Highway Safety Program Evaluation and Statistical Crash Table Development

Final Report

by

Joseph E. Hummer, Ph.D., P.E.
Associate Professor
Department of Civil Engineering, North Carolina State University
Raleigh, North Carolina 27695-7908
T: (919) 515-7733 E: hummer@eos.ncsu.edu

Rynal Stephenson
Research Assistant
Department of Civil Engineering, North Carolina State University
Raleigh, North Carolina 27695-7908
E: rgstephe@eos.ncsu.edu

Daniel Carter
Research Assistant
Department of Civil Engineering, North Carolina State University
Raleigh, North Carolina 27695-7908
E: dlcarter@eos.ncsu.edu

For the
Traffic Safety Systems Management Unit
North Carolina Department of Transportation

July 8, 2000

Technical Report Documentation Page

1. Report No. FHWA/NC/2002-021	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Highway Safety Program Evaluation and Statistical Crash Table Development		5. Report Date July 2000
		6. Performing Organization Code
7. Author(s) Joseph E. Hummer, Rynal Stephenson, Daniel Carter		8. Performing Organization Report No.
9. Performing Organization Name and Address Department of Civil Engineering North Carolina State University Campus Box 7908 Raleigh, NC 27695-7908		10. Work Unit No. (TRAIS)
		11. Contract or Grant No.
12. Sponsoring Agency Name and Address North Carolina Department of Transportation Research and Analysis Group 1 South Wilmington Street Raleigh, NC 27601		13. Type of Report and Period Covered FINAL January 2000 – June 2000
		14. Sponsoring Agency Code HWY-0850
15. Supplementary Notes		

16. Abstract

The objectives of this project were to assist the NCDOT in incorporating countermeasure evaluation and crash model applications into their safety management system. In particular, the project team recommended software that will allow the NCDOT to regularly perform statistically-valid evaluations of countermeasures that the NCDOT is installing, and wrote software to execute crash prediction models to judge the relative hazardousness of different highway sites. For each of the two areas of inquiry, the project team delivered a recommended or new piece of software to perform the necessary calculations, software documentation, instructions on how the NCDOT should obtain data and use the software, and a demonstration of the recommended or new software using NCDOT data.

17. Key Words Safety, crash, model, countermeasure, software	18. Distribution Statement		
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 78	22. Price

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

Disclaimer

The contents of this report reflect the views of the author(s) and not necessarily the views of the University. The author(s) 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 either the North Carolina Department of Transportation or the Federal Highway Administration at the time of publication. This report does not constitute a standard, specification, or regulation.

Acknowledgements

The authors would like to thank the North Carolina Department of Transportation and the Traffic Safety Systems Management Unit for their support of this project. The members of this unit provided training with the TEAAS software, account arrangement, and program installation at North Carolina State University.

The authors thank McPherson Engineered Solutions for providing the site locations and intersection data used in the countermeasure evaluation phase of the project. McPherson Engineered Solutions was also responsible for designing the database to contain traffic data and writing the crash prediction software.

Table of Contents

I. Introduction

A major goal of the North Carolina Department of Transportation (NCDOT) is to reduce the number and severity of traffic crashes on public roads within the State. To aid in that effort, the NCDOT has developed a Safety Management System (SMS) to plan, develop, establish, and continuously update systematic procedures and processes to ensure that highway issues and improvement opportunities are identified, evaluated, and implemented whenever and wherever appropriate.

Recently, the Traffic Safety Systems Management Unit (TSSMU), which leads the NCDOT's efforts in maintaining an SMS, identified two specific areas in which they could achieve immediate improvements in the SMS. These are areas where significant advances in the state of the art have occurred recently.

The first of these areas is a methodology for periodically evaluating the effectiveness of countermeasures being implemented. The TSSMU wants to be able to evaluate how many collisions, if any, were eliminated through installing a single type of countermeasure at a number of similar sites. Until recently, there was no feasible way for evaluators to adjust for selection bias inherent in this kind of highway safety study. Now, several statistical adjustments are available in the literature, which do not require random site selection or other practically impossible steps. Improved evaluations of previous countermeasures will allow NCDOT officials to optimize future countermeasure selections.

The second area that has seen recent advances is the capability to identify sites with abnormally high crash rates and patterns using existing crash prediction models. An intuitive notion concerning crash characteristics at highway intersection locations is that similar locations relative to configuration, signalization, and traffic volumes should, on average, have similar crash types, rates, and patterns. If crash characteristics for a particular intersection deviate from the normal or "expected" values for other similar locations, those crash characteristics are said to be abnormal. The NCDOT does not currently have a standard for statistically identifying abnormal crash types, frequencies, or patterns at specific highway intersections. Abnormal crash characteristics can be determined by various statistical analyses.

Two well-known techniques that have been used to identify abnormal crash patterns are cluster and expected value analysis. Cluster analysis has been widely used to identify crash types in safety studies. However, there are some problems in using cluster analysis because it depends upon the skills, knowledge, and judgments of the engineer conducting the analysis. This reliance on the technical subjectivity of the engineer may result in different study conclusions for different analysts using the same database. Expected value analysis is a statistically based method for identifying abnormal crash patterns. The expected value for a particular crash characteristic or type is the mean of the probability distributions for the characteristic or type. The probability distribution of a discrete random variable is a graph, table, or formula that specifies the probability associated with each value the random variable can assume. The difficulty with expected value analysis is the large amount of sites and data needed to find reliable mean values for each plausible crash characteristic or type.

Recently, the Federal Highway Administration (FHWA) and others have built statistical models of the relationship between traffic crashes and intersection geometric design, traffic control, and traffic volume variables for different types of roadway sites. To overcome the difficulties with cluster and expected value analysis, the TSSMU would like to use those models to judge which sites are performing abnormally and may need attention.

This project, using left-turn lane installation as a collision countermeasure for at-grade intersections, is contributing to a larger study entitled "Safety Evaluation of Intersection Design Improvements" currently being conducted by the Midwest Research Institute (MRI). The Principal Investigator of the project is Douglas Harwood at Midwest Research Institute. This is a pooled-fund study with participation from the following eight states: Iowa, Illinois, Louisiana, Minnesota, North Carolina, Nebraska, Oregon, and Virginia. Members of the research team include Penn State University along with Professor Ezra Hauer. The proposal was submitted to the Federal Highway Administration in response to Request for Proposal No. DTFH61-96-R-00055. The official project due date is September 30, 2000, but a time extension may be requested.

The objective of the MRI research is to determine the safety effectiveness of selected intersection design improvements or combination of improvements. After having a meeting with representatives from each state, it was decided to focus on the safety effectiveness of installing left-turn and right-turn lanes at intersections. Each state identified selected projects constructed between 1992 and 1996 in which at least one turn lane was installed as part of the project. Then a member of the research team visited the respective states to gather more information on the selected projects. For each intersection approach the following information was gathered: number of through lanes, number of left turn lanes, number of right turn lanes, posted speed, ADT, and type of traffic control. In addition, the number of left-turn and right-turn lanes was identified along with the intersection approach location of each. Any other improvements that were made to the intersection in conjunction with the construction were identified. From this information, the list of treated intersections was narrowed down to those in which only turn lanes were installed and maybe other minor improvements were made (signalization, sight distance improvements, etc). In addition, locations were cut from the list of intersections that had other improvements made at later dates or where other major modifications occurred (installation of another leg, construction of a shopping center, etc). The objective of site selection was to find "typical" intersections in which only left-turn lane or right-turn lanes were installed possibly with a change in the type of signal control.

For each treatment intersection, a comparison intersection was identified with similar geometric and traffic characteristics. For each treated intersection and comparison intersection it was necessary to gather more detailed information such as lane widths, horizontal alignment, vertical alignment, type of signing, ADT history, etc. These sites composed the prototype database required by Task 3 of the project proposal.

Participating in the pooled-fund study offered an opportunity to select and test software to evaluate countermeasure effectiveness of left-turn lane installation at intersections in North Carolina using data gathered as part of that study.

Objectives

The objectives of this project were to assist the TSSMU in incorporating the two advances described above into their SMS and their regular activities. In particular, the project team:

1. recommended software that will allow the TSSMU to regularly perform statistically valid evaluations of countermeasures that the NCDOT is installing, and
2. wrote software to execute crash prediction models to judge the relative hazardousness of different highway sites.

Scope

Currently, there are not many software programs available to evaluate countermeasures. Statistical software such as SAS can determine statistical significance of data, but does not account for the problem of regression-to-the mean present in many crash analysis studies. Several older safety analysis packages from the FHWA also have this same problem. Software used to evaluate countermeasure effectiveness should account for the regression-to-the mean bias and also be user-friendly.

As a test, the selected countermeasure evaluation software was applied to intersections throughout North Carolina to evaluate collision reduction due to left-turn lane installation. Left-turn lane installation was the only countermeasure evaluated using the selected software due to time required for site selection and data collection.

There have been attempts in the past to formulate crash prediction models, but no particular model has been widely accepted in the traffic safety field. For this project, software was developed to predict crashes based on the crash prediction models recently developed by Bauer and Harwood in an earlier research project sponsored by FHWA (Bauer and Harwood 1996 and 1998). The software developed may be used to predict crashes over a three-year period at intersections and interchanges in North Carolina.

Deliverables

Products of the study for the two areas of inquiry (countermeasure evaluation and executing crash prediction models) that the project team delivered to the TSSMU include two volumes and one software package. The three delivered products are:

- User's Guide - assistance with each software program used to complete the objectives,
- Final Report - documents the procedures and results of the project (this document), and
- Transportation Safety Evaluation Database System – software package with five program links and database forms, including
  BEATS software for countermeasure evaluation and
  TCPM software using models to predict crashes.

The User's Guide contains a description of the Transportation Safety Evaluation Database System package that includes the two recommended software programs for countermeasure evaluation (BEATS) and crash prediction (TCPM). These two recommended software programs and supplemental software programs used to accomplish the two objectives are discussed in detail and an example using each program is provided. The User's Guide contains information on the following software programs:

- TSEDS – database system
- BEATS – countermeasure evaluation
- TCPM – crash prediction
- TEAAS – crash data collection

This Final Report contains:

- Methodology for selecting and using the software
- Recommendations on how the TSSMU may use the software,
- Demonstration of the recommended or new software using NCDOT data.
- Description of expected value tables spreadsheet

The software written and designed specifically for this project (TSEDS and TCPM) in July of 2000 was delivered in a "first release" version. McPherson Engineered Solutions anticipates future upgraded releases of these programs.

Organization of this Report

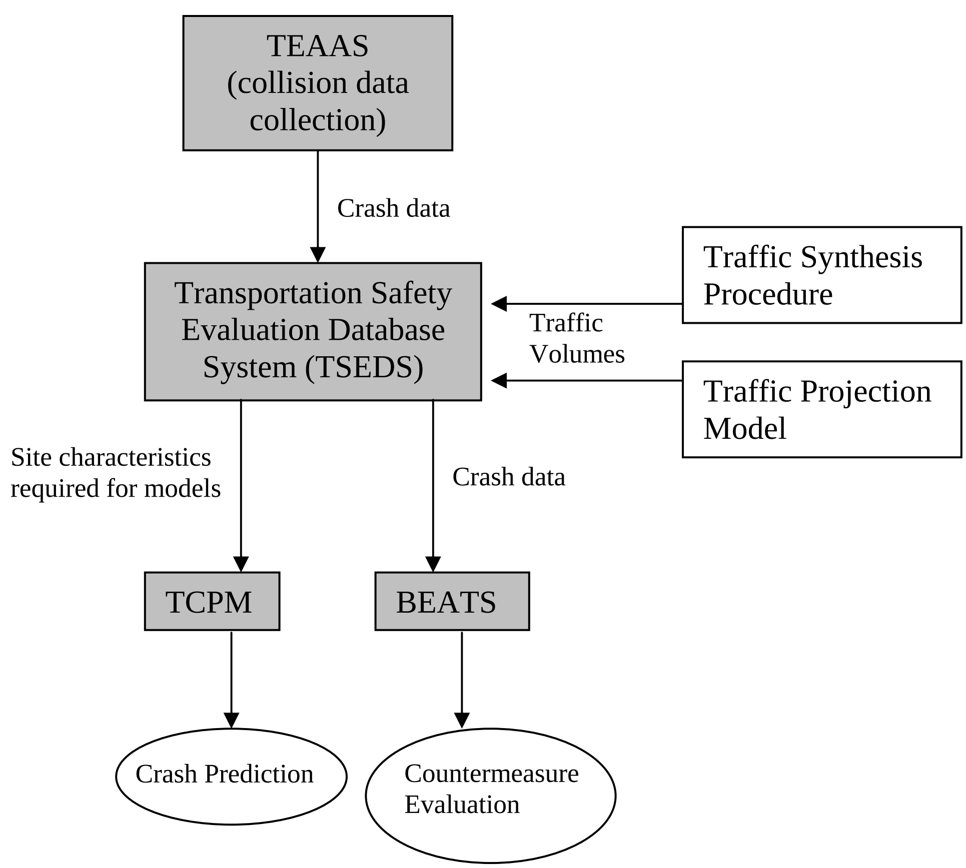
This report includes two major sections, and extensive Appendices. The first section is a description of the countermeasure evaluation process and the BEATS software that was recommended to perform this evaluation. The second section is a description of crash prediction models and the TCPM software that was designed to execute these models.

Following these two sections is a conclusions and recommendations section. This section contains a summary of the report, possible future modifications to the software, potential uses of the software, and information on model recalibration.

Program Relationships

This flowchart shows the intended flow of traffic data and the programs linked to the TSEDS database that collect or use that data.

NOTE: Programs in shaded boxes have instructions contained in the User's Guide.



II. Determining Countermeasure Effectiveness

Regression-to-the-mean is common in safety studies. In many cases, traffic engineers may select a site or group of sites that need safety improvements due to high crash frequencies. Countermeasures are then installed at these sites and the sites experience a reduction in crashes. Part of the reduction in crashes may be due to the countermeasure, but regression-to-the-mean may be responsible for the remaining reduction in crashes. This prevents the traffic engineer from obtaining a true evaluation of the countermeasure installed at the sites. Task 5 of the project proposal was to recommend a methodology to conduct a statistically valid evaluation. The BEATS software corrects the regression-to-the-mean bias in the data and produces a more accurate evaluation of the countermeasure effects than an uncorrected study produces.

Countermeasures are installed to reduce the number of crashes or potentially hazardous events occurring on the transportation infrastructure. Some common examples of types of countermeasures are warning signs, pavement markings, and stoplight installation.

Determining the effectiveness of countermeasures is important to traffic engineers responsible for providing safe transportation to the public while operating with a limited budget. Due to the limited funding it is important that countermeasures selected for installation provide the maximum benefit to highway users and engineers by reducing crashes. In some situations, expensive countermeasures are installed that yield very little reduction in crashes while inexpensive countermeasures may yield large reductions. Determining countermeasure effectiveness will help the traffic engineers to select the most appropriate countermeasure to maximize safety for the highway user.

Determining countermeasure effectiveness is part of the Safety Management System (SMS) to plan, develop, establish, and continuously update systematic procedures and processes to ensure that improvement opportunities are identified, evaluated, and implemented whenever and wherever appropriate. Major steps in a traditional safety improvement process include:

1. Identify hazardous sites,
2. Select countermeasure,
3. Countermeasure installation, and
4. Countermeasure evaluation.

Countermeasure evaluation may include other tasks in addition to determining countermeasure effectiveness. Some other factors in evaluating countermeasures may be ease of installation or implementation, user cost, installation cost, and availability of the countermeasure. All of these factors with countermeasure safety effectiveness would contribute to evaluating countermeasures.

For this project, countermeasure effectiveness is determined by the change in crashes from before the countermeasure was installed to after the countermeasure was installed. Currently, a before and after study is the usual method to determine countermeasure effectiveness. Using crash data as the measure of effectiveness typically requires three years of crash data before countermeasure installation and three years after countermeasure installation to produce meaningful results. In most cases there should also be some time allowed immediately after the countermeasure is installed in which no crash data are recorded to allow a transition period.

The ultimate result for traffic engineers would be a table of possible countermeasures to consider installing at identified high locations with a rating of effectiveness for each countermeasure. The countermeasure rating may contain multiple values such as a cost/benefit value and a collision reduction value. Although specific values may be calculated for each of these factors or ratings, it would be more appropriate to report effectiveness ratings as a range of values.

Current NCDOT Practice

The Traffic Engineering Branch of the NCDOT currently operates a safety program through the Traffic Safety Systems Management Unit. The Unit is made of four sections – a Safety Evaluation Section, a Highway Safety Improvement Program Section, Safety Information Management and Support Section, and a Highway Safety Section. The highway safety improvement program section identifies and ranks high hazard locations for bridges, intersections, and sections. High hazard locations must meet certain warrant criteria established by the NCDOT and found in the Highway Safety Improvement Program (HSIP).

The HSIP section of NCDOT works closely with the Safety Evaluation section to develop crash modification factors by analyzing the effects of installed countermeasures. The TSSMU accepts recommendations for countermeasures or improvements to high hazard locations from area traffic engineers or division engineers in the NCDOT.

BEATS Overview

The Bayesian Estimation of Accidents in Transportation Studies (BEATS) software program was selected for this project to determine countermeasure effectiveness. BEATS is a MS-DOS based program that will run on virtually any computer. Memory limitations are controlled by the size of the input file. BEATS may be used for the following purposes:

1. Estimation of effectiveness of highway safety treatment (countermeasures),
2. Identification and ranking of high crash locations, and

3. Combining safety treatment effects from multiple sources.

The software program applies the Empirical Bayesian (EBEST) methodology for purposes 1 and 2 above. The Meta-Analysis Combining Estimates of Safety Treatments (MACEST) method is used for purpose 3.

BEATS may be simply defined as a software program that uses Bayesian statistics to evaluate before-after study input data. For this project, BEATS was demonstrated by evaluating the effectiveness of left-turn lane installation at intersections in North Carolina. A before and after study design with comparison sites was selected to determine the countermeasure effectiveness. BEATS input included traffic volumes and crash frequencies for both the before treatment period and after treatment period.

BEATS could be used to rank the high crash locations of the input data, but the objective of our demonstration was to evaluate left-turn lane installation, which requires a study of the sample size as a whole. The ranking process in BEATS would only rank the high crash locations relative to the group of input sites, which was not very helpful in this project.

A link to the BEATS software is included on the switchboard screen of the Transportation Safety Evaluation Database System. The button is labeled "Regression-to-the-Mean Adjustment Procedure."

BEATS Literature Review

Task 1 in the project proposal was to review and critique Safety Evaluation Methodologies. However, literature on software used for countermeasure evaluation was limited. Several documents were reviewed containing information on the EBEST method and the BEATS program. The main resource for the BEATS program was a three-volume report sponsored by the FHWA (Pendleton 1991) entitled "Application of New Accident Analysis Methodologies."

Volume I of the report discusses the general EBEST methodology used in the software. The Empirical Bayesian Estimation of Safety in Transportation (EBEST) method models crashes as a non-normal distribution, accounts for the confounding effect of time on the treatment, and adjusts the data for regression-to the mean.

EBEST uses the odds ratio or cross-product ratio as the test statistic in a before-after with comparison group study design. The EBEST method uses the method of maximum likelihood to calculate the estimates and uses a measure of exposure to allow each site to be individually evaluated and weighted (Pendleton 1991, Volume I).

Two methods are used to estimate the treatment effect – the empirical Bayesian method (discussed above) and the frequentist method. The frequentist method uses only the treatment site data to estimate the treatment effect.

Volume II of the report is a users manual for the BEATS software program. Much of the important information in this document is summarized below. This document was written as a user's manual and should answer some questions about the operation of the software not discussed below.

Volume III presents the statistical theoretical developments in the empirical Bayesian methodology for estimating treatment effect on safety. The equations used in the empirical Bayesian method calculations are presented along with an explanation of the variables.

Dr. Olga Pendleton, author or co-author of each of the three-volume reports discussing BEATS, published a report entitled "A Systemwide Methodology for Evaluating Highway Safety Studies" that discusses the EBEST method (Pendleton 1992). A portion of Chapter 1 of the report explains regression-to the mean, the EBEST method and the differences between the EBEST method and the previously used empirical Bayes approach. Chapter 2 applies presents the three stages of a study, which shows how the EBEST method is applied to traffic data.

Chapter 12 of Ezra Hauer's book *Observational Before and After Studies* discusses accident models. The chapter is a theoretical discussion of accident models and the methodology used to develop the models. Hauer also discusses the maximum likelihood estimation and the Empirical Bayes approach, which is used to estimate the accident frequency at a location for any year using the entire accident history of the location and adjust the accident frequency to remove regression-to the mean bias (Hauer 1997).

Program Data

Program Input

Task 2 in the project proposal required that an analysis be done of the data required for safety evaluation. The following section details the data that BEATS requires and how it should be entered.
BEATS requires a separate text file for input. The input file should be in a table format with data columns separated by tabs (Figure 1). From left to right, the columns should correspond to the following headings:

Identification Number or Site Number – a numeric variable 5 digits or less. The identification number or site number may only contain numbers.

Treatment Period – 1 before treatment, 2 after treatment. NOTE: Even though the reference sites did not receive treatment, entering '2' for a reference site refers to data collected after the treatment was installed in the treatment sites.

Group Type – 1 if treatment, 2 if reference group, 3 if comparison group. A site is labeled as a treatment site if the countermeasure has been recently installed and is being evaluated.

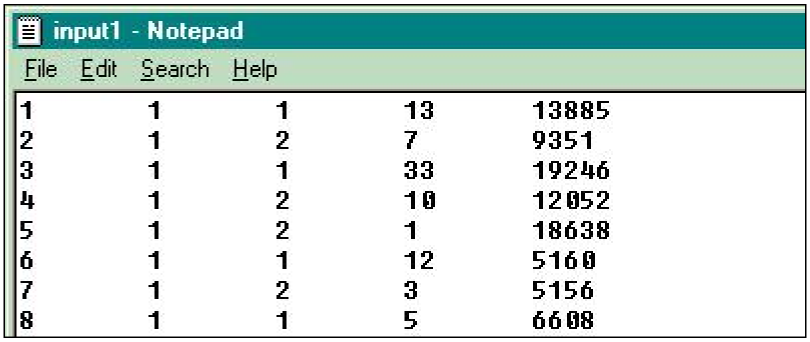
The reference group is the population of all treatable sites. It includes the selected treatment sites and all other sites that could have been selected as treatment sites. It is not required for the reference site to have similar traffic or geometric characteristics as any treatment site.

The comparison group is the population of treatable sites with similar traffic and geometric characteristics to the treatment sites. The comparison group is a small group that is included in the reference group, but a comparison site must have similar traffic and geometric characteristics to a treatment site.

The EBEST method of calculating output requires reference groups. In the evaluation of left-turn lanes in this project, sites selected as comparison sites for data collection purposes were labeled as reference sites in the input file for BEATS. This was done as an example, because comparison sites could not be used for this method, only reference sites. The label "reference" or "comparison" determines whether the user inputs a 2 or 3, and determines the manner in which BEATS calculates its statistics.

Number of crashes – crash frequency
The crash frequency may be reported over any time period, provided the exposure is calculated over the same time period. In this project, the crash frequency was recorded for 3, 4 or 5 years for each intersection location.

Exposure – exposure of site; may only be one number
The most common form of exposure is AADT volume. The exposure units may be scaled to any desired dimensions such as AADT/10,000 miles, etc. For our demonstration involving intersections, the ADT volumes of both roads were summed to have an intersection ADT. To account for differing number of years of crash data collection, the ADT for the intersection was summed over the same number of years. For example, the first row in Figure 1 would represent Site # 1, before treatment was completed, treatment group, crash frequency of 13 from 1/1/1990 to 12/31/1992 (3 years), and an exposure of 13885. The exposure is the sum of intersection ADT volumes for 1990, 1991, and 1992.



input1 - Notepad

File Edit Search Help

1	1	1	13	13885
2	1	2	7	9351
3	1	1	33	19246
4	1	2	10	12052
5	1	2	1	18638
6	1	1	12	5160
7	1	2	3	5156
8	1	1	5	6608

Figure 1. Example of BEATS input text file.

The input file used in the evaluation of left-turn lane installation is shown in Appendix B. There was a total of 30 treatment sites and 26 reference sites. Crash data could not be obtained for four of the 30 comparison sites identified originally by MRI; therefore, the four sites were omitted. All treatment sites were included in the input to use all available data in an effort to most accurately predict the left-turn lane installation effect. It is good practice to have at least as many reference sites as treatment sites, though not necessary to have an equal number of each.

Two methods were considered in the calculation of the exposure number. Only one number could be used for each intersection site. One method was to add the ADT's of both intersecting roads as the exposure for the intersection. The second method was to multiply the two ADT's of the intersecting roads to produce the potential conflicts at the intersection. The method of summing both ADT volumes was used in this project. The multiplicative method seemed less appropriate for rural locations where the ADT volumes were relatively low. At these locations, there is less chance that these potential conflicts will occur due to the low ADT volumes.

Program Output

Using the input data, BEATS generates an output that includes:

- A narrative summary and interpretation of the results,
- Descriptive statistics about the study data,
- Data listings and rankings as requested by the user, and
- Statistical details of the parameter estimates and test statistics which is the basis for the narrative.

The narrative BEATS output states:

> Three methods were used to estimate the treatment effect: the EBEST method using the comparison group (if available), the frequentist or cross product ratio, and the EBEST method using the reference group in place of the comparison group, if available.

The frequentist method does not adjust for regression-to-the-mean. This method assumes the total number of accidents in the treatment group after treatment is equal to the total number of accidents in the treatment group before treatment. The method then reports any difference in the treatment groups before and after treatment and tests for statistical significance.

Both EBEST methods adjust for regression-to-the-mean in the data.

A sample BEATS output is included in Appendix C of the User's Guide. Page 4 of the BEATS output contains summary statistics and results using the methods described above. The bottom row of numbers contains the summary treatment effect and test statistics used to evaluate the treatment. The "trteEBr" value and "trtefr" value are the percent changes in total accidents in the treatment group due to the treatment using the EBEST method and frequentist method, respectively.

Selecting the appropriate method's results to report depends on several factors. One indication is the B_{avg} value shown on page 4 of the output. If this number is close to 1.00, regression-to-the-mean was present in the data and one of the EBEST methods should be used. If this number is close to 0.00, then regression-to-the-mean is not high in the data and the frequentist method may be used. Another factor to consider is that the frequentist method may be easier to explain to decision-makers and the public.

Problems and Limitations

The BEATS software is an MS-DOS based program that may appear ancient compared to today's Windows operating system environments. BEATS does not require many steps to produce an output, but the steps may be confusing to beginning users. Below are some of the problems experienced with the BEATS software during our demonstration on the left-turn lane countermeasure.

- BEATS will not function correctly if there are blank spaces in the input file.
- BEATS does not offer any suggestions or information on how to organize the data used in the program. The user must decide which years of data to use and the appropriate procedure for collecting and inputting the data into the BEATS software. BEATS is a program that may be used to evaluate countermeasure effectiveness, but the results are only as meaningful as the input data.
- Using BEATS to evaluate countermeasures requires a good amount of data. Each treatment site should be accompanied by at least one comparison site and a reasonable sample size (>10 treatment sites) should be used.
- BEATS offers no guide as to the best way to enter exposure data. A user may sum multiple-year ADT, multiply major and minor road ADT, etc.
- Crash reports may not contain a valid milepost. These reports are represented in the Fiche output of TEAAS as having a milepost value of 999.999 as seen in Figure 7 in

the Users Guide. This was not a great concern in this demonstration because study sites were intersections and could be identified by the two road names. However, an invalid milepost location would present a problem if two intersections had the same intersecting road names such as in situations with loop roads. One loop road situation was present in this study, but the crash locations were identified using a road present at one intersection and not the other. One intersection had three approaches and the other had four approaches where the fourth approach had a different road name.

Demonstration of BEATS

As stated earlier, BEATS was demonstrated on an evaluation of the left-turn lane installation countermeasure at intersections across North Carolina. Task 4 of the project proposal was to analyze a prototype database. The following sections describe the process of evaluating crash reduction at intersections due to left-turn lane installation using BEATS. BEATS was only used for statistical analysis of the data to determine if there was a significant reduction in crashes at treatment sites.

A before-after study was selected as the study design. This study design may also be used in BEATS to evaluate countermeasure effectiveness.

Site Location and Countermeasure Selection

As part of the pooled-fund study mentioned in the Introduction, left-turn lane installation was selected as a countermeasure to evaluate in North Carolina. This project uses the BEATS software to evaluate the countermeasure effectiveness at intersections. Thirty-three intersections across North Carolina were originally selected where a left-turn lane had been installed between early 1994 and late 1996. This allowed a minimum three-year period before installation and three-year period after installation for collision data collection.

One comparison site location was selected for each treatment site, with the comparison sites having similar geometrical design, traffic volumes, and intersection characteristics. Many comparison sites selected were located near their corresponding treatment site.

Thirty treatment sites and 26 reference sites were included in the input file used to generate the output and determine countermeasure effectiveness. The remaining 3 original treatment sites and 7 original reference sites were not included in the final input file due to missing collision data or incomplete site information. Twenty-six treatment sites had corresponding reference sites in the final input file. The remaining comparison sites did not have corresponding treatment sites but were included to provide as large of data sample as possible. The BEATS input file is shown in Appendix B and a complete table including all data for these locations are shown in Appendix A.

Data Collection

Data collection was moderately labor intensive due to the lack of a centralized database containing the required data. For each treatment site and comparison site intersection location, data were collected for major road ADT, minor road ADT, before treatment crashes, and after treatment crashes. It was also necessary to determine the year in which the left-turn lane was installed at the treatment sites. The traffic data and crash data collected and used in the BEATS input is shown in Appendix A.

Traffic Volumes

Traffic volumes were collected from traffic count maps maintained by the GIS Unit of NCDOT and available on the NCDOT website. Any additional years of traffic volumes for a particular location were estimated using a linear traffic projection model. A base year traffic volume obtained from NCDOT and a linear growth rate of 2 to 4 percent was used in the traffic projection model. This traffic projection program is included in the software delivered with this project. If the growth rate is unknown, a value of 3 percent may be used for the traffic volume estimation. The assumption that traffic volumes change linearly was valid for this project due to the limited time range of the required traffic data. Traffic data from 1990 to 1999 were used in the left-turn lane evaluation.

If traffic volumes were not available for the intersection roads, The Highway Emulator (THE) was used to calculate the volumes (Bromage, 1988). THE projected the volumes by building a highway network containing the desired roads and other roads with known traffic volumes. Traffic volumes that were available near the site from the NCDOT were input into the network and used to estimate traffic volumes on other links in the network.

Crash Data

For the demonstration, the project team used TEAAS to collect crash data for each selected intersection during both the before treatment period and after treatment period. At the time of this project, the TEAAS program contained crash data from January 1, 1990 to December 31, 1999.

Crash data were collected for different time periods for each intersection. For every treatment and comparison site, the before treatment crash data were collected starting on 1/1/1990 and extending until the year in which the treatment was installed. For every treatment and comparison site, the after treatment crash data were collected starting on January 1 of the year after the treatment was installed and continued until 12/31/1999. The year that treatment was completed varied from 1994 to 1996 which caused some sites to have more than three years of crash data in one of the time periods.

The intersection boundaries used in the crash data collection extended to 150 feet on each approach away from the intersection.

Input

The project team created a BEATS input file after volume and crash data were collected. A table was established using MS-EXCEL to allow easy data sorting and organization. The input data table had five columns with the headings and information as described above. When the input file was complete, only the numerical data were selected and copied into a text editor. The final input file was saved in the text editor and used in the BEATS program. Appendix B shows a copy of the input file.

Column 1 of the input file is a site identification number assigned to each entry. Site identification numbers were assigned beginning with 1 and increasing to 56 down the column to sites in the before treatment period. Each site corresponding to the before treatment period was assigned a number from 1 to 56. The sites in the after treatment period were assigned numbers from 1 to 56 also.

BEATS requires reference group data to use the EBEST method. Therefore, the non-treated intersections were labeled as reference sites in the input file. These sites are indicated by a number ‘2’ in the third column of the data entry. There were no sites labeled as a comparison site in the BEATS input file. The sites labeled as reference sites in the BEATS input had similar traffic and geometric characteristics to the treatment sites. There was an equal number of before sites and after sites (56), but different numbers of treatment sites (30) and reference sites (26).

Column 5 of the BEATS input is the exposure for the intersection. The intersection exposure value was calculated for each required year of data. If data were available, each treatment and comparison intersection had two rows in the input file – one for the before treatment period and one for the after treatment period. The number of years of available data varied at each intersection. Therefore, it was important that the crash data and exposure data were reported for the same time period. The crash frequency was reported for the before treatment time period from 1/1/1990 to the last day of the year prior to treatment completion. The crash frequency for the after treatment period was reported from the beginning of the year after treatment was completed to 12/31/1999.

The exposure data were summed over the corresponding number years of crash data. For example, if crash data were recorded from 1/1/1990 to 12/31/1993, the four annual ADT volumes for the intersection over the four years was added to give the exposure.

For each intersection (treatment and comparison) data from the year that treatment was completed was omitted from the input file. This “warm-up” period allowed drivers an opportunity to become familiar with the new geometric configuration and traffic operations at the treated intersections.

Output

In the project to evaluate left-turn lane installation, BEATS used two methods in the statistical output calculations - the frequentist method and the EBEST method using reference sites. Appendix C contains BEATS output for this demonstration. Both the frequentist method and EBEST method with reference group showed a statistically significant reduction in crashes after the NCDOT installed left-turn lanes at the treatment sites. According to BEATS there was a 52.1 percent reduction in crashes using the frequentist method and a 50.7 percent reduction in crashes using the EBEST method with reference group. Since the treatment sites averaged 16 crashes per year in the before period, a 50.7 percent reduction means a savings of about 8 crashes per intersection per year at those sites. The two methods produced similar results due to the relatively small amount of regression-to-the-mean in the data. The B_{avg} value for the data group was 0.18, which indicates that very little regression-to-the mean was present in the data.

Potential Use of BEATS in NCDOT

BEATS allows the traffic engineer to quickly evaluate countermeasures with given data. The data collection required for input is the labor-intensive phase of this countermeasure evaluation process. If crashes are the measure of effectiveness, then approximately threes years of crash data should be available for both the before and after periods in the study. The time period required for an adequate sample size of crash data is the bottleneck in evaluating countermeasures.

A possible application of the BEATS software for NCDOT is to follow a similar procedure as described in the evaluation of left-turn lane installation at intersections to evaluate other countermeasures. A particular countermeasure may be selected for evaluation and determination of crash reduction. Locations across the state may be randomly selected where the particular countermeasure was installed between 1993 and 1996. These locations would be the treatment sites. This time period allows a minimum of three years of crash data in the before treatment period and after treatment period. The TEAAS software developed for NCDOT contains crash data beginning in 1990. Comparison sites would also be selected in a similar manner as in the left-turn lane evaluation. At these locations, the countermeasure should have been in installed for a long period of time and the site characteristics should be as similar as possible to the treatment site. Data collection for countermeasure evaluation would consist of crash frequencies and exposure over the specific time periods.

ADT volumes may be used for exposure and may be obtained from the Traffic Surveys Unit of NCDOT or may be calculated using The Highway Emulator. An ADT volume would be required for each year of the before treatment period and after treatment period. ADT volumes may be calculated for any year using the Linear Traffic Projection model submitted with this report given the measured ADT volume and year.

Discussion

According to the results of the study, left-turn lane installation during 1994-1996 significantly reduced crashes at some intersections in North Carolina. If the NCDOT could find more intersections like those, it should expect to achieve similar collision savings due to left-turn lane installation. Regression-to-the-mean bias was miniscule in the data, which suggests the sites were not selected after a spike in collision frequency.

In this project, only crash frequency was used in BEATS to evaluate crash reduction due to left-turn lane installation. BEATS is used for statistical analysis and therefore any number representing crash data (any type of crash) at the site may be used in column four of the input file. For example, the NCDOT currently uses a combination of crash severity, crash frequency, and collision type to determine a crash index for use in identifying high hazard locations. This number would be sufficient for use in the BEATS input file, provided that all entries followed this procedure. The disadvantage to this method is the additional time required to complete the calculation of an index or number to represent the crash data for each site included in the BEATS input file.

Countermeasure effectiveness should include other factors in addition to crash frequency reduction. Some other factors for determining countermeasure effectiveness are cost/benefit analysis, installation cost, and crash severity.

Advantages of BEATS include correction of regression-to-the-mean bias in data, quick software installation, fast running time, little training time, and understandable results. To the authors' knowledge, BEATS is currently the most efficient software program for statistically evaluating the reduction of crashes due to countermeasure installation and correcting regression-to-the-mean bias. FHWA efforts to produce other software for this task have not borne fruit to this point.

BEATS quickly calculates the statistical data used to determine countermeasure effectiveness. The time consuming phase of countermeasure evaluation is site selection and data collection. One reason for the success of this demonstration is likely the painstaking care used by the MRI team to select the sites. The MRI team consulted frequently with NCDOT headquarters and division staffs and made site visits before selecting the sites. The inventory and GIS data available to NCDOT personnel are not sufficient to select sites without field visits and many consultations at this point. Unless NCDOT procedures change (to require identification of a comparison site at the time a treatment is funded, for example) or the quality of inventory and/or GIS data improves dramatically, site identification will remain cumbersome and time consuming.

One disadvantage of BEATS is the inconvenience caused by creating a text only input file. This makes data organization more difficult and could cause errors in data entry. It is recommended that an MS-EXCEL spreadsheet be used to store data and then either copy the required data to a text file or create only the required data in a spreadsheet and save the spreadsheet as a text file.

III. Determining Relative Hazardousness of Sites

In the pursuit of safer highways, the identification of hazardous sites is an undisputed necessity. With a finite budget, but an infinite number of places to spend this money, a prioritization system of some type is needed. Current methods of identifying hazardous sites can be divided into crash-based and non-crash-based methods. Crash-based methods include Frequency, Rate, Frequency-Rate, Rate Quality Control, and Severity. Non-crash based methods include Hazardous Roadway Features Inventory, Public Service Requests, and reports by the agency or enforcement officials (Parker 1991).

The current North Carolina Department of Transportation (NCDOT) method of identifying hazardous sites is detailed in the Spring 2000 issue of the Highway Safety Improvement Program (HSIP) report (Braam 2000). This method prioritizes intersections, sections, and bridges based on a weighting factor. This weighting factor is made up of several smaller factors for individual warrants designed to take into account different types of crashes or events (i.e. frontal impact, run off road crashes).

The method of hazardous site identification presented in this report uses crash prediction models formed by Bauer and Harwood in FHWA-RD-97-106 and FHWA-RD-96-125 (Bauer and Harwood 1996 and 1998) to give an expected number of crashes over a three-year period for a particular intersection, interchange, or rail crossing. This predicted number of crashes then allows the engineer to see if the site deviates from the safety "expected" of a site in its category. Deviant sites may be easier to treat with cost-effective countermeasures than sites performing as expected.

This crash prediction model method differs from the current HSIP method in several ways:

- **Basis –** The current method is based on the recorded number and type of crashes. The crash prediction method is based on models that input site characteristics such as traffic volumes, number of lanes, and roadway channelization and output a predicted number of crashes.
- **Comparison –** Models provide an explicit standard against which to compare each site, while the current method has no such standard.
- **Time –** Evaluation of a particular site using the current method requires time to collect sufficient crash data. The results from the prediction models are available immediately assuming necessary data is available; however, crashes to compare these predictions against will take time to collect.
- **Safety –** The current method evaluation is based on past crash records. The crash prediction method compares actual crashes against an expected future number of crashes for a particular site.
- **Calibration –** The models were calibrated in California and Washington, and one must assume in using them that conditions in North Carolina are similar to conditions in these states.

Crash Prediction Software Overview

The crash prediction software developed for this project was named Transportation Crash Prediction Models (TCPM). This software was developed to fulfill Task 8 in the project proposal. This program is designed as an interface for the crash prediction models. It is a Windows-based program developed by McPherson Engineered Solutions for use in the NCDOT Traffic Safety Systems Management Unit.

TCPM requires the following system components:

- Windows 95/98/NT
- 1 Mb disk space available
- 4 Mb RAM

Of the models reported in the Bauer and Harwood documents, the software contains the interfaces for five intersection models and eight interchange models. The software also contains three railway-crossing models. Upon selection of a model, the software provides the user an input screen with a listing of the characteristics needed for the particular model. The following models are programmed into the software:

Intersections (total multiple-vehicle accidents)

1. Rural, four-leg, STOP-controlled intersection,
2. Rural, three-leg, STOP-controlled intersection,
3. Urban, four-leg, STOP-controlled intersection,
4. Urban, three-leg, STOP-controlled intersection, and
5. Urban, four-leg, signalized intersection.

Interchanges (total accidents of all types)

1. Ramp proper segments,
2. Ramp proper segments on off-ramps (rear-end crashes excluded),
3. Entire ramps,
4. Entire off-ramps (rear-end crashes excluded),
5. Acceleration lanes,
6. Deceleration lanes,
7. Entire ramp plus adjacent speed change lanes (10% significance level), and
8. Entire ramp plus adjacent speed change lanes (20% significance level).

Railway crossings

1. Passive warning device,
2. Light warning device, and
3. Gate warning device.

Each site characteristic that the user inputs becomes a variable (X_x) in the model. The program contains all the model coefficients (β_x) and provides an on-screen crash prediction. The crash prediction calculation follows the format below:

$$Y = \exp(\beta_0)(ADT_{\text{major road}})^{\beta 1}(ADT_{\text{crossroad}})^{\beta 2} \exp(\beta_3 X_{i3}) \exp(\beta_4 X_{i4}) \bullet \ldots \bullet \exp(\beta_q X_{iq})$$

Where Y = Prediction of multiple-vehicle crashes for a three-year period
X = Input variables
β = Model regression coefficients

Model Literature Review

A literature review was conducted to find information on crash prediction models that may be selected for further study and used to develop crash prediction software. This followed Task 6 in the project proposal, which required a review and critique of crash prediction models. The two main resources used in the literature search were the TRIS database and literature reviews in previously published reports. Information concerning accident models was available; however, literature concerning intersection or interchange models was scarce. Most of the literature discussed models for two-lane highways and variations of accident models for roadway segments.

The Transportation Research Information Services (TRIS) database was queried to search for documentation pertaining to collision models. The TRIS Database is the world's largest and most comprehensive bibliographic resource on transportation information. TRIS is produced and maintained by the Transportation Research Board. The search produced approximately 40 titles that seemed to warrant further inspection.

Only five titles concerning accident models were selected from the TRIS search for mention in this literature review. Among the relevant literature returned by the search were two reports by Lau that developed accident prediction models for signalized intersections and unsignalized intersections. The first report (Lau 1989) is entitled "Accident Prediction Model Development for Unsignalized Intersections." This report discusses the methodology of developing injury, PDO, and fatal accident models for unsignalized intersections based on the Traffic Accident Surveillance and Analysis System (TASAS) in California. The significant variables for injury and PDO accidents were: traffic intensity, proportion of cross-street traffic, intersection type, control type, number of lanes, left-turn arrangements, traffic flow arrangements, and environmental locations. The significant variables in the fatal accident model include traffic intensity, percentage of cross-street traffic, and design speed.

Lau also authored a report entitled "Accident Prediction Model Development: Signalized Intersections" (Lau 1988). This report follows a similar process as the above-mentioned report. Lau developed injury, PDO, and fatal accident models for signalized intersections based on the TASAS in California. Significant variables in the injury and PDO models were traffic intensity, proportion of cross-street traffic, intersection type, signal type, number of lanes, and left-turn arrangements. In the fatal accidents model the significant variables were traffic intensity, intersection type, and design speed.

Although the methodologies differ, some of the same variables found significant in Lau's models were significant in the Bauer and Harwood models for signalized and unsignalized intersections. Design speed, number of lanes, left-turn arrangements, and traffic intensity were variables that were significant in both Lau's model and Bauer and Harwood's models for unsignalized intersections. Intersection type, control type, and traffic flow arrangements were not independent variables in Bauer and Harwood's models, but were accounted for in the model development.

In both models for signalized intersections, traffic intensity, number of lanes, left-turn arrangements and design speed were significant independent variables. The other factors such as intersection type and signal type were accounted for in the Bauer and Harwood signalized intersection models.

Bared and Vogt (1998) have recently developed accident models for segments and intersections on rural, two-lane roads based on data collected from Washington and Minnesota. The segments are on two-lane roads and the intersections are two-lane roads with three-leg and four-leg approaches and stop-controlled minor roads. Variables collected include accident counts, traffic exposure, surface and shoulder width, Roadside Hazard Rating, number of driveways, channelization, horizontal and vertical alignments, intersection angles, speed limits, and commercial traffic percentage. The most significant variables were exposure and traffic counts, but surface width, shoulder width, roadside conditions, and alignments were also significant.

Earlier research by Bared and Vogt developed a model to predict collision frequency in a five-year period on highway segments with an ADT of 5,000 or less. The equation of the model is:

$$AC = (L)\exp[-5.2513+1.0794\log(ADT)-0.0774(TW)-0.0809(SW)+0.0457(RHR)+0.0061(DD)+0.0355(H)+0.0275(V)]$$

where:
L = segment length, in miles,
ADT = average daily traffic on segment,
TW = travel lane width, in feet,
SW = shoulder width, in feet,
RHR = roadside hazard rating,
DD = driveway density, in driveways per mile,
H = horizontal curve index, and
V = vertical curve index.

The R^2 for the equation is 0.65. The H and V curve indices are calculated using simple formulas and the roadside hazard rating is a qualitative value from 1 to 7. The variables in the model are usually available from a state DOT database (Bared and Vogt 1997).

Dzbik and Persaud have developed accident prediction models for freeways. Macroscopic and microscopic data are used to develop generalized linear models to estimate a freeway section's accident potential. Empirical Bayesian procedures are used to refine the accident estimates (Dzbik and Persaud 1993).

A study by Kalokota and Seneviratne used geometric design variables to model accidents on two-lane rural highways. Data were collected for selected geometric variables on highways in northern Utah. It was shown that exposure in terms of distance traveled was the most significant variable. Horizontal curvature and cross-section were found to have very little effect on accident frequency in the two-lane models (Kalokota and Seneviratne 1994).

Several research reports were reviewed in an attempt to find crash prediction models that may be applied to North Carolina. Relevant literature discussing highway crash prediction models was limited. The researchers were looking for models that were easy to use, contained relevant and pertinent variables, and produced meaningful results.

A brief review of several previously developed crash prediction models was documented in a report by Hummer et al. Previous research has produced crash prediction models for a variety of highway elements such as horizontal curves, vertical curves, bridges, and two-lane sections (Hummer et al. 1999).

As of this writing, the best available model for predicting crash rates on or near bridges in rural areas was developed by Turner (1984). The model used for calculating the crash rates is:

$$Y = 0.4949 - 0.0612(RW) + 0.0022(RW)^2$$

where Y is the predicted number of collisions per million vehicle miles, and
RW is the bridge width minus the approach roadway width, in feet.

This model was based on 2,849 bridge crashes during a four-year period on rural, two-lane highways in Texas. The model's goodness-of-fit value for R^2 is 0.81, which is relatively high for safety studies.

One model developed by Hadi et al. (1995) attempts to predict total crash frequency for a four-year period on two-lane, rural mid-block segments. The model is based on four years of collision data from Florida. The model assumed a Poisson distribution and used Negative Binomial regression. The equation for the model is:

$$N = \exp[-10.26+0.8249(Llen)+0.8783(Ladt)-0.0857(Lw)-0.0130(Sp)+0.0589(Is)-0.0150(Ts)]$$

where
Llen = log(1,000 x section length in miles),
Ladt = log(ADT),
Lw = lane width, in feet,
Sp = posted speed limit, in mph,
Is = number of intersections, and
Ts = total shoulder width, in feet.

The lack of any reported goodness-of-fit measures prevents the potential user from knowing the accuracy of the model and justifying the use of this model.

Some of the same independent variables are used in this model as in the Bauer and Harwood models used for predicting crash frequencies at intersections. Just like in the Bauer and Harwood intersection models, the most sensitive variable is the ADT volume.

Various models have been developed for highway segments. A model developed by Zegeer predicts collisions on horizontal curves of highway segments (Zegeer 1991). Glennon et a1. developed an equation to predict collision frequency on highway segments accounting for horizontal curves (Glennon et al. 1985). Neuman's model predicts annual collision frequency on highway segments with vertical curves (Neuman et al. 1984).

To the authors' knowledge, none of these models are currently widely used for collision predictions in the highway safety field. Some criticisms or reasons for lack of acceptance associated with these models are low goodness-of-fit measures, incorrect data distribution assumptions, lack of data availability for independent variables, and lack of software to implement the models.

The software developed during this project implements crash prediction models for at-grade intersections and interchange and speed change lanes (Bauer and Harwood 1996 and 1998). A detailed summary of both reports follows. The intersection model developed in the report and used for this project predicts only multiple-vehicle collisions. Bauer and Harwood have since written an addendum to the intersection report that includes models based on all collision types including both multiple-vehicle and single-vehicle collisions.

The paragraphs below summarize the Bauer and Harwood crash prediction models for each intersection or interchange type in the report. The at-grade intersection models and interchange and speed change lane models developed by Bauer and Harwood and discussed below are the models used in the crash prediction software.

Statistical Models of At-Grade Intersection Accidents

Bauer and Harwood used data from the Caltrans database to develop models of the relationship between traffic crashes and roadway geometric elements for five types of at-grade intersections. Equations were developed for total multiple-vehicle accidents and fatal and injury multiple-vehicle accidents for each of five intersection types.

North Carolina compares reasonably well to California in terms of topography, land use development, and traffic characteristics. Although California is approximately 3 times larger than North Carolina by land area and 4 times larger by population, the topography of both states includes coastal regions, mountainous regions and flat or level regions. However, Table 1 shows that North Carolina has different highway safety patterns from California, particularly for fatal collisions.

Table 1. Crash Statistics for California, North Carolina, and the USA

State	CA	NC	USA
1997 Licensed Drivers (Thousands)	20,385	5,399	182,709
Fatalities Per 100,000 Drivers	18.09	27.47	22.99
1997 Registered Vehicles (Thousands)	25,399	5,856	203,568
Fatalities per 100,000 Registered Vehicles	14.52	25.32	20.64
1997 Population (Thousands)	32,182	7,431	267,744
Fatalities per 100,000 Population	11.46	19.96	15.69
1997 Total Killed	3,688	1,483	42,013

Applicability to NCDOT

One major limitation of applying the models to NCDOT is data collection and organization. To the authors' knowledge, NCDOT does not record data for some of the input variables required by the crash prediction models. If these data are stored by NCDOT, they is not stored in a central location, which makes it difficult to acquire all data necessary for the crash prediction models.

Data Collection

The California Department of Transportation (Caltrans) database was used to gather all relevant variables pertaining to geometric design, traffic control, and traffic volume. Table 1 below shows the complete list of variables. Crash information variables were derived from the Caltrans Accident File. Collision data from 1990, 1991, and 1992 were used.

Table 2. Variables Available in the Existing Caltrans Data Base

Geometric Design
- Intersection configuration (i.e., three-leg, four-leg, multileg,etc.)
- Number of lanes on major road
- Number of lanes on crossroad
- Presence of median on major road (i.e., divided / undivided)
- Median width on major road
- Average lane width on major road
- Shoulder width on major road
- Design speed of major road
- Functional classification of major road
- Presence of left-turn channelization on major road (i.e., separate left-turn lane)
- Presence of left-turn channelization on crossroad (i.e., separate left-turn lane)
- Presence of right-turn channelization on major road (i.e., separate roadway for free right turns)
- Presence of right-turn channelization on crossroad (i.e., separate roadway for free right turns)
- Presence of access control on major road (none / partial)

Traffic Control
- Type of intersection traffic control (STOP sign, traffic signal, etc.)
- One-way vs. two-way operation on major road
- Left-turn prohibition from major road
- Left-turn prohibition from crossroad
- Presence of right-turn channelization on major road (i.e., separate roadway for free right turns)
- Presence of right-turn channelization on crossroad (i.e., separate roadway for free right turns)
- Signal timing (i.e., pretimed / semiactuated / fully actuated)
- Signal phasing (i.e., two-phase / multiphase)

Traffic Volume Data
- Average daily traffic (ADT) of major road (veh/day)
- Average daily traffic (ADT) of crossroad (veh./day)

Other Related Data
- Rural / urban
- Terrain
- Presence of intersection lighting

Intersections were classified into five categories for statistical modeling:
1) rural, four-leg, STOP-controlled – 1,434 study intersections,
2) rural, three-leg, STOP-controlled – 2,692 study intersections,
3) urban, four-leg, STOP-controlled – 1,342 study intersections,
4) urban, three-leg, STOP-controlled – 3,057 study intersections, and
5) urban, four-leg, signalized – 1,306 study intersections.

In addition to the samples collected through the computer, listed above, field data were manually collected for 198 randomly selected urban, four-leg, signalized intersections for the purposes of collecting additional variables not included in the Caltrans database and verifying and updating the existing geometric and traffic data. Variables collected for these intersections are listed in Table 2.

Table 3. Variables collected during field study for 198 urban, four-leg, signalized intersections

Geometric Design
- Number of through lanes on each approach
- Number of exclusive left-turn lanes on each approach
- Number of exclusive right-turn lanes on each approach
- Type of left-turn treatment on each approach
- Type of right-turn treatment on each approach
- Horizontal alignment of each approach
- Approach grades on each approach
- Presence of crest/sag vertical curve on each approach
- Total through lane width on each approach (ft)
- Total left-turn lane width on each approach (ft)
- Presence of median on each approach (divided / undivided)
- Type of median on each approach
- Median width on each approach (ft)
- Number of driveways within 250 ft of the intersection on each approach
- Type of driveways on each approach
- Angle between intersecting approaches
- Curb return radius (ft) in intersection quadrant to the right of each approach

Traffic Control
- Presence of left-turn prohibition on each approach
- One-way vs. two-way operation on each approach
- Curb parking within 250 ft of the intersection on each approach
- Number of signal faces for each approach
- Signal head mounting for each approach
- Left-turn phasing for each approach
- Presence of pedestrian signals for crossing each approach
- Presence of advance warning signs for each approach
- Posted speed limit for each approach

Traffic Volume Data
- Turning movement volumes for all approaches by 15-min periods for 2-hr morning peak and 2-hr evening peak
- Level of pedestrian activity

Other Related Data
- Presence of intersection lighting
- Character of surrounding development

Regression Model Types

Using the Caltrans database, along with the collected field data in one case, the input variable values were determined and placed into a general equation. The general multiplicative form of the crash prediction model equations is:

$$\text{Function}(\mu_i) = \exp(\beta_0)(ADT_{major\ road})^{\beta 1}(ADT_{crossroad})^{\beta 2}\ \exp(\beta_3 X_3)\ \exp(\beta_4 X_4) \bullet \ldots \bullet \exp(\beta_q X_q)$$

Where X = input variables
β = model regression coefficients

Models were developed for total multiple-vehicle accidents and fatal and injury multiple-vehicle accidents. The resulting predicted collisions were for a three-year period, which corresponds to a three-year period of input crash data (1990-1992).

Loglinear

A loglinear model may be considered when the average number of intersection crashes becomes small. The two types of loglinear models used for at-grade intersections are the Poisson and negative binomial models.

Poisson distribution

The Poisson model assumes that the expected number of crashes, Y, follows a Poisson distribution. The Poisson distribution contains the limitation that the mean is equal to the variance of the data. The Poisson model is not appropriate in cases of overdispersion - where the variance, or dispersion, exceeds the estimated mean of the distribution.

Negative binomial distribution

The negative binomial model may be used in cases when overdispersion causes the Poisson distribution to be violated. The sites that used negative binomial distribution were the rural, 4-leg, stop-controlled; rural, 3-leg, stop-controlled; and urban, 3-leg, stop-controlled intersections.

Lognormal

The lognormal model assumes that the natural logarithm of Y, the expected number of crashes, follows a normal distribution. The regression coefficients, β, are estimated by the least-squares method. The sites that used the lognormal distribution were the urban, 4-leg, stop-controlled and the urban, 4-leg, signalized intersections.

Crashes occur randomly, which makes it difficult to model crash data. The goodness-of-fit measures used for the models are deviance/(n-p), Pearson chi-square/(n-p), R^2, and R^2_{FT} and root mean squared error. The (n-p) term is the degrees of freedom associated with the statistic. The R^2_{FT} parameter is based on the Freman-Tukey variance stabilizing transformation of variables discussed in Fridstrom et al (p.40).

The acceptable goodness-of-fit values for these models are listed below.

deviance/(n-p) - tends asymptotically towards 1.00
Pearson chi-square/(n-p) – between 0.8 and 1.2 generally indicates that the model appropriately fits the data
*R^2 – ideal fit is 100 percent
*R^2_{FT} – ideal fit is 100 percent

Published highway safety models have R^2 and R^2_{FT} values as low as 20 percent.

Bauer and Harwood developed models for the five categories of intersections described below.

Rural, Four-leg, STOP-controlled Intersections

The Poisson regression model including all 14 independent variables was first used to fit the data, but was considered inappropriate due to overdispersion. The Poisson regression model was reduced to include only statistically significant variables at the 10 percent significance level, but overdispersion remained present.

The negative binomial regression model corrected the overdispersion and was a better fit to the data than the reduced Poisson regression equation. The independent variables that were significant at the 90 percent confidence level and remained in the reduced negative binomial equation for the total multiple-vehicle accident model and the fatal and injury multiple-vehicle model are shown below.

Total Multiple-Vehicle Accidents

$$Y = e^{-11.246}(X_1)^{0.586}(X_2)^{0.797}\exp(0.463X_3)\exp(0.013X_4)\exp(0.244X_5)\exp(0.241X_6)\exp(0.268X_7)\exp(0.155X_8)\exp(-0.101X_9)\exp(0.091X_{10})\exp(0.313X_{11})$$

where:
Y = predicted number of crashes for 3 years,
X_1 = Crossroad ADT (log),
X_2 = Major-road ADT (log),
X_3 = Number of lanes on major road (1 if number of lanes is 3 or less, 0 otherwise)
X_4 = Design speed on major road (mph) = posted speed limit + 10 mph,
X_5 = 1 if functional class of major road is minor arterial, 0 otherwise,
X_6 = 1 if functional class of major road is major collector, 0 otherwise,
X_7 = 1 if major road has no access control, 0 otherwise,
X_8 = 1 if the terrain is flat, 0 otherwise,
X_9 = 1 if the terrain is mountainous, 0 otherwise,
X_{10} = 1 if major road has no left-turn lane, 0 otherwise, and
X_{11} = 1 if major road has a curbed left-turn lane, 0 otherwise.

Fatal & Injury Multiple-Vehicle Accidents:

$$Y = e^{-11.116}(X_1)^{0.602}(X_2)^{0.674}\exp(0.509X_3)\exp(0.016X_4)\exp(0.254X_5)\exp(-0.185X_6)\exp(0.250X_7)\exp(0.154X_8)\exp(-0.045X_9)\exp(0.424X_{10})\exp(0.191X_{11})\exp(0.190X_{12})$$

where:
Y = predicted number of crashes for 3 years,
X_1 = Crossroad ADT (log),
X_2 = Major-road ADT (log),
X_3 = Number of lanes on major road (1 if number of lanes is 3 or less, 0 otherwise),
X_4 = Design speed on major road,
X_5 = 1 if the terrain is flat, 0 otherwise,
X_6 = 1 if the terrain is mountainous, 0 otherwise,

X_7 = 1 if functional class of major road is minor arterial, 0 otherwise,
X_8 = 1 if functional class of major road is major collector, 0 otherwise,
X_9 = 1 if major road has no left-turn lane, 0 otherwise,
X_{10} = 1 if major road has a curbed left-turn lane, 0 otherwise,
X_{11} = 1 if there is no lighting at the intersection, and
X_{12} = 1 if major road has no access control, 0 otherwise.

Table 4. The goodness-of-fit measures for the reduced negative binomial regression:

	Total Multiple-Vehicle Accidents (3 years)	**Fatal & Injury Multiple-Vehicle Accidents (3 years)**
Deviance/(n-p)	1.01	1.00
Pearson chi-square/(n-p)	1.01	1.04
R^2 (%)	38.16	32.17
R^2_{FT} (%)	40.51	31.35

Rural, Three-Leg, STOP-Controlled Intersections

The statistical approach for this group was identical to that used for rural, four-leg, STOP-controlled intersections. The Poisson regression equation was used with 14 independent variables, and then reduced to include only 8 independent variables, but both equations resulted in overdispersion. The negative binomial regression equation was used with the independent variables that were significant at the 90 % confidence level for both accident models.

Total Multiple-Vehicle Accidents:
$Y = e^{-11.364}(X1)^{0.987} (X_2)^{0.429} \exp(0.249X_3) \exp(-0.071X_4) \exp(0.201X_5) \exp(0.196X_6) \exp(0.242X_7)$
where:
Y = predicted number of crashes for 3 years,
X_1 = Major-road ADT (log),
X_2 = Crossroad ADT (log),
X_3 = 1 if major road has no left-turn lane, 0 otherwise,
X_4 = 1 if major road has a curbed left-turn lane, 0 otherwise,
X_5 = 1 if functional class of major road is minor arterial, 0 otherwise,
X_6 = 1 if functional class of major road is major collector, 0 otherwise, and
X_7 = 1 if major road has no access control, 0 otherwise.

Fatal & Injury Multiple-Vehicle Accidents

The model for fatal & injury multiple-vehicle accidents was not available at the time of this report.

Table 5. The goodness-of-fit measures for the reduced negative binomial regression equation:

	Total Multiple-Vehicle Accidents (3-year)	**Fatal & Injury Multiple-Vehicle Accidents (3-year)**
Deviance/(n-p)	1.01	1.00
Pearson chi-square/(n-p)	1.17	1.30
R^2 (%)	35.16	27.82
R^2_{FT} (%)	36.26	25.92

Urban, Four-Leg, STOP-controlled Intersections

The Poisson regression model was initially used with 16 independent variables, but the goodness-of-fit results were poor due to overdispersion. A full lognormal regression model was used with 16 independent variables, which produced relatively similar goodness-of-fit results. The independent variables were then tested at the 90 % confidence level, and eight variables remained in both the total multiple-vehicle accident model and fatal and injury multiple-vehicle accident model; however, each model retained a different set of eight variables. The eight significant variables in the reduced lognormal equation for each accident model are as follows:

Total Multiple-Vehicle Accidents:

$$Y = e^{-5.073}(X_1)^{0.635}(X_2)^{0.294}\exp(-0.969X_3)\exp(-0.518X_4)\exp(-0.091X_5)\exp(0.340X_6)\exp(0.087X_7)\exp(-0.331X_8)\exp(-0.175X_9)$$

where:

Y = predicted number of crashes for 3 years,
X_1 = Major-road ADT (log),
X_2 = Crossroad ADT (log),
X_3 = 1 if left turns are prohibited on the major road, 0 otherwise,
X_4 = 1 if major road has no access control, 0 otherwise,
X_5 = Average lane width on major road (ft),
X_6 = 1 if number of lanes on major road is 3 or less, 0 otherwise,
X_7 = 1 if number of lanes on major road is 4 or 5, 0 otherwise,
X_8 = 1 if there are no free right turns on the crossroad, and
X_9 = 1 if there is no lighting, 0 otherwise.

Fatal & Injury Multiple-Vehicle Accident:

$$Y = e^{-4.745}(X_1)^{0.573}(X_2)^{0.216}\exp(-0.768X_3)\exp(-0.398X_4)\exp(-0.081X_5)\exp(0.234X_6)\exp(0.044X_7)\exp(-0.019X_8)\exp(-0.284X_9)$$

where:

Y = predicted number of crashes for 3 years,
X_1 = Major-road ADT (log),
X_2 = Crossroad ADT (log),
X_3 = 1 if left turns are prohibited on the major road, 0 otherwise,
X_4 = 1 if major road has no access control, 0 otherwise,

X_5 = Average lane width on major road (ft),
X_6 = 1 if number of lanes on major road is 3 or less, 0 otherwise,
X_7 = 1 if number of lanes on major road is 4 or 5, 0 otherwise,
X_8 = Outside shoulder width on major road (ft), and
X_9 = 1 if there are no free right turns on the crossroad, 0 otherwise.

Table 6. The goodness-of-fit measures using the reduced lognormal regression equation:

	Total Multiple-Vehicle Accidents (3-year)		**Fatal & Injury Multiple-Vehicle Accidents (3-year)**	
	Full	**Reduced**	**Full**	**Reduced**
Deviance/(n-p)	1.00	1.00	0.80	0.80
Pearson chi-square/(n-p)	1.00	1.00	0.80	0.80
R^2 (%)	20.54	20.58	18.17	18.06
Root mean square error	1.00	1.00	0.89	0.90

Urban, Three-leg, STOP-controlled Intersections

Similar to the previous intersection types, the Poisson regression model was initially used with 18 independent variables. Overdispersion was present in the data, therefore, the negative binomial model was used and the independent variables tested at the 90 % confidence level. A different set of eight variables remained in each of the accident models, which are listed below.

Total Multiple-Vehicle Accidents:
$Y = e^{-6.808}(X_1)^{0.775}(X_2)^{0.266}\exp(-0.478X_3)\exp(-0.601X_4)\exp(0.012X_5)\exp(0.192X_6)\exp(-0.006X_7)\exp(-0.160X_8)\exp(-0.030X_9)$
where:
Y = predicted number of crashes for 3 years,
X_1 = Major-road ADT (log),
X_2 = Crossroad ADT (log),
X_3 = 1 if left turns are prohibited on the major road, 0 otherwise,
X_4 = 1 if there are no free right turns on the crossroad, 0 otherwise,
X_5 = 1 if there is no left-turn lane on the major road, 0 otherwise,
X_6 = 1 if there is a curbed left-turn lane on the major road, 0 otherwise,
X_7 = Design speed of major road,
X_8 = 1 if the median of the major road is divided, 0 otherwise, and
X_9 = Average lane width on major road.

Fatal & Injury Multiple-Vehicle Accidents:
$Y = e^{-7.358}(X_1)^{0.766} (X_2)^{0.254} \exp(-0.458X_3) \exp(-0.575X_4) \exp(-0.055X_5) \exp(0.194X_6) \exp(-0.187X_7) \exp(-0.042X_8) \exp(-0.234X_9)$
where:
Y = predicted number of crashes for 3 years
X_1 = Major-road ADT (log)
X_2 = Crossroad ADT (log)
X_3 = 1 if left turns are prohibited on the major road, 0 otherwise
X_4 = 1 if there are no free right turns on the crossroad, 0 otherwise
X_5 = 1 if there is no left-turn lane on the major road, 0 otherwise
X_6 = 1 if there is a curbed left-turn lane on the major road, 0 otherwise
X_7 = 1 if the median of the major road is divided, 0 otherwise
X_8 = Average lane width on major road (ft)
X_9 = 1 if there is no access control on the major road

Table 7. The goodness-of-fit measures for the reduced negative binomial regression equation:

	Total Multiple-Vehicle Accidents (3-year)	**Fatal & Injury Multiple-Vehicle Accidents (3-year)**
Deviance/(n-p)	1.00	1.01
Pearson chi-square/(n-p)	1.13	1.07
R^2 (%)	16.07	16.30
R^2_{FT} (%)	17.64	16.38

Urban, Four-leg, Signalized Intersections

The lognormal regression model was initially used based on high crash frequencies and the shape of the crash frequency distribution. Nineteen independent variables were included in the full lognormal regression model for both accident models. Eight and seven independent variables remained significant at the 90% confidence level for the total multiple-vehicle accidents model and the fatal and injury multiple-vehicle accidents model, respectively. The significant variables are listed below for each accident model.

Total Multiple-Vehicle Accidents:
$Y = e^{-3.744}(X_1)^{0.234} (X_2)^{0.517} \exp(0.032X_3) \exp(0.636X_4) \exp(-0.312X_5) \exp(-0.221X_6) \exp(-0.134X_7) \exp(-0.051X_8) \exp(-0.240X_9) \exp(-0.146X_{10}) \exp(-0.119X_{11})$
where:
Y = predicted number of crashes for 3 years,
X_1 = Crossroad ADT (log),
X_2 = Major-road ADT (log),
X_3 = 1 if intersection signal timing is pretimed, 0 otherwise,
X_4 = 1 if intersection signal timing is fully actuated, 0 otherwise,
X_5 = 1 if there is no access control on the major road, 0 otherwise,

X_6 = 1 if the signal phasing is multiphase, 0 otherwise,
X_7 = 1 if the number of lanes on the crossroad is 3 or less, 0 otherwise,
X_8 = Average lane width on major road (ft),
X_9 = 1 if the number of lanes on the major road is 3 or less, 0 otherwise,
X_{10} = 1 if the number of lanes on the major road is 4 or 5, 0 otherwise, and
X_{11} = 1 if there are no free right turns on the major road, 0 otherwise.

Fatal & Injury Multiple-Vehicle Accidents:
$Y = e^{-5.845}(X_1)^{0574}(X_2)^{0.219}\exp(-0.073X_3)\exp(0.389X_4)\exp(-0.247X_5)\exp(-0.153X_6)$
$\exp(-0.265X_7)\exp(-0.186X_8)\exp(-0.168X_9)\exp(0.005X_{10})$
where:
Y = predicted number of crashes for 3 years,
X_1 = Major-road ADT (log),
X_2 = Crossroad ADT (log),
X_3 = 1 if intersection signal timing is pre-timed, 0 otherwise,
X_4 = 1 if intersection signal timing is fully actuated, 0 otherwise,
X_5 = 1 if the signal phasing is multiphase, 0 otherwise,
X_6 = 1 if the number of lanes on the crossroad is 3 or less, 0 otherwise,
X_7 = 1 if there is no access control on the major road, 0 otherwise,
X_8 = 1 if the number of lanes on the major road is 3 or less, 0 otherwise,
X_9 = 1 if the number of lanes on the major road is 4 or 5, 0 otherwise, and
X_{10} = Design speed on major road.

Table 8. The goodness-of-fit measures for the reduced lognormal regression equation:

	Total Multiple-Vehicle Accidents (3-year)	**Fatal & Injury Multiple-Vehicle Accidents (3-year)**
Deviance/(n-p)	0.72	0.70
Pearson chi-square/(n-p)	0.72	0.70
R^2 (%)	25.08	24.31
R^2_{FT} (%)	na	na
Root mean square error	0.85	0.84

Urban, Four-leg, Signalized Intersections – Sample of 198 Intersections

A pilot study including 198 randomly selected urban, four-leg, signalized intersections was conducted to collect additional geometric, traffic control, traffic volume variables and determine if the additional data improved the goodness-of-fit of the models.

The data collected at each of the 198 intersections during the field study were shown in Table 2 above. A negative binomial model form was used with the variables to develop an equation to estimate total multiple-vehicle accidents and fatal and injury multiple-vehicle accidents. Any variable that was not significant at the 90 percent confidence level was omitted from the reduced equation.

The sample of 198 intersections was tested using the following three methods:

1. The model was estimated using independent variables in the Caltrans database as in the previous cases.
2. The model was estimated using independent variables with updated values from the field data collection where available.
3. The model was estimated using only variables from the field data collection.

The following are the significant variables included in the reduced negative binomial models for each equation.

<u>Method 1</u> – using Caltrans database

Total Multiple-Vehicle Accidents:
$Y = e^{-5.775}(X_1)^{0.258} (X_2)^{0.670} \exp(-0.500X_3) \exp(-0.287X_4)$
where:
Y = predicted number of crashes for 3 years,
X_1 = Crossroad ADT (log),
X_2 = Major-road ADT (log),
X_3 = 1 if the number of lanes on the major road is 3 or less, 0 otherwise, and
X_4 = 1 if the number of lanes on the major road is 4 or 5, 0 otherwise.

Fatal & Injury Multiple-Vehicle Accidents:
$Y = e^{-4.406}(X_1)^{0.189} (X_2)^{0.470} \exp(-0.308X_3) \exp(-0.262X_4) \exp(0.008X_5)$
where:
Y = predicted number of crashes for 3 years,
X_1 = Crossroad ADT (log),
X_2 = Major-road ADT (log),
X_3 = 1 if the number of lanes on the major road is 3 or less, 0 otherwise,
X_4 = 1 if the number of lanes on the major road is 4 or 5, 0 otherwise, and
X_5 = Design speed on major road.

<u>Method 2</u> – Using field data to update Caltrans database

Total Multiple-Vehicle Accidents:
$Y = e^{-7.740}(X_1)^{0.909} (X_4)^{0.167} \exp(0.475X_2) \exp(-0.176X_3) \exp(-0.332X_5) \exp(0.005X_6)$
$\exp(0.368X_7) \exp(-0.200X_8)$
where:
Y = predicted number of crashes for 3 years,
X_1 = Major-road ADT (log),
X_2 = 1 if there is no left-turn lane on the major road, 0 otherwise,
X_3 = 1 if there is a curbed left-turn lane on the major road, 0 otherwise,
X_4 = Crossroad ADT (log),
X_5 = 1 if there is no left-turn lane on the crossroad, 0 otherwise,
X_6 = 1 if there is a curbed left-turn lane on the crossroad, 0 otherwise,
X_7 = 1 if intersection signal timing is fully actuated, 0 otherwise, and

X_8 = 1 if the number of lanes on the major road is 1, 0 otherwise.

Fatal & Injury Multiple-Vehicle Accidents:
$Y = e^{-5.977}(X_1)^{0.642} (X_2)^{0.191}$
where:
Y = predicted number of crashes for 3 years,
X_1 = Major-road ADT (log), and
X_2 = Crossroad ADT (log).

<u>Method 3</u> – Using field data only

Total Multiple-Vehicle Accidents:
$Y = e^{-7.206}(X_{i1})^{0.836} (X_2)^{0.214} \exp(-0.394X_3) \exp(0.346X_4) \exp(-0.234X_5)$
where:
Y = predicted number of crashes for 3 years,
X_1 = Major-road ADT (log),
X_2 = Crossroad ADT (log),
X_3 = 1 if there is no left-turn lane on the crossroad, 0 otherwise,
X_4 = 1 if there is no left-turn lane on the major road, 0 otherwise, and
X_5 = 1 if the angle of the intersection is less than 90 degrees, 0 otherwise.

Fatal & Injury Multiple-Vehicle Accidents:
$Y = e^{-5.838}(X_{i1})^{0.625} (X_2)^{0.185} \exp(0.214X_3) \exp(-0.224X_4)$
where:
Y = predicted number of crashes for 3 years,
X_1 = Major-road ADT (log),
X_2 = Crossroad ADT (log),
X_3 = 1 if there is curbed parking on the major road, 0 otherwise, and
X_4 = 1 if the angle of the intersection is less than 90 degrees, 0 otherwise.

Table 9. The goodness-of-fit measures for the reduced equations using each of the three above methods:

	Method 1 Caltrans data	**Method 2** Updated Caltrans data	**Method 3** Field data only
Total Multiple-Vehicle Acc.			
Deviance/(n-p)	0.993	0.990	1.009
Pearson chi-square/(n-p)	0.888	0.838	0.868
R^2 (%)	31.88	37.56	36.28
R^2_{FT} (%)	35.09	40.66	38.79
Fatal & Injury Multi-veh. Acc.			
Deviance/(n-p)	1.005	1.002	0.997
Pearson chi-square/(n-p)	0.886	0.865	0.865
R^2 (%)	26.58	24.51	28.44
R^2_{FT} (%)	28.65	25.52	29.28

Statistical Models of Accidents on Interchange Ramps and Speed-Change Lanes

Data Collection

The Washington State Department of Transportation database was used to collect interchange and speed change lane data for the variables listed below. The Washington database included 2,046 ramps.

Geometric Design Features:
- Ramp type (on-ramp / off-ramp)
- Number of lanes
- Surface width (ft)
- Right shoulder type
- Right shoulder width (ft)
- Left shoulder type
- Left shoulder width (ft)
- Ramp or speed-change lane segment length (mi) for segments with homogeneous cross sections
- Ramp length (mi)
- Speed-change lane length (mi)

Traffic Volume Data:
- Annual average daily traffic (veh/day) for ramp or speed-change lane
- Annual average daily traffic (veh/day) for adjacent mainline freeway

Other Related Data:
- Area type (rural / urban)

A manual review of Washington State DOT interchange diagrams was necessary to obtain the additional variable information shown below.

Geometric Design Features
- Ramp segment type
- Ramp configuration
- Traveled way width (ft)
- Average lane width (ft)
- Right shoulder width (ft)
- Left shoulder width (ft)

The additional variables shown below were obtained from existing records of the Washington State DOT.

Geometric Design Features
- Minimum radius of any horizontal curve on the ramp
- Horizontal alignment index (curviness) for the ramp (based on equation)
- Horizontal alignment index (curviness) for the ramp (based on modified equation)

- General grade of ramp (upgrade / downgrade)

Traffic Volume Variables:

- Annual average daily traffic volume of the mainline freeway section adjacent to speed-change lane (veh/day)

Statistical Modeling

The two dependent variables used most in the modeling were 1) total accidents of all severity levels that occurred during the 3-year study period and 2) fatal and injury accidents in the 3-year period.

The loglinear regression models were used on the accident data, although, the accident frequency distribution did not ideally fit a loglinear distribution model. Poisson and negative binomial regression models were the loglinear models used for the data. A loglinear model may be considered when the average number of crashes is small.

Poisson distribution

The Poisson model assumes the number of crashes, Y, follows a Poisson distribution. The Poisson distribution contains the limitation that the mean is equal to the variance of the data. The Poisson model is not appropriate in cases of overdispersion - where the variance, or dispersion, exceeds the estimated mean of the distribution.

Negative binomial distribution

The negative binomial model may be used in cases when overdispersion causes the Poisson distribution to be violated.

The general form of the multiplicative model relating the expected accidents and the independent variables is:

$$\text{function}(\mu_i) = \exp(\beta_0)(AADT_{Ramp})^{\beta 1}\ \exp(\beta_2 X_2)\ \exp(\beta_3 X_3) \bullet \ldots \bullet \exp(\beta_q X_q)$$

where β_0, β_1, β_2, …, β_q are the model regression coefficients estimated using the maximum likelihood method.

The goodness-of-fit measures used for the models are deviance/(n-p), Pearson chi-square/(n-p), R^2, and R^2_{FT} and root mean squared error. The (n-p) term is the degrees of freedom associated with the statistic. The R^2_{FT} parameter is based on the Freman-Tukey variance stabilizing transformation of variables discussed in Fridstrom et al (p.40).

The acceptable goodness-of-fit values for these models are listed below.
deviance/(n-p) - tends asymptotically towards 1.00
Pearson chi-square/(n-p) – between 0.8 and 1.2 generally indicates that the model appropriately fits the data

R^2 – ideal fit is 100 percent
R^2_{FT} – ideal fit is 100 percent

Bauer and Harwood developed models for the seven categories of ramps described below.

Ramp Proper Segments (Including All Accidents)

The negative binomial regression model with nine independent variables was first used to fit the data. The variables were tested at the 90 % confidence level to determine which variables were significant. Only the significant variables were then used in the negative binomial regression model. The significant variables for each accident model (total and fatal and injury) are shown below.

Total Multiple-Vehicle Accidents:
$Y = e^{-9.81}(AADT_{Ramp})^{0.93} \exp(5.78X_2) \exp(0.78X_3) \exp(0.77X_4) \exp(0.56X_5) \exp(0.66X_6) \exp(1.09X_7) \exp(0.72X_8) \exp(0.29X_9) \exp(-0.05X_{10})$
where:
Y = predicted number of crashes for 3 years,
$AADT_{Ramp}$ = Ramp AADT (log),
X_2 = Segment length (miles),
X_3 = Ramp type - 1 if off-ramp, 0 otherwise,
X_4 = Number of lanes - 1 if one lane, 0 otherwise,
X_5 = 1 if ramp is an off-ramp, diamond configuration, 0 otherwise,
X_6 = 1 if ramp is an off-ramp, loop configuration, 0 otherwise,
X_7 = 1 if ramp is an off-ramp, outer connection, 0 otherwise,
X_8 = 1 if ramp is an on-ramp, diamond configuration, 0 otherwise,
X_9 = 1 if ramp is an on-ramp, loop configuration, 0 otherwise, and
X_{10} = 1 if ramp is an on-ramp, outer connection, 0 otherwise.

Fatal and InjuryMultiple-Vehicle Accidents
$Y = e^{-12.33}(AADT_{Ramp})^{1.04} \exp(1.45X_2) \exp(5.20X_3) \exp(0.78X_4) \exp(-0.81X_5) \exp(-0.39X_6) \exp(2.24X_7) \exp(0.99X_8) \exp(0.68X_9) \exp(-1.62X_{10}) \exp(0.07X_{11})$
where:
Y = predicted number of crashes for 3 years,
$AADT_{Ramp}$ = Ramp AADT (log),
X_2 = 1 if ramp is an off-ramp, 0 otherwise,
X_3 = Segment length (miles),
X_4 = Number of lanes: 1 if one lane, 0 otherwise,
X_5 = 1 if ramp is an off-ramp, diamond configuration, 0 otherwise,
X_6 = 1 if ramp is an off-ramp, loop configuration, 0 otherwise,
X_7 = 1 if ramp is an off-ramp, outer connection, 0 otherwise,
X_8 = 1 if ramp is a diamond configuration, 0 otherwise,
X_9 = 1 if ramp is a loop configuration, 0 otherwise,

X_{10} = 1 if ramp is an outer connection, 0 otherwise, and
X_{11} = Right shoulder width (feet).

Table 10. The goodness-of-fit measures for the reduced negative binomial regression model:

	Total Accidents (3-year)	Fatal & Injury Accidents (3-year)
Deviance/(n-p)	1.01	1.01
Pearson chi-square/(n-p)	1.14	1.62
R^2 (%)	14	14
R^2_{FT} (%)	13	15

Ramp Proper Segments, Off-Ramps Only (Excluding Rear-End Accidents)

It was determined that most rear-end crashes on off-ramps were related to the operation of the crossroad ramp terminal and not to the ramp geometrics. Therefore, a dependent variable was used that excluded rear-end collisions and other collisions related to the crossroad ramp terminal.

The full negative binomial model was initially used with eight independent variables. Five independent variables were found to be significant at the 90% confidence level for both accident models. The significant variables are listed below.

Total Multiple-Vehicle Accidents:
$Y = e^{-5.50}(AADT_{Ramp})^{0.62}\exp(1.03X_2)\exp(-0.15X_3)\exp(-0.01X_4)\exp(0.63X_5)$
$\exp(4.41X_6)\exp(-0.06X_7)$
where:
Y = predicted number of crashes for 3 years,
$AADT_{Ramp}$ = Ramp AADT (log),
X_2 = Number of lanes - 1 if one lane, 0 otherwise,
X_3 = 1 if ramp is a diamond configuration, 0 otherwise,
X_4 = 1 if ramp is a loop configuration, 0 otherwise,
X_5 = 1 if ramp is a outer connection, 0 otherwise,
X_6 = Ramp segment length (miles), and
X_7 = Average lane width (feet).

Fatal and Injury Multiple-Vehicle Accidents:
$Y = e^{-6.20}(AADT_{Ramp})^{0.68}\exp(1.20X_{i2})\exp(-0.67X_{i3})\exp(-0.54X_{i4})\exp(0.16X_{i5})$
$\exp(-0.08X_{i6})\exp(2.98X_{i7})$
where:
Y = predicted number of crashes for 3 years,
$AADT_{Ramp}$ = Ramp AADT (log),
X_{i2} = Number of lanes - 1 if one lane, 0 otherwise,
X_{i3} = 1 if ramp is a diamond configuration, 0 otherwise,

X_{i4} = 1 if ramp is a loop configuration, 0 otherwise,
X_{i5} = 1 if ramp is a outer connection, 0 otherwise,
X_{i6} = Average lane width (feet), and
X_{i7} = Ramp segment length (miles).

Table 11. The goodness-of-fit measures for the reduced negative binomial regression model:

	Total Accidents (3-year)	**Fatal & Injury Accidents (3-year)**
Deviance/(n-p)	1.00	1.00
Pearson chi-square/(n-p)	1.22	1.73
R^2 (%)	18	18
R^2_{FT} (%)	16	10

Entire Ramps (Including All Accidents)

Crashes on the entire ramp were analyzed rather than crashes on a particular ramp segment. A full negative binomial regression model was used with five independent variables and one interaction. When tested at the 90% confidence level, the same four independent variables remained in both accident models. The variables are listed below.

Total Multiple-Vehicle Accidents:
$$Y = e^{-6.32}(AADT_{Ramp})^{0.72}\exp(0.62X_2)\exp(1.18X_3)\exp(0.15X_4)\exp(0.89X_5)\exp(0.50X_6)\exp(-0.35X_7)$$
where:
Y = predicted number of crashes for 3 years,
$AADT_{Ramp}$ = Ramp AADT (log),
X_2 = 1 if ramp is a diamond configuration, 0 otherwise,
X_3 = 1 if ramp is a parclo loop configuration, 0 otherwise,
X_4 = 1 if ramp is a free-flow loop configuration, 0 otherwise,
X_5 = 1 if ramp is a outer connection, 0 otherwise,
X_6 = 1 if ramp is an off-ramp, 0 otherwise, and
X_7 = 1 if area type is rural, 0 otherwise.

Fatal and Injury Multiple-Vehicle Accidents:
$$Y = e^{-7.87}(AADT_{Ramp})^{0.85}\exp(0.54X_2)\exp(1.22X_3)\exp(0.01X_4)\exp(0.80X_5)\exp(0.55X_6)\exp(-0.34X_7)$$
where:
Y = predicted number of crashes for 3 years,
$AADT_{Ramp}$ = Ramp AADT (log),
X_2 = 1 if ramp is a diamond configuration, 0 otherwise,
X_3 = 1 if ramp is a parclo loop configuration, 0 otherwise,
X_4 = 1 if ramp is a free-flow loop configuration, 0 otherwise,

X_5 = 1 if ramp is a outer connection, 0 otherwise,
X_6 = 1 if ramp is an off-ramp, 0 otherwise, and
X_7 = 1 if area type is rural, 0 otherwise.

Table 12. The goodness-of-fit measures for the reduced negative binomial regression equation:

	Total Accidents (3-year)	**Fatal & Injury Accidents (3-year)**
Deviance/(n-p)	1.00	1.01
Pearson chi-square/(n-p)	0.98	1.35
R^2 (%)	21	19
R^2_{FT} (%)	22	19

Entire Ramps, Off-Ramps Only (Excluding Rear-End Accidents)

To reduce the influence of the crossroad ramp terminals and focus on crashes that are potentially related to the ramp geometrics, rear-end crashes and other crashes related to the cross-road terminal were omitted from this analysis.

The full negative binomial regression model included four independent variables and was reduced to two variables after testing the variables at the 90 % confidence level. The two variables were the same for the two accident models.

Total Multiple-Vehicle Accidents:
$Y = e^{-3.97}(AADT_{Ramp})^{0.54} \exp(-0.16X_2) \exp(0.11X_3) \exp(-0.23X_4) \exp(0.56X_5)$
where:
Y = predicted number of crashes for 3 years,
$AADT_{Ramp}$ = Ramp AADT (log),
X_2 = 1 if ramp is a diamond configuration, 0 otherwise,
X_3 = 1 if ramp is a parclo loop configuration, 0 otherwise,
X_4 = 1 if ramp is a free-flow loop configuration, 0 otherwise, and
X_5 = 1 if ramp is a outer connection, 0 otherwise.

Fatal and Injury Multiple-Vehicle Accidents:
$Y = e^{-5.21}(AADT_{Ramp})^{0.61} \exp(-0.56X_2) \exp(0.01X_3) \exp(-0.69X_4) \exp(0.28X_5)$
where:
Y = predicted number of crashes for 3 years,
$AADT_{Ramp}$ = Ramp AADT (log),
X_2 = 1 if ramp is a diamond configuration, 0 otherwise,
X_3 = 1 if ramp is a parclo loop configuration, 0 otherwise,
X_4 = 1 if ramp is a free-flow loop configuration, 0 otherwise, and
X_5 = 1 if ramp is a outer connection, 0 otherwise.

Table 13. The goodness-of-fit measures for the reduced negative binomial regression equation:

	Total Accidents (3-year)	**Fatal & Injury Accidents (3-year)**
Deviance/(n-p)	1.00	0.99
Pearson chi-square/(n-p)	0.96	1.33
R^2 (%)	17	16
R^2_{FT} (%)	14	10

Speed-Change Lanes

Speed-change lanes include both acceleration and deceleration lanes. The full negative binomial regression model included seven independent variables for both acceleration and deceleration lanes. The number of independent variables in the model was reduced for acceleration and deceleration lanes when each variable was tested at the 90 % confidence level. The significant variables are listed below. Fatal and injury accidents included only diamond ramps due to modeling difficulties resulting from inclusion of all five ramp configurations.

Total Multiple-Vehicle Accidents

Acceleration lanes:
$Y = e^{-12.84}(AADT_{Ramp})^{0.98}\exp(6.88X_2)\exp(-0.59X_3)\exp(0.32X_4)$
where:
Y = predicted number of crashes for 3 years,
$AADT_{Ramp}$ = Ramp AADT (log),
X_2 = Acceleration lane length (miles),
X_3 = 1 if area type is rural, 0 otherwise, and
X_4 = Mainline freeway AADT (log).

Deceleration lanes:
$Y = e^{-9.73}(AADT_{Ramp})^{1.04}\exp(-1.21X_2)\exp(0.09X_3)$
where:
Y = predicted number of crashes for 3 years,
$AADT_{Ramp}$ = Ramp AADT (log),
X_2 = 1 if the area type is rural, 0 otherwise, and
X_3 = Right shoulder width (ft).

Fatal and Injury Multiple-Vehicle Accidents for Diamond On-Ramps Only:

$Y = e^{-15.81}(AADT_{Ramp})^{0.99}\ \exp(5.32X_2)\ \exp(0.56X_3)$
where:
Y = predicted number of crashes for 3 years,
$AADT_{Ramp}$ = Ramp AADT (log),
X_2 = Acceleration lane length (ft), and
X_3 = Mainline freeway AADT (log).

Table 14. The goodness-of-fit measures for the reduced negative binomial regression equation:

	Total Accidents (3-year)		**Fatal & Injury Accidents (3-year) for diamond ramps only**	
	Acceleration Lanes	**Deceleration Lanes**	**Acceleration Lanes**	**Deceleration Lanes**
Deviance/(n-p)	1.01	1.00	1.00	--
Pearson chi-square/(n-p)	1.36	2.81	1.35	--
R^2 (%)	36	15	34	--
R^2_{FT} (%)	38	16	34	--

* Diamond on-ramps only

Entire Ramps and Adjacent Speed-Change Lanes Combined

The aforementioned models considered the ramp and speed-change lane as independently with separate models, while this analysis considers both in a single model. The negative binomial regression model was initially used with eight independent variables. The model was reduced to include significant variables at the 90 % confidence level. Because most of the independent variables in the full model were significant at the 80 % confidence level and the variables were important to the research objectives, the negative binomial model included the significant variables at the 80 % confidence level. The significant variables are listed below for each accident model.

Total Multiple-Vehicle Accidents:
$Y = e^{-5.75}(AADT_{Ramp})^{0.80}\ \exp(-0.47X_2)\ \exp(0.41X_3)\ \exp(0.70X_4)\ \exp(-0.18X_5)$
$\exp(-0.66X_6)$
where:
Y = predicted number of crashes for 3 years,
$AADT_{Ramp}$ = Ramp AADT (log),
X_2 = 1 if area type is rural, 0 otherwise,
X_3 = ramp is a diamond configuration, 0 otherwise,
X_4 = 1 if ramp is a parclo loop configuration, 0 otherwise,
X_5 = 1 if ramp is a free-flow loop configuration, 0 otherwise, and

X_6 = 1 if ramp is a outer connection, 0 otherwise.

Fatal and Injury Multiple-Vehicle Accidents:
$Y = e^{-10.68}(AADT_{Ramp})^{0.91}\exp(-4.55X_2)\exp(2.90X_3)\exp(0.49X_4)\exp(0.29X_5)$
where:
Y = predicted number of crashes for 3 years,
$AADT_{Ramp}$ = Ramp AADT (log),
X_2 = Length of speed-change lane (ft),
X_3 = Ramp length (ft),
X_4 = 1 if ramp is an off-ramp, 0 otherwise, and
X_5 = Mainline freeway AADT (log).

Table 15. The goodness-of-fit measures for both equations using a 90% confidence level and 80% confidence level:

	Total Accidents (3-year)		**Fatal & Injury Accidents (3-year)**	
	90% Confidence	**80% Confidence**	**90% Confidence**	**80% Confidence**
Deviance/(n-p)	1.00	1.00	1.00	1.00
Pearson chi-square/(n-p)	0.94	0.96	1.11	1.11
R^2 (%)	37	38	36	36
R^2_{FT} (%)	42	43	37	37

Selected Urban Off-Ramps

This analysis focused on one ramp type and configuration rather than considering numerous ramp types and configurations as the previous interchange models. Almost all data collected were for urban diamond off-ramps; therefore, a model was developed for this ramp type and configuration.

An objective of this analysis was to test the accident relationship of three alternative measures for categorizing the ramp horizontal alignment. The alternative measures of the horizontal curvature or curviness of each ramp were considered in statistical analyses. These were:

- Alternative 1 – The smallest radius of all horizontal curves on the ramp.
- Alternative 2 – The horizontal alignment index (curviness) of the ramp based on the following equations used in previous work by Bared and Vogt:

$$H = (1/L_h)\,(\Sigma(D_i)^{1.5}\, lh_i)$$

Where:
H = horizontal alignment index
L_h = total length of ramp, including horizontal curves and tangents, in hundreds of feet,

D_i = degree of curvature for the i^{th} horizontal curve [change in angular heading per 31m (100 ft)], and
lh_i = length of i^{th} horizontal curve (in hundreds of feet).

- Alternative 3 – The same equation for horizontal alignment index, but with the coefficient of the Di term set equal to 1.0, rather than 1.5.

Table 16. Goodness of fit measures for varying measures of horizontal curvature

	Measure of horizontal curvature considered					
	Minimum radius		Horizontal alignment index from equation		Horizontal alignment index from equation (modified)	
	Poisson	NB	Poisson	NB	Poisson	NB
Total accidents (3-year period)						
Deviance/(n-p)	15.38	1.00	15.16	1.00	15.10	1.00
Pearson chi-square/(n-p)	15.38	0.65	15.16	0.64	15.10	0.64
R^2 (%)	10	NC[a]	11	NC	11	NC
R^2_{FT} (%)	4	NC	5	NC	5	NC
Fatal and injury accidents (3-year period)						
Deviance/(n-p)	4.79	1.01	4.77	1.00	4.74	1.01
Pearson chi-square/(n-p)	4.79	0.84	4.77	0.84	4.74	0.84
R^2 (%)	10	3	10	6	10	5
R^2_{FT} (%)	5	1	5	4	5	3

[a] Not calculated. Statistic could not be estimated (negative value found).

Rail/Highway Crossing Models

The model equations used for the rail/highway crossings were developed in the NCDOT Rail Division, Engineering Safety Branch. These models have been used as part of the Railway/Highway Crossing Hazard Elimination Program. The models included in this software were for crossings with:

- Passive warning device,
- Light warning device, and
- Gate warning device.

The original form of these models can be found in the Railway/Highway Grade Crossing Handbook (FHWA-TS-86-215). The models were modified by the NCDOT and are recalibrated annually. Those interested in the most up-to-date model equations should contact A.R. (Drew) Thomas, P.E. in the NCDOT Rail Division.

Procedure of Software Use

Task 7 in the project proposal required that an analysis be conducted of the data required to apply the crash models. This section lists and describes the input required by TCPM. The input consists of intersection, interchange, or railway crossing characteristics. However, if a group of sites is being analyzed, the source(s) for these data may be scattered. In this case, a certain characteristic may be unavailable and the user may wish to use the suggested default values listed below each variable. All AADT's are considered two-direction volumes, and number of lanes does not include turning bays or lanes at intersections. It is recommended that the default AADT's be used only as a last resort. Collecting actual AADT information should be the highest priority.

***General procedure for using the models programmed into TCPM*:**

1. Click on the appropriate heading at the top of the screen for the type of road segment to be analyzed.
2. Choose the applicable model.
3. Input the requested data in the blanks. Note especially the units on the input data.
4. Click the button at the bottom of the screen to see the crash prediction.
5. To print the prediction, click the print button. The entire page will be printed, including all the input variables.

Input for TCPM Intersection Models

1. *Model for rural, four-legged, stop-controlled intersections*

X1 – AADT of the minor road (vehicles/day)
The range of values for this variable used in this model is 100-9500 vehicles/day.
Suggested default: 1000 veh/day
X2 – AADT of the major road (vehicles/day)
The range of values for this variable used in this model is 400-72000 vehicles/day.
Suggested default: 7,000 veh/day
X3 – Number of lanes on major road
Suggested default: 3 or fewer
X4 – Major road design speed (mph)
The range of values for this variable used in this model is 25-70 miles/hour.
Suggested default: 45 mph
X5, X6 – Functional class of major road
To determine this variable, the user may have to refer to the city or state thoroughfare plans.
Suggested default: Minor arterial
X7 – Access control on major road (Yes/No)
Suggested default: No access control

X8, X9 – Terrain
This variable is fairly subjective and is a judgement call for the analyst.
Suggested default: Flat
X10, X11 – Major road left-turn channelization
If left-turn lane is a painted turn lane, enter '0' for both.
Suggested default: No left-turn lane

2. Model for rural, three-legged, STOP-controlled intersections

X1 – AADT of the major road (vehicles/day)
The range of values for this variable used in this model is 400-72,000 vehicles/day.
Suggested default: 7,000 veh/day
X2 – AADT of the minor road (vehicles/day)
The range of values for this variable used in this model is 100-10,000 vehicles/day.
Suggested default: 500 veh/day
X3, X4 - Major road left-turn channelization
If left-turn lane is a painted turn lane, enter '0' for both.
Suggested default: No left-turn lane
X5, X6 – Functional class of major road
To determine this variable, the user may have to refer to the city or state thoroughfare plans.
Suggested default: Minor arterial
X7 – Access control on major road (Yes/No)
Suggested default: No access control

3. Model for urban, four-legged, STOP-controlled intersections

X1 – AADT of major road (vehicles/day)
The range of values for this variable used in this model is 1100-79,000 vehicles/day.
Suggested default: 22,000 veh/day
X2 – AADT of minor road (vehicles/day)
The range of values for this variable used in this model is 100-17,000 vehicles/day.
Suggested default: 1000 veh/day
X3 – Major road left-turn prohibition (Yes/No)
Suggested default: No, left turns not prohibited
X4 – Access control on major road (Yes/No)
Suggested default: No access control
X5 – Major road average lane width
The range of values for this variable used in this model is 8-15 feet.
Suggested default: 12 ft
X6, X7 – Number of lanes on major road
Suggested default: 3 or fewer
X8 – Major road right-turn channelization (Yes/No)
i.e. Separate roadway for free right turns

Suggested default: No free right turns
X9 – Lighting (Yes/No)
Suggested default: Yes, lighting present

4. Model for urban, three-legged, STOP-controlled intersections

X1 – AADT of major road (vehicles/day)
The range of values for this variable used in this model is 520-97,000 vehicles/day.
Suggested default: 25,000 veh/day
X2 – AADT of minor road (vehicles/day)
The range of values for this variable used in this model is 100-22,000 vehicles/day.
Suggested default: 1000 veh/day
X3 – Major road left-turn prohibition (Yes/No)
Suggested default: No, left turns not prohibited
X4 – Minor road right-turn channelization
Suggested default: No free right turns
X5, X6 – Major road left-turn channelization
If left-turn lane is a painted turn lane, enter '0' for both.
Suggested default: No left-turn lane
X7 – Major road design speed
The range of values for this variable used in this model is 25-70 miles/hour.
Suggested default: 50 mph
X8 – Presence of median on major road (Yes/No)
Suggested default: Median present
X9 – Average lane width on major road
The range of values for this variable used in this model is 8-15 feet.
Suggested default: 12 ft

5. Model for urban, four-legged, signalized intersections

X1 – AADT of major road (vehicles/day)
The range of values for this variable used in this model is 2400-79,000 vehicles/day.
Suggested default: 31,000 veh/day
X2 – AADT of minor road (vehicles/day)
The range of values for this variable used in this model is 100-48,000 vehicles/day.
Suggested default: 5,000 veh/day
X3, X4 – Signal timing
The choices presented are pre-timed and fully actuated. If signal is semi-actuated, enter '0' for both X3 and X4.
Suggested default: Pre-timed
X5 – Access control on major road (Yes/No)
Suggested default: No access control
X6 – Signal phasing
If phasing is multiphase, enter '1'. If phasing is two-phase, enter '0'.

Suggested default: Two-phase
X7 – Number of lanes on MINOR road
Suggested default: 3 or fewer
X8 – Average lane width on major road
The range of values for this variable used in this model is 8-15 feet.
Suggested default: 12 ft
X9, X10 – Number of lanes on major road
Suggested default: 4 or 5
X11 – Major road right-turn channelization (Yes/No)
Suggested default: No free right turns

Input for TCPM Interchange Models

1. Model for Ramp Proper Segments

X1 – AADT of the ramp segment (vehicles/day)
The range of values for this variable used in this model is 27-24,000 vehicles/day.
Suggested default: 5,000 veh/day
X2 – Length of the ramp segment (miles)
The range of values for this variable used in this model is 0.01-0.69 miles.
Suggested default: 0.12 miles
X3 – Ramp type
Suggested default: Off-ramp
X4 – Number of lanes
Suggested default: 1 lane
X5, X6, X7, X8, X9, X10 – Ramp configuration
Parclo and free-flow loops are grouped together under “loop ramps”.
Suggested default: Diamond

2. Model for Ramp Proper Segments on Off-Ramps (Rear-end Crashes Excluded)

X1 – AADT of the ramp segment (vehicles/day)
The range of values for this variable used in this model is 27-24,000 vehicles/day.
Suggested default: 5,000 veh/day
X2 – Number of lanes
Suggested default: 1 lane
X3, X4, X5 – Ramp configuration
Parclo and free-flow loops are grouped together under “loop ramps”.
Suggested default: Diamond
X6 – Length of the ramp segment (miles)
The range of values for this variable used in this model is 0.01-0.5 miles.
Suggested default: 0.12 miles

X7 – Average lane width (ft)
The range of values for this variable used in this model is 10-36 feet.
Suggested default: 14 ft

3. Model for Entire Ramps

X1 – AADT of the ramp (vehicles/day)
The range of values for this variable used in this model is 27-24,000 vehicles/day.
Suggested default: 4,000 veh/day
X2, X3, X4, X5 – Ramp configuration
If the ramp is direct or semi-direct connection, enter '0' for all the configuration variables.
Suggested default: Diamond
X6 – Ramp type
Suggested default: Off-ramp
X7 – Area type
Suggested default: Urban

4. Model for Entire Off-Ramps (Rear-end Crashes Excluded)

X1 – AADT of the ramp (vehicles/day)
The range of values for this variable used in this model is 27-24,000 vehicles/day.
Suggested default: 4,000 veh/day
X2, X3, X4, X5 – Ramp configuration
If the ramp is direct or semi-direct connection, enter '0' for all the configuration variables.
Suggested default: Diamond

5. Model for Acceleration Lanes

X1 – AADT of the ramp (vehicles/day)
The range of values for this variable used in this model is 54-21,000 vehicles/day.
Suggested default: 4,000 veh/day
X2 – Length of acceleration lane (miles)
The range of values for this variable used in this model is 0.02-0.38 miles. This length includes the taper.
Suggested default: 0.25 miles
X3 – Area type
Suggested default: Urban
X4 – Mainline freeway AADT (vehicles/day)
The range of values for this variable used in this model is 4,000-100,000 vehicles/day.
Suggested default: 30,000 veh/day

6. *Model for Deceleration Lanes*

> **NOTE: The crash prediction software (TCPM) delivered with this report contains computational errors in this Deceleration Lane model. It is recommended NOT to use this model until future versions of TCPM are released.**

X1 – AADT of the ramp (vehicles/day)
The range of values for this variable used in this model is 54-24,000 vehicles/day.
Suggested default: 3,000 veh/day
X2 – Area type
Suggested default: Urban
X3 – Right shoulder width (ft)
The range of values for this variable used in this model is 0-16 feet.
Suggested default: 8 ft

7. *Model for Entire Ramp plus Adjacent Speed Change Lanes (10-Percent Significance Level)*

X1 – AADT of the ramp (vehicles/day)
The range of values for this variable used in this model is 27-24,000 vehicles/day.
Suggested default: 3,000 veh/day
X2 – Area type
Suggested default: Urban
X3, X4, X5, X6 – Ramp configuration
Suggested default: Diamond

8. *Model for Entire Ramp plus Adjacent Speed Change Lanes (20-Percent Significance Level)*

X1 – AADT of the ramp (vehicles/day)
The range of values for this variable used in this model is 27-24,000 vehicles/day.
Suggested default: 3,000 veh/day
X2 – AADT of the mainline freeway (vehicles/day)
The range of values for this variable used in this model is 2300-106,000 vehicles/day.
Suggested default: 30,000 veh/day
X3, X4, X5, X6 – Ramp configuration
Suggested default: Diamond
X7 – Area type
Suggested default: Urban
X8 – Ramp type
Suggested default: Off-ramp

X9 – Speed-change lane length (miles)
The range of values for this variable used in this model is 0.04-0.5 miles.
Suggested default:	0.18 miles

X10 – Ramp length (miles)
The range of values for this variable used in this model is 0.14-0.82 miles.
Suggested default:	0.33 miles

Example of Software Use

For the purposes of this example, an engineer is given the following information about a rural, three-legged, stop-controlled intersection:

Major road AADT = 10000 veh/day
Minor road AADT = 4000 veh/day

The engineer must arrive at an estimate of crashes in the next three years. Information is not provided about left-turn channelization, functional classification of the major road, or access control on the major road. The reader will assume that these data are unavailable or would take too much time to find. Therefore, the engineer must use the defaults found in the previous section (i.e. no left-turn lane, minor arterial, and no access control). Figure 2 is a screenshot of this data being entered into TCPM.

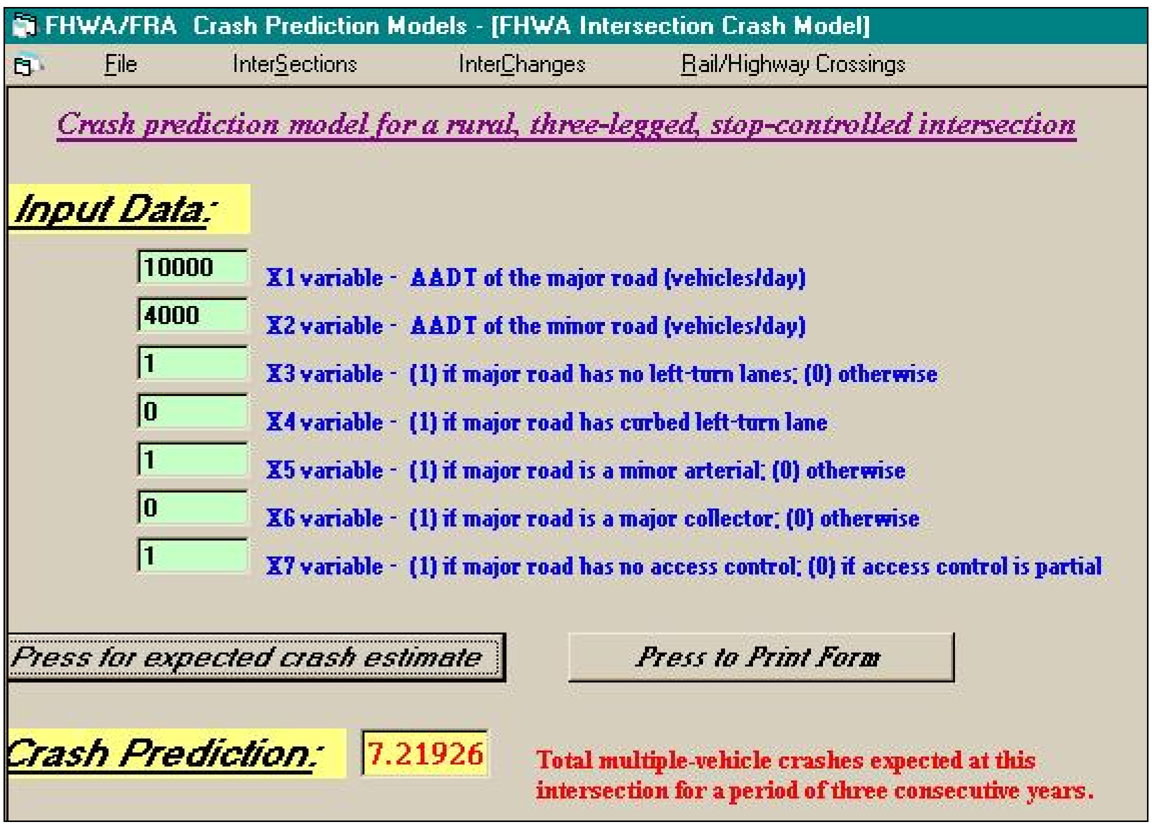


Figure 2. TCPM

The program has predicted **7.2** multiple-vehicle crashes over the next three years. This prediction can now be compared with the actual crash records of the intersection to determine if the site is abnormally hazardous.

Validity of Results

Tasks 9 and 10 in the project proposal required that the model be tested and its suitability determined. In order to test how well the crash predictions related to the actual North Carolina crash records, crash frequencies were predicted for a group of 24 intersections where the NCDOT installed left-turn lanes (treatment sites) and 24 intersections comparable to the 24 treatment sites (reference sites). The following is a breakdown of the types of intersections used in the validation study:

- Ten rural, three-legged, STOP-controlled intersections,
- Five rural, four-legged, STOP-controlled intersections,
- Two urban, three-legged, STOP-controlled intersections, and
- Seven urban, four-legged, signalized intersections.

The treatment sites were selected for the left-turn lane countermeasure because they previously had been judged as hazardous. The reference sites were selected only on the basis of their geometrical similarity to their partner site.

In collecting information about the intersections in the study, some of the site characteristics were not accessible or impractical to locate. To be able to supply all the variables for the models, default values were used in these situations. A list of the default values used by the NCSU team in the crash predictions can be found in Appendix G.

To compare the predictions against the actual crash records, two adjustments had to be made. Of the actual crash data collected, only the **multiple-vehicle** crashes were compared with the predicted crashes. This is because the intersection models are designed only to predict for multiple-vehicle crashes. For more information on what types of crashes are considered multiple-vehicle, see Appendix H.

TCPM outputs a crash estimate for a **three-year period**. In order to have an equal comparison, it was necessary to have the actual crashes in three-year increments. However, when the crash data were collected, the number of years varied from three to six. To adjust for this, the number of crashes was divided by the number of years that it spanned, to get crashes per year, and then multiplied by three. This gave the equivalent three-year crashes. For example, if the number of crashes collected was 10, and the years spanned was 5, the adjusted three-year average was (10/5)*3 or 6.

Analysis of Hazardous Site Rankings

To gain more insight into how the model will be used to judge the relative hazardousness of a site, it is helpful to look at how the "hazardous" sites according to the predictions compare with the hazardous sites according to the actual crash records. Tables 17-19 list the top five selections from both.

Both rankings of hazardousness are based solely on crash frequency. Ranking sites based on number of collisions can be biased towards high volume locations. The sites selected as hazardous by this method may not be the sites where countermeasures may be used most effectively.

On comparing the rankings, it is more revealing to examine separate group of sites rather than the set as a whole. The first group is the **treatment sites before the treatment** was installed (Table 17). These sites are not randomly selected; rather, as stated before, were selected for treatment because they have met the requirements for a hazardous site. For these sites, the prediction selections have three sites in common with the selections of the actual crashes (common sites shown in these tables in bold face).

Table 17. Five most hazardous treatment sites before treatment

Rank	*Predicted (Before)*	*Actual (Before)*
1	**078**	**065**
2	046	045
3	**013**	003
4	032	**013**
5	**065**	**078**

The second group to examine is the **treatment sites after treatment** was installed (Table 18). These sites are the same sites as in the previous section – therefore not randomly selected. Again, the same three sites are common to both columns.

Table 18. Five most hazardous treatment sites after treatment

Rank	*Predicted (After)*	*Actual (After)*
1	**078**	**065**
2	046	**078**
3	**013**	**013**
4	**065**	003
5	032	004

The third group is the **reference sites** (Table 19). Since these sites were chosen for this study only because of their geometric similarities to the treatment sites, it can be assumed that they are random sites with respect to hazardousness. In the ranking comparison, four of the five most hazardous sites are common to both columns.

Table 19. Five most hazardous reference sites

Rank	*Predicted*	*Actual*
1	078A	**046A**
2	**046A**	**065A**
3	**065A**	**013A**
4	**013A**	032A
5	**087A**	**087A**

Statistical Analysis

To further examine the predictions, it is useful to investigate the statistical characteristics of the groups of sites (Table 20). The mean, variance, and skewness of each column of actual crash records are higher than those for the predicted columns. The difference is greater for the **treatment sites** where the crash frequencies are greater and more varied. For the **reference sites**, where crash frequencies are lower, the difference in these statistics is smaller. In contrast to the treatment site predictions, the reference site predictions are over-predicting the number of crashes for each intersection.

Another way of examining the sites is to look at the correlation between the predicted and actual crashes. Correlation determines whether two ranges of data move together – that is, whether large values of one set are associated with large values of the other (positive correlation near one), whether small values of one set are associated with large values of the other (negative correlation near negative one), or whether values in both sets are unrelated (correlation near zero). The least correlation is that of the **treatment sites before treatment** (0.44). The greatest correlation is that of the **reference sites** (0.80).

Table 20. Statistical characteristics of sites

Treatment Sites before treatment

	Predicted (Before)	*Actual (Before)*
Mean	5.48	10.06
Variance	16.78	54.02
Skewness	1.52	0.68
Correlation	0.44	

Treatment Sites after treatment

	Predicted (After)	*Actual (After)*
Mean	5.03	8.51
Variance	15.11	71.89
Skewness	1.37	1.36
Correlation	0.58	

Reference Sites

	Predicted	*Actual*
Mean	3.96	3.38
Variance	9.16	7.64
Skewness	1.06	0.71
Correlation	0.80	

See Appendices D, E, and F for correlation plots for all three groups.

Problems and limitations

- The information required by the models is not always available or easily attained. The information on some features (functional classification, presence of lighting) is not stored on a centralized database in North Carolina.
- The intersection crash prediction is given for only multiple-vehicle crashes. If this prediction is used as an expected value for a given intersection, the actual number of crashes may be higher, due to the inclusion of unpredicted single-vehicle crashes.
- The crash prediction is given for a three-year period. If this number is to be compared with actual crash records, the user must take care to ensure that the actual crash data is in three-year increments or some average three-year increment.

Potential Use of the Software

A potential use of the crash prediction models programmed into TCPM is in creating an expected values table. This table would give the engineer or analyst a general expected value for a roadway site, given the major and minor traffic volumes.

Appendix I contains a sample expected values table. Since traffic volumes are the most important factor in predicting crashes, these traffic volumes could be projected to a given year. The engineer could then use these projected volumes along with known or default intersection characteristics to arrive at an expected number of crashes for some future time period. These values would be useful in such procedures as a benefit/cost analysis or a prioritization of future projects.

The expected values table in Appendix I is read using the major road ADT along the vertical left edge and the minor road ADT along the horizontal top edge. Follow the row and column of that corresponds to the traffic volumes to the intersecting block. That value is the expected number of total multiple-vehicle crashes for a three-year period.

A spreadsheet named “Expected Values Table” is included with the deliverables for this project. A description of this spreadsheet is given in the Future Software Applications section near the end of this report.

Discussions

Discussion of Models

The goodness-of-fit measures for the at-grade intersection models do not support a strong relationship between the model and the Caltrans data. The interchange and speed-change lane models had even weaker relationships between the models and the data. The relationship between the variables and predicted crashes is not very strong for any model according to the goodness-of-fit measures. The goodness-of-fit measures were very poor for all urban diamond off-ramp models. Although the models do not exhibit a strong relationship with the Caltrans data, these models seem to fit acceptably well with the test intersection data run by the project team.

Discussion of Validity Comparison

In this comparison, the models seemed to fit the reference sites better than the treatment sites. The abnormality of the treatment sites (more hazardous, higher frequency) restricted the accuracy of the predictions for these sites. The reference sites represent a fairly random sample of North Carolina intersections because of the way in which they were chosen for this study. Therefore, to analyze the performance of the crash prediction model on North Carolina data, one should place more trust in the reference sites.

From the ranking analysis, the crash predictions for the reference sites seem to be fairly accurate. However, a possible problem with this is that the sites that were selected as most hazardous are the ones with the highest traffic volumes. Therefore, if the engineer were ranking hazardous sites by crash frequency alone, he would select these sites with or without the model. The correlation coefficient is also important to this process. Its value shows how well the two trends move together. The coefficient of 0.80 for the reference sites indicates that the predicted and actual crash trends move together fairly closely.

For a visual representation of the accuracy of the crash predictions, one can look at the Correlation graphs in Appendices D, E, and F. To produce these graphs, the Predicted crashes were plotted against the Actual crashes. The ideal scenario would be that the Predicted crashes would match the Actual crashes exactly, and the graph would form a line with slope equal to one. This is shown on the graphs by the dashed line. Looking at the regression line fitted to the plotted points and comparing it to the ideal line can tell us much.

From Appendix D (treatment sites before treatment), we can see that the two slopes are fairly close. A Student’s T-test was run to compare the two slope values. The conclusion from the t-test was that the two slopes were not significantly different. This would seem to say that the regression line slope is statistically similar to the ideal slope.

However, there is a gap between the two lines. This shows that the actual crashes are higher than the predicted ones for this group of sites. Again, this is to be expected since these sites are the most hazardous sites, thereby necessitating the treatment.

From Appendix E (treatment sites after treatment), we can see that the slopes have about the same difference. Again, a t-test showed that the two slopes were not significantly different. In addition, the gap is a bit narrower in the low-crash region. This shows that the predictions are more accurate for low crash frequency intersections.

From Appendix F (reference sites), we can see that the data points are even closer together. The regression line is closer to the ideal line and intersects the ideal line. However, the t-test shows that the two slopes are significantly different.

These comparisons show that the crash predictions followed the general trend of the actual crashes. However the predictions are far from being a perfect fit with the actual data. A recalibration using North Carolina data would have good chances of improving the accuracy of the predictions. A cost analysis for this recalibration is presented in the Cost Analysis section.

Discussion of Software

TCPM is relatively easy to use. The only inputs are the values for the variables listed on the screen. The definition of each variable and directions for inputting data are placed next to the variable. The output is available on-screen at the click of a button.

The difficulties are presented in locating the appropriate information on the characteristics of a site. Many of the variables that were contained in the Caltrans database are not stored in a centralized database in North Carolina. Examples of these more obscure characteristics are the functional classification of the major road and the presence of lighting at an intersection.

One solution to this problem is the use of default values for characteristics that prove too difficult or time-consuming to find. Future users of this software may make use of the default values provided in this report or develop their own set of default values based on experience with the area under analysis.

Discussion of Results

The output of TCPM is an estimate of expected crashes over a three-year period. The intersection models are designed to estimate total **multiple-vehicle** crashes. If the user wishes to compare the crash predictions with actual intersection crash data, he will need to sort out single-vehicle crashes and only compare with actual multiple-vehicle crashes.

There are certain issues concerning usage of the model and the output. These issues are broached in the following paragraphs.

- **Upper and Lower 90% Confidence Levels**
  The rural, four-legged, STOP-controlled intersection model has two additional output options. These are the upper and lower 90% confidence limits of the regression coefficients. These procedures use different coefficients for each of the variables in the model, and thus show the range of estimates that would be possible if the upper or lower 90% confidence level were used. One will quickly see that the range of crash estimates is beyond practical use, and therefore is only included in the program to demonstrate what the output would be if these limits were used.

- **Significant figures**
  Although the Crash Prediction estimate is given to several decimal places, it should not be read past one or two significant figures. The estimate is not precise enough to predict beyond that.

- **The difference between 10 and 20 percent significance in the "Entire Ramp plus Adjacent Speed-change Lanes" models**

  The percent significance has mainly to do with which variables are included in the model. When a model is created, it is run with many variables at first. In the analysis afterwards, if the variables are not *statistically significant*, then they are taken out of the model. At 20 percent significance (a more relaxed standard), the user can see that more variables remained in the model than at the 10 percent significance level.

  The FHWA document recommends that "…caution be exercised in including variables that are not statistically significant at the 10 percent significance level in a predictive model…" (Bauer and Harwood 1998). There is less risk in predicting crashes with the 10 percent significance level model.

IV. Conclusions and Recommendations

Summary

The two objectives of this project involved software programs. The first objective was to recommend software for statistically valid countermeasure evaluation. The software BEATS was recommended to account for regression-to-the-mean bias in the data and to provide an evaluation on the performance of the countermeasure. This software is moderately easy to use, but the user will need to undergo some training in order to learn how to what type of data is needed, how it is to be entered into BEATS, and how to interpret the output. We have also recommended other software (and included it in our delivery to NCDOT) to estimate traffic volumes at the sites of interest during the years of interest.

The second objective was to write software to execute crash prediction models. The models chosen for intersections and interchanges were from FHWA-RD-97-106 and FHWA-RD-96-125 by Bauer and Harwood. The models chosen for railway crossings were taken from the NCDOT Rail Division office. The software TCPM was written to execute these models. The TCPM software is easy to use and the output is easy to understand. The difficulties with this program are in gathering all the data necessary for the models. However, default values may be substituted for site characteristics at the user's discretion.

Desired Software Modifications

The delivered software is an initial effort to combine and coordinate transportation safety programs and procedures into one software package. At the time of submission of this project, the TSEDS software is a base or foundation for transportation safety software that can be expanded or developed to best serve the NCDOT. The following is a list of desired or recommended software modifications or improvements for future consideration.

- *Link software program input data to database*
  In the current version of TSEDS, the database functions as a data organization feature. Most of the data contained therein must be extracted and then copied or re-created in an input file for use in the desired program. It would be more efficient if the software program input was automatically extracted directly from the database and used in the desired software program.

- *Report expected value range*
  The Crash Prediction Software in the current version of TSEDS may be used to prepare an expected values table. The results of the crash prediction models are reported as one number. Given the randomness of highway collisions and the less

than ideal goodness-of-fit values associated with the models, it would be more appropriate to report the expected value for crash frequency as a range of values.

- *Include interactive HELP feature and/or tutorial feature for each software program*
  A help button has been included on the switchboard underneath a few of the procedure buttons. However, these help buttons only open up the section of the User's Guide pertaining to that procedure. A more ideal help function would be an interactive type help, where the user could ask questions, or look up topics in an index.

- *Organize switchboard if future software programs are added*
  If more software programs are added to the switchboard, the numerous buttons could create clutter. A possible organization would be to include a button on the switchboard for a general task such as Countermeasure Evaluation and include buttons or links to all the relevant software programs for that task on a separate screen.

- *Incorporate other crash prediction models*
  The crash prediction program TCPM could be expounded to include other crash prediction models. As mentioned in the literature review in the Crash prediction section of the final report, other models currently exist. NCDOT currently categorizes high hazard locations as bridges, sections, or intersections. This report includes crash prediction models for certain intersection and interchange types. Other models may be included in this section of the software. This would require more data collection and input into the database. The database is included in this software for the purpose of accumulating and storing data for model input.

- *Add more software procedure links to the TSEDS switchboard*
  One of the most useful aspects of the TSEDS database software is that it contains links to other transportation programs such as The Highway Emulator and Safety Resource Allocation Program. This usefulness could be increased by adding links to other traffic programs. Two desirable types of programs are highway capacity software (i.e. Highway Capacity Software) and traffic simulation software (i.e. CORSIM).

Automated Expected Values Table

Instead of having numerous hard-copy pages of expected values tables covering every combination of site characteristics, another option is to have a type of automated expected values table on a spreadsheet. A sample spreadsheet is included with the deliverables for this project. This spreadsheet is provided as an example of an automated expected values table, and the user is advised to modify it according to the particular The spreadsheet is named "Expected Values Table" and was made with Microsoft® Excel. It contains seven sheets. The first two are for the user interaction and the last five are the individual expected value tables.

The **Model Inputs** sheet uses the same manner of input as the TCPM software. The first sheet of the spreadsheet is where the user enters the site characteristic data. The shaded gray cells are the only cells that the user can manipulate. All other cells on the sheets are locked against changes. The coefficients listed under each model are only for the spreadsheet calculations. If a model has a '0' for a coefficient next to a certain characteristic, it doesn't include that characteristic.

The **Suggested Defaults** sheet contains the suggested default characteristics for use if the user does not have access to one or more of the characteristics. These defaults were chosen by the NCSU team as the most likely characteristics for North Carolina intersections, however, the user should feel free to develop new default values for the area under analysis.

The remaining five sheets (**R-4-STOP**…**U-4-Sig**) contain the automated expected value tables. Once the values are entered into the gray cells of the Model Inputs sheet, the values in the table are updated automatically. To find an expected crash value for a particular set of traffic volumes, simply trace down the column from the appropriate minor road ADT and find the cell that intersects with the row from the appropriate major road ADT.

Each table is ready to be printed as soon as the data is entered into the Model Inputs sheet. These tables are convenient because the appropriate table can be printed out on a single sheet of paper to include with a report. Also, with the input table included in the spreadsheet, the expected value tables can be tailored to a specific intersection.

Software Users

Traffic Safety Systems Management Unit, NCDOT

The TSSMU would be the primary user of the software. The current version of the software could be used to evaluate countermeasures and predict crashes at intersections and interchanges.

Project Development & Environmental Analysis Branch, NCDOT

Crash prediction may also be of interest to the Project Development and Environmental Analysis Branch (PD & EA) of the NCDOT, which is responsible for early project planning. The branch currently produces planning documents for projects that includes crash history information. The capability to quickly predict crash frequencies would be of value to the Branch in the discussion of safety impacts in planning documents for new construction or 3R projects.

An expected value table may also be of use to the Branch in discussing safety impacts for projects. This would eliminate the need for the crash prediction software mentioned above. With additional good crash prediction models, expected values tables may be stored together used for the appropriate situation. The expected values tables may change too often to provide the PD & EA branch with hardcopies of the tables. Most likely, the TSSMU should be responsible for creating, updating and maintaining the expected values tables.

Rail Division, NCDOT

The Crash Prediction Software program of the TSEDS currently contains crash prediction models for highway/railroad crossings. This may be of particular interest to the Rail Division and the section of the Traffic Engineering Branch that monitors highway/railroad crossings. This software may be used in the future to create expected values tables or to identify high hazard highway/railroad crossings.

Recalibration of Bauer and Harwood Models

The models created by Bauer and Harwood were calibrated on California intersections and Washington interchanges. To use these models in North Carolina, one must assume that conditions here are similar to conditions in those states. In the event that a user of these models wishes to calibrate the coefficients specifically for North Carolina intersections, it will be necessary to collect data on the effect of certain characteristics on crash frequency. Recalibrating the models using North Carolina data would give the models more credibility when used by the NCDOT even if the goodness-of-fit measures do not improve.

The recalibration effort would not require a complete redevelopment of the models. The existing variables for each model would be used to eliminate the need to collect data and determine the significance of variables that may be excluded in the final model. Independent variables in the current models would just as adequately represent North Carolina data as California data.

Relative Importance of Variables

As this would be a large data collection effort, some guidance on the important characteristics would be helpful. Tables 21 and 22 rank the model variables by their effect on the final crash prediction. The ranking of the variables is based on the *value of their term* in the model. As this method of ranking may be unclear, the following example is provided.

Recall that the general prediction equation is such:

$$Y = \exp(\beta_0)(ADT_{\text{major road}})^{\beta 1}(ADT_{\text{crossroad}})^{\beta 2} \exp(\beta_3 X_{i3}) \exp(\beta_4 X_{i4}) \bullet \ldots \bullet \exp(\beta_q X_{iq})$$

Where Y = Prediction of multiple-vehicle crashes for a three-year period
X = Input variables

β = Model regression coefficients

If the following is given:

Major road ADT = 20,000 veh/day

$\beta 1$ = 0.797 (for a rural, 4-legged, STOP-controlled intersection)

Its value in the model equation would be:

Term value = $(ADT_{major\ road})^{\beta 1} = (20{,}000)^{0.797}$ = **2678.68**

This term value shows the effect that the coefficient has on the final outcome, basically how important the variable is compared to the other variables. Appendix K shows the term values for all of the models. Appendix K shows that the ADT's for the major and minor roads are by far the most important factor in the equation. This concurs with the literature and with the common sense notion that traffic volumes are the largest factor in predicting crashes. With the information in these tables, it is expected that the intersection characteristics can be prioritized and gathered at the analyst's discretion.

Data Collection

The most time consuming and costly process in model recalibration is data collection. Data must be gathered for each independent variable at randomly selected intersections and interchanges throughout the state. A sample size of 1500 intersections and 1500 ramps were used to estimate the cost of data collection. Bauer and Harwood used approximately the same sample size in the calibration of their models using Caltrans data. A list of independent variables for which data must be gathered are listed in Tables 23 (intersections) and 24 (interchanges) along with an estimate of the time and cost required to collect the data for one site.

Because the NCDOT inventory database is not as developed as the databases that Bauer and Harwood used, data collection through mining inventory databases would be substantially less productive for North Carolina sites than it was for California and Washington. For this reason, a field visit to each site was deemed a more cost-effective method of collecting site data.

The characteristics used by the models are not ones that require special instruments or tedious measurements. Most of the site characteristics can be determined by a simple glance at the site. Taking into account the travel time between sites, each intersection field visit was estimated to require two hours and each ramp visit one hour.

Task 11 in the project proposal required an estimation of the cost involved in recalibration. A cost estimate of $15.00 per hour was used to calculate the cost of data collection for all variables. The time estimated for data collection could vary greatly depending on how well the data collection effort is planned and organized and how experienced the data collectors are.

The cost estimate was divided into two sets – intersection models and interchange models. An estimated value of 1500 sites was used for both cost estimates. As shown in Tables 23 and 24, the total preliminary cost of data collection for intersection recalibration is $64,000, and the preliminary cost for interchange recalibration is $38,000. This reflects a total data collection time of 4200 hours for intersections and 2500 hours for the interchanges. For an accurate estimation, the final cost estimate should take into account a 20% increase in cost due to benefits, and a 47% increase due to overhead. The total cost for intersection recalibration is $112,000 and the total cost for interchange recalibration is $68,000.

Data Analysis

After the data are collected, they must be analyzed to determine the model coefficients. The total time estimated to analyze the collected data was 2 months for each set of sites. Considering that the analysis will require a professional statistician, the total cost of data analysis for recalibration was estimated to be $13,000 for each set of sites, for a total cost of $26,000. Adding the 20% increase for benefits and 47% increase for overhead, the final total cost for analysis is $45,800.

Table 23. Intersection Data Collection Preliminary Cost Analysis

		Number of Intersections =		1500
		Hours per		
Variables	**Data Location**	**Intersection**	**Cost per hour**	**Total Cost**
ADT of crossroad (veh/day)	Traffic Survey Unit	0.25	$15.00	$5,625.00
ADT of major road (veh/day)	Traffic Survey Unit	0.25	$15.00	$5,625.00
Number of lanes on Major Road	Field visit			
Number of lanes on Minor Road	Field visit			
Major road average lane width (ft)	Field visit			
Design speed on major road (mph)	Field visit			
Functional Classification of Major Road	Field visit			
Access control on major road	Field visit			
Terrain	Field visit			
Presence of divided median on major road	Field visit			
Presence of free right turns	Field visit			
Left-turn lane prohibition	Field visit			
Left turn channelization on major road	Field visit			
Lighting at intersection	Field visit			
Signal timing	Field visit			
Signal phasing	Field visit			
	Total field visit time =	2.00	$15.00	$45,000.00
Crash data	TEAAS	0.33	$15.00	$7,425.00
	Total hours/intersection =	**2.83**		
	Total hours =	**4245**	**Total Cost =**	**$63,675.00**

Table 24. Interchange Data Collection Preliminary Cost Analysis

		Number of Ramps =		1500
		Hours per		
Variables	**Data Location**	**Ramp**	**Cost per hour**	**Total Cost**
Ramp AADT	Traffic Survey Unit	0.25	$15.00	$5,625.00
Freeway AADT	Traffic Survey Unit	0.125	$15.00	$2,812.50
Ramp Segment length (mi)	Field visit			
Speed Change Lane Length	Field visit			
Right shoulder width (ft)	Field visit			
Ramp number of lanes	Field visit			
Ramp Configuration	Field visit			
Ramp Type	Field visit			
Average lane width (ft)	Field visit			
Area type	Field visit			
	Total field visit time =	1.00	$15.00	$22,500.00
Crash Data	TEAAS	0.33	$15.00	$7,425.00
	Total hours/intersection =	**1.71**		
	Total hours =	**2558**	**Total Cost =**	**$38,362.50**

V. References

Bared, J.G. and A.Vogt, "Accident Models for Two-Lane Rural Roads: Segments and Intersections," Federal Highway Administration Final Report, FHWA-RD-98-133, October 1998.

Bared, J.G. and A.Vogt, "Highway Safety Evaluation System for Planning and Preliminary Design of Two-Lane Rural Highways," Presented at 76th Annual Meeting of the Transportation Research Board, Washington, D.C., January 1997.

Bauer, K.M. and D.W. Harwood, "Statistical Models of At-Grade Intersection Accidents," Final Report, FHWA-RD-96-125, McLean, VA, November 1996.

Bauer, K.M. and D.W. Harwood, "Statistical Models of At-Grade Intersection Accidents-Addendum," Final Report, FHWA-RD-99-094, McLean, VA, March 2000.

Bauer, K.M. and D.W. Harwood, "Statistical Models of Accidents on Interchange Ramps and Speed-Change Lanes," Final Report, FHWA-RD-97-106, McLean, VA, June 1998.

Braam, A.C., R.J. Jaeger, "North Carolina Highway Safety Improvement Program Spring 2000," Traffic Safety Systems Management Unit, North Carolina Department of Transportation, February 2000, rev. March 2000.

Bromage, E., "The Highway Emulator," Release 3.0, October 1988.

Dzbik, L. and B. Persaud, "Accident Prediction Models for Freeways," Transportation Research Record 1401, Transportation Research Board, Washington, D.C., 1993.

Glennon, J.C., T.R. Neuman and J.E. Leisch, "Safety and Operational Considerations for Design of Rural Highway Curves," FHWA-RD-86-035, Washington, DC., December 1985.

Hadi, M.A., J. Aruldhas, L.F. Chow, and J.A. Wattleworth, "Estimating Safety Effects of Cross-Section Design for Various Highway Types Using Negative Binomial Regression." Transportation Research Record 1500, TRB, 1995, p. 169.

Hauer, E., "Observational Before-After Studies In Road Safety," Pergamon, New York, 1997.

Hummer, J. et al, "Accidents on Secondary Highways and Countermeasures, Phase II: Identification of Promising Sites on Rural, Two-Lane Highways Using Inventory Data," Final Report, North Carolina State University, July 1999.

Kalokota, K.R. and P.N. Seneviratne, "Accident Prediction Models for Two-Lane Rural Highways," MPC Report No. 94-32, May 1994.

Lau, M.Y., A.D. May, "Accident Prediction Model Development for Unsignalized Intersections: Final Report," California Department of Transportation, FHWA/CA/TE/89-3, May 1989.

Lau, M.Y., A.D. May, "Accident Prediction Model Development for Signalized Intersections: Final Report," California Department of Transportation,
ITS-UCB-RR-88-7, May 1988.

Neuman, T.R., J.C. Glennon and J.E. Leisch, "Functional Analysis of Stopping Sight Distance Requirements," Transportation Research Record 923, TRB, 1984, p. 57.

Parker, M. et al, "Highway Safety Engineering Studies Procedural Guide," U.S. Department of Transportation, Federal Highway Administration, March 1991.

Pendleton, O.J., "Application of New Accident Analysis Methodologies, Volume I – General Methodology," FHWA-RD-90-091, McLean, VA, September 1991.

Pendleton, O.J., O. Gonzalez, and H. Duarte, "Application of New Accident Analysis Methodologies, Volume II – A Users Manual for BEATS," FHWA-RD-91-014, McLean, VA, September 1991.

Pendleton, O.J., O. Gonzalez, and H. Duarte, "Application of New Accident Analysis Methodologies, Volume III – A Users Manual for BEATS," FHWA-RD-91-014, McLean, VA, September 1991.

Pendleton, O., "A Systemwide Methodology for Evaluating Highway Safety Studies," FHWA-RD-92-049, McLean, VA, October 1992.

Turner, D.S., "Prediction of Bridge Accident Rates," Journal of Transportation Engineering, Volume 110, Number 1, January 1984, p.45.

Zegeer, C. et al., "Cost-Effective Geometric Improvements for Safety Upgrading of Horizontal Curves," Final Report, FHWA-RD-90-021, McLean, VA, October 1991.