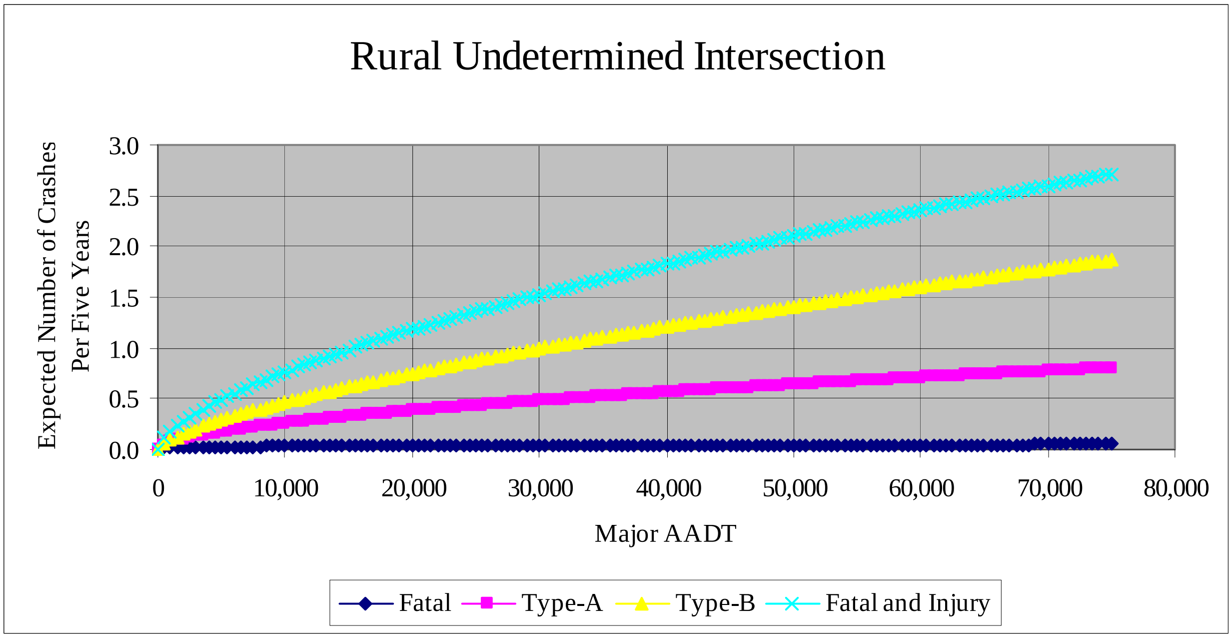
APPENDIX F GRAPHICAL REPRESENTATIONS OF INTERSECTION SPFS

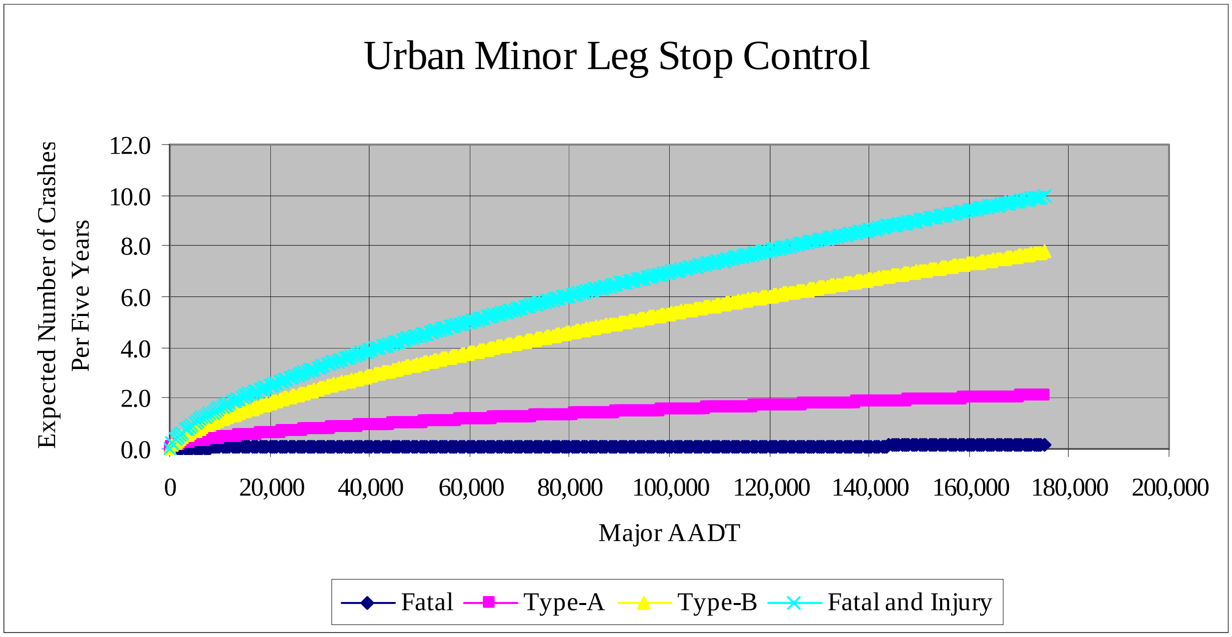
Note: The minor Route AADTs are set at 5000 vehicles per day

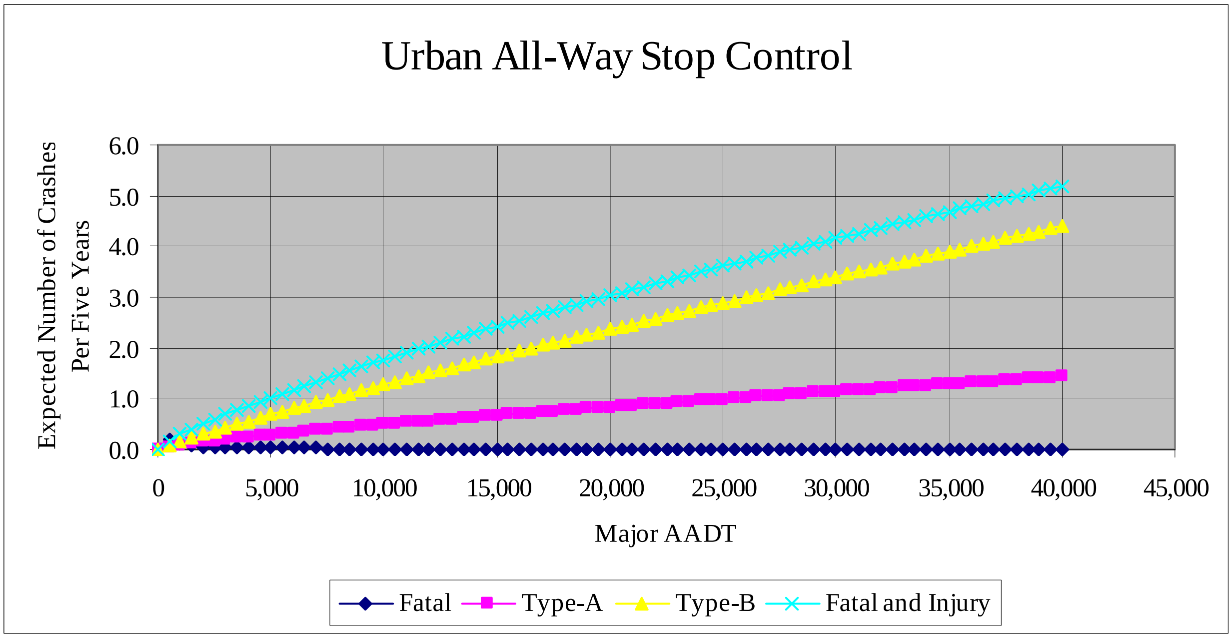


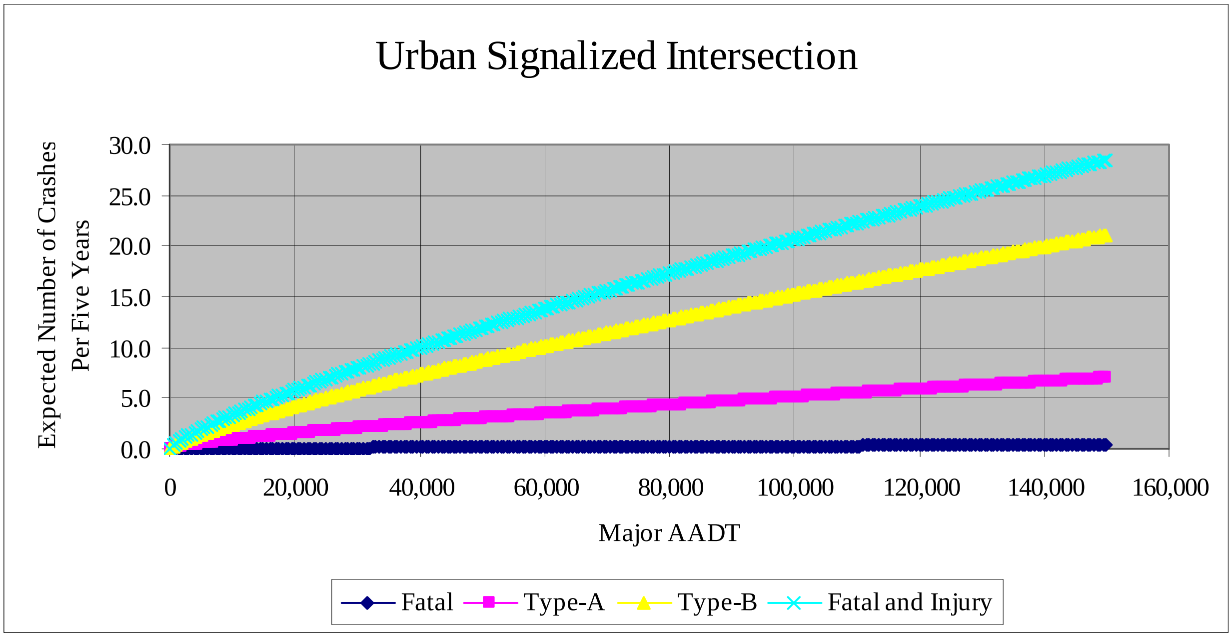


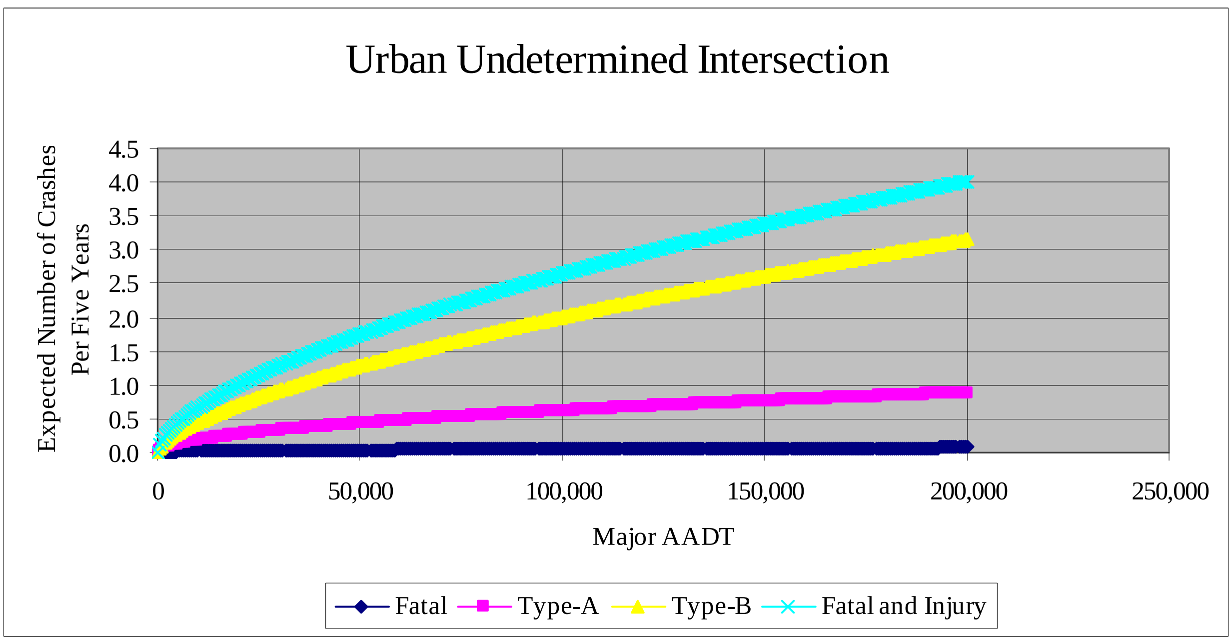

Rural Signalized Intersection
Expected Number of Crashes Per Five Years
16.0
14.0
12.0
10.0
8.0
6.0
4.0
2.0
0.0
0
5,000
10,000
15,000
20,000
25,000
30,000
35,000
Major AADT
Fatal
Type-A
Type-B
Fatal and Injury


Rural Undetermined Intersection
Expected Number of Crashes Per Five Years
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0
10,000
20,000
30,000
40,000
50,000
60,000
70,000
80,000
Major AADT
Fatal
Type-A
Type-B
Fatal and Injury


Urban Minor Leg Stop Control
Expected Number of Crashes Per Five Years
12.0
10.0
8.0
6.0
4.0
2.0
0.0
0
20,000
40,000
60,000
80,000
100,000
120,000
140,000
160,000
180,000
200,000
Major AADT
Fatal
Type-A
Type-B
Fatal and Injury


Urban All-Way Stop Control
Expected Number of Crashes Per Five Years
6.0
5.0
4.0
3.0
2.0
1.0
0.0
0
5,000
10,000
15,000
20,000
25,000
30,000
35,000
40,000
45,000
Major AADT
Fatal
Type-A
Type-B
Fatal and Injury


Urban Signalized Intersection
Expected Number of Crashes Per Five Years
30.0
25.0
20.0
15.0
10.0
5.0
0.0
0
20,000
40,000
60,000
80,000
100,000
120,000
140,000
160,000
Major AADT
Fatal
Type-A
Type-B
Fatal and Injury


Urban Undetermined Intersection
Expected Number of Crashes Per Five Years
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0
50,000
100,000
150,000
200,000
250,000
Major AADT
Fatal
Type-A
Type-B
Fatal and Injury

APPENDIX G FATALITY OUTPUT

The GENMOD Procedure

Model Information

Data Set	FATAL.RURAL_2LANE_HIGHWAY
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_K_Mile

Number of Observations Read	25834
Number of Observations Used	25834

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	26E3	1512.8573	0.0586
Scaled Deviance	26E3	1512.8573	0.0586
Pearson Chi-Square	26E3	39675.3365	1.5359
Scaled Pearson X2	26E3	39675.3365	1.5359
Log Likelihood		-1942.6972	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-7.2492	0.7888	-8.7952	-5.7032	84.46
LogAADT	1	0.5205	0.0980	0.3284	0.7127	28.19
Dispersion	1	77.8856	4.9798	68.1254	87.6459	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	<.0001
LogAADT	<.0001
Dispersion	

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

The GENMOD Procedure

Model Information

Data Set	FATAL.RURAL_MULTI_UNDIV_HIGHWAY
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_K_Mile

Number of Observations Read	200
Number of Observations Used	200

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	198	18.8174	0.0950
Scaled Deviance	198	18.8174	0.0950
Pearson Chi-Square	198	219.5294	1.1087
Scaled Pearson X2	198	219.5294	1.1087
Log Likelihood		-25.0341	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-3.9246	11.3202	-26.1119	18.2627	0.12
LogAADT	1	0.0681	1.3214	-2.5218	2.6581	0.00
Dispersion	1	25.7763	22.6208	-0.3000	70.1124	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	0.7288
LogAADT	0.9589
Dispersion	

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

The GENMOD Procedure

Model Information

Data Set	FATAL.RURAL_MULTI_DIV_HIGHWAY
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_K_Mile

Number of Observations Read	1940
Number of Observations Used	1940

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	1938	75.7951	0.0391
Scaled Deviance	1938	75.7951	0.0391
Pearson Chi-Square	1938	1192.9012	0.6155
Scaled Pearson X2	1938	1192.9012	0.6155
Log Likelihood		-132.7766	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-9.6934	5.0445	-19.5805	0.1937	3.69
LogAADT	1	0.7248	0.5650	-0.3825	1.8321	1.65
Dispersion	1	136.2171	36.2208	65.2256	207.2085	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	0.0547
LogAADT	0.1995
Dispersion	

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

The GENMOD Procedure

Model Information

Data Set	FATAL.RURAL_FREEWAY_3OR4LANES
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_K_Mile

Number of Observations Read	2996
Number of Observations Used	2996

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	2994	553.8482	0.1850
Scaled Deviance	2994	553.8482	0.1850
Pearson Chi-Square	2994	2647.0049	0.8841
Scaled Pearson X2	2994	2647.0049	0.8841
Log Likelihood		-722.8792	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-10.5749	2.8930	-16.2450	-4.9048	13.36
LogAADT	1	0.8809	0.2959	0.3009	1.4609	8.86
Dispersion	1	20.6747	2.2481	16.2685	25.0810	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	0.0003
LogAADT	0.0029
Dispersion	

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

The GENMOD Procedure

Model Information

Data Set	FATAL.RURAL_FREEWAY_5PLUSLANES
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_K_Mile

Number of Observations Read	138
Number of Observations Used	138

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	136	6.2612	0.0460
Scaled Deviance	136	6.2612	0.0460
Pearson Chi-Square	136	90.0532	0.6622
Scaled Pearson X2	136	90.0532	0.6622
Log Likelihood		-10.9080	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-33.4026	89.3307	-208.487	141.6823	0.14
LogAADT	1	2.9240	8.7171	-14.1612	20.0092	0.11
Dispersion	1	87.3216	91.7972	-0.2900	267.2407	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	0.7085
LogAADT	0.7373
Dispersion	

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

The GENMOD Procedure

Model Information

Data Set	FATAL.URBAN_2LANE_ARTERIAL
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_K_Mile

Number of Observations Read	10069
Number of Observations Used	10069

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	1E4	397.3433	0.0395
Scaled Deviance	1E4	397.3433	0.0395
Pearson Chi-Square	1E4	8184.0440	0.8130
Scaled Pearson X2	1E4	8184.0440	0.8130
Log Likelihood		-577.3243	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-6.8991	2.1725	-11.1572	-2.6410	10.08
LogAADT	1	0.4232	0.2398	-0.0468	0.8932	3.11
Dispersion	1	146.0430	16.3122	114.0718	178.0142	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	0.0015
LogAADT	0.0776
Dispersion	

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

The GENMOD Procedure

Model Information

Data Set	FATAL.URBAN_ONEWAY_ARTERIAL
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_K_Mile

Number of Observations Read	1229
Number of Observations Used	1229

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	1227	13.6006	0.0111
Scaled Deviance	1227	13.6006	0.0111
Pearson Chi-Square	1227	409.3717	0.3336
Scaled Pearson X2	1227	409.3717	0.3336
Log Likelihood		-15.5363	

WARNING: The relative Hessian convergence criterion of 0.0002527515 is greater than the limit of 0.0001. The convergence is questionable.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-82.1558	44.0095	-168.413	4.1013	3.48
LogAADT	1	8.3859	4.7053	-0.8362	17.6081	3.18
Dispersion	1	278.1296	148.4867	-0.0900	569.1582	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	0.0619
LogAADT	0.0747
Dispersion	

NOTE: The negative binomial dispersion parameter was estimated by maximum

The GENMOD Procedure

Model Information

Data Set	FATAL.URBAN_MULTI_UNDIV_HIGHWAY
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_K_Mile

Number of Observations Read	4252
Number of Observations Used	4251
Missing Values	1

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	4249	138.0914	0.0325
Scaled Deviance	4249	138.0914	0.0325
Pearson Chi-Square	4249	2486.6634	0.5852
Scaled Pearson X2	4249	2486.6634	0.5852
Log Likelihood		-195.2793	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-7.7066	5.4319	-18.3530	2.9397	2.01
LogAADT	1	0.4753	0.5548	-0.6120	1.5626	0.73
Dispersion	1	195.9627	35.4872	126.4091	265.5163	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	0.1560
LogAADT	0.3916
Dispersion	

NOTE: The negative binomial dispersion parameter was estimated by maximum

The GENMOD Procedure

Model Information

Data Set	FATAL.URBAN_MULTI_DIV_HIGHWAY
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_K_Mile

Number of Observations Read	9095
Number of Observations Used	9095

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	9093	246.4229	0.0271
Scaled Deviance	9093	246.4229	0.0271
Pearson Chi-Square	9093	6327.9710	0.6959
Scaled Pearson X2	9093	6327.9710	0.6959
Log Likelihood		9.5127	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-8.8651	2.8620	-14.4746	-3.2557	9.59
LogAADT	1	0.6060	0.2886	0.0405	1.1716	4.41
Dispersion	1	271.8734	34.6092	204.0407	339.7062	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	0.0020
LogAADT	0.0357
Dispersion	

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

The GENMOD Procedure

Model Information

Data Set	FATAL.URBAN_FREEWAY_3OR4LANES
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_K_Mile

Number of Observations Read	2183
Number of Observations Used	2183

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	2181	373.1859	0.1711
Scaled Deviance	2181	373.1859	0.1711
Pearson Chi-Square	2181	1830.5529	0.8393
Scaled Pearson X2	2181	1830.5529	0.8393
Log Likelihood		-504.9168	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-16.2561	2.3791	-20.9191	-11.5932	46.69
LogAADT	1	1.3713	0.2296	0.9212	1.8213	35.66
Dispersion	1	20.2823	2.7172	14.9566	25.6080	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	<.0001
LogAADT	<.0001
Dispersion	

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

The SAS System 13

The GENMOD Procedure

Model Information

Data Set	FATAL.URBAN_FREEWAY_5OR6LANES
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_K_Mile

Number of Observations Read	1424
Number of Observations Used	1424

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	1422	313.7880	0.2207
Scaled Deviance	1422	313.7880	0.2207
Pearson Chi-Square	1422	809.3433	0.5692
Scaled Pearson X2	1422	809.3433	0.5692
Log Likelihood		-292.1530	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-6.9265	2.7719	-12.3594	-1.4937	6.24
LogAADT	1	0.4988	0.2409	0.0267	0.9710	4.29
Dispersion	1	23.3607	2.8517	17.7715	28.9499	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	0.0125
LogAADT	0.0384
Dispersion	

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

The GENMOD Procedure

Model Information

Data Set	FATAL.URBAN_FREEWAY_7PLUSLANES
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_K_Mile

Number of Observations Read	418
Number of Observations Used	418

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	416	111.1357	0.2672
Scaled Deviance	416	111.1357	0.2672
Pearson Chi-Square	416	178.0971	0.4281
Scaled Pearson X2	416	178.0971	0.4281
Log Likelihood		-7.1522	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-15.8550	12.4638	-40.2837	8.5737	1.62
LogAADT	1	1.2473	1.0191	-0.7501	3.2448	1.50
Dispersion	1	21.9431	4.1708	13.7686	30.1176	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	0.2033
LogAADT	0.2210
Dispersion	

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

The GENMOD Procedure

Model Information

Data Set	FATAL.RURAL_MINOR_LEG_STOP
Distribution	Poisson
Link Function	Log
Dependent Variable	Num_K_Int

Number of Observations Read	14933
Number of Observations Used	14933

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	15E3	2233.3998	0.1496
Scaled Deviance	15E3	2233.3998	0.1496
Pearson Chi-Square	15E3	15385.6113	1.0305
Scaled Pearson X2	15E3	15385.6113	1.0305
Log Likelihood		-1392.6095	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-7.7382	0.7397	-9.1880	-6.2884	109.43
Log_Major_Aadt	1	0.2152	0.0884	0.0420	0.3885	5.93
Log_Minor_Aadt	1	0.3550	0.0498	0.2575	0.4525	50.91
Scale	0	1.0000	0.0000	1.0000	1.0000	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	<.0001
Log_Major_Aadt	0.0149
Log_Minor_Aadt	<.0001
Scale	

Data Set	FATAL.RURAL_ALL_WAY_STOP
Distribution	Poisson
Link Function	Log
Dependent Variable	Num_K_Int

Number of Observations Read	351
Number of Observations Used	351

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	349	43.7010	0.1252
Scaled Deviance	349	43.7010	0.1252
Pearson Chi-Square	349	294.5042	0.8439
Scaled Pearson X2	349	294.5042	0.8439
Log Likelihood		-28.8505	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-25.4636	7.7069	-40.5688	-10.3583	10.92
Log_Major_Aadt	1	2.5200	0.8682	0.8183	4.2216	8.42
Scale	0	1.0000	0.0000	1.0000	1.0000	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	0.0010
Log_Major_Aadt	0.0037
Scale	

NOTE: The scale parameter was held fixed.

The GENMOD Procedure

Model Information

Data Set	FATAL.RURAL_SIGNAL
Distribution	Poisson
Link Function	Log
Dependent Variable	Num_K_Int

Number of Observations Read	199
Number of Observations Used	199

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	197	60.9408	0.3093
Scaled Deviance	197	60.9408	0.3093
Pearson Chi-Square	197	189.5647	0.9623
Scaled Pearson X2	197	189.5647	0.9623
Log Likelihood		-40.0841	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-16.6907	5.9600	-28.3721	-5.0094	7.84
Log_Major_Aadt	1	1.5012	0.6361	0.2544	2.7479	5.57
Scale	0	1.0000	0.0000	1.0000	1.0000	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	0.0051
Log_Major_Aadt	0.0183
Scale	

NOTE: The scale parameter was held fixed.

The GENMOD Procedure

Model Information

Data Set	FATAL.RURAL_UNDETERMINED
Distribution	Poisson
Link Function	Log
Dependent Variable	Num_K_Int

Number of Observations Read	5579
Number of Observations Used	5579

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	5576	685.4096	0.1229
Scaled Deviance	5576	685.4096	0.1229
Pearson Chi-Square	5576	5555.2134	0.9963
Scaled Pearson X2	5576	5555.2134	0.9963
Log Likelihood		-423.3185	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-7.2881	1.2710	-9.7793	-4.7969	32.88
Log_Major_Aadt	1	0.2399	0.1524	-0.0589	0.5386	2.48
Log_Minor_Aadt	1	0.1866	0.0745	0.0405	0.3327	6.27
Scale	0	1.0000	0.0000	1.0000	1.0000	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	<.0001
Log_Major_Aadt	0.1156
Log_Minor_Aadt	0.0123
Scale	

The GENMOD Procedure

Model Information

Data Set	FATAL.URBAN_MINOR_LEG_STOP
Distribution	Poisson
Link Function	Log
Dependent Variable	Num_K_Int

Number of Observations Read	12121
Number of Observations Used	12121

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	12E3	2650.1439	0.2187
Scaled Deviance	12E3	2650.1439	0.2187
Pearson Chi-Square	12E3	13254.3472	1.0938
Scaled Pearson X2	12E3	13254.3472	1.0938
Log Likelihood		-1676.5270	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-9.3293	0.8303	-10.9566	-7.7020	126.26
Log_Major_Aadt	1	0.3855	0.0789	0.2309	0.5401	23.89
Log_Minor_Aadt	1	0.3054	0.0567	0.1942	0.4166	28.99
Scale	0	1.0000	0.0000	1.0000	1.0000	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	<.0001
Log_Major_Aadt	<.0001
Log_Minor_Aadt	<.0001
Scale	

The GENMOD Procedure

Model Information

Data Set	FATAL.URBAN_ALL_WAY_STOP
Distribution	Poisson
Link Function	Log
Dependent Variable	Num_K_Int

Number of Observations Read	132
Number of Observations Used	132

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	130	22.0155	0.1693
Scaled Deviance	130	22.0155	0.1693
Pearson Chi-Square	130	131.0491	1.0081
Scaled Pearson X2	130	131.0491	1.0081
Log Likelihood		-14.0077	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	3.5178	8.4224	-12.9899	20.0254	0.17
Log_Major_Aadt	1	-0.8385	0.9803	-2.7599	1.0829	0.73
Scale	0	1.0000	0.0000	1.0000	1.0000	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	0.6762
Log_Major_Aadt	0.3924
Scale	

NOTE: The scale parameter was held fixed.

The GENMOD Procedure

Model Information

Data Set	FATAL.URBAN_SIGNAL
Distribution	Poisson
Link Function	Log
Dependent Variable	Num_K_Int

Number of Observations Read	4311
Number of Observations Used	4311

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	4308	1619.7687	0.3760
Scaled Deviance	4308	1619.7687	0.3760
Pearson Chi-Square	4308	4558.2478	1.0581
Scaled Pearson X2	4308	4558.2478	1.0581
Log Likelihood		-1081.3149	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-13.3795	1.1469	-15.6273	-11.1317	136.10
Log_Major_Aadt	1	0.8901	0.1066	0.6811	1.0991	69.68
Log_Minor_Aadt	1	0.2131	0.0442	0.1265	0.2996	23.28
Scale	0	1.0000	0.0000	1.0000	1.0000	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	<.0001
Log_Major_Aadt	<.0001
Log_Minor_Aadt	<.0001
Scale	

The GENMOD Procedure

Model Information

Data Set	FATAL.URBAN_UNDETERMINED
Distribution	Poisson
Link Function	Log
Dependent Variable	Num_K_Int

Number of Observations Read	4250
Number of Observations Used	4250

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	4248	812.6113	0.1913
Scaled Deviance	4248	812.6113	0.1913
Pearson Chi-Square	4248	4866.8103	1.1457
Scaled Pearson X2	4248	4866.8103	1.1457
Log Likelihood		-507.6016	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-7.8378	1.0573	-9.9100	-5.7656	54.96
Log_Major_Aadt	1	0.4288	0.1059	0.2213	0.6364	16.40
Scale	0	1.0000	0.0000	1.0000	1.0000	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	<.0001
Log_Major_Aadt	<.0001
Scale	

NOTE: The scale parameter was held fixed.

APPENDIX H TYPE-A INJURY OUTPUT

The SAS System 1

The GENMOD Procedure

Model Information

Data Set	TYPE_A.RURAL_2LANE_HIGHWAY
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_A_Mile

Number of Observations Read	25831
Number of Observations Used	25831

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	26E3	4991.0773	0.1932
Scaled Deviance	26E3	4991.0773	0.1932
Pearson Chi-Square	26E3	34699.5268	1.3434
Scaled Pearson X2	26E3	34699.5268	1.3434
Log Likelihood		-2272.3404	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-5.1935	0.4211	-6.0188	-4.3682	152.13
LogAADT	1	0.4723	0.0525	0.3695	0.5752	81.06
Dispersion	1	26.5689	0.8159	24.9697	28.1681	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	<.0001
LogAADT	<.0001
Dispersion	

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

The GENMOD Procedure

Model Information

Data Set	TYPE_A.RURAL_MULTI_UNDIV_HIGHWAY
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_A_Mile

Number of Observations Read	196
Number of Observations Used	196

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	194	27.5525	0.1420
Scaled Deviance	194	27.5525	0.1420
Pearson Chi-Square	194	161.4453	0.8322
Scaled Pearson X2	194	161.4453	0.8322
Log Likelihood		-37.4661	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-16.8554	11.5119	-39.4183	5.7076	2.14
LogAADT	1	1.6296	1.3277	-0.9725	4.2318	1.51
Dispersion	1	17.0609	11.0732	-0.3400	38.7639	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	0.1431
LogAADT	0.2197
Dispersion	

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

The SAS System 3

The GENMOD Procedure

Model Information

Data Set	TYPE_A.RURAL_MULTI_DIV_HIGHWAY
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_A_Mile

Number of Observations Read	1942
Number of Observations Used	1942

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	1940	300.1244	0.1547
Scaled Deviance	1940	300.1244	0.1547
Pearson Chi-Square	1940	1202.1771	0.6197
Scaled Pearson X2	1940	1202.1771	0.6197
Log Likelihood		62.2247	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-6.3267	2.8279	-11.8693	-0.7841	5.01
LogAADT	1	0.5826	0.3176	-0.0398	1.2050	3.37
Dispersion	1	42.1276	4.8532	32.6156	51.6396	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	0.0253
LogAADT	0.0666
Dispersion	

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

The SAS System 4

The GENMOD Procedure

Model Information

Data Set	TYPE_A.RURAL_FREEWAY_3OR4LANES
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_A_Mile

Number of Observations Read	3004
Number of Observations Used	3004

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	3002	1718.9175	0.5726
Scaled Deviance	3002	1718.9175	0.5726
Pearson Chi-Square	3002	3275.9326	1.0913
Scaled Pearson X2	3002	3275.9326	1.0913
Log Likelihood		168.7394	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-9.3388	1.3230	-11.9319	-6.7458	49.83
LogAADT	1	0.9518	0.1353	0.6865	1.2170	49.46
Dispersion	1	6.7641	0.3409	6.0960	7.4321	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	<.0001
LogAADT	<.0001
Dispersion	

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

The SAS System 5

The GENMOD Procedure

Model Information

Data Set	TYPE_A.RURAL_FREEWAY_5PLUSLANES
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_A_Mile

Number of Observations Read	137
Number of Observations Used	137

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	135	48.4096	0.3586
Scaled Deviance	135	48.4096	0.3586
Pearson Chi-Square	135	64.3724	0.4768
Scaled Pearson X2	135	64.3724	0.4768
Log Likelihood		-7.1271	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-3.9833	7.5852	-18.8499	10.8834	0.28
LogAADT	1	0.3579	0.7468	-1.1059	1.8216	0.23
Dispersion	1	15.0070	4.3738	6.4345	23.5795	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	0.5995
LogAADT	0.6318
Dispersion	

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

The SAS System 6

The GENMOD Procedure

Model Information

Data Set	TYPE_A.URBAN_2LANE_ARTERIAL
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_A_Mile

Number of Observations Read	10071
Number of Observations Used	10071

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	1E4	1679.6800	0.1668
Scaled Deviance	1E4	1679.6800	0.1668
Pearson Chi-Square	1E4	10072.1118	1.0003
Scaled Pearson X2	1E4	10072.1118	1.0003
Log Likelihood		3413.9556	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-5.9792	1.0817	-8.0993	-3.8590	30.55
LogAADT	1	0.5709	0.1195	0.3367	0.8050	22.83
Dispersion	1	42.3541	1.9871	38.4595	46.2487	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	<.0001
LogAADT	<.0001
Dispersion	

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

The GENMOD Procedure

Model Information

Data Set	TYPE_A.URBAN_ONEWAY_ARTERIAL
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_A_Mile

Number of Observations Read	1264
Number of Observations Used	1264

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	1262	83.7905	0.0664
Scaled Deviance	1262	83.7905	0.0664
Pearson Chi-Square	1262	598.4296	0.4742
Scaled Pearson X2	1262	598.4296	0.4742
Log Likelihood		1284.9533	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-3.3978	3.1807	-9.6318	2.8362	1.14
LogAADT	1	0.3015	0.3530	-0.3904	0.9934	0.73
Dispersion	1	154.0989	29.7093	95.8697	212.3280	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	0.2854
LogAADT	0.3931
Dispersion	

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

The SAS System 8

The GENMOD Procedure

Model Information

Data Set	TYPE_A.URBAN_MULTI_UNDIV_HIGHWAY
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_A_Mile

Number of Observations Read	4260
Number of Observations Used	4259
Missing Values	1

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	4257	748.3245	0.1758
Scaled Deviance	4257	748.3245	0.1758
Pearson Chi-Square	4257	3059.4114	0.7187
Scaled Pearson X2	4257	3059.4114	0.7187
Log Likelihood		1901.6894	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-7.7459	1.9385	-11.5453	-3.9465	15.97
LogAADT	1	0.7277	0.1978	0.3401	1.1154	13.54
Dispersion	1	41.6618	2.8843	36.0088	47.3148	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	<.0001
LogAADT	0.0002
Dispersion	

NOTE: The negative binomial dispersion parameter was estimated by maximum

The GENMOD Procedure

Model Information

Data Set	TYPE_A.URBAN_MULTI_DIV_HIGHWAY
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_A_Mile

Number of Observations Read	9087
Number of Observations Used	9087

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	9085	1914.5526	0.2107
Scaled Deviance	9085	1914.5526	0.2107
Pearson Chi-Square	9085	6656.4086	0.7327
Scaled Pearson X2	9085	6656.4086	0.7327
Log Likelihood		10978.1018	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-6.6669	1.0756	-8.7751	-4.5587	38.42
LogAADT	1	0.6588	0.1085	0.4461	0.8715	36.85
Dispersion	1	36.9572	1.5480	33.9233	39.9912	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	<.0001
LogAADT	<.0001
Dispersion	

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

The GENMOD Procedure

Model Information

Data Set	TYPE_A.URBAN_FREEWAY_3OR4LANES
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_A_Mile

Number of Observations Read	2148
Number of Observations Used	2148

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	2146	1237.4183	0.5766
Scaled Deviance	2146	1237.4183	0.5766
Pearson Chi-Square	2146	1523.6551	0.7100
Scaled Pearson X2	2146	1523.6551	0.7100
Log Likelihood		1751.7883	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-10.0453	1.1235	-12.2473	-7.8432	79.94
LogAADT	1	1.0128	0.1096	0.7979	1.2277	85.34
Dispersion	1	7.5816	0.4296	6.7397	8.4235	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	<.0001
LogAADT	<.0001
Dispersion	

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

The SAS System 12

The GENMOD Procedure

Model Information

Data Set	TYPE_A.URBAN_FREEWAY_5OR6LANES
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_A_Mile

Number of Observations Read	1376
Number of Observations Used	1376

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	1374	1154.2094	0.8400
Scaled Deviance	1374	1154.2094	0.8400
Pearson Chi-Square	1374	823.5612	0.5994
Scaled Pearson X2	1374	823.5612	0.5994
Log Likelihood		8615.5735	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-7.9097	1.2222	-10.3052	-5.5142	41.88
LogAADT	1	0.8153	0.1064	0.6067	1.0239	58.70
Dispersion	1	4.8675	0.2610	4.3559	5.3791	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	<.0001
LogAADT	<.0001
Dispersion	

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

The SAS System 13

The GENMOD Procedure

Model Information

Data Set	TYPE_A.URBAN_FREEWAY_7PLUSLANES
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_A_Mile

Number of Observations Read	361
Number of Observations Used	361

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	359	330.6076	0.9209
Scaled Deviance	359	330.6076	0.9209
Pearson Chi-Square	359	158.8551	0.4425
Scaled Pearson X2	359	158.8551	0.4425
Log Likelihood		5446.0992	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-12.9062	4.0299	-20.8047	-5.0077	10.26
LogAADT	1	1.2260	0.3306	0.5780	1.8741	13.75
Dispersion	1	4.5481	0.4368	3.6921	5.4041	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	0.0014
LogAADT	0.0002
Dispersion	

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

The SAS System 1

The GENMOD Procedure

Model Information

Data Set	TYPE_A.RURAL_MINOR_LEG_STOP
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_A_Int

Number of Observations Read	14933
Number of Observations Used	14933

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	15E3	6062.5116	0.4061
Scaled Deviance	15E3	6062.5116	0.4061
Pearson Chi-Square	15E3	14980.8187	1.0034
Scaled Pearson X2	15E3	14980.8187	1.0034
Log Likelihood		-5601.3231	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-8.5737	0.3351	-9.2305	-7.9170	654.71
Log_Major_Aadt	1	0.6006	0.0394	0.5235	0.6777	232.93
Log_Minor_Aadt	1	0.2933	0.0236	0.2470	0.3397	154.03
Dispersion	1	2.1389	0.1680	1.8096	2.4682	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	<.0001
Log_Major_Aadt	<.0001
Log_Minor_Aadt	<.0001
Dispersion	

The GENMOD Procedure

Model Information

Data Set	TYPE_A.RURAL_ALL_WAY_STOP
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_A_Int

Number of Observations Read	351
Number of Observations Used	351

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	349	174.6615	0.5005
Scaled Deviance	349	174.6615	0.5005
Pearson Chi-Square	349	344.0677	0.9859
Scaled Pearson X2	349	344.0677	0.9859
Log Likelihood		-171.6617	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-6.0949	1.9840	-9.9835	-2.2064	9.44
Log_Major_Aadt	1	0.5437	0.2387	0.0758	1.0116	5.19
Dispersion	1	2.1528	0.8923	0.4040	3.9016	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	0.0021
Log_Major_Aadt	0.0227
Dispersion	

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

The GENMOD Procedure

Model Information

Data Set	TYPE_A.RURAL_SIGNAL
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_A_Int

Number of Observations Read	199
Number of Observations Used	199

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	197	167.0922	0.8482
Scaled Deviance	197	167.0922	0.8482
Pearson Chi-Square	197	200.1064	1.0158
Scaled Pearson X2	197	200.1064	1.0158
Log Likelihood		-145.5344	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-11.2428	2.4285	-16.0025	-6.4831	21.43
Log_Major_Aadt	1	1.1899	0.2647	0.6710	1.7088	20.20
Dispersion	1	1.2029	0.3380	0.5405	1.8654	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	<.0001
Log_Major_Aadt	<.0001
Dispersion	

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

The GENMOD Procedure

Model Information

Data Set	TYPE_A.RURAL_UNDETERMINED
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_A_Int

Number of Observations Read	5579
Number of Observations Used	5579

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	5576	2089.0592	0.3747
Scaled Deviance	5576	2089.0592	0.3747
Pearson Chi-Square	5576	6103.2676	1.0946
Scaled Pearson X2	5576	6103.2676	1.0946
Log Likelihood		-1959.8903	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-7.1319	0.5468	-8.2036	-6.0602	170.12
Log_Major_Aadt	1	0.5646	0.0641	0.4390	0.6902	77.60
Log_Minor_Aadt	1	0.0669	0.0336	0.0010	0.1329	3.96
Dispersion	1	2.4825	0.3353	1.8254	3.1396	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	<.0001
Log_Major_Aadt	<.0001
Log_Minor_Aadt	0.0465
Dispersion	

The GENMOD Procedure

Model Information

Data Set	TYPE_A.URBAN_MINOR_LEG_STOP
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_A_Int

Number of Observations Read	12121
Number of Observations Used	12121

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	12E3	9181.6581	0.7577
Scaled Deviance	12E3	9181.6581	0.7577
Pearson Chi-Square	12E3	12621.4567	1.0415
Scaled Pearson X2	12E3	12621.4567	1.0415
Log Likelihood		-8645.3411	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-7.7951	0.2881	-8.3598	-7.2304	732.04
Log_Major_Aadt	1	0.5562	0.0269	0.5035	0.6089	427.80
Log_Minor_Aadt	1	0.2141	0.0213	0.1723	0.2558	100.88
Dispersion	1	0.9785	0.0555	0.8697	1.0874	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	<.0001
Log_Major_Aadt	<.0001
Log_Minor_Aadt	<.0001
Dispersion	

The GENMOD Procedure

Model Information

Data Set	TYPE_A.URBAN_ALL_WAY_STOP
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_A_Int

Number of Observations Read	132
Number of Observations Used	132

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	130	110.1738	0.8475
Scaled Deviance	130	110.1738	0.8475
Pearson Chi-Square	130	142.6293	1.0971
Scaled Pearson X2	130	142.6293	1.0971
Log Likelihood		-96.0124	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-7.8248	2.5345	-12.7923	-2.8574	9.53
Log_Major_Aadt	1	0.7721	0.2801	0.2232	1.3210	7.60
Dispersion	1	0.3371	0.3535	-0.2500	1.0299	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	0.0020
Log_Major_Aadt	0.0058
Dispersion	

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

The GENMOD Procedure

Model Information

Data Set	TYPE_A.URBAN_SIGNAL
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_A_Int

Number of Observations Read	4311
Number of Observations Used	4311

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	4308	4575.2726	1.0620
Scaled Deviance	4308	4575.2726	1.0620
Pearson Chi-Square	4308	4792.3405	1.1124
Scaled Pearson X2	4308	4792.3405	1.1124
Log Likelihood		-1810.4690	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-9.3836	0.3515	-10.0725	-8.6948	712.86
Log_Major_Aadt	1	0.7651	0.0325	0.7013	0.8289	552.67
Log_Minor_Aadt	1	0.2589	0.0141	0.2313	0.2866	336.84
Dispersion	1	0.6950	0.0336	0.6291	0.7608	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	<.0001
Log_Major_Aadt	<.0001
Log_Minor_Aadt	<.0001
Dispersion	

The SAS System 8

The GENMOD Procedure

Model Information

Data Set	TYPE_A.URBAN_UNDETERMINED
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_A_Int

Number of Observations Read	4250
Number of Observations Used	4250

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	4248	2502.2037	0.5890
Scaled Deviance	4248	2502.2037	0.5890
Pearson Chi-Square	4248	4669.2625	1.0992
Scaled Pearson X2	4248	4669.2625	1.0992
Log Likelihood		-2337.1962	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-6.4556	0.4059	-7.2512	-5.6600	252.92
Log_Major_Aadt	1	0.5193	0.0408	0.4393	0.5994	161.74
Dispersion	1	1.4751	0.1612	1.1592	1.7911	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	<.0001
Log_Major_Aadt	<.0001
Dispersion	

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

APPENDIX I TYPE-B INJURY OUTPUT

The SAS System 1

The GENMOD Procedure

Model Information

Data Set	TYPE_B.RURAL_2LANE_HIGHWAY
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_B_Mile

Number of Observations Read	25829
Number of Observations Used	25829

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	26E3	7093.4977	0.2747
Scaled Deviance	26E3	7093.4977	0.2747
Pearson Chi-Square	26E3	35479.2000	1.3737
Scaled Pearson X2	26E3	35479.2000	1.3737
Log Likelihood		974.2657	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-5.0390	0.3707	-5.7655	-4.3124	184.77
LogAADT	1	0.5229	0.0462	0.4323	0.6136	127.87
Dispersion	1	19.0545	0.4688	18.1357	19.9734	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	<.0001
LogAADT	<.0001
Dispersion	

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

The GENMOD Procedure

Model Information

Data Set	TYPE_B.RURAL_MULTI_UNDIV_HIGHWAY
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_B_Mile

Number of Observations Read	157
Number of Observations Used	157

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	155	43.8150	0.2827
Scaled Deviance	155	43.8150	0.2827
Pearson Chi-Square	155	134.8906	0.8703
Scaled Pearson X2	155	134.8906	0.8703
Log Likelihood		64.3369	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-17.9548	9.9340	-37.4251	1.5154	3.27
LogAADT	1	2.0656	1.1783	-0.2439	4.3750	3.07
Dispersion	1	20.5642	6.1446	8.5210	32.6074	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	0.0707
LogAADT	0.0796
Dispersion	

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

The GENMOD Procedure

Model Information

Data Set	TYPE_B.RURAL_MULTI_DIV_HIGHWAY
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_B_Mile

Number of Observations Read	1979
Number of Observations Used	1979

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	1977	529.7411	0.2680
Scaled Deviance	1977	529.7411	0.2680
Pearson Chi-Square	1977	1803.0113	0.9120
Scaled Pearson X2	1977	1803.0113	0.9120
Log Likelihood		1833.6941	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-7.2932	1.8092	-10.8392	-3.7472	16.25
LogAADT	1	0.7948	0.2024	0.3981	1.1916	15.42
Dispersion	1	25.3344	2.0739	21.2696	29.3991	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	<.0001
LogAADT	<.0001
Dispersion	

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

The GENMOD Procedure

Model Information

Data Set	TYPE_B.RURAL_FREEWAY_3OR4LANES
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_B_Mile

Number of Observations Read	3057
Number of Observations Used	3057

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	3055	2266.5474	0.7419
Scaled Deviance	3055	2266.5474	0.7419
Pearson Chi-Square	3055	2620.4604	0.8578
Scaled Pearson X2	3055	2620.4604	0.8578
Log Likelihood		1171.2365	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-4.8969	1.0197	-6.8956	-2.8983	23.06
LogAADT	1	0.5479	0.1043	0.3434	0.7524	27.57
Dispersion	1	4.6377	0.1992	4.2472	5.0282	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	<.0001
LogAADT	<.0001
Dispersion	

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

The GENMOD Procedure

Model Information

Data Set	TYPE_B.RURAL_FREEWAY_5PLUSLANES
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_B_Mile

Number of Observations Read	134
Number of Observations Used	134

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	132	63.1121	0.4781
Scaled Deviance	132	63.1121	0.4781
Pearson Chi-Square	132	93.6187	0.7092
Scaled Pearson X2	132	93.6187	0.7092
Log Likelihood		133.4486	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-4.2743	6.5595	-17.1307	8.5821	0.42
LogAADT	1	0.4515	0.6459	-0.8145	1.7175	0.49
Dispersion	1	11.6784	2.7956	6.1991	17.1577	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	0.5146
LogAADT	0.4846
Dispersion	

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

The SAS System 6

The GENMOD Procedure

Model Information

Data Set	TYPE_B.URBAN_2LANE_ARTERIAL
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_B_Mile

Number of Observations Read	10065
Number of Observations Used	10065

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	1E4	2721.0811	0.2704
Scaled Deviance	1E4	2721.0811	0.2704
Pearson Chi-Square	1E4	8695.5052	0.8641
Scaled Pearson X2	1E4	8695.5052	0.8641
Log Likelihood		11816.4941	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-6.3543	0.8089	-7.9396	-4.7689	61.71
LogAADT	1	0.6975	0.0893	0.5225	0.8726	60.98
Dispersion	1	26.2031	0.9324	24.3756	28.0306	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	<.0001
LogAADT	<.0001
Dispersion	

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

The SAS System 7

The GENMOD Procedure

Model Information

Data Set	TYPE_B.URBAN_ONEWAY_ARTERIAL
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_B_Mile

Number of Observations Read	1248
Number of Observations Used	1248

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	1246	128.0348	0.1028
Scaled Deviance	1246	128.0348	0.1028
Pearson Chi-Square	1246	573.7497	0.4605
Scaled Pearson X2	1246	573.7497	0.4605
Log Likelihood		1162.7844	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-5.5627	2.2805	-10.0324	-1.0930	5.95
LogAADT	1	0.5507	0.2525	0.0558	1.0456	4.76
Dispersion	1	88.2121	14.1045	60.5678	115.8563	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	0.0147
LogAADT	0.0292
Dispersion	

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

The SAS System 8

The GENMOD Procedure

Model Information

Data Set	TYPE_B.URBAN_MULTI_UNDIV_HIGHWAY
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_B_Mile

Number of Observations Read	4254
Number of Observations Used	4253
Missing Values	1

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	4251	1255.7789	0.2954
Scaled Deviance	4251	1255.7789	0.2954
Pearson Chi-Square	4251	2735.7928	0.6436
Scaled Pearson X2	4251	2735.7928	0.6436
Log Likelihood		11106.8753	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-6.5509	1.6072	-9.7010	-3.4009	16.61
LogAADT	1	0.7141	0.1641	0.3925	1.0357	18.94
Dispersion	1	26.3981	1.3315	23.7884	29.0079	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	<.0001
LogAADT	<.0001
Dispersion	

NOTE: The negative binomial dispersion parameter was estimated by maximum

The GENMOD Procedure

Model Information

Data Set	TYPE_B.URBAN_MULTI_DIV_HIGHWAY
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_B_Mile

Number of Observations Read	9084
Number of Observations Used	9084

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	9082	2993.3191	0.3296
Scaled Deviance	9082	2993.3191	0.3296
Pearson Chi-Square	9082	7778.3041	0.8565
Scaled Pearson X2	9082	7778.3041	0.8565
Log Likelihood		36086.8920	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-7.0622	0.8541	-8.7362	-5.3883	68.38
LogAADT	1	0.7799	0.0862	0.6110	0.9488	81.90
Dispersion	1	23.6353	0.7648	22.1364	25.1342	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	<.0001
LogAADT	<.0001
Dispersion	

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

The SAS System 11

The GENMOD Procedure

Model Information

Data Set	TYPE_B.URBAN_FREEWAY_3OR4LANES
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_B_Mile

Number of Observations Read	2146
Number of Observations Used	2146

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	2144	1660.8757	0.7747
Scaled Deviance	2144	1660.8757	0.7747
Pearson Chi-Square	2144	1526.9187	0.7122
Scaled Pearson X2	2144	1526.9187	0.7122
Log Likelihood		9864.5945	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-9.6281	0.8736	-11.3402	-7.9159	121.47
LogAADT	1	1.0521	0.0853	0.8849	1.2194	152.03
Dispersion	1	5.2772	0.2429	4.8011	5.7533	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	<.0001
LogAADT	<.0001
Dispersion	

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

The SAS System 12

The GENMOD Procedure

Model Information

Data Set	TYPE_B.URBAN_FREEWAY_5OR6LANES
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_B_Mile

Number of Observations Read	1372
Number of Observations Used	1372

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	1370	1434.3442	1.0470
Scaled Deviance	1370	1434.3442	1.0470
Pearson Chi-Square	1370	1118.4738	0.8164
Scaled Pearson X2	1370	1118.4738	0.8164
Log Likelihood		58991.7681	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-11.5665	0.9324	-13.3940	-9.7391	153.89
LogAADT	1	1.2354	0.0812	1.0764	1.3945	231.73
Dispersion	1	3.3532	0.1503	3.0586	3.6478	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	<.0001
LogAADT	<.0001
Dispersion	

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

The GENMOD Procedure

Model Information

Data Set	TYPE_B.URBAN_FREEWAY_7PLUSLANES
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_B_Mile

Number of Observations Read	380
Number of Observations Used	380

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	378	452.6572	1.1975
Scaled Deviance	378	452.6572	1.1975
Pearson Chi-Square	378	239.6382	0.6340
Scaled Pearson X2	378	239.6382	0.6340
Log Likelihood		63325.0153	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-12.0600	2.5787	-17.1141	-7.0058	21.87
LogAADT	1	1.3011	0.2115	0.8865	1.7157	37.83
Dispersion	1	2.2105	0.1638	1.8894	2.5316	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	<.0001
LogAADT	<.0001
Dispersion	

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

The SAS System 1

The GENMOD Procedure

Model Information

Data Set	TYPE_B.RURAL_MINOR_LEG_STOP
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_B_Int

Number of Observations Read	14933
Number of Observations Used	14933

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	15E3	8423.3936	0.5642
Scaled Deviance	15E3	8423.3936	0.5642
Pearson Chi-Square	15E3	15497.6709	1.0380
Scaled Pearson X2	15E3	15497.6709	1.0380
Log Likelihood		-7608.7497	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-9.2195	0.2631	-9.7352	-8.7039	1228.14
Log_Major_Aadt	1	0.7636	0.0308	0.7032	0.8241	613.66
Log_Minor_Aadt	1	0.2645	0.0183	0.2286	0.3003	209.19
Dispersion	1	1.3381	0.0848	1.1720	1.5043	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	<.0001
Log_Major_Aadt	<.0001
Log_Minor_Aadt	<.0001
Dispersion	

The GENMOD Procedure

Model Information

Data Set	TYPE_B.RURAL_ALL_WAY_STOP
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_B_Int

Number of Observations Read	351
Number of Observations Used	351

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	348	246.5393	0.7084
Scaled Deviance	348	246.5393	0.7084
Pearson Chi-Square	348	389.6096	1.1196
Scaled Pearson X2	348	389.6096	1.1196
Log Likelihood		-242.4554	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-5.9267	1.6174	-9.0968	-2.7565	13.43
Log_Major_Aadt	1	0.4559	0.1801	0.1030	0.8089	6.41
Log_Minor_Aadt	1	0.1772	0.1147	-0.0475	0.4019	2.39
Dispersion	1	1.9074	0.4379	1.0491	2.7656	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	0.0002
Log_Major_Aadt	0.0113
Log_Minor_Aadt	0.1223
Dispersion	

The GENMOD Procedure

Model Information

Data Set	TYPE_B.RURAL_SIGNAL
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_B_Int

Number of Observations Read	199
Number of Observations Used	199

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	196	198.2452	1.0115
Scaled Deviance	196	198.2452	1.0115
Pearson Chi-Square	196	193.7977	0.9888
Scaled Pearson X2	196	193.7977	0.9888
Log Likelihood		-36.6389	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-14.3894	2.2801	-18.8583	-9.9204	39.83
Log_Major_Aadt	1	1.4815	0.2372	1.0165	1.9464	39.00
Log_Minor_Aadt	1	0.1703	0.0721	0.0290	0.3115	5.58
Dispersion	1	1.1028	0.2204	0.6709	1.5348	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	<.0001
Log_Major_Aadt	<.0001
Log_Minor_Aadt	0.0182
Dispersion	

The GENMOD Procedure

Model Information

Data Set	TYPE_B.RURAL_UNDETERMINED
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_B_Int

Number of Observations Read	5579
Number of Observations Used	5579

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	5576	2833.0289	0.5081
Scaled Deviance	5576	2833.0289	0.5081
Pearson Chi-Square	5576	5939.4837	1.0652
Scaled Pearson X2	5576	5939.4837	1.0652
Log Likelihood		-2686.4446	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-7.5473	0.4298	-8.3896	-6.7049	308.39
Log_Major_Aadt	1	0.6901	0.0506	0.5908	0.7894	185.66
Log_Minor_Aadt	1	0.0503	0.0270	-0.0026	0.1032	3.47
Dispersion	1	1.8587	0.1890	1.4882	2.2292	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	<.0001
Log_Major_Aadt	<.0001
Log_Minor_Aadt	0.0623
Dispersion	

The GENMOD Procedure

Model Information

Data Set	TYPE_B.URBAN_MINOR_LEG_STOP
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_B_Int

Number of Observations Read	12121
Number of Observations Used	12121

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	12E3	11649.1488	0.9613
Scaled Deviance	12E3	11649.1488	0.9613
Pearson Chi-Square	12E3	13611.7060	1.1233
Scaled Pearson X2	12E3	13611.7060	1.1233
Log Likelihood		-8704.1170	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-8.3063	0.2202	-8.7378	-7.8747	1423.17
Log_Major_Aadt	1	0.6698	0.0204	0.6299	0.7097	1082.27
Log_Minor_Aadt	1	0.2655	0.0162	0.2337	0.2972	268.33
Dispersion	1	0.9355	0.0285	0.8796	0.9914	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	<.0001
Log_Major_Aadt	<.0001
Log_Minor_Aadt	<.0001
Dispersion	

The SAS System 3

The GENMOD Procedure

Model Information

Data Set	TYPE_B.URBAN_ALL_WAY_STOP
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_B_Int

Number of Observations Read	132
Number of Observations Used	132

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	130	127.6393	0.9818
Scaled Deviance	130	127.6393	0.9818
Pearson Chi-Square	130	158.4145	1.2186
Scaled Pearson X2	130	158.4145	1.2186
Log Likelihood		-97.1917	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-7.9816	2.1165	-12.1298	-3.8335	14.22
Log_Major_Aadt	1	0.8931	0.2352	0.4321	1.3541	14.42
Dispersion	1	0.9893	0.3073	0.3871	1.5915	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	0.0002
Log_Major_Aadt	0.0001
Dispersion	

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

The GENMOD Procedure

Model Information

Data Set	TYPE_B.URBAN_SIGNAL
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_B_Int

Number of Observations Read	4311
Number of Observations Used	4311

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	4308	4945.4535	1.1480
Scaled Deviance	4308	4945.4535	1.1480
Pearson Chi-Square	4308	4295.6276	0.9971
Scaled Pearson X2	4308	4295.6276	0.9971
Log Likelihood		16615.8830	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-8.6608	0.2772	-9.2041	-8.1176	976.34
Log_Major_Aadt	1	0.8009	0.0258	0.7504	0.8514	966.22
Log_Minor_Aadt	1	0.2535	0.0113	0.2313	0.2757	500.01
Dispersion	1	0.6490	0.0213	0.6073	0.6907	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	<.0001
Log_Major_Aadt	<.0001
Log_Minor_Aadt	<.0001
Dispersion	

The SAS System 4

The GENMOD Procedure

Model Information

Data Set	TYPE_B.URBAN_UNDETERMINED
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_B_Int

Number of Observations Read	4250
Number of Observations Used	4250

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	4248	3458.9531	0.8143
Scaled Deviance	4248	3458.9531	0.8143
Pearson Chi-Square	4248	4955.5341	1.1666
Scaled Pearson X2	4248	4955.5341	1.1666
Log Likelihood		-2907.5896	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-6.7745	0.3074	-7.3770	-6.1720	485.68
Log_Major_Aadt	1	0.6493	0.0310	0.5885	0.7101	437.49
Dispersion	1	1.4154	0.0836	1.2516	1.5792	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	<.0001
Log_Major_Aadt	<.0001
Dispersion	

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

APPENDIX J FATAL AND INJURY OUTPUT

The SAS System 1

The GENMOD Procedure

Model Information

Data Set	FI.RURAL_2LANE_HIGHWAY
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_FI_Mile

Number of Observations Read	25885
Number of Observations Used	25885

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	26E3	9745.4597	0.3765
Scaled Deviance	26E3	9745.4597	0.3765
Pearson Chi-Square	26E3	42708.4213	1.6501
Scaled Pearson X2	26E3	42708.4213	1.6501
Log Likelihood		13909.4934	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-4.4354	0.3094	-5.0419	-3.8290	205.50
LogAADT	1	0.5248	0.0386	0.4492	0.6005	184.86
Dispersion	1	14.3861	0.2859	13.8258	14.9464	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	<.0001
LogAADT	<.0001
Dispersion	

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

The GENMOD Procedure

Model Information

Data Set	FI.RURAL_MULTI_UNDIV_HIGHWAY
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_FI_Mile

Number of Observations Read	188
Number of Observations Used	188

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	186	55.5571	0.2987
Scaled Deviance	186	55.5571	0.2987
Pearson Chi-Square	186	104.1086	0.5597
Scaled Pearson X2	186	104.1086	0.5597
Log Likelihood		-41.5279	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-3.0048	8.9319	-20.5111	14.5014	0.11
LogAADT	1	0.2588	1.0474	-1.7941	2.3118	0.06
Dispersion	1	17.0161	4.8500	7.5104	26.5219	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	0.7366
LogAADT	0.8048
Dispersion	

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

The GENMOD Procedure

Model Information

Data Set	FI.RURAL_MULTI_DIV_HIGHWAY
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_FI_Mile

Number of Observations Read	2022
Number of Observations Used	2022

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	2020	762.0457	0.3773
Scaled Deviance	2020	762.0457	0.3773
Pearson Chi-Square	2020	2167.8554	1.0732
Scaled Pearson X2	2020	2167.8554	1.0732
Log Likelihood		6574.4987	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-7.7671	1.4669	-10.6422	-4.8921	28.04
LogAADT	1	0.9228	0.1637	0.6021	1.2436	31.80
Dispersion	1	18.2732	1.1979	15.9254	20.6210	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	<.0001
LogAADT	<.0001
Dispersion	

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

The SAS System 4

The GENMOD Procedure

Model Information

Data Set	FI.RURAL_FREEWAY_3OR4LANES
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_FI_Mile

Number of Observations Read	3116
Number of Observations Used	3116

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	3114	2763.3805	0.8874
Scaled Deviance	3114	2763.3805	0.8874
Pearson Chi-Square	3114	6585.2666	2.1147
Scaled Pearson X2	3114	6585.2666	2.1147
Log Likelihood		13290.4240	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-6.6871	0.8369	-8.3274	-5.0467	63.84
LogAADT	1	0.8039	0.0855	0.6363	0.9714	88.41
Dispersion	1	3.9335	0.1387	3.6616	4.2054	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	<.0001
LogAADT	<.0001
Dispersion	

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

The GENMOD Procedure

Model Information

Data Set	FI.RURAL_FREEWAY_5PLUSLANES
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_FI_Mile

Number of Observations Read	138
Number of Observations Used	138

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	136	87.8815	0.6462
Scaled Deviance	136	87.8815	0.6462
Pearson Chi-Square	136	125.5775	0.9234
Scaled Pearson X2	136	125.5775	0.9234
Log Likelihood		1310.8639	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-7.4058	5.6745	-18.5276	3.7160	1.70
LogAADT	1	0.8717	0.5584	-0.2227	1.9661	2.44
Dispersion	1	9.6706	1.7762	6.1893	13.1518	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	0.1919
LogAADT	0.1185
Dispersion	

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

The GENMOD Procedure

Model Information

Data Set	FI.URBAN_2LANE_ARTERIAL
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_FI_Mile

Number of Observations Read	10089
Number of Observations Used	10089

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	1E4	3427.5785	0.3398
Scaled Deviance	1E4	3427.5785	0.3398
Pearson Chi-Square	1E4	19484.7011	1.9317
Scaled Pearson X2	1E4	19484.7011	1.9317
Log Likelihood		37374.1391	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-3.5572	0.6998	-4.9288	-2.1856	25.84
LogAADT	1	0.4623	0.0773	0.3108	0.6138	35.77
Dispersion	1	22.5425	0.6828	21.2043	23.8808	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	<.0001
LogAADT	<.0001
Dispersion	

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

The SAS System 7

The GENMOD Procedure

Model Information

Data Set	FI.URBAN_ONEWAY_ARTERIAL
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_FI_Mile

Number of Observations Read	1264
Number of Observations Used	1264

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	1262	178.5460	0.1415
Scaled Deviance	1262	178.5460	0.1415
Pearson Chi-Square	1262	767.3162	0.6080
Scaled Pearson X2	1262	767.3162	0.6080
Log Likelihood		9480.1525	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-5.4945	2.0711	-9.5539	-1.4352	7.04
LogAADT	1	0.6886	0.2294	0.2390	1.1381	9.01
Dispersion	1	78.2419	9.9419	58.7562	97.7277	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	0.0080
LogAADT	0.0027
Dispersion	

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

The GENMOD Procedure

Model Information

Data Set	FI.URBAN_MULTI_UNDIV_HIGHWAY
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_FI_Mile

Number of Observations Read	4285
Number of Observations Used	4284
Missing Values	1

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	4282	1545.4172	0.3609
Scaled Deviance	4282	1545.4172	0.3609
Pearson Chi-Square	4282	18498.8588	4.3201
Scaled Pearson X2	4282	18498.8588	4.3201
Log Likelihood		67488.6620	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-4.8758	1.6985	-8.2048	-1.5468	8.24
LogAADT	1	0.6471	0.1734	0.3072	0.9870	13.92
Dispersion	1	25.5562	1.0922	23.4155	27.6969	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	0.0041
LogAADT	0.0002
Dispersion	

NOTE: The negative binomial dispersion parameter was estimated by maximum

The GENMOD Procedure

Model Information

Data Set	FI.URBAN_MULTI_DIV_HIGHWAY
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_FI_Mile

Number of Observations Read	9118
Number of Observations Used	9118

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	9116	3695.0805	0.4053
Scaled Deviance	9116	3695.0805	0.4053
Pearson Chi-Square	9116	15935.0440	1.7480
Scaled Pearson X2	9116	15935.0440	1.7480
Log Likelihood		102399.1023	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-6.2056	0.8180	-7.8089	-4.6024	57.55
LogAADT	1	0.7607	0.0826	0.5989	0.9225	84.90
Dispersion	1	20.4350	0.5735	19.3111	21.5590	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	<.0001
LogAADT	<.0001
Dispersion	

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

The GENMOD Procedure

Model Information

Data Set	FI.URBAN_FREEWAY_3OR4LANES
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_FI_Mile

Number of Observations Read	2215
Number of Observations Used	2215

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	2213	2008.8522	0.9078
Scaled Deviance	2213	2008.8522	0.9078
Pearson Chi-Square	2213	3844.6741	1.7373
Scaled Pearson X2	2213	3844.6741	1.7373
Log Likelihood		50550.7740	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-10.3691	0.7229	-11.7859	-8.9523	205.76
LogAADT	1	1.2008	0.0704	1.0628	1.3388	290.76
Dispersion	1	4.5938	0.1779	4.2452	4.9424	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	<.0001
LogAADT	<.0001
Dispersion	

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

The GENMOD Procedure

Model Information

Data Set	FI.URBAN_FREEWAY_5OR6LANES
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_FI_Mile

Number of Observations Read	1453
Number of Observations Used	1453

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	1451	1635.3194	1.1270
Scaled Deviance	1451	1635.3194	1.1270
Pearson Chi-Square	1451	1502.1765	1.0353
Scaled Pearson X2	1451	1502.1765	1.0353
Log Likelihood		197005.1704	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-13.0215	0.8603	-14.7076	-11.3355	229.12
LogAADT	1	1.4250	0.0748	1.2785	1.5715	363.36
Dispersion	1	3.0319	0.1197	2.7972	3.2665	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	<.0001
LogAADT	<.0001
Dispersion	

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

The GENMOD Procedure

Model Information

Data Set	FI.URBAN_FREEWAY_7PLUSLANES
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_FI_Mile

Number of Observations Read	437
Number of Observations Used	437

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	435	527.9795	1.2137
Scaled Deviance	435	527.9795	1.2137
Pearson Chi-Square	435	927.1767	2.1314
Scaled Pearson X2	435	927.1767	2.1314
Log Likelihood		190234.2988	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-10.5076	2.7756	-15.9478	-5.0674	14.33
LogAADT	1	1.2281	0.2269	0.7833	1.6728	29.29
Dispersion	1	2.5856	0.1681	2.2562	2.9151	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	0.0002
LogAADT	<.0001
Dispersion	

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

The SAS System 1

The GENMOD Procedure

Model Information

Data Set	FI.RURAL_MINOR_LEG_STOP
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_FI_Int

Number of Observations Read	14933
Number of Observations Used	14933

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	15E3	10237.5874	0.6857
Scaled Deviance	15E3	10237.5874	0.6857
Pearson Chi-Square	15E3	15707.8622	1.0521
Scaled Pearson X2	15E3	15707.8622	1.0521
Log Likelihood		-9432.8178	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-8.0049	0.2206	-8.4373	-7.5725	1316.67
Log_Major_Aadt	1	0.6736	0.0260	0.6227	0.7245	672.89
Log_Minor_Aadt	1	0.2719	0.0157	0.2411	0.3026	300.04
Dispersion	1	1.4289	0.0621	1.3071	1.5507	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	<.0001
Log_Major_Aadt	<.0001
Log_Minor_Aadt	<.0001
Dispersion	

The GENMOD Procedure

Model Information

Data Set	FI.RURAL_ALL_WAY_STOP
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_FI_Int

Number of Observations Read	351
Number of Observations Used	351

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	348	284.7191	0.8182
Scaled Deviance	348	284.7191	0.8182
Pearson Chi-Square	348	374.5248	1.0762
Scaled Pearson X2	348	374.5248	1.0762
Log Likelihood		-262.3340	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-5.9072	1.4081	-8.6671	-3.1473	17.60
Log_Major_Aadt	1	0.5065	0.1571	0.1987	0.8143	10.40
Log_Minor_Aadt	1	0.1705	0.1003	-0.0261	0.3671	2.89
Dispersion	1	1.6376	0.3177	1.0149	2.2603	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	<.0001
Log_Major_Aadt	0.0013
Log_Minor_Aadt	0.0891
Dispersion	

The GENMOD Procedure

Model Information

Data Set	FI.RURAL_SIGNAL
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_FI_Int

Number of Observations Read	199
Number of Observations Used	199

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	196	210.0987	1.0719
Scaled Deviance	196	210.0987	1.0719
Pearson Chi-Square	196	187.3530	0.9559
Scaled Pearson X2	196	187.3530	0.9559
Log Likelihood		140.9352	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-13.5015	2.1069	-17.6311	-9.3720	41.06
Log_Major_Aadt	1	1.4426	0.2185	1.0144	1.8708	43.60
Log_Minor_Aadt	1	0.1509	0.0657	0.0220	0.2797	5.27
Dispersion	1	1.1028	0.1896	0.7312	1.4745	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	<.0001
Log_Major_Aadt	<.0001
Log_Minor_Aadt	0.0217
Dispersion	

The GENMOD Procedure

Model Information

Data Set	FI.RURAL_UNDETERMINED
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_FI_Int

Number of Observations Read	5579
Number of Observations Used	5579

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	5576	3444.2753	0.6177
Scaled Deviance	5576	3444.2753	0.6177
Pearson Chi-Square	5576	6435.7775	1.1542
Scaled Pearson X2	5576	6435.7775	1.1542
Log Likelihood		-3388.3619	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-6.6375	0.3700	-7.3627	-5.9122	321.75
Log_Major_Aadt	1	0.6314	0.0436	0.5460	0.7168	209.97
Log_Minor_Aadt	1	0.0651	0.0234	0.0193	0.1109	7.75
Dispersion	1	1.9901	0.1385	1.7187	2.2615	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	<.0001
Log_Major_Aadt	<.0001
Log_Minor_Aadt	0.0054
Dispersion	

The GENMOD Procedure

Model Information

Data Set	FI.URBAN_MINOR_LEG_STOP
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_FI_Int

Number of Observations Read	12121
Number of Observations Used	12121

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	12E3	12394.1609	1.0228
Scaled Deviance	12E3	12394.1609	1.0228
Pearson Chi-Square	12E3	13926.6636	1.1493
Scaled Pearson X2	12E3	13926.6636	1.1493
Log Likelihood		-5697.3787	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-7.5800	0.1977	-7.9675	-7.1926	1470.10
Log_Major_Aadt	1	0.6391	0.0183	0.6031	0.6750	1213.28
Log_Minor_Aadt	1	0.2541	0.0146	0.2254	0.2827	302.36
Dispersion	1	0.9045	0.0235	0.8583	0.9506	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	<.0001
Log_Major_Aadt	<.0001
Log_Minor_Aadt	<.0001
Dispersion	

The SAS System 1

The GENMOD Procedure

Model Information

Data Set	FI.URBAN_ALL_WAY_STOP
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_FI_Int

Number of Observations Read	132
Number of Observations Used	132

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	130	133.5714	1.0275
Scaled Deviance	130	133.5714	1.0275
Pearson Chi-Square	130	161.5205	1.2425
Scaled Pearson X2	130	161.5205	1.2425
Log Likelihood		-64.8210	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-6.5960	1.9213	-10.3616	-2.8304	11.79
Log_Major_Aadt	1	0.7779	0.2144	0.3578	1.1981	13.17
Dispersion	1	1.1255	0.2784	0.5799	1.6712	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	0.0006
Log_Major_Aadt	0.0003
Dispersion	

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

The GENMOD Procedure

Model Information

Data Set	FI.URBAN_SIGNAL
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_FI_Int

Number of Observations Read	4311
Number of Observations Used	4311

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	4308	4967.2059	1.1530
Scaled Deviance	4308	4967.2059	1.1530
Pearson Chi-Square	4308	4333.3961	1.0059
Scaled Pearson X2	4308	4333.3961	1.0059
Log Likelihood		31848.8552	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-8.2481	0.2648	-8.7672	-7.7291	969.99
Log_Major_Aadt	1	0.7926	0.0246	0.7444	0.8409	1035.20
Log_Minor_Aadt	1	0.2517	0.0110	0.2302	0.2732	528.01
Dispersion	1	0.6641	0.0197	0.6254	0.7028	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	<.0001
Log_Major_Aadt	<.0001
Log_Minor_Aadt	<.0001
Dispersion	

The GENMOD Procedure

Model Information

Data Set	FI.URBAN_UNDETERMINED
Distribution	Negative Binomial
Link Function	Log
Dependent Variable	Num_FI_Int

Number of Observations Read	4250
Number of Observations Used	4250

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	4248	3847.1612	0.9056
Scaled Deviance	4248	3847.1612	0.9056
Pearson Chi-Square	4248	5142.4168	1.2106
Scaled Pearson X2	4248	5142.4168	1.2106
Log Likelihood		-2705.3293	

Algorithm converged.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Chi-Square
Intercept	1	-5.9577	0.2722	-6.4911	-5.4242	479.15
Log_Major_Aadt	1	0.6023	0.0276	0.5482	0.6564	476.15
Dispersion	1	1.2847	0.0646	1.1581	1.4113	

Analysis Of Parameter Estimates

Parameter	Pr > ChiSq
Intercept	<.0001
Log_Major_Aadt	<.0001
Dispersion	

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

APPENDIX K LOGIC FOR CRASH DATA CONVERSION AND CRASH-ROADWAY DATA MATCHING

Crash Data Conversion

Given the current IDOT databases, a few spatial discrepancies exist between the crash data and roadway data. As a result, a pre-processing step is required to convert the crash data locations and also to match them with roadway sites. IDOT's Data Mart System will help prepare the input data according to the following logic.
As show in Figure K1, the crash and roadway data in the IDOT GIS databases currently use different positioning and roadway systems when they represent the crash locations and roadway network. For the roadway representation, the crash data uses the TS route number system, which provides a numeric representation of the route, while the roadway data utilizes inventory number, a string of numbers that represent the route name and other roadway characteristics through a sequence of numbers. In case of marking positions along the given route, the crash data uses mileposts, which act as an absolute numbering system over the entire state, while the roadway data utilizes stationing, which provides a localized representation of position.

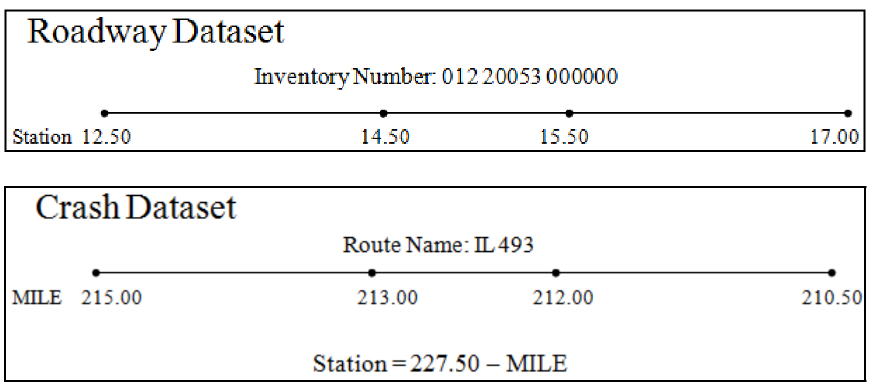


Figure K1: Different GIS database systems between crash data and roadway data.

In order to solve the discrepancy problem, the pre-processing step for making two different datasets into one consistent system is required to match the crash data with roadway data. This pre-processing step is to convert TS route number and milepost system of the crash data into the inventory number and stationing system that the roadway data are utilizing; see Figure K2.

The preprocessing step is not incorporated into the functionality of Excel VBA software. IDOT translation tables or SQL queries in an independent database tool (e.g. Microsoft Access) can be used for the crash data conversion.

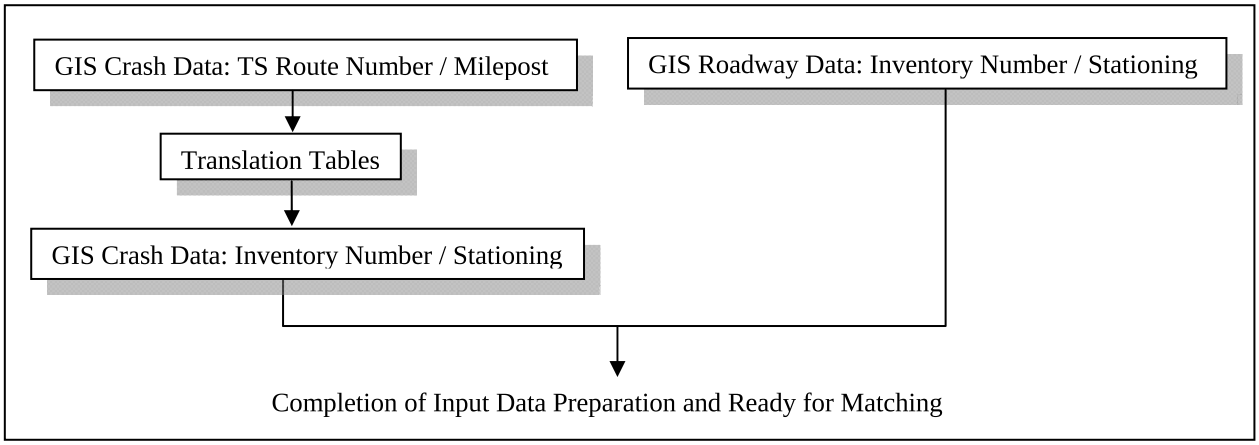


Figure K2: Logic of pre-processing step for crash data conversion.

Crash-Roadway Data Matching

Appendix A describes the methodology to convert crash location information from the current TS route number and milepost format to the inventory number and station format. This appendix B further describes how converted crash data can be matched with the roadway data.
The converted crash data should at least contain information on the inventory number, station (instead of the TS route number and milepost), and severity of the crashes. Figure L1 illustrates the desired crash data input in an Excel spreadsheet.

	A	B	C
1	INVENTORY	Station	Severit
2	016 10055 000000	0.1	A
3	016 10055 000000	0.2	A
4	016 10055 000000	0.3	A
5	016 10055 000000	0.4	A
6	016 10055 000000	0.5	A
7	016 10055 000000	0.5	K
8	016 10055 000000	0.11	K
9	016 10055 000000	0.12	K
10	016 10055 000000	0.15	K
11	016 10055 000000	0.16	K
12	016 10055 000000	0.16	K
13	016 10055 000000	0.3	PD
14	016 10055 000000	0.5	B
15	016 10055 000000	0.11	B

Figure L1: Sample crash data after conversion.

Figure L2 illustrates a sample of IRIS roadway database.

	A	B	C	D	E	F	G	H
1	OBJECTID_1	OBJECTID	INVENTORY	BEG_STA	END_STA	AADT	AADT_YR	ACC_CNTL
2	1	37730	016 10055 000000	0	0.17	171700	2005	0
3	2	37729	016 10055 000000	0.17	0.2	171700	2005	0
4	3	41518	016 10055 000000	0.2	0.24	177300	2005	0
5	4	53750	016 10055 000000	0.24	0.35	181200	2005	0
6	5	76494	016 10055 000000	0.35	0.52	158600	2005	0
7	6	76096	016 10055 000000	0.52	0.57	158600	2005	0
8	7	41586	016 10055 000000	0.57	0.75	146500	2005	0
9	8	41585	016 10055 000000	0.75	0.76	160600	2005	0
10	9	41585	016 10055 000000	0.76	0.77	160600	2005	0
11	10	41584	016 10055 000000	0.77	0.94	148100	2005	0
12	11	41587	016 10055 000000	0.94	1.03	162800	2005	0
13	12	41583	016 10055 000000	1.03	1.21	162800	2005	0
14	13	41581	016 10055 000000	1.21	1.26	151800	2005	0
15	14	42259	016 10055 000000	1.26	1.38	151800	2005	0
16	15	41580	016 10055 000000	1.38	1.5	151800	2005	0
17	16	41588	016 10055 000000	1.5	1.56	151800	2005	0
18	17	41535	016 10055 000000	1.56	1.68	151800	2005	0
19	18	41539	016 10055 000000	1.68	1.7	151800	2005	0
20	19	41538	016 10055 000000	1.7	1.76	151800	2005	0

Figure L2: Sample roadway data input (extracted from IRIS).

Once the crash and roadway data input are completely prepared, the crash data with the inventory number and station information is merged with the roadway data based on the inventory number, beginning station, and ending station of the roadway data.

First, we shall identify an inventory number of the first road segment and then searches for crashes with the same inventory number in the crash data so as to select crashes occurred on the first road segment. Second, crashes are matched to the road segment if the stations of selected crashes are located between beginning and ending station of the road segment. If a station of crash is exactly on the boundary between two consecutive road segments, the crash is allocated to the longer segment. If the length of two successive segments is same each other, the crash is matched to the former segment. Figure L3 illustrates the crash data matching algorithm.

Crash Dataset

Crash Number	Inventory	Station	Severity
1	1002004356	12.5	A
2	1002004356	18.3	B
3	1002004356	21.2	A
:	:	:	:
:	:	:	:

Roadway Dataset

Segment Number	Inventory	Beginning Station	Ending Station	AADT
1	1002004356	18.1	18.9	16000
2	1002004356	20.8	21.4	17500
3	1002004356	21.4	21.9	17700
:	:	:	:	:
:	:	:	:	:

Figure L3: Crash data matching algorithm.

APPENDIX L PEER GROUP DEFINITION FOR ROADWAY SEGMENTS AND INTERSECTIONS

Segments

Group 1: Rural Two-Lane Highway
if (Fc <= 60) and (Lns <= 2) and (Urban = 0) then classification = 1;

Group 2: Rural Multilane Undivided Highway
if (Fc = 30 or Fc = 40 or Fc = 50 or Fc = 55 or Fc = 60) and (Lns > 2) and (Med_typ = 0) and (Urban = 0) then classification = 2;

Group 3: Rural Multilane Divided Highway
if (Fc = 30 or Fc = 40 or Fc = 50 or Fc = 55 or Fc = 60) and (Lns > 2) and (Med_typ ne 0) and (Urban = 0) then classification = 3;

Group 4: Rural Freeway, 4 Lanes
if (Fc <= 20) and (Urban = 0) and (Lns = 4 or Lns = 3) then classification = 4;

Group 5: Rural Freeway, 6+ Lanes
if (Fc <= 20) and (Lns >= 5) and (Urban = 0) then classification = 5;

Group 6: Urban Two-Lane Highway
if (Fc <= 30 or Fc >= 70) and (Lns <= 2)and (Urban ne 0) and (Op_1_2_way = 2) then classification = 6;

Group 7: Urban One-Way Arterial
if (Fc <= 30 or Fc >= 70) and (Urban ne 0) and (Op_1_2_way = 1) then classification = 7;

Group 8: Urban Multilane Undivided Highway
if (Fc = 30 or Fc = 70 or Fc = 80 or Fc = 90) and (Lns > 2) and (Med_typ = 0) and (Urban ne 0) and (Op_1_2_way = 2) then classification = 8;

Group 9: Urban Multilane Divided Highway
if (Fc = 30 or Fc = 70 or Fc = 80 or Fc = 90) and (Lns > 2) and (Med_typ ne 0) and (Urban ne 0) and (Op_1_2_way = 2) then classification = 9;

Group 10: Urban Freeway, 4 Lanes
if (Fc <= 20) and (Lns = 3 or Lns = 4) and (Urban ne 0) and (Op_1_2_way = 2) then classification = 10;

Group 11: Urban Freeway, 6 Lanes
if (Fc <= 20) and (Lns = 5 or Lns = 6) and (Urban ne 0) and (Op_1_2_way = 2) then classification = 11;

Group 12: Urban Freeway, 8+ Lanes
if (Fc <= 20) and (Lns >= 7) and (Urban ne 0) and (Op_1_2_way = 2) then classification = 12

Note: If roadway segments cannot be fallen in one of the peer groups defined above due to lack of information, the road segments are categorized as “Peer Group 0”. Also, the software tool does not consider these road segments.

Intersections

Group 1: Rural Minor Leg Stop Control
```
if (Urban_Class = 0) and (Traf_Control = '1') then Classification=1;
if (Urban_Class = 0) and (Traf_Control = '3') then Classification=1;
if (Urban_Class = 0) and (Traf_Control = 'B') then Classification=1;
if (Urban_Class = 0) and (Traf_Control = 'A') then Classification=1;
```

Group 2: Rural All-Way Stop Control
```
if (Urban_Class = 0) and (Traf_Control = '2') then Classification=2;
if (Urban_Class = 0) and (Traf_Control = '4') then Classification=2;
```

Group 3: Rural Signalized Intersection
```
if (Urban_Class = 0) and (Traf_Control = '5') then Classification=3;
if (Urban_Class = 0) and (Traf_Control = '6') then Classification=3;
if (Urban_Class = 0) and (Traf_Control = '7') then Classification=3;
if (Urban_Class = 0) and (Traf_Control = '8') then Classification=3;
```

Group 4: Rural Undetermined
```
if (Urban_Class = 0) and (Traf_Control = '9') then Classification=4;
if (Urban_Class = 0) and (Traf_Control = '0') then Classification=4;
if (Urban_Class = 0) and (Traf_Control = 'N') then Classification=4;
if (Urban_Class = 0) and (Traf_Control = '') then Classification=4;
```

Group 5: Urban Minor Leg Stop Control
```
if (Urban_Class ne 0) and (Traf_Control = '1') then Classification=5;
if (Urban_Class ne 0) and (Traf_Control = '3') then Classification=5;
if (Urban_Class ne 0) and (Traf_Control = 'B') then Classification=5;
if (Urban_Class ne 0) and (Traf_Control = 'A') then Classification=5;
```

Group 6: Urban All-Way Stop Control
```
if (Urban_Class ne 0) and (Traf_Control = '2') then Classification=6;
if (Urban_Class ne 0) and (Traf_Control = '4') then Classification=6;
```

Group 7: Urban Signalized Intersection
```
if (Urban_Class ne 0) and (Traf_Control = '5') then Classification=7;
if (Urban_Class ne 0) and (Traf_Control = '6') then Classification=7;
if (Urban_Class ne 0) and (Traf_Control = '7') then Classification=7;
if (Urban_Class ne 0) and (Traf_Control = '8') then Classification=7;
```

Group 8: Urban Undetermined
```
if (Urban_Class ne 0) and (Traf_Control = '9') then Classification=8;
if (Urban_Class ne 0) and (Traf_Control = '0') then Classification=8;
if (Urban_Class ne 0) and (Traf_Control = 'N') then Classification=8;
if (Urban_Class ne 0) and (Traf_Control = '') then Classification=8;
```