2010-09 Final Report

DEVELOPMENT OF SAFETY PERFORMANCE FUNCTIONS FOR NORTH CAROLINA

Raghavan Srinivasan

Phone: 919-962-7418; Email: srini@hsrc.unc.edu

Daniel Carter

Phone: 919-962-8720; Email: carter@hsrc.unc.edu

Performing Agency:

University of North Carolina Highway Safety Research Center

730 Martin Luther King Jr Blvd

Chapel Hill, NC 27599-3430

Submitted to North Carolina Department of Transportation

December 6, 2011

Technical Report Documentation Page

1. Report No. FHWA/NC/2010-09	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Development of Safety Performance Functions for North Carolina		5. Report Date December 6, 2011
		6. Performing Organization Code
7. Author(s) Raghavan Srinivasan and Daniel Carter		8. Performing Organization Report No.
9. Performing Organization Name and Address University of North Carolina Highway Safety Research Center 730 Martin Luther King Jr Blvd Chapel Hill NC 27599-3430		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 Raney Building, 104 Fayetteville Street Raleigh, North Carolina 27601		13. Type of Report and Period Covered Final Report August 16, 2009 – June 30, 2011
		14. Sponsoring Agency Code 2010-09
Supplementary Notes:		

16. Abstract

The objective of this effort is to develop safety performance functions (SPFs) for different types of facilities in North Carolina and illustrate how they can be used to improve the decision making process. The prediction models in Part C of the Highway Safety Manual were calibrated using data from North Carolina. SPFs were estimated for 9 crash types for 16 roadway types using statewide data from North Carolina – these SPFs just include AADT and can be used for network screening. In addition, SPFs for rural two lane roads were developed using AADT and other site characteristics including shoulder width/type and terrain. Using examples, this report outlines how the different SPFs can be used for network screening, project level analysis, and before-after evaluation using the empirical Bayes method. Finally, there is a discussion of how NCDOT can use the SPFs in the future by either calibrating the SPFs developed in this effort and/or develop SPFs using negative binomial regression. Along with the final report are Excel files that can be used by NCDOT to calibrate the SPFs in the future and to implement the HSM prediction methodology.

17. Key Words *Highway Safety Manual; Calibration; Safety Performance Function; Negative binomial regression; SafetyAnalyst; Empirical Bayes; Network screening; Before-after*		18. Distribution Statement	
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 88	22. Price

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

DISCLAIMER

ACKNOWLEDGMENTS

The authors would like to thank the North Carolina Department of Transportation for funding this effort. The authors thank the members of the Steering and Implementation Committee for their guidance throughout the process:

- Brian Murphy
- Tony Houser
- Brian Mayhew
- Shawn Troy
- Terry Hopkins
- Kevin Lacy
- Christopher Oliver
- Missy Pair
- Anthony Wyatt
- Ernest Morrison

Special thanks to Brian Murphy, the Chair of the Committee for all his help. We thank Nrupen Mandava, graduate student from North Carolina State University who worked diligently to compile the data from aerial photographs and GIS files that were necessary for the calibration. We also thank Yusuf Mohamedshah for his help in providing the roadway inventory data from HSIS.

SUMMARY

The objective of this project is to develop safety performance functions (SPFs) for different types of facilities in North Carolina and illustrate how they can be used to improve the decision making process. SPFs are essentially mathematical equations that relate site characteristics of a road segment or intersection to the number of predicted crashes at that site.

The AASHTO Highway Safety Manual (HSM) has predictive methods that make use of SPFs. In order to use these prediction methods in North Carolina, they need to be calibrated using data from North Carolina. This project calibrated the HSM predictive models for the following facility types:

Roadway Segments

- Rural 4 Lane Divided
- Urban 2 Lane Undivided (2U)
- Urban 2 Lane with TWLTL (3T)
- Urban 4 Lane divided (4D)
- Urban 4 Lane Undivided (4U)
- Urban 4 Lane with TWLTL (5T)

Intersections

- Rural 2 Lane, minor road stop controlled 3-leg (3ST)
- Rural 2 Lane, signalized 4-leg (4SG)
- Rural 2 Lane, minor road stop controlled 4-leg (4ST)
- Rural 4 Lane, signalized 4-leg (4SG)
- Urban arterial, signalized 3-leg (3SG)
- Urban arterial, minor road stop controlled 3-leg (3ST)
- Urban arterial, signalized 4-leg (4SG)
- Urban arterial, minor road stop controlled 4-leg (4ST)

To calibrate the prediction models, roadway, roadside, traffic, and crash data were compiled for a sample of sites in North Carolina. The intent was to identify a sufficient number of sites that will provide at least 100 crashes per year for each type of facility. Data were compiled from aerial photographs, GIS files, roadway inventory files from the Highway Safety Information System (HSIS), and NCDOT's TEAAS database. The HSM prediction methods were used to compute the calibration factor for total crashes for the facility types mentioned above.

In addition to calibrating the HSM prediction models using sample data for North Carolina, the project also developed SPFs for roadway segments using statewide data from North Carolina. This was possible for roadway segments since the data for these segments are available through HSIS. Using data from TEAAS, segments within the influence of at grade intersections and railroad grade crossings (250 feet on either side of at grade intersections or railroad grade crossings) were removed. In addition, freeway segments within 0.5 miles of an interchange were designated as within the influence of interchanges. SPFs with just AADT as the independent variable (type 1 SPFs) were estimated for the following 16 roadway types:

- Rural Two Lane Roads
- Rural Freeways - 4 lanes - outside the influence of interchanges
- Rural Freeways – 6+ lanes - outside the influence of interchanges
- Rural Freeways - 4 lanes - within the influence of interchanges
- Rural Freeways – 6+ lanes - within the influence of interchanges
- Rural Multilane Divided Roads
- Rural Multilane Undivided Roads
- Urban Two Lane Roads
- Urban Freeway - 4 lanes - outside the influence of interchanges
- Urban Freeway - 6 lanes - outside the influence of interchanges
- Urban Freeway - 8+ lanes - outside the influence of interchanges
- Urban Freeway - 4 lanes - within the influence of interchanges
- Urban Freeway - 6 lanes - within the influence of interchanges
- Urban Freeway - 8+ lanes - within the influence of interchanges
- Urban Multilane Divided Roads
- Urban Multilane Undivided Roads

The SPFs were estimated for 9 different crash types that were identified to be of primary importance to NCDOT. The type 1 SPFs can be used for network screening. In addition, SPFs for rural two lane roads were estimated by including other site characteristics such as shoulder width/type and terrain (type 2 SPFs).

By providing examples, the report shows how the different SPFs can be used for network screening, project level analysis to compare the safety implications of different design decisions, and evaluation of the safety effect of engineering treatments using before-after empirical Bayes methods.

The report concludes with a discussion of how NCDOT can update the SPFs in the future by either developing SPFs using negative binomial regression or calibrating the existing SPFs. Along with the final report are Excel files that can be used by NCDOT to calibrate the SPFs in the future and to implement the HSM prediction methodology.

Contents

1. INTRODUCTION

The objective of this project is to develop safety performance functions (SPFs) for different types of facilities in North Carolina and illustrate how they can be used to improve the decision making process. This section of the report starts with a discussion of the meaning of SPFs, followed by a discussion of the different applications of SPFs, statistical issues related to the development of SPFs, a brief examination of issues regarding the development of state specific SPFs versus calibration of existing SPFs, and a brief overview of SPFs that have been developed in previous studies using data from North Carolina. The last part of this section outlines the structure of the rest of the report.

What are Safety Performance Functions (SPFs)?

SPFs are essentially mathematical equations that relate the number of crashes of different types to site characteristics. Depending on the application and the facility type, examples of site characteristics include traffic volume (AADT), lane width, shoulder width, radius/degree of horizontal curves, presence of turn lanes (at intersections), and traffic control (at intersections). Following is an example of a crash prediction model for non-intersection accidents on rural two lane roads developed by Vogt and Bared (1998) using data from Washington and Minnesota. This model was estimated based on detailed information on the location of horizontal and vertical curves within each section.

$$
\begin{aligned}
&Y = EXPO_m \exp(0.165)\exp(-0.278 LW_m - 0.194 SHW_m + 0.0668 RHR + 0.0135 DD_m + 0.139 STATE) \\
&* (\textstyle\sum_i WH\{i\}\exp(0.0137 DEG_m\{i\}))(\textstyle\sum_j WV\{j\}\exp(0.142 V_m\{j\}))(\textstyle\sum_k WG\{k\}\exp(0.105 GR\{k\}))
\end{aligned}
$$

where,

Y = predicted mean number of non-intersection accidents on the segment

$EXPO_m$ = traffic exposure in millions of vehicle kilometers

LW_m = lane width in meters

SHW_m = average of left and right shoulder widths in meters

RHR = average roadside hazard rating along segment

DD_m = driveway density in driveways per kilometer

STATE = 0 for Minnesota, 1 for Washington

$DEG_m\{i\}$ = degree of curve in degrees per hundred meters of the i-th horizontal curve that overlaps the segment

WH{i} = fraction of the total segment length occupied by the i-th horizontal curve

$V_m\{j\}$ = absolute change in grade in percent per hundred meters of the j-vertical crash curve that overlaps the segment
WV{j} = the fraction of the total segment length occupied by the j-th vertical crest curve
GR{k} = absolute grade in percent of the k-th uniform grade section that overlaps the segment
WG{k} = fraction of the total segment length occupied by the k-th uniform grade section

Applications of SPFs

The Highway Safety Manual (HSM, 2010) outlines at least three different ways in which SPFs can be used by jurisdictions to make better safety decisions. One application is to use SPFs to determine the safety impacts of design changes at the project level. The second application is to use SPFs as part of network screening to identify sections that may have the best potential for improvements. The third application is the use of SPFs as part of an empirical Bayes before-after study to evaluate the safety effects of engineering treatments. Here is a brief discussion of these applications:

Determining the Expected Safety Impacts of Design Changes at the Project Level

Part C of the HSM provides prediction methods for estimating the *average expected crash frequency* of a site/project. The prediction methods can be used for estimating the average expected crash frequency for existing conditions, alternatives to existing conditions, or proposed new roadways. For roadway sections, Part C of the HSM provides prediction methods for the following road types:

- Rural two lane roads
- Rural four-lane divided and undivided roads
- Two lane, three lane (with center TWLTL), four lane divided, four lane undivided, and five lane roads (with center TWLTL) in urban and suburban arterials

For intersections, prediction methods are available for:

- Three and four leg minor road stop controlled and four leg signalized intersections on rural two lane roads
- Three and four leg minor road stop controlled and four leg signalized intersections on rural four lane roads
- Three and four leg minor road stop controlled and signalized intersections on urban and suburban arterials

Identifying Locations with Promise (Network Screening)

Another application of SPFs is to identify locations with promise, i.e., locations that may benefit the most by the application of some treatment. This is also called network screening. In this application, SPFs can be used to estimate the average number of crashes for a particular traffic volume for a particular facility type. This average can then be compared with the actual number of crashes at a particular site or the empirical Bayes estimate of the expected number of crashes at that site to determine if that site (compared to other sites) should be identified as a site with promise (SPFs can be used as part of the empirical Bayes procedure to estimate the expected number of crashes).

Evaluation of the Effect of Engineering Treatments

Most safety researchers agree that before-after studies provide more reliable estimates of the safety effect of engineering treatments compared to cross-sectional comparisons of locations with and without a particular treatment. However, since many engineering treatments are implemented at locations that may have a higher than normal accidents, before-after studies need to account for potential bias due to regression to the mean. The empirical Bayes procedure developed by Hauer (1997) is designed to do that. SPFs are an integral part of this empirical Bayes procedure.

Statistical Issues Related to the Development of SPFs

Earlier, SPFs were developed by trying to relate crash rates (ratio of crash frequency with exposure) to other site characteristics apart from traffic volume such as lane and shoulder width. Using crash rates as the dependent variable implicitly assumes that crash frequency and exposure are linearly related. Although it is possible that crash frequency and exposure are linearly related in some situations, in many cases the relationship is non-linear. By including exposure as one of the independent variables in the SPF, the non-linear relationship between crash frequency and exposure can be appropriately addressed (Shankar et al., 1995).

Some of the earlier SPFs used conventional linear regression to model the relationship between crash frequency and site characteristics. Linear regression assumes that the dependent variable (i.e., crash frequency) is normally distributed. Many studies have shown that this is not a correct assumption. In addition, linear regression can predict values that are negative, which are obviously inconsistent with crash counts that are either zero or positive. Crashes are *count data* (i.e., non-negative integers), and they are properly modeled using certain types of methods among which poisson and negative binomial regression are the most popular (Washington et al., 2011). In a poisson regression model, the probability of site i having y_i accidents per year (where y_i is a non-negative integer) is given as follows:

$$P(y_i) = \frac{\exp(-\lambda_i)\lambda_i^{\,y_i}}{y_i!} \tag{1}$$

where $\lambda_i = f(\beta X_i)$ (2)

X_i is the set of site characteristics, β represents the set of coefficients that need to be estimated, and *f* is a function that relates the site characteristics to the poisson parameter λ_i (expected number of crashes per year at the site). The limitation of the poisson distribution is that the mean and variance are considered equal. Most often with crash data, the variance has been found to exceed the mean. This phenomenon is called overdispersion. Negative binomial regression is able to account for this overdispersion by allowing the variance to differ from the mean as follows:

$$Var(y_i) = E(y_i) + k[E(y_i)]^2 \tag{3}$$

where k is the overdispersion parameter, Var is the variance and E is the expected value (i.e., mean). Negative binomial regression has become the most common method for developing SPFs and is also the recommended modeling approach in the HSM.

In both poisson and negative binomial regression, the most common function *f* is the exponential function. In other words, the relationship can be written as follows:

$$\lambda_i = \exp(\beta X_i) \tag{4}$$

The exponential function implies a log-linear relationship between site characteristics and the expected number of crashes per year at the site (i.e., $\ln(\lambda_i) = \beta X_i$, where ln represents the natural logarithm). The log-linear relationship has become common because it allows the poisson and negative binomial regression models to be estimated using a technique called generalized linear models (McCullagh and Nelder, 1989). In fact, all the predictive models reported in Part C of the HSM assumed a log-linear relationship between site characteristics and the expected number of crashes. More recently, some researchers have argued that other functional forms (other than the log-linear relationship) need to be investigated since the log-linear model may not always be the most appropriate relationship between the expected number of crashes and site characteristics. For example, Hauer (2004) discusses a method that involves the identification of the appropriate functional form for each independent variable (i.e., site characteristic) one at a time as they are included in a model. The chosen functional form for a particular independent variable is then revised in an iterative process as other variables are included because the independent variables may be correlated with each other. Compared to the traditional log-linear approach, this new approach is more time consuming and has had limited application so far (e.g., Baek and Hummer, 2008). Other methods that have tried to make use of more flexible functional forms for the relationship between crash frequency and site characteristics include generalized additive models (GAM) (Xie and Zhang, 2008) and use of neural networks (Kononov et al., 2008). Again, these new methods have had limited application probably because of concerns about the generalizability and the complexity of these types of models (Xie et al., 2007).

Development of North Carolina Specific SPFs versus Calibration of Existing SPFs

There are two ways to come up with SPFs for a particular facility. One approach is to develop state specific (in our case, North Carolina) SPFs using state of the art methods (i.e., negative binomial regression). The second approach is to calibrate SPFs developed by others (especially the HSM) using data from North Carolina. The second approach would involve the calculation of calibration factors to adjust the SPF-predicted crashes to be more accurate for North Carolina sites. North Carolina specific SPFs are expected to provide more accurate results. However, the sample of sites and the number of crashes necessary to develop North Carolina specific SPFs is much higher than the sample that is necessary for calibrating the SPFs using data from North Carolina. The ultimate aim should be to develop North Carolina specific SPFs. However, this is not possible for some of the facility types because data may not be available for a sufficient sample of sites.

The decision on the appropriate approach may also depend on the specific application of the SPF. For network screening, the aim is to have an SPF that has only the most critical information about the site (in most cases this is the AADT). In fact, the SPFs in Module 1 of SafetyAnalyst[1] (for network screening) include only AADT. Since the majority of North

[1] SafetyAnalyst provides analytical tools for use in the decision-making process to identify and manage a systemwide program of site-specific improvements to enhance highway safety. SafetyAnalyst includes a network screening tool, diagnosis tool, countermeasure selection tool, economic appraisal tool, priority ranking tool, and a countermeasure evaluation tool (www.safetyanalyst.org). SafetyAnalyst is available from AASHTO as a licensed AAHSTOWare product.

Carolina's state roadway system is part of FHWA's Highway Safety Information System (HSIS), developing SPFs with just AADT for different types of roadway segments was feasible within this project.

On the other hand, if an SPF will be used for crash prediction at a project level, it needs to include other site characteristics apart from AADT to provide the best possible prediction of the expected number of crashes. For example, according to the HSM, in order to accurately predict the expected number of crashes on minor road stop controlled 3 leg intersections on rural two lane roads, the following variables are critical:

- Major road AADT
- Minor road AADT
- Intersection skew angle
- Presence/absence of left turn lanes on the major road
- Presence/absence of right turn lanes on the major road
- Presence/absence of intersection lighting

Most states (including North Carolina) do not have this level of information for a sufficient number of stop controlled intersections in an electronic database in order to allow the development of state specific SPFs. Hence, in this situation, calibration of existing models would be appropriate.

Safety Performance Functions Developed in Other Studies Using North Carolina Data

Following is a list of SPFs that have been developed or calibrated using recent data from North Carolina.

Signalized Intersections in Urban Areas

Srinivasan et al., (2008a) estimated SPFs using data from urban signalized intersections in Winston-Salem. Some of these intersections may have been on state roads. Data from 1991 to 2004 for 60 signalized intersections were utilized to develop an SPF for total crashes (this sample had a total of 4,235 crashes). Information on percentage of angle, nighttime, nighttime angle, and left turns on major roads, were used to develop the SPFs for these crash types as well. The independent variables in the SPFs included major road AADT, minor road AADT, and the number of legs (3 or 4). These SPFs were used as part of the empirical Bayes method to evaluate the safety impacts of selected treatments including changes in left turn phasing, conversion of nighttime to regular phasing, conversion of 8-inch to 12-inch signal heads, and introduction of dual red-signal lenses. This effort was funded by NCHRP Project 17-25 and the final report was published as NCHRP Report 617 (Harkey et al., 2008).

Council et al., (2005) developed SPFs using combined data from Charlotte, NC, and Baltimore, MD. An SPF was developed for total crashes. Information on percentage of injury, right angle, and rear end crashes, were used to develop the SPFs for these crash types as well. Seventy intersections from Charlotte and 86 intersections from Baltimore, MD, were used for developing the SPFs. The SPFs were used as part of the empirical Bayes before-after evaluation of introducing red light cameras.

Harwood et al., (2007) developed SPFs for single vehicle and multiple vehicle crashes using combined data from Charlotte, NC, and Minnesota. These models have been included in Chapter 12 of Part C of the HSM. The independent variables in the SPFs included major and minor road AADT. Separate SPFs were developed three levels of severity: all, injury and fatal, and PDO. SPFs were developed using data from 42 3-leg intersections and 44 4-leg signalized intersections on urban and suburban arterials.

Harwood et al., (2008) developed SPFs for pedestrian crashes using combined data from Charlotte, NC, and Toronto, Canada. These models have been included in Chapter 12 of Part C of the HSM. The independent variables in the SPF included major and minor road AADT, pedestrian crossing volumes, and the maximum number of lanes to be crossed by a pedestrian at a leg after accounting for refuge islands. For Charlotte, data from 1997 to 2005 for 84 3-leg intersections and 267 4-leg intersections were utilized for the analysis. The 3-leg intersections experienced 47 pedestrian accidents and the 4-leg intersections experienced 294 crashes during the study period. Harwood et al., (2007) and Harwood et al., (2008) were both funded by NCHRP Project 17-26.

Srinivasan et al. (2011) developed SPFs for total, injury and fatal, left turn opposing through, and rear end crashes as part of NCHRP Project 17-35. These SPFs were used as part of the empirical Bayes before-after study to evaluate the safety of changes in left turn phasing. Data from 49 signalized intersections in North Carolina were used for estimating the SPFs.

Minor Road Stop Controlled Intersections in Urban Areas

As mentioned earlier, Harwood et al., (2007) developed SPFs for single vehicle and multiple vehicle crashes using combined data from Charlotte, NC, and Minnesota. Models were developed for minor road stop controlled intersections as well. Forty seven 3-leg intersections and 48 4-leg intersections from Charlotte were used for these models.

Minor Road Stop Controlled 4-leg Intersections on Rural Two Lane Roads

Srinivasan et al., (2008b) developed SPFs for total crashes, angle crashes, injury and fatal crashes, and rear-end crashes for minor road stop controlled 4-leg intersections on rural two lane roads. There were a total of 231 intersections in the sample that was used for developing the SPFs. Data from 1990 to 2004 were included. These intersections experienced a total of 3,792 crashes during the study period. Independent variables included major road AADT, minor road AADT (if available), and area type (rural versus urban/suburban). The SPFs were used as part of an empirical Bayes before-after study to evaluate the safety of overhead and stop sign mounted flashing beacons. This effort was funded by FHWA as part of the low cost pooled fund effort.

Stop Controlled Intersections on Rural Multilane Roads

Dr. Joseph Hummer and his colleagues at NC State University evaluated the safety of super street intersections on rural multilane roads (Hummer et al., 2010b). This effort was funded by NCDOT. These intersections were controlled by stop signs on the minor roads before the super street design was implemented. As part of their safety analysis of superstreets, they calibrated predictive models from the Highway Safety Manual for North Carolina roads. Specifically, they developed calibration factors for rural multilane stop-controlled intersections with three and four legs. They reported a calibration factor of 1.57 for three-leg intersections (total crashes) and a factor of 1.39 for four-leg intersections (total crashes). They developed these calibration factors using data from 2004 to 2009.

Rural two lane roads

Hummer et al. (2010a) examined curve crash characteristics, developed a manual field investigation procedure for curves, developed GIS methods for finding key curve parameters, and developed a calibration factor of 1.33 for the HSM crash prediction model for rural two lane roads.

Suburban multilane roads

Baek and Hummer (2008) developed crash prediction models for suburban multilane roads as part of their study to evaluate the safety of curbs. This work was funded by NCDOT and the Southeastern Transportation Center. Data from 2001 to 2003 on 191.9 miles of four-lane road segments were utilized. This work used the method recommended by Hauer (2004) (discussed earlier) to estimate negative binomial regression models with a more flexible functional form. The independent variables considered in the models included shoulder type, shoulder width, median type, median width, posted speed limit, lane width, and number of access points.

Phillips et al., (2005) developed crash prediction models for suburban 4 lane divided and 5 lane roads in North Carolina. This effort was funded by NCDOT. The crash prediction models were used to compare the safety performance of 4 lane divided roads (with a raised median) with 5 lane roads. Only sections longer than 0.25 miles with AADT equal to or greater than 20,000 and speed limit between 35 and 45 mph were included. A total of 143 mid-block segments were utilized for the crash prediction models. Three years of crash data from October 1, 1999 to October 1, 2002, were used. The independent variables in the models included AADT, driveway density, and type of land use (residential, industrial, office, business).

Freeways

As part of NCHRP Project 17-30, UNC HSRC and Texas A&M University developed SPFs for freeway sections in North Carolina, California, Washington, and Ohio (Ullman et al., 2008). The intent of this project was to evaluate the safety of daytime and nighttime construction zones. SPFs were developed for total, injury and fatal, and PDO crashes for day and night time periods. SPFs were used as part of an empirical Bayes before-during evaluation of construction zones. Data from 1995 to 2004 were included in developing the SPFs. All types of limited access freeways were included, but the freeway sections were limited to counties where selected construction projects had occurred. The independent variables in the SPFs included traffic volume, shoulder width, area type (rural versus urban), and whether the freeway segment was within the influence on an interchange.

Council et al., (2007) developed SPFs for injury and fatal truck-non-truck crashes on interstates in North Carolina. The intent was to use these SPFs to identify freeway sections with higher than expected crashes for North Carolina's TACT (Ticketing Aggressive Cars and Trucks) program. This effort was funded by FHWA and FMCSA. Data from 2000 to 2004 was included. This study used a beta version of SafetyAnalyst to identify freeway corridors with promise. To be consistent with the SPFs used in SafetyAnalyst, the only independent variable in these SPFs was AADT. Separate SPFs were developed for different categories of truck percentage.

Influence Area of Ramps

Moon and Hummer (2009) developed crash prediction models for the influence area of ramps in freeways. One objective was to study the safety of left hand ramps that had not been adequately addressed in previous work. A total of 158 sites were used for the modeling among which 33 sites were left-hand ramps. Data from 2002, 2003, and 2004 were used. Models were estimated for total collisions, fatal and injury collisions, total collisions involving merging and diverging, and fatal and injury merging- or diverging-related collisions.

Structure of the Report

Section 2 of the report gives an overview of the HSM prediction methodology for roadway segments and intersections. Section 3 discusses the calibration of the prediction models using data from North Carolina. Section 4 is an overview of the approach that was used to develop state-specific SPFs using data from North Carolina. Section 5 provides examples on how the SPFs (both calibrated and state-specific ones) can be used for network screening, determining the expected safety impacts at the project level, and for evaluating the safety effect of engineering treatments. Section 6 is a discussion of different approaches that NCDOT can use to develop SPFs in the future.

2. OVERVIEW OF THE HIGHWAY SAFETY MANUAL PREDICTION METHODOLOGY FOR ROADWAY SEGMENTS AND INTERSECTIONS

Part C of the Highway Safety Manual (HSM) provides models to predict average expected crash frequency of a site. The HSM provides prediction methods for the following road types:

Roadway Segments

- Rural two lane roads
- Rural four-lane divided and undivided roads
- Two lane, three lane (with center TWLTL), four lane divided, four lane undivided, and five lane roads (with center TWLTL) in urban and suburban arterials

Intersection Types

- Three and four leg minor road stop controlled and four leg signalized intersections on rural two lane roads
- Three and four leg minor road stop controlled and four leg signalized intersections on rural four lane roads
- Three and four leg minor road stop controlled and signalized intersections on urban and suburban arterials

The predictive method in Part C of the HSM is an 18-step procedure to estimate the average expected crash frequency at a site. A site in the HSM is defined as an intersection or a homogenous roadway segment. The predictive method utilizes crash prediction models that were developed from observed crash data for a number of similar sites. The method uses three types of components to predict the average expected crash frequency at a site – the base model, called a safety performance function (SPF); crash modification factors (CMFs) to adjust the estimate for additional site specific conditions; and a calibration factor to adjust the estimate for accuracy in the state or local area. These components are used in the general form below:

$$N_{predicted} = N_{spf} \text{ x } (CMF_{1x} \text{ x } CMF_{2x} \text{ x } \ldots \text{ x } CMF_{yz}) \text{ x } C_x \quad (5)$$

Where:

$N_{predicted}$ = predicted average crash frequency for a specific year for site type x;

N_{spf} = predicted average crash frequency determined for base conditions of the SPF developed for site type x;

CMF_{nx} = crash modification factors specific to SPF for site type x; and

C_x = calibration factor to adjust SPF for local conditions for site type x.

As indicated, each predictive model is specific to a facility or site type (e.g., urban four-lane divided segments) and a specific year. The HSM stresses that the advantage of using these predictive models is that the user will obtain a value for long-term expected average crash frequency rather than short-term observed crash frequency. This will minimize the error due to selecting sites for treatment that look hazardous based on short term observations, or in other terms, a bias called regression-to-the-mean. It should also be noted that the predictive method can be used to predict crashes for past years based on observed AADT or for future years based on forecast AADT.

The steps for the predictive method are presented in detail in section C.5. of Volume 2 of the HSM. In short, they are:

- Decide which facilities and roads will be used in the predictive process and for what period of time (Steps 1 and 2)
- Identify homogenous sites and assemble geometric conditions, crash data, and AADT data for the sites to be used (Steps 3 through 8)
- Apply the safety performance function, any applicable crash modification factors, and a calibration factor if available (Steps 9 through 11)
- Apply site- or project-specific empirical Bayes method if applicable (Steps 12 through 15)
- Repeat for all sites and years, sum, and compare results (Steps 16 through 18)

An example of how to apply an SPF, CMFs, and calibration factor for the predictive method is shown below.

Example of calculating average expected crash frequency using HSM predictive method

This example will demonstrate how to use the HSM predictive method to calculate the expected average crash frequency for a rural four-lane divided roadway segment with the following characteristics:

- 1.0 mile length
- 12 foot lane width
- 6 foot paved right shoulder
- 15,000 AADT
- 80 foot traversable median with no barrier
- No roadway lighting
- No automated enforcement

All table, equation, and page numbers in the example below refer to Chapter 11 of the HSM.

Steps 1 through 8

Since this example is directed at applying the predictive method to a single pre-selected segment with existing data, steps 1 through 8 are not necessary.

Step 9: Apply the appropriate safety performance function (SPF)

The SPF for a rural divided roadway segment is presented in Equation 11-9 (p. 11-18) in the HSM with coefficients listed in Table 11-5.

$N_{spf\ rd} = e^{(a + b \times \ln(AADT) + \ln(L))}$

Where:

$N_{spf\ rd}$ =base total number of roadway segment crashes per year;
AADT = annual average daily traffic (vehicles/day) on roadway segment;
L = length of roadway segment; and
a, b = regression coefficients (appropriate values to be selected from Table 11-5)

Using the SPF for this example yields the following base number:

$N_{spf\ rd} = e^{(-9.025 + 1.049 \times \ln(15000) + \ln(1.0))}$
= 2.892 crashes per year

Step 10: Apply the appropriate crash modification factors

The HSM procedure for rural divided roadways involves five CMFs.

Lane Width (CMF_{1rd})
Based on Table 11-16 for a lane width of 12 feet, CMF_{1rd} = 1.0.

Right Shoulder Width (CMF_{2rd})
Based on Table 11-17 for a right shoulder width of 6 feet, CMF_{2rd} = 1.04.

Median Width (CMF_{3rd})
Based on Table 11-18 for a median width of 80 feet, CMF_{3rd} = 0.95.

Lighting (CMF_{4rd})
Since there is no roadway lighting at this location, CMF_{4rd} = 1.0 (the base condition for CMF_{4rd} is absence of lighting).

Automated Enforcement (CMF_{5rd})
Since there is no automated enforcement at this location, CMF_{5rd} = 1.0 (the base condition for CMF_{5rd} is absence of automated enforcement).

<u>Combined CMF</u>

The combined CMF value is calculated below.

CMF_{comb} = 1.0 x 1.04 x 0.95 x 1.0 x 1.0 = 0.99

Step 11: Apply a calibration factor if available

For this example, the calibration factor (C_r) for the local area is assumed to be 0.96.

Calculation of Average Expected Crash Frequency

$N_{predicted\ rd}$ = $N_{spf\ rd}$ x CMFcomb x C_r

= 2.892 x 0.99 x 0.96 = 2.75 crashes per year

3. CALIBRATION OF THE HSM PREDICTION MODELS WITH DATA FROM NORTH CAROLINA

Why is Calibration Needed?

The HSM predictive models were developed using data from many states in the country. The HSM recommends that these predictive models be calibrated using data from a jurisdiction where these models will be applied. Calibration is important because "the general level of crash frequencies may vary substantially from one jurisdiction to another for a variety of reasons including crash reporting thresholds and crash reporting system procedures" (HSM, page C-18). The development and use of calibration factors will assist NCDOT personnel in arriving at crash predictions that are more accurate for North Carolina sites. As discussed in Section 2 of this report, the calibration factor is used by multiplying the predicted crashes from the HSM model by the specific calibration factor for that facility type.

Calibration Process

The process of developing calibration factors for the Part C predictive models is laid out in Appendix A of Part C (Volume 2) of the HSM. The steps are as follows:

1. Identify facility types for which the applicable Part C predictive model is to be calibrated
2. Select sites for calibration of the predictive model for each facility type
3. Obtain data for each facility type applicable to a specific calibration period
4. Apply the applicable Part C predictive model to predict total crash frequency for each site during the calibration period as a whole
5. Compute calibration factors for use in Part C predictive model

The sections below discuss how each step was executed in the development of the North Carolina calibration factors.

Step 1 – Identify facility types for which the applicable Part C predictive model is to be calibrated

There are predictive models in the HSM for eight types of roadway segments and ten types of intersections. Calibration factors were developed for six of the roadway types and eight of the intersection types. The remaining four models were not involved in the development of calibration factors, as explained in the below table.

Facility types for which calibration factors were developed are as follows:

Roadway Segments

- Rural 4 Lane Divided
- Urban 2 Lane Undivided (2U)
- Urban 2 Lane with TWLTL (3T)
- Urban 4 Lane divided (4D)
- Urban 4 Lane Undivided (4U)
- Urban 4 Lane with TWLTL (5T)

Intersections

- Rural 2 Lane, minor rd stop controlled 3-leg (3ST)
- Rural 2 Lane, signalized 4-leg (4SG)
- Rural 2 Lane, minor rd stop controlled 4-leg (4ST)
- Rural 4 Lane, signalized 4-leg (4SG)
- Urban arterial, signalized 3-leg (3SG)
- Urban arterial, minor rd stop controlled 3-leg (3ST)
- Urban arterial, signalized 4-leg (4SG)
- Urban arterial, minor rd stop controlled 4-leg (4ST)

Facility types for which calibration factors were <u>NOT</u> developed are as follows:

Roadway Segments

- Rural 2 Lane (REASON: Calibration factor has already been developed in a recent NCDOT project - see Section 1 of this report)
- Rural 4 Lane Undivided (REASON: Insufficient mileage in state to prioritize this facility type)

Intersections

- Rural 4 Lane, Minor Road Stop Controlled 3-leg (3ST) (REASON: Calibration factor has already been developed in a recent NCDOT project - see Section 1 of this report)
- Rural 4 Lane, Minor Road Stop Controlled 4-leg (4ST) (REASON: Calibration factor has already been developed in a recent NCDOT project - see Section 1 of this report)

Step 2 – Select sites for calibration of the predictive model for each facility type

The calibration process requires detailed data on each site. Hence, the calibration process must be based on a sample of miles or intersections for which detailed data can be collected. The selection of this sample is important. The sites must be selected in as random a manner as possible, so as not to bias the calibration process. The HSM instructs that sites should not be selected so as to limit the sample only to either high or low crash frequencies. The size of the sample is also important. The HSM recommends that the desired minimum sample size for each facility type is 30 to 50 sites and that the entire group of the sample for each facility type should represent at least 100 crashes per year in order for the calibration to be reliable.

The site selection process in this effort started with obtaining a list of all North Carolina road segments from the Highway Safety Information System (HSIS). HSIS maintains an archived database of roadway inventory, traffic volumes, and crash data for nine states, including North Carolina. The team used the HSIS data for each segment on the number of lanes, type of median division, population density, and town limits to classify each segment as one of the HSM facility types (e.g., Rural 2 Lane Undivided, Urban 4 Lane Divided, etc.). The team also classified the segments as belonging to one of three geographic areas (coast, piedmont, or mountain) based on the county of location (see Appendix D for a list of counties by geographic area).

After all segments were classified as a particular facility type, the team selected a group of segments within that facility type on which the data collector would obtain the detailed data necessary for calibration. Ideally, the team would have selected segments randomly from the entire group. However, the data collection procedure was much more efficient if sites were adjacent to each other. For this reason, the team selected entire routes and collect data on all segments on that route. In order to minimize route selection bias, the team would typically select all routes in a single county or multiple counties if additional sample size was needed. This resulted in a good mix of road classes in the sample. The team also made sure to select roughly equal groups from each of the three geographic areas of the state.

The HSM segment-based predictive models predict only non-intersection crashes, so it was important to make sure that segments did not include intersection influence areas. To address this issue, when an intersection was encountered on a route, the adjacent segments were redefined so as to exclude 250 feet on either side of the intersection.

Intersections were collected as part of the segment data collection. When an intersection matching of the facility types from the HSM models was encountered on a route, it was added to the intersection sample and the appropriate data were collected for it.

Step 3 – Obtain data for each facility type applicable to a specific calibration period

The data collection step involved obtaining data on geometric and cross-sectional characteristics, traffic volumes, and crash data for each site. The sources of characteristics and volume data were HSIS, NCDOT GIS files of the road network and traffic volume points, and Google online aerial and Streetview imagery (Streetview is a way of viewing photos shot from a unidirectional camera mounted on a vehicle). Crash data were obtained through NCDOT from the Traffic Engineering Accident Analysis System (TEAAS). The data obtained from NCDOT GIS files and Google imagery were collected with the assistance of a civil engineering graduate student. Tables 1 and 2 show the data elements collected for segments and intersections and the source of each element.

Table 1. Data sources for roadway segments

Data Element	Source
Number of through traffic lanes	HSIS (verified visually)
Low-speed vs. intermediate or high speed	HSIS
Median presence and width	HSIS (median presence verified visually)
Presence of center two-way left-turn lane	HSIS (verified visually)
Shoulder width	HSIS
Number of driveways by land-use type	Aerial/Streetview Imagery
Presence of automated speed enforcement	Aerial/Streetview Imagery
Presence of lighting	Aerial/Streetview Imagery
On-street parking presence and type	Aerial/Streetview Imagery
Roadside fixed object density	Aerial/Streetview Imagery
Traffic volume	HSIS and NCDOT GIS
Crash data	NCDOT TEAAS

Table 2. Data sources for intersections

Data Element	Source
Intersection skew angle	Aerial/Streetview Imagery
Number of intersection legs	Aerial/Streetview Imagery
Type of traffic control	Aerial/Streetview Imagery
Number of approaches with left-turn lanes	Aerial/Streetview Imagery
Number of approaches with right-turn lanes	Aerial/Streetview Imagery
Presence of lighting	Aerial/Streetview Imagery
Presence of left-turn phasing	Aerial/Streetview Imagery
Type of left-turn phasing	Aerial/Streetview Imagery
Use of right-turn-on-red signal operation	Aerial/Streetview Imagery

Use of red-light cameras	Aerial/Streetview Imagery
Maximum number of lanes crossed by pedestrians on any approach	Aerial/Streetview Imagery
Average annual daily traffic (AADT) for major road	NCDOT GIS
Average daily traffic (AADT) for minor road	NCDOT GIS
Crash data	NCDOT TEAAS

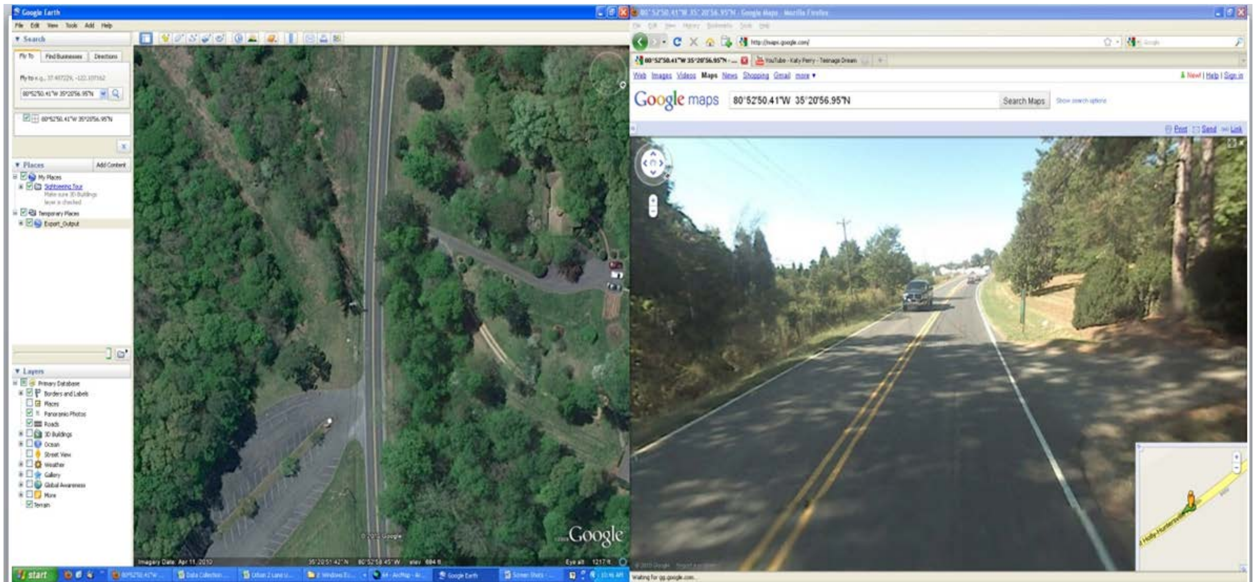
The sections below describe how data were collected for roadway segments and intersections.

Segment characteristics data collection

In order to accurately track mileposts and collect the required data, it was necessary for the data collector to track along the route in both the GIS environment and the Google imagery. To accomplish this, he would delineate each segment in the GIS line layer (using the indicated begin and end mileposts), then export that layer to a file that could be read into Google Earth. Since the segments were selected from the HSIS list according to entire routes, the data collector could track along the route, collecting data on each segment sequentially. This method greatly improved the efficiency of data collection, as opposed to jumping around to randomly selected segments, which would take considerably more time.

The first task for the data collector on each segment was to confirm that it was indeed the correct facility type indicated in HSIS (e.g., rural four-lane divided) and confirm that the beginning and ending mileposts were correct. Sometimes it was the case that a road would be a different facility type than was indicated in HSIS, either due to miscoding in the initial NCDOT Universe file, or due to the fact that the road had been upgraded since its initial entry in the inventory system. When confirming segment end points, it was often the case that the beginning or ending milepost of a segment had to be redefined due to the fact that the segment as defined in HSIS encompassed two or more non-homogenous sections (e.g., the median was discontinued partway through the indicated segment). Additionally, if there was an intersection in the segment, the segment would be broken into two new segments, with the beginning or ending points of the new segments defined to exclude 250 feet on either side of the intersection. The locations of these intersections would be noted and they would be collected separately for the intersection sample.

Once each segment was confirmed and accurately defined, the data collector would collect the necessary geometric and cross-section characteristics using a combination of aerial and Streetview imagery. Figure 1 shows an example image of the two types of views and indicates below the images which elements were collected from each.



Segments (aerial)
- Mileposts and coordinates of endpoints
- Length of on-street parking

Intersections (aerial)
- Coordinates of intersections
- Turn lanes
- Number lanes

Segments (Streetview)
- Driveway count
- Lighting
- Presence of on-street parking
- Speed enforcement
- Fixed object density and offset
- Verify roadway characteristics

Intersections (Streetview)
- Traffic control type
- Anything not visible in aerial

Figure 1. Examples of aerial and Streetview imagery with lists of data elements collected from each

Intersection characteristics data collection

Intersection data were collected in a similar manner to the segment data. Geometric data, traffic control, configuration, and other characteristics were collected through viewing the Google aerial and Streetview imagery. Traffic volumes were obtained from the GIS file. All identifying route names and numbers were collected for both the major and minor roads for use in obtaining crash data. Additionally, the latitude and longitude of the intersection were recorded to allow for quick locating of the intersection if needed in the future.

Traffic volumes data collection

Traffic volumes (AADT) were obtained from HSIS for roadway segments and from NCDOT GIS data for intersections. For segments, HSIS provided yearly AADT for each roadway segment in the initial list that was used for site selection, so the AADT information was easily obtained. For intersections, AADT for the major and minor roads was obtained from the GIS data made available by the NCDOT GIS Unit. The shapefiles used consisted of two types of files. First was a line layer of the road network that had one AADT value and year for each segment. Second was a group of point layers that represented traffic volume count points around the state. The point layer was used to estimate AADT if there was not a value available in the line layer.

Crash data collection

Crash data were obtained from NCDOT. Mr. Brian Murphy ran queries on the TEAAS database to obtain the crash data for 2007-2009 for the segments and intersections.

Step 4 – Apply the applicable Part C predictive model to predict total crash frequency for each site during the calibration period as a whole

The predictive models were applied for each facility type following the HSM predictive method (as discussed in Section 2 of this report). Microsoft Excel™ spreadsheets were developed to run the predictive models for the entire group of sample sites. These spreadsheets will be delivered with this report to allow NCDOT to develop new calibration factors in future years (as discussed in Section 6 of this report).

Step 5 – Compute calibration factors for use in Part C predictive model

The calibration factor for each facility type was calculated as indicated in the HSM, by the following method:

sample with fewer than 100 crashes. This is because there is very little mileage of this facility type in North Carolina; the sample used in the calibration consisted of all the mileage available for this facility type.

Table 3. Calibration Factors for Segment Models

Segment Facility Types	**2007**	**2008**	**2009**	**3 yr avg**
Rural 4 Lane Divided	***0.96***	***0.95***	***0.99***	***0.97***
Urban 2 Lane Undivided (2U)	***1.58***	***1.66***	***1.36***	***1.54***
Urban 2 Lane with TWLTL (3T)	***4.30***	3.65	2.90	3.62
Urban 4 Lane Divided (4D)	***3.90***	***4.25***	***3.45***	***3.87***
Urban 4 Lane Undivided (4U)	***4.10***	***4.45***	***3.57***	***4.04***
Urban 4 Lane with TWLTL (5T)	***1.75***	***1.72***	***1.68***	***1.72***

Table 4. Calibration Factors for Intersection Models

Intersection Facility Types	**2007**	**2008**	**2009**	**3 yr avg**
Rural 2 Lane, minor rd stop controlled 3-leg (3ST)	0.57	0.58	0.57	0.57
Rural 2 Lane, signalized 4-leg (4SG)	***1.14***	***1.13***	0.84	***1.04***
Rural 2 Lane, minor rd stop controlled 4-leg (4ST)	0.77	0.64	0.65	0.68
Rural 4 Lane, signalized 4-leg (4SG)	***0.55***	***0.47***	***0.46***	***0.49***
Urban arterial, signalized 3-leg (3SG)	***2.86***	***2.46***	***2.09***	***2.47***
Urban arterial, minor rd stop controlled 3-leg (3ST)	1.75	***2.03***	1.38	1.72
Urban arterial, signalized 4-leg (4SG)	***3.01***	***2.85***	***2.51***	***2.79***
Urban arterial, minor rd stop controlled 4-leg (4ST)	1.61	1.13	1.22	1.32

The tables above also show three individual years of calibration factors for 2007 through 2009, and it can be seen that the calibration factors do not vary significantly from year to year. This is true even for the highest and lowest calibration factors. For instance, the factors for urban four-lane undivided facilities show yearly factors of 4.10, 4.45, and 3.57, indicating that the combined three-year factor of 4.04 was not skewed by one anomalous year. The team also developed calibration factors separately for each geographic area. These factors are presented in Appendix E.

The large magnitude of the calibration factors for the urban 3- and 4-lane road segments bears some discussion. The team has investigated to determine if there is a clear reason why these facilities would have such high factors. The site selection was done appropriately. The crash data, obtained from NCDOT TEAAS, would not have included intersection crashes, since the segments were defined so as to exclude intersection influence areas. The team used locally derived (North Carolina) values for the proportions of nighttime crashes and fixed object crashes in the predictive method (the HSM procedure allows for the use of default values). Since the HSM segment models were developed with data from Washington State, the team looked at comparisons between Washington and North Carolina data. A recent effort by FHWA to compare fatality rates across the nation showed a difference in the rural fatality rates (per VMT) between the two states; North Carolina had approximately 50% higher fatality crash rate than Washington in 2008 for rural roads. However, there does not appear to be a full explanation of the magnitude of the factors on the urban segments.

Special note about calibration of pedestrian collision model

It should be noted that the calibration of urban signalized intersection models was based on vehicle-vehicle and vehicle-bicycle crashes only. The process did not involve pedestrian crashes in the calibration. This was for two reasons. One, the models to predict pedestrian crashes required detailed data on the number of bus stops, schools, and alcohol sales establishments within 1,000 feet of the intersection. The labor to acquire this data would have been extensive. Two, the HSM models to predict pedestrian crashes were developed by an NCHRP project (also performed by this project team) that used Charlotte and Toronto data. During the course of the NCHRP project, the City of Charlotte provided the team with GIS files that indicated the locations of bus stops, schools, and alcohol sales establishments; these data were subsequently used in the development of the predictive model. Thus, since the pedestrian models were developed using North Carolina data, the need to calibrate these models was minimal compared to the rest of the calibration effort.

4. DEVELOPMENT OF STATE-SPECIFIC SAFETY PERFORMANCE FUNCTIONS FOR ROADWAY SEGMENTS IN NORTH CAROLINA

This section describes the approach that was used to develop state-specific SPFs for North Carolina. As mentioned in the introduction, state-specific SPFs can be used for network screening and for the evaluation of engineering treatments using the EB method. State-specific SPFs may not be the best tool for project level analysis unless they can be developed using data sets that have information about a large number of site characteristics (apart from AADT). Most states do not have such detailed information in electronic form for a large number of sites.

Since the majority of North Carolina's state roadway system is part of FHWA's Highway Safety Information System (HSIS), developing SPFs with just AADT for different types of roadway segments was feasible within this project. Two types of SPFs were developed for roadway segments: type 1 SPFs just include AADT and type 2 SPFs include other site characteristics apart from AADT (type 2 SPFs were estimated only for rural 2 lane roads). SafetyAnalyst allows only type 1 SPFs to be included, and the format and functional form of the type 1 SPFs that were estimated in this project is consistent with the requirements for SafetyAnalyst.

Data

The project team utilized the following data sources to estimate the SPFs for roadway segments:

- Roadway inventory files from HSIS
- Crash data from Traffic Engineering Accident Analysis System (TEAAS)
- A file from NCDOT with information about the location of at grade intersections, railroad grade crossings, and interchanges (this is also based on TEAAS)

The roadway inventory file from HSIS (based on NCDOT's 'universe' file) provides information about the characteristics of each segment including segment length, AADT, shoulder width, etc. Roadway segments within 250 feet of an at-grade intersection or a railroad grade crossing were excluded for this analysis. In addition, segments with AADT less than or equal to 500 were excluded as well, because NCDOT staff indicated that the AADT for such segments have not been found to be very reliable. Data from 2004 to 2008[2] were utilized for the analysis.

---

[2] Roadway inventory files for 2009 were not available from HSIS when the analysis was conducted

NCDOT was interested in SPFs for the following 9 types of crashes:

- Total crashes
- Injury and fatal crashes (K, A, B, C)
- Injury and fatal crashes (K, A, B)
- PDO crashes
- Lane departure crashes – This included crashes with the First Harmful Event = (1) Ran off road – right, or (2) Ran off road – left, or (3) Ran off road – straight, or (19) Fixed object, or (27) Head on, or (29) Sideswipe, opposite direction
- Single vehicle crashes (includes animal crashes)
- Multi vehicle crashes
- Wet crashes – This included crashes with Road Surface Condition = (2) Wet or (3) Water (standing, moving)
- Night crashes – This included crashes with Ambient Light = (4) Dark – lighted roadway, or (5) Dark – roadway not lighted, or (6) Dark – unknown lighting

NCDOT provided information about these crash types from the TEAAS database.

Roadway Types

Since one possible application of these SPFs is being able to use them within SafetyAnalyst, data were compiled for the following 16 roadway types as defined in SafetyAnalyst:

- Rural Two Lane Roads
- Rural Freeways - 4 lanes - outside the influence of interchanges
- Rural Freeways – 6+ lanes - outside the influence of interchanges
- Rural Freeways - 4 lanes - within the influence of interchanges
- Rural Freeways – 6+ lanes - within the influence of interchanges
- Rural Multilane Divided Roads
- Rural Multilane Undivided Roads
- Urban Two Lane Roads
- Urban Freeway - 4 lanes - outside the influence of interchanges
- Urban Freeway - 6 lanes - outside the influence of interchanges
- Urban Freeway - 8+ lanes - outside the influence of interchanges
- Urban Freeway - 4 lanes - within the influence of interchanges
- Urban Freeway - 6 lanes - within the influence of interchanges
- Urban Freeway - 8+ lanes - within the influence of interchanges
- Urban Multilane Divided Roads
- Urban Multilane Undivided Roads

A segment was classified as rural if TOWN (a variable in NCDOT's roadway inventory file) was equal to zero or null, or if the city population was less than or equal to 5,000. Otherwise, the segment was classified as urban. Segments were classified as freeway, multilane divided, multilane undivided, or two lane based on the *roadway class* variable in HSIS. Freeway segments within 0.5 miles of either side of an interchange were considered *within the influence of interchanges*; otherwise, they were considered *outside the influence of interchanges*.

Safety Performance Functions for Roadway Segments

Consistent with the state of the art, the relationship between the dependent variable (i.e., the number of crashes) and the independent variables was log-linear. The SPFs were estimated using negative binomial (NB) regression. In addition, to be consistent with the functional form that is compatible with SafetyAnalyst, the relationship between the number of crashes and AADT in the type 1 SPFs was as follows:

$$Y = L^* \exp\{\alpha + \beta * \ln(AADT)\} = L^* (e^{\alpha})(AADT^{\beta}) \qquad (6)$$

Where, Y is the expected number of crashes per year, ln represents the natural logarithm, L is the length of a section, and AADT is the average annual daily traffic. α (also called the intercept) and β are parameters (i.e., coefficients) estimated as part of the negative binomial regression model.

In a NB model, the variance is related to the mean as follows:

$$Var(y_i) = E(y_i) + k(E(y_i))^2 \qquad (7)$$

where:

$Var(y_i)$ is the variance,

$E(y_i)$ is the expected value, and

k is the overdispersion parameter (some studies, including some sections of the HSM use the inverse of the overdispersion parameter (φ) instead of the overdispersion parameter (k), where φ = 1/k).

Many previous studies assumed k to be a constant value while estimating the NB models for roadway segments. Hauer (2001) argued that assuming k as a constant provides too much weight to shorter sections and not enough weight to longer sections. He suggested estimating k that applies to a unit length of road, and this was the approach that was followed in this study while estimating the SPFs for roadway segments. In doing this, k is replaced by $\frac{k_1}{L}$, where k_1 is the overdispersion parameter for a 1 mile section, and L is the length of a section[3].

Type 1 SPFs were estimated for the 9 crash types for all the 16 roadway types, i.e., a total of 144 SPFs. In addition, type 2 SPFs were estimated for rural two lane roads, the roadway type that is of primary interest to NCDOT. In type 2 SPFs, which include other variables in addition to AADT, the relationship between crash frequency and the independent variables was again log-linear (again, consistent with the state of the art):

$$Y = L^* \exp\{\alpha + \beta_1 * f_1(AADT) + \beta_2 f_2(X_2) + \beta_3 f_3(X_3) + \beta_4 f_4(X_4) + ...\} \qquad (8)$$

Where f_1, f_2, f_3, and f_4, represent functions of the independent variables AADT, X_2, X_3, and X_4. For a log-linear model to be estimated using generalized linear modeling techniques, the functions are typically either identity or natural logarithm. Again, α, β_1, β_2, β_3, β_4, are parameters (coefficients) estimated as part of the NB model.

Appendix A shows the type 1 SPFs and Appendix B shows the type 2 SPFs for rural two lane roads. The Appendices show the coefficients, their standard errors, the overdispersion parameter (k_1), the observed number of crashes that were used for the estimation, the number of crashes predicted by the model, and goodness of fit (GOF) statistics. The two GOF statistics that are included are: Freeman-Tukey R^2 (Fridstrom et al., 1995) and the Pseudo R^2 (Miaou, 1996). Unlike linear regression, the traditional R^2 is rarely used in negative binomial regression, and there is no universally accepted GOF measure. If any of the coefficients are not statistically different from zero at the 5% significance level, they are shown in italics.

The Appendices also show the annual factors for each year of data that was used in the analysis. The annual factor for a particular year is defined as the ratio of observed crashes to predicted crashes from the SPF for that year. Annual factors are used to account for the effect of changes in factors such as weather, crash reporting practices, demography, and others (that

[3] It is important to note that there are other ways of estimating the overdispersion parameter. For example, in some of the models in the HSM, the overdispersion parameter was estimated as a function of segment length as follows: $k = \frac{1}{\exp(c + \ln(L))}$, where, c is a parameter to be estimated.

are not explicitly considered in the model), over time. Since the annual factors are the ratio of observed to predicted crashes, a sufficient sample of crashes are needed in each year to get a reliable factor. So, annual factors based on less than 150 observed crashes per year (on average) (i.e., 750 crashes in 5 years) are shown in italics and should be used with caution.

Finally, for each roadway type, summary statistics about the data are shown as well. The statistics include the minimum, maximum, and average AADT, the minimum, maximum, and average segment length, the number of observations used in the development of the SPF, and the number of mile-years in the data.

5. APPLICATIONS OF SAFETY PERFORMANCE FUNCTIONS

As discussed earlier, SPFs can be used for network screening, project level analysis, and for evaluation of effect of engineering treatments. Following is a discussion of how SPFs can be used in these three situations.

Network Screening

The intent of network screening is to identify roadway segments and/or intersections *with promise* for an engineering treatment/intervention. There are many ways of screening the network. Chapter 4 of the HSM discusses the pros and cons of the different methods for screening the network. The HSM indicates that being able to address the possible bias due to regression to the mean (RTM) and accounting for the effect of traffic volume are two important considerations to use in order to select an appropriate method for screening the network. Among the different methods discussed in the HSM, following are two methods that account for the effect of traffic volume and also address the possible bias due to RTM: (1) Expected average crash frequency with EB adjustments, and (2) Excess expected average crash frequency with EB adjustment. Both these methods require the use of SPFs.

In order to illustrate the steps involved in computing the EB expected crash frequency for a segment, here are some hypothetical data for 5 segments that are assumed to be part of a rural two lane road:

Table 5. Hypothetical data for the illustration of network screening application

Segment	Length (miles)	Crashes					AADT				
		2004	2005	2006	2007	2008	2004	2005	2006	2007	2008
1	1.3	0	2	3	1	1	1500	1500	1600	1650	1675
2	0.2	1	0	0	0	2	6100	6200	6300	6300	6700
3	0.4	0	0	1	0	1	3200	3000	2900	2900	2700
4	0.7	2	1	0	1	6	1100	1200	1600	2000	2100
5	0.9	5	0	0	1	1	8000	8000	8300	8400	8500

Here are the steps involved in computing the EB expected crash frequency for each segment:

Step 1: Using the SPFs for rural two lane roads compute the predicted number of crashes for each segment (call this P).

Based on Appendix A, the SPF for rural two lane roads is

Y = L* $(e^{-4.0852})(AADT^{0.5830})$

Based on the data in the table and SPF, for segment 1 in year 2004, the predicted value = 1.3* $(e^{-4.0852})(1500^{0.5830})$=1.554. This number needs to be multiplied by the annual factor for 2004 (which is 1.058) to get 1.554*1.058 = 1.644. This approach is repeated for each year. For segment 1, here are the values for each year:

- 2004: 1.644
- 2005: 1.498
- 2006: 1.574
- 2007: 1.668
- 2008: 1.666

The total predicted number of crashes in segment 1 is the sum of all these values, and that comes out to 8.050.

Step 2: Compute the total number of observed crashes for each segment (call this X).

From the Table, the total number of observed crashes in segment 1 during the same time period is 7.

Step 3: Compute the EB expected number of crashes for each segment.

The EB expected crashes is a weighted average of the predicted number of crashes and the observed number of crashes (i.e., weighted average of P and X). The EB expected crashes is the estimate of the *long-term crash frequency* after accounting for possible bias due to RTM. The EB expected crashes is calculated as follows:

EB_expected = P**w*+X*(1-*w*) (9)

Where, *w* is a function of the overdispersion parameter that was estimated as part of the NB model, and the predicted number of crashes. If the overdispersion parameter was estimated as a constant value k, then *w* is computed as follows:

$$w = \frac{1}{1 + k * P} \tag{10}$$

If the overdispersion parameter was estimated such that it applies to a unit length of road (as was done in this study), then *w* is computed as follows:

$$w = \frac{1}{1 + \left(\frac{k_1}{L}\right) * P} \tag{11}$$

From Appendix A, k_1 for rural two lane roads is 0.3110. For segment 1, $\frac{k_1}{L}$ is 0.3110/1.3 = 0.2390. Hence, *w* for segment 1 will be:

$$w = \frac{1}{1 + 0.2390 * 8.050} = 0.342 \tag{12}$$

The EB expected crash frequency for segment 1 will be:

EB_expected = 0.342*8.050 + (1-0.342)*7 = 7.3590.

Step 4: Compute the EB excess expected number of crashes for each segment.

The EB excess expected number of crashes = EB_expected – P. For segment 1, the EB excess expected will be:

EB_excess_expected = 7.359 – 8.050 = -0.691.

Step 5: Compute EB expected and EB excess expected per mile per year

Since 5 years of data are being used, by dividing the EB expected and EB excess expected by the product of 5 and the section length, we can get the EB expected and EB excess expected per mile per year. For segment 1, these numbers will be as follows:

EB_expected_per_mile_per_year = 7.359/(5*1.3) = 1.132
EB_excess_expected_per_mile_per_year = -0.691/(5*1.3) = -0.106.

This procedure can be repeated for all the segments in the network (see Table 6) and segments with higher expected crashes or excess expected crashes can be selected for further review.

Table 6. Results of EB computations for each segment

Segment	P	X	k_1	k_1/L	*w*	Empirical Bayes estimates			
						Expected	Excess	Expected/mi/year	Excess/mi/year
1	8.050	7	0.311	0.239	0.342	7.359	-0.691	1.132	-0.106
2	2.774	3	0.311	1.555	0.188	2.957	0.183	2.957	0.183
3	3.552	2	0.311	0.778	0.266	2.413	-1.140	1.206	-0.570
4	4.323	10	0.311	0.444	0.342	8.056	3.733	2.302	1.067
5	14.573	7	0.311	0.346	0.166	8.255	-6.318	1.834	-1.404

The same procedure can be repeated for individual crash types. For example, the state may be interested in screening the network based on lane departure crashes and severe injury crashes in addition to total crashes.

Whether the EB expected value or the EB excess expected value should be used has been a subject of debate within the research community without any clear consensus. There are advantages and disadvantages for both these methods. The use of expected collisions is embedded in the concept of Crash Modification Factors (CMFs) since the benefit of a treatment can be expressed as the product of the expected collisions with (CMF – 1). On the other hand, there is no way to directly apply CMFs to excess expected collisions. However, using excess expected is attractive and intuitive because it “rests on the belief that if a site has more collisions than what is normal at similar sites, there must be site-specific causes that explain the excess, and that if causes are identified, they could be remedied, and the excess reduced” (Hauer et al., 2002). For this reason, SafetyAnalyst allows the user to select either method for network screening.

Although in the example the calculations were done at the segment level, the procedure is rarely implemented at the segment level in this manner. One common approach is the use of a *sliding window* where a window of fixed length moves in defined increments and all the calculations (discussed above) are performed at each window location. Each segment is then characterized by the maximum value calculated at any window position within or overlapping the beginning of an adjacent segment. In so doing, there is an increased chance of detecting a high risk site at the screening stage if the collision problem manifests itself in a window overlapping the adjacent site.

The second is the *peak search* approach[4]. This approach makes use of incrementally growing window lengths that are selected so no windows span multiple roadway segments. The window starts at the left boundary of a road segment and increases in length incrementally until it reaches the end. At each increment, we have a specific window where an estimated collision count can be calculated. For example, a segment of 0.5 mile can produce windows with lengths of 0.1, 0.2, 0.3, 0.4, and 0.5 miles assuming an increment length of 0.1 mile. The window with the largest value of the estimate of expected or expected excess collisions per mile is then tested for statistical significance. The test of significance is the coefficient of variation (CV), equal to the standard error of the estimate divided by the estimate. A limiting value of the CV is specified by the analyst, and values of CV below the limiting value pass the test. If the window passes the test then the entire road segment is ranked by the largest value of the estimate per mile. If the test is not passed then the window size is increased (say, to 0.2 mile from 0.1 mile) and the process starts again for the road segment. The advantage of this method is that localized safety problems are not overlooked by using too large a window yet the statistical test ensures that they are in fact reliable estimates and not due to some randomness in the data.

Project Level Analysis

For individual projects, SPFs can be used to determine the average expected crash frequency for the site of interest. The process would be similar to the procedure above for network screening, except that the analyst would be able to use a more detailed SPF with greater data requirements that would have greater accuracy for the specific site. Network screening SPFs, such as those used in SafetyAnalyst and described in the above section, predict crashes based solely on AADT. Project level SPFs, such as those presented in the Highway Safety Manual Part C, start with a model that uses AADT to make a base prediction and then use crash modification factors to modify that predicted value based on site-specific characteristics such as lane width, median width, lighting, traffic control, and exclusive turn lanes. A calibration factor, if available, would also be used to adjust the estimate for local conditions.

If appropriate, the EB method should be used to arrive at EB expected frequencies, using the procedure described above in Steps 3 through 5. In order to apply the EB method, the analyst will need to make use of the overdispersion parameter (k) for each of the prediction models (these are available from the HSM). Most situations would be appropriate for the EB method. Only if the predictive method is being used to predict future years of crashes AND if the site is

[4] Draft Functional Specification for Module 1 - Network Screening (SafetyAnalyst), May 2003, Contract No. GS-23F-0379K, Task No. DTFH61-01-F-00096.

expected to undergo significant modifications would the EB method be inappropriate, since the history of observed crashes would not accurately reflect the anticipated future conditions.

The HSM talks about four methods for estimating the change in expected average crash frequency of a proposed project or project design alternative (HSM, page C-19). Three of these methods make use of SPFs and are discussed below:

Method 1: Apply the HSM Part C prediction methodology to estimate the expected average crash frequency of both the existing and proposed conditions.
Section 2 of this report provided an example for calculating the expected number of crashes by following the HSM procedure. Suppose the intent is to replace a section of a rural two lane undivided road with a rural four lane divided road. First, we can use the HSM procedure to estimate the expected number of crashes for rural two lane undivided roads. Next, we will use the HSM procedure to estimate the expected number of crashes for rural four lane divided road. In both cases, we should make use of calibration factors that have been developed using North Carolina data for these two facility types.

Method 2: Apply the HSM Part C Predictive method to estimate the expected average crash frequency of the existing condition and apply the project CMF from Part D of the HSM.
Suppose the intent is to determine the change in crashes that would occur if shoulder rumble strips are introduced in a rural two lane divided road. In this case, first we could use the HSM procedure to estimate the expected number of crashes for rural two lane divided roads. One could then estimate the EB expected number of crashes for this piece of rural two lane divided road using the procedure described above in Steps 3 through 5. Following this, using the CMF for shoulder rumble strips from Part D of the HSM it is possible to estimate the expected number of crashes if shoulder rumble strips are installed.

Method 3: If the HSM Part C predictive method is not available, but a Safety Performance Function (SPF) applicable to the existing roadway is available, use that SPF and a CMF from Part D or other sources.
As discussed earlier, predictive methods in the HSM are currently available only for certain types of facilities. For example, predictive methods are not available in the current edition of the HSM for freeway sections. So, if the intent is to determine the expected change in crashes due to some improvements in a freeway section, one option is to use the state-specific SPFs developed in this project using North Carolina data (from Appendix A of this report) to estimate the expected number of crashes, estimate the EB expected number of crashes (using steps 3 through 5 discussed earlier under Network Screening), and apply the appropriate CMF from

either Part D of the HSM or other sources (e.g., the CMF clearinghouse) to obtain the expected number of crashes.

It should be noted that NCHRP project 17-38 produced a set of spreadsheets to facilitate use of the HSM predictive method. These spreadsheets allow the user to enter all applicable data on the site (e.g., AADT, cross-section characteristics, other site-specific characteristics) and the spreadsheet will calculate the expected average crash frequency based on the SPFs and CMFs presented in the HSM. For NCDOT convenience, these spreadsheets will be delivered along with this final report.

Before-After Evaluation

SPFs can be used as part of an EB before-after evaluation to estimate the safety effectiveness of engineering treatments/interventions. The EB method is able to account for possible bias due to RTM when sites may be selected for treatment based on high crash rates during a relatively short period of time. The EB method estimates the expected crashes that would have occurred in the after period without the treatment (π), compares that with the number of reported crashes in the after period (λ), and estimates an index of effectiveness, also called the crash modification factor (CMF). The following steps can be used to estimate the CMF:

1. Identify a reference group of untreated sites that is otherwise similar to the treatment group. For example, if the treatment being evaluated is the introduction of 3 foot paved shoulders on rural two lane roads that did not have any shoulders, the reference group will be a group of rural two lane roads without shoulders.

2. Use the data from the reference group to estimate SPFs as a function of traffic volumes and other site characteristics. The SPF estimation process will provide an overdispersion parameter. From the SPF, compute the annual factors for each year by taking the ratio of the observed crashes to predicted crashes. If the reference group is small, then it may not be possible to estimate reliable SPFs. In that case, one option is to calibrate the SPFs that have been developed in this project for a particular type of road by using the data from the reference group that has been identified. The calibration procedure is as discussed earlier for calibrating the HSM models with North Carolina data.

3. Use the SPFs, the annual factors, and the data on traffic volumes and site characteristics for each year in the *before* period for each treatment site to estimate the number of crashes that would be predicted in each year of the *before* period for each treatment site. This is the same as the procedure used in *Step 1 in the network screening illustration* with the main difference being that it is done using data for the treatment site before the treatment was implemented. The sum of these predictions for each site can be called as P_b, where the subscript b is added to denote that it is the predicted number of crashes for the before period.

4. Compute the EB expected number of crashes in the before period (called as EB_b) as the weighted average of P_b and X_b (the actual number of crashes in the before period). The approach here is the same as *Step 3 in the network screening illustration*, but with the main difference being that it is done using only the data before the treatment was implemented. The equation for EB_b is the following:

$$EB_b = P_b * w + X_b * (1-w) \tag{13}$$

In the case of SPFs for intersections or for SPFs for roadway segments where the overdispersion was assumed to be a constant, *w* is calculated as follows:

$$w = \frac{1}{1 + k * P_b} \tag{14}$$

If the overdispersion parameter was estimated such that it applies to a unit length of road for a roadway segment SPF, then *w* is computed as follows:

$$w = \frac{1}{1 + \left(\frac{k_1}{L}\right) * P_b} \tag{15}$$

5. Use the SPFs, the annual factors, and the data on traffic volumes and site characteristics for each year in the *after* period for each treatment site to estimate the number of crashes that would be predicted in each year of the *after* period for each treatment site. This is the same as Step 3 above, except that it is done with the data from the after period. The sum of these predictions for each site can be called as P_a, where the subscript a is added to denote that it is the predicted number of crashes for the after period.

6. The expected crashes that would have occurred in the after period without the treatment (π) is calculated as follows for each treatment site:

$$\pi = EB_b \left(\frac{P_a}{P_b}\right) \tag{16}$$

7. The variance of π is calculated as follows for each treatment site:

$$Var(\pi) = \pi \left(\frac{P_a}{P_b}\right)(1 - w) \tag{17}$$

8. The number of reported crashes in the after period (λ), π , and Var(π) are used to estimate the index of effectiveness (θ) (also called as the crash modification factor (CMF)) and the standard deviation of θ for each treatment site as follows:

$$\theta = \frac{\lambda / \pi}{1 + \left(Var(\pi) / \pi^2 \right)} \tag{18}$$

$$StDev(\theta) = \sqrt{Var(\theta)} = \sqrt{\frac{\theta^2 \left(\frac{1}{\lambda} + \frac{Var(\pi)}{\pi^2} \right)}{\left(1 + \frac{Var(\pi)}{\pi^2} \right)^2}} \tag{19}$$

The summation of π (π_{sum}) and its variance (Var(π_{sum})) are then used, along with the summation of crash counts after treatment (λ_{sum}), to estimate the overall (mean) index of effectiveness ($\overline{\theta}$), (i.e., the mean crash modification factor (CMF)), and the standard error of the mean CMF). This is done by replacing π by π_{sum}, Var(π) by Var(π_{sum}), and λ by λ_{sum} in equations 18 and 19 shown above.

If the CMF is greater than 1, it implies that the treatment can lead to an increase in crashes; whereas, if the CMF is less than 1, the treatment can lead to a reduction in crashes. The standard error of the mean value of the CMF makes it possible to determine if the CMF is statistically different from 1.0 for a specific level of significance. The percent change in crashes is 100(1–CMF); thus a value of CMF = 0.6 with a standard error of 0.12 indicates a 40 percent reduction in crashes with a standard error of 12 percent. The 95% confidence interval for the mean value of the CMF will be (0.6-1.96*0.12, 0.6+1.96*0.12), which will be (0.3648, 0.8352). Further description of this method is available in Hauer (1997).

Example

Following is an illustration of the before-after EB evaluation of a treatment that was evaluated recently as part of NCHRP Project 17-35 (Srinivasan et al., 2011). The treatment was the change from permissive to protected-permissive left turn phasing at signalized intersections in North Carolina. Data from 12 locations were used in this evaluation. A reference group of 49 signalized intersections was identified for the development of SPFs. The analysis looked at total intersection crashes, injury and fatal crashes, rear end crashes, and left turn opposing through (LTOPP) crashes. For illustrating the procedure, only the data for LTOPP crashes will be used.

The SPF for LTOPP crashes was:

LTOPP/intersection/year = $e^{-0.3696}\,(MajAADT/10000)^{0.5564}\,e^{0.6585*(MinAADT/10000)}$

Where, MajAADT is the major road AADT and the MinAADT is the minor road AADT. The overdispersion parameter (k) was 0.5641.

Table 7 shows the results of the EB analysis. Each site represents data for one intersection.

Table 7. Illustration of EB before-after analysis using data from NC intersections

Site	X_b	P_b	*w*	EB_b	P_a	λ	π	**Var(π)**
1	10	5.535	0.243	8.917	11.391	14	18.350	28.603
2	2	1.829	0.492	1.916	3.224	6	3.377	3.023
3	5	6.473	0.215	5.317	9.873	9	8.109	9.710
4	12	7.721	0.187	11.201	8.407	8	12.197	10.801
5	1	11.003	0.139	2.388	11.074	2	2.403	2.083
6	8	8.985	0.165	8.162	5.695	1	5.174	2.739
7	22	5.571	0.241	18.034	3.610	14	11.686	5.744
8	3	7.724	0.187	3.882	7.523	8	3.781	2.995
9	8	14.827	0.107	8.729	2.729	0	1.606	0.264
10	21	9.757	0.154	19.271	6.231	18	12.307	6.651
11	7	3.981	0.308	6.070	7.345	4	11.200	14.298
12	31	9.334	0.160	27.542	14.147	31	41.743	53.170
All	130	92.740		121.429	91.249	115	131.933	140.080

For example, in the first site, the observed number of crashes in the before period was 10 (X_b), and the predicted number of crashes from the SPF in the before period was 5.535 (P_b). *w* is then equal to $\frac{1}{1+0.5641*5.535}=0.243$

The EB estimate of the crashes in the before period (EB_b) = 5.535*0.243 + 10*(1-0.243) = 8.917.

The predicted number of crashes from the SPF in the after period was 11.391 (P_a). Hence, the EB expected number of crashes in the after period had the treatment not been implemented (π) is equal to $8.917 * \frac{11.391}{5.535} = 18.350$.

Var(π) = $18.350 * \left(\frac{11.391}{5.535}\right) * (1 - 0.243) = 28.603$

The last row in the Table shows total values. By comparing the total for X_b with the total for EB_b, it is possible to get an idea of the magnitude of the bias due to regression to the mean had we done a simple before-after comparison instead of a before-after comparison using the EB method.

From the total values for the last 3 columns in the Table, it is possible to calculate the CMF and the standard error of the CMF as follows:

$$CMF = \frac{\frac{115}{131.933}}{(1 + \frac{140.080}{131.933^2})} = 0.865$$

Standard error of the CMF = $\sqrt{\frac{0.865^2\left(\frac{1}{115} + \frac{140.080}{131.933^2}\right)}{\left(1 + \frac{140.080}{131.933^2}\right)^2}} = 0.111$

So, based on this dataset, the 95% confidence interval for the CMF is (0.865-1.96*0.111, 0.865+1.96*0.111), which is (0.647, 1.083).

Evaluating Systemwide Improvements/Treatments

In some cases, treatments may be installed systemwide for a particular type of facility. For example, a jurisdiction may decide to increase the retroreflectivity of all their stop signs (Persaud et al, 2007). Since sites are not specifically selected based on their crash history, there is no risk of bias due to regression to the mean. However, it is still necessary to account for changes in traffic volume and other trends. To evaluate the safety of such installations, a reference group is not necessary, but a *comparison group* is necessary in order to account for trends. SPFs can be estimated using the before-data from the treatment sites and these SPFs can be used to account for changes in traffic volumes. In addition, SPFs could be estimated for a group of comparison sites and the annual factors from these SPFs can be used to account for trends. Further details about such evaluations can be found in Bahar et al., (2004) and Persaud et al., (2007).

6. DEVELOPING SAFETY PERFORMANCE FUNCTIONS IN THE FUTURE

As vehicle technology, engineering treatments, reporting practices, and other things change in future years, the safety performance functions and calibration factors developed in this effort will become less accurate at predicting expected crash frequencies on North Carolina roads. It will be beneficial for NCDOT to use the most recent years of data to re-develop or re-calibrate the SPFs. The two products of this project, state-specific SPFs and calibration factors for HSM SPFs, can be updated as described below.

Updating state-specific SPFs developed with NC data

There are two main options for updating the SPFs that were developed specifically for North Carolina in this effort.

First, if sufficient expertise is available, the SPFs may be re-developed. This would involve an analyst repeating the process used in this effort – assembling a dataset of all applicable roadway segments, obtaining crash and volume information for all segments, and performing the negative binomial regression to develop new SPFs. This process would be facilitated by using the SAS code used in this effort, which is provided in Appendix F. The analyst may wish to assemble a completely new set of roadway segments, or he or she may wish to use the dataset used in this effort and simply update the AADT information.

Second, the NC-specific SPFs developed in this effort may be used in future years by simply calculating a calibration factor for each future year, similar to the way in which calibration factors were calculated for the HSM SPFs. That process would involve the following steps:

1. Identify the roadway segments to use in the calibration. The data necessary to calibrate the type 1 SPFs is minimal (the only characteristics needed for each segment would be segment length and AADT), so it would be beneficial to use all available segments state-wide rather than a small sample.
2. For the year of calibration, assign that year's AADT and count of observed crashes to each segment.
3. Apply the appropriate NC-specific SPF to each segment to predict the average crash frequency for the year of calibration.

4. For each facility type, divide the sum of observed crashes by the sum of predicted crashes to calculate the calibration factor.

Following is a hypothetical example to illustrate the steps discussed above for estimating the calibration factor for 2010 for the SPF for total crashes for rural 4 lane freeways within the influence of interchanges. Let us assume that there are 1099 roadway segments of this type. As mentioned in steps 1 and 2 above, data are needed on the length of the segment, AADT, and count of observed crashes in 2010 (see first 4 columns in Table 8).

From Appendix A, the predicted number of total crashes per year for a segment of length L is:
=L*exp{-7.9146+0.9811*ln(AADT)}

For site number 1, the predicted crashes will be:
= 0.3*exp{-7.9146+0.9811*ln(34000)} = 3.06. This number is shown in the last column.

Similar calculations need to be done for each segment. The sum of observed crashes and predicted crashes is then calculated. In this hypothetical example, the sum of observed crashes is 3302 and the sum of predicted crashes is 3155.4. The ratio of 3302 and 3155.4 is the calibration factor for 2010.

Table 8. Illustration of Steps to Calibrate NC Specific Models using Hypothetical Data

Site Number	AADT	Length (miles)	Observed Crash Count	Predicted Crash Count
1	34000	0.3	2	3.06
2	18000	0.7	1	3.83
3	11000	1.0	5	3.37
…	…	…	…	…
…	…	…	…	…
…	…	…	…	…
…	…	…	…	…
1099	101000	0.8	31	23.74
			SUM = 3302	SUM = 3155.4
		Calibration Factor = 3302/3155.4 = 1.046		

When conducting a network screening analysis, the calibration factors would be used by multiplying the predicted crashes for a particular year by the calibration factor for that year to arrive at the average expected crash frequency for a particular site.

Updating calibration factors for HSM SPFs

The calibration factors for HSM SPFs that were developed in this effort can be re-calculated for any future year. The spreadsheets used in the development of the calibration factors were designed to be used repeatedly and updated easily. At a minimum, each sample site would have to have the following values updated for the intended year of calibration:

- AADT. For segments, this will be a single AADT value; for intersections, this will be the AADT on both the major and minor roads.
- Observed crashes. This will be the total observed crashes on the segment or intersection for the year of calibration.

A more thorough effort would also make sure that no site characteristics have changed significantly, such as road widening or an expansion of town limits and land development so that a site which was previously rural is now urban. Both of these example changes would change the facility type of the road, so that it would no longer be appropriate for the sample in which it was previously included. Some of the minor characteristics may change through the years as well, such as the presence of lighting, and it would be beneficial from time to time to check each sample site for modifications such as these and update the spreadsheet accordingly.

Once the AADT values are updated in the calibration spreadsheet, the CMF values and SPF predictions in the subsequent columns will update automatically (i.e., they contain live formulas). The calibration factor can then be recalculated by dividing the sum of the observed crashes by the sum of the predicted crashes. If NCDOT desires to add additional sites to any calibration spreadsheet, it will simply be necessary to make sure all the CMF and SPF formulas are copied down for the additional rows.

REFERENCES

Baek, J. and Hummer, J.E. (2008), Collision Models for Multilane Highway Segments to Examine Safety of Curbs, *Transportation Research Record* 2083, pp. 128-136.

Bahar, G., Mollett, C., Persaud, B., Lyon, C., Smiley, A., Smahel, T., and McGee, H. (2004), Safety Evaluation of Permanent Raised Pavement Markers, *NCHRP Report* 518, Transportation Research Board, Washington, D.C.

Council, F.M., B. Persaud, K. Eccles, C. Lyon, and M.S. Griffith (2005), *Safety Evaluation of Red-Light Cameras*, Federal Highway Administration, Report # FHWA-HRT-05-048.

Council, F.M., R. Srinivasan, B. Hejazi (2007), *Identification of High Car-Truck Crash Corridors on North Carolina Interstate Roadways*, Submitted to FMCSA and FHWA, March 2007.

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

Harkey, D., et al. (2008), Crash Reduction Factors for Traffic Engineering and ITS Improvements, *NCHRP Report* 617, National Cooperative Highway Research Program.

Harwood, D. W., K. M. Bauer, K. R. Richard, D. K. Gilmore, J. L. Graham, I. B. Potts, D. J. Torbic, and E. Hauer (2007), Methodology to Predict the Safety Performance of Urban and Suburban Arterials, *NCHRP Web Document No. 129: Phases I and II*, National Cooperative Research Program.

Harwood, D. W., D. J. Torbic, D.K. Gilmore, C.D. Bokenkroger, J.M. Dunn, C.V. Zegeer, R. Srinivasan, D. Carter, C. Raborn, C. Lyon, and B. Persaud (2008), Pedestrian Safety Prediction Methodology, *NCHRP Web Document No. 129: Phase III*, National Cooperative Research Program.

Hauer, E. (1997), *Observational Before-After Studies in Road Safety*, Elsevier Science, Oxford, U.K.

Hauer, E. (2001), Overdispersion in Modeling Accidents on Road Sections and in Empirical Bayes Estimation, *Accident Analysis and Prevention*, Vol. 33(6), pp. 799-808.

Hauer, E., Kononov, J., Allery, B., and M.S. Griffith (2002), Screening the Road Network for Sites with Promise, *Transportation Research Record* 1784, pp. 27-32.

Hauer, E. (2004), Statistical Road Safety Modeling, Transportation Research Record 1897, pp. 81-87.

HSM (2010), *Highway Safety Manual*, AASHTO.

Hummer, J., et al., (2010a), *Procedure for curve warning signing, delineation, and advisory speeds for horizontal curves*, Report FHWA/NC/2009-07, North Carolina Department of Transportation.

Hummer, J. et al., (2010b), *Superstreet Benefits and Capacities*, Report FHWA/NC/2009-06, North Carolina Department of Transportation.

Kononov, J., Bailey, B., and Allery, B.K. (2008), Relationships between Safety and Both Congestion and Number of Lanes on Urban Freeways, *Transportation Research Record* 2083, pp. 26-39.

McCullagh, P., and Nelder, J.A. (1989), *Generalized Linear Models*, Second Edition, Chapman and Hall, London.

Miaou, S.P. (1996). *Measuring the Goodness-of-fit of Accident Prediction Models* (FHWA-RD-96-040). Oak Ridge, TN: Oak Ridge National Laboratory.

Moon, J.-P. and J.E. Hummer (2009), Development of Safety Prediction Models for Influence Areas of Ramps in Freeways, *Journal of Transportation Safety & Security*, Vol. 1, pp. 1-17.

Persaud, B., Craig, L., Eccles, K., Lefler, N., and Amjadi, R. (2007), *Safety Evaluation of Increasing Retroreflectivity of STOP Signs*, FHWA-HRT-08-041, Federal Highway Administration, December 2007.

Phillips, S., Carter, D., Hummer, J., and Foyle, R. (2005), Safety Comparison of Four-Lane Median Divided and Five-Lane and TWLTL Segments, *3rd International Symposium on Highway Geometric Design*, Chicago, June 29-July 1, 2005.

Shankar, V., Mannering, F., and Barfield, W. (1995), Effect of Roadway Geometric and Environmental Factors on Rural Freeway Accident Frequencies, *Accident Analysis and Prevention*, Vol. 27(3), pp. 371-389.

Srinivasan, R., Council, F., Lyon, C., Gross, F., Lefler, N., and Persaud, B. (2008a), Safety Effectiveness of Selected Treatments at Urban Signalized Intersections, *Transportation Research Record* 2056, pp. 70-76.

Srinivasan, R., Carter, D., Persaud, B., Eccles, K., and Lyon, C. (2008b), Safety Evaluation of Flashing Beacons at Stop-Controlled Intersections, *Transportation Research Record* 2056, pp. 77-86.

Srinivasan, R., F. Gross, C. Lyon, B. Persaud, K. Eccles, A. Hamidi, J. Baek, S. Smith, N. Lefler, C. Sundstrom, and D. Carter (2011), Evaluation of Safety Strategies at Signalized Intersections, *NCHRP Report* 705, National Cooperative Highway Research Program.

Ullman, G.L., Finley, M.D., Bryden, J.E., Srinivasan, R., and Council, F.M. (2008), Traffic Safety Evaluation of Nighttime and Daytime Work Zones, *NCHRP Report* 627, National Cooperative Highway Research Program.

Vogt, A. and Bared, J.G. (1998), *Accident Models for Two-lane Rural Roads: Segments and Intersections*, Federal Highway Administration, Report # FHWA-RD-98-133.

Washington, S.P., Karlaftis, M.W., and Mannering, F.L. (2011), *Statistical and Econometric Methods for Transportation Data Analysis*, Second Edition, Chapman and Hall/CRC, Boca Raton, FL.

Xie, Y., Lord, D., and Zhang, Y. (2007), Predicting Motor Vehicle Crashes using Bayesian Neural Network Models, *Accident Analysis and Prevention*, Vol. 39(5), pp. 922-933.

Xie, Y. and Zhang, Y. (2008), Crash Frequency Analysis with Generalized Additive Models, *Transportation Research Record* 2061, pp. 39-45.

APPENDIX A. TYPE 1 SAFETY PERFORMANCE FUNCTIONS FOR 16 ROADWAY TYPES IN NORTH CAROLINA

This Appendix shows the type 1 SPFs developed for 9 crash types for 16 roadway types in North Carolina. As mentioned earlier, following is the functional form of the SPFs:

Y = L*(Annual Factor)* $\exp\{\alpha + \beta * \ln(AADT)\}$ = L*(Annual Factor)* $(e^{\alpha})(AADT^{\beta})$

Where, Y is the predicted number of crashes per year for a segment with length L (in miles), α is the intercept, β is the coefficient for ln(AADT), and Annual Factor is the Annual Calibration Factor for a particular year.

For each SPF, the tables show the coefficients, their standard errors, the overdispersion parameter (k_1), the observed number of crashes that were used for the estimation, the number of crashes predicted by the model, and goodness of fit (GOF) statistics. The two GOF statistics that are included are: Freeman-Tukey R^2 (R^2 FT) and the Pseudo R^2.

The Appendices also show the annual factors for each year of data that was used in the analysis. The annual factor for a particular year is defined as the ratio of observed crashes to predicted crashes from the SPF for that year. Annual factors are used to account for the effect of changes in factors such as weather, crash reporting practices, demography, and others (that are not explicitly considered in the model), over time. Since the annual factors are the ratio of observed to predicted crashes, a sufficient sample of crashes are needed in each year to get a reliable factor. So, annual factors based on less than 150 observed crashes per year (on average) (i.e., 750 crashes in 5 years) are shown in italics and should be used with caution.

Finally, for each roadway type, summary statistics about the data are shown as well. The statistics include the minimum, maximum, and average AADT, the minimum, maximum, and average segment length, the number of observations used in the development of the SPF, and the number of mile-years in the data.

Here is an example on an SPF can be used to estimate the average predicted number of crashes for a roadway segment:

For rural two lane roads, for total crashes, α is -4.0852 and β is 0.5830.

Hence, the average predicted number of total crashes in 2005 for a two lane rural road segment with length 2 miles and AADT of 3000 is:

Y = 2.0*(0.964)* $\exp\{-4.0852 + 0.5830 * \ln(3000)\}$ = 3.452

Similarly, for the same segment, the average predicted number of injury and fatal (KABC) crashes in 2005 is:

Y = 2*(0.990)* $\exp\{-5.2717 + 0.6071 * \ln(3000)\}$ = 1.313

Variable		Total		KABC		KAB		PDO		Lane Departure	
		Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
Intercept		-4.0852	0.0286	-5.2717	0.0431	-5.3024	0.0615	-4.7167	0.0349	-3.1631	0.0412
ln(AADT)		0.5830	0.0036	0.6071	0.0054	0.5018	0.0077	0.5978	0.0044	0.3436	0.0052
k_1		0.3110		0.4193		0.4123		0.3410		0.3833	
R^2_{FT}		0.287		0.156		0.090		0.221		0.182	
R^2 Pseudo		0.378		0.390		0.334		0.394		0.181	
Observed Crashes		155420		57570		23886		93102		57755	
Predicted Crashes		154858.8		57443.1		23877.3		92807.7		57767.3	
Annual Factors	2004	1.058		1.102		1.084		1.025		1.107	
	2005	0.964		0.990		0.992		0.943		0.988	
	2006	0.975		0.991		0.996		0.965		0.962	
	2007	1.015		1.008		1.009		1.023		0.971	
	2008	1.006		0.921		0.922		1.060		0.971	
		Single Vehicle		Multi vehicle		Wet		Night			
Intercept		-2.5289	0.0329	-8.3039	0.0497	-6.0011	0.0621	-3.8198	0.0399		
ln(AADT)		0.3217	0.0042	0.9861	0.0061	0.5963	0.0077	0.4318	0.0050		
k_1		0.2487		0.7324		0.5850		0.2927			
R^2_{FT}		0.286		0.103		0.082		0.202			
R^2 Pseudo		0.195		0.544		0.364		0.290			
Observed Crashes		91577		63843		25418		60196			
Predicted Crashes		91637.1		62736.3		25382.5		60193.9			
Annual Factors	2004	1.014		1.132		1.089		0.996			
	2005	0.936		1.012		0.990		0.939			
	2006	0.961		1.005		1.007		0.970			
	2007	1.030		1.001		0.792		1.037			
	2008	1.056		0.939		1.125		1.058			
AADT	Min	504									
	Average	2980									
	Max	20000									
Segment Length	Min	0.01									
	Average	0.36									
	Max	18.98									
Observations		264731									

SPFs for Rural Two Lane Roads

Variable		Total		KABC		KAB		PDO		Lane Departure	
		Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
Intercept		-7.9741	0.2280	-10.4511	0.3447	-10.1753	0.5080	-8.0595	0.2529	-5.0797	0.3082
ln(AADT)		0.9420	0.0221	1.0608	0.0332	0.9417	0.0489	0.9151	0.0245	0.5882	0.0300
k_1		0.2339		0.2467		0.2739		0.2478		0.3905	
R^2_{FT}		0.725		0.590		0.421		0.685		0.593	
R^2 Pseudo		0.510		0.601		0.535		0.495		0.219	
Observed Crashes		20018		5793		2199		13887		9190	
Predicted Crashes		19954.8		5785.0		2202.7		13847.1		9202.5	
Annual Factors	2004	1.250		1.242		1.174		1.240		1.322	
	2005	1.015		1.044		1.099		1.004		1.008	
	2006	0.961		0.980		0.990		0.951		0.956	
	2007	0.923		0.940		0.951		0.918		0.880	
	2008	0.890		0.830		0.804		0.922		0.850	
		Single vehicle		Multi vehicle		Wet		Night			
Intercept		-3.943	0.270	-17.753	0.367	-6.053	0.418	-6.762	0.317		
ln(AADT)		0.505	0.026	1.783	0.035	0.624	0.041	0.719	0.031		
k_1		0.313		0.299		0.774		0.265			
R^2_{FT}		0.656		0.641		0.445		0.593			
R^2 Pseudo		0.197		0.755		0.173		0.383			
Observed Crashes		12191		7827		5017		6632			
Predicted Crashes		12207.4		7743.3		5013.8		6635.1			
Annual Factors	2004	1.255		1.252		1.286		1.201			
	2005	1.018		1.022		0.944		1.021			
	2006	0.944		0.992		1.021		0.953			
	2007	0.897		0.956		0.781		0.915			
	2008	0.896		0.871		0.989		0.924			
AADT	Min	3500									
	Average	29843									
	Max	90000									
Segment Length	Min	0.01									
	Average	0.80									
	Max	8.76									
Observations		4454									
Mile-years		3555.4									

SPFs for 4 lane Rural Freeways outside the influence of interchanges

Variable		Total		KABC		KAB		PDO		Lane Departure	
		Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
Intercept		-4.2583	0.7094	-5.9223	0.9884	-4.3124	1.4734	-4.6207	0.7704	0.0000	
ln(AADT)		0.6229	0.0637	0.6599	0.0885	0.4136	0.1321	0.6234	0.0692	0.1514	0.0034
k_1		0.1827		0.2045		0.2737		0.1901		0.2595	
R^2_{FT}		0.620		0.501		0.371		0.593		0.552	
R^2 Pseudo		0.178		0.206		0.113		0.178		0.001	
Observed Crashes		4506		1294		411		3154		1651	
Predicted Crashes		4501.1		1291.3		410.9		3151.0		1651.4	
Annual Factors	2004	1.075		1.172		*1.139*		1.033		1.123	
	2005	1.013		0.981		*1.182*		1.025		1.140	
	2006	0.908		0.895		*0.999*		0.908		0.946	
	2007	1.039		1.015		*0.897*		1.058		0.904	
	2008	0.975		0.961		*0.845*		0.980		0.926	
		Single vehicle		Multi vehicle		Wet		Night			
Intercept		0.0000		-10.6728	0.9736	-3.7862	1.1851	-4.8954	0.9287		
ln(AADT)		0.1726	0.0030	1.1371	0.0870	0.4768	0.1065	0.5718	0.0832		
k_1		0.1990		0.2723		0.4102		0.1536			
R^2_{FT}		0.614		0.481		0.439		0.539			
R^2 Pseudo		0.000		0.323		0.063		0.193			
Observed Crashes		2088		2418		1405		1344			
Predicted Crashes		2088.9		2385.2		1408.5		1343.9			
Annual Factors	2004	1.095		1.078		1.002		0.880			
	2005	1.142		0.916		1.129		1.040			
	2006	0.955		0.885		0.984		0.961			
	2007	0.901		1.165		0.828		1.041			
	2008	0.940		1.001		1.063		1.057			
AADT	Min	9300									
	Average	71412									
	Max	134000									
Segment Length	Min	0.01									
	Average	0.50									
	Max	4.02									
Observations		610									
Mile-years		307.6									

SPFs for 6+ lane Rural Freeways outside the influence of interchanges

Variable		Total		KABC		KAB		PDO		Lane Departure	
		Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
Intercept		-7.9146	0.2569	-10.4048	0.3844	-10.0498	0.5663	-8.0833	0.2845	-4.4462	0.3288
ln(AADT)		0.9811	0.0248	1.0960	0.0367	0.9581	0.0540	0.9642	0.0274	0.5588	0.0318
k_1		0.1883		0.1914		0.1239		0.2073		0.2181	
R^2_{FT}		0.482		0.328		0.179		0.437		0.319	
R^2 Pseudo		0.398		0.489		0.503		0.388		0.173	
Observed Crashes		16243		4498		1492		11481		6223	
Predicted Crashes		16082.6		4470.9		1492.4		11373.6		6229.0	
Annual Factors	2004	1.269		1.335		1.301		1.237		1.320	
	2005	1.029		1.037		1.073		1.022		1.043	
	2006	0.951		0.935		0.926		0.955		0.952	
	2007	0.927		0.919		0.968		0.932		0.878	
	2008	0.891		0.829		0.754		0.916		0.822	
		Single vehicle		Multi vehicle		Wet		Night			
Intercept		-3.1612	0.2899	-16.0548	0.3910	-5.9298	0.4134	-6.8271	0.3530		
ln(AADT)		0.4622	0.0281	1.6806	0.0372	0.6517	0.0399	0.7630	0.0340		
k_1		0.1870		0.2681		0.3467		0.1958			
R^2_{FT}		0.382		0.442		0.212		0.322			
R^2 Pseudo		0.139		0.601		0.192		0.321			
Observed Crashes		8246		7997		3729		4854			
Predicted Crashes		8252.1		7793.6		3717.0		4845.5			
Annual Factors	2004	1.243		1.310		1.187		1.252			
	2005	1.011		1.057		1.031		1.002			
	2006	0.963		0.951		1.051		0.962			
	2007	0.895		0.960		0.776		0.919			
	2008	0.897		0.881		0.985		0.888			
AADT	Min	3900									
	Average	32226									
	Max	102000									
Segment Length	Min	0.01									
	Average	0.31									
	Max	1.00									
Observations		5461									

SPFs for 4 lane Rural Freeways within the influence of interchanges

Variable		Total		KABC		KAB		PDO		Lane Departure	
		Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
Intercept		-6.4026	0.7609	-7.5315	1.0559	-5.4835	1.5481	-6.8502	0.8250	0.0000	
ln(AADT)		0.8599	0.0677	0.8399	0.0937	0.5435	0.1374	0.8720	0.0734	0.1766	0.0033
k_1		0.1414		0.1647		0.1449		0.1541		0.1938	
R^2_{FT}		0.559		0.434		0.278		0.534		0.396	
R^2 Pseudo		0.184		0.202		0.145		0.184		0.008	
Observed Crashes		6503		1676		456		4763		1767	
Predicted Crashes		6489.5		1672.5		455.8		4751.7		1767.5	
Annual Factors	2004	1.031		1.043		*1.148*		1.020		1.170	
	2005	0.929		1.040		*1.098*		0.894		1.062	
	2006	1.053		1.054		*0.920*		1.050		0.958	
	2007	1.016		0.904		*0.914*		1.064		0.916	
	2008	0.980		0.983		*0.957*		0.978		0.930	
		Single vehicle		Multi vehicle		Wet		Night			
Intercept		0.0000		-12.3514	0.9870	-2.5140	1.2321	-6.1662	1.0089		
ln(AADT)		0.2005	0.0029	1.3451	0.0876	0.3950	0.1098	0.7208	0.0896		
k_1		0.1442		0.1976		0.3089		0.1449			
R^2_{FT}		0.475		0.510		0.337		0.462			
R^2 Pseudo		0.008		0.282		0.029		0.174			
Observed Crashes		2311		4192		1663		1704			
Predicted Crashes		2312.0		4162.2		1665.2		1702.8			
Annual Factors	2004	1.156		0.971		0.936		1.104			
	2005	1.071		0.850		0.962		1.003			
	2006	0.921		1.129		1.256		1.004			
	2007	0.937		1.068		0.778		1.006			
	2008	0.945		0.997		1.052		0.908			
AADT	Min	9300									
	Average	78306									
	Max	141000									
Segment Length	Min	0.01									
	Average	0.25									
	Max	1.00									
Observations		997									
Mile-years		244.3									

SPFs for 6+ lane Rural Freeways within the influence of interchanges

Variable		Total		KABC		KAB		PDO		Lane Departure	
		Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
Intercept		-5.8986	0.1819	-7.2034	0.2727	-5.9773	0.3794	-6.6241	0.2099	-4.1869	0.2856
ln(AADT)		0.7673	0.0188	0.7943	0.0281	0.5584	0.0392	0.7954	0.0217	0.4332	0.0296
k_1		0.3484		0.5047		0.452		0.3631		0.393	
R^2_{FT}		0.186		0.099		0.091		0.146		0.200	
R^2 Pseudo		0.191		0.195		0.141		0.212		0.103	
Observed Crashes		24590		8620		2944		15636		5256	
Predicted Crashes		24258.9		8555.8		2941.1		15441.9		5250.7	
Annual Factors	2004	1.116		1.191		1.156		1.067		1.172	
	2005	0.989		1.015		1.057		0.969		1.072	
	2006	0.988		0.977		0.984		0.987		0.967	
	2007	1.020		1.007		0.984		1.025		0.934	
	2008	0.959		0.857		0.836		1.014		0.872	
		Single vehicle		Multi vehicle		Wet		Night			
Intercept		-2.4450	0.2131	-9.5870	0.2662	-7.7941	0.3411	-5.0018	0.2492		
ln(AADT)		0.3115	0.0222	1.0942	0.0274	0.7903	0.0351	0.5567	0.0258		
k_1		0.2606		0.7037		0.5737		0.3189			
R^2_{FT}		0.307		-0.009		0.074		0.166			
R^2 Pseudo		0.068		0.241		0.221		0.140			
Observed Crashes		9272		15318		4597		7691			
Predicted Crashes		9269.4		14861.5		4559.1		7676.6			
Annual Factors	2004	1.046		1.167		1.149		1.065			
	2005	0.965		1.010		1.044		0.954			
	2006	0.967		1.010		1.021		0.957			
	2007	1.001		1.042		0.856		1.025			
	2008	1.021		0.930		0.981		1.007			
AADT	Min	560									
	Average	15850									
	Max	60000									
Segment Length	Min	0.01									
	Average	0.23									
	Max	7.99									
Observations		23779									
Mile-years		5423.3									

SPFs for Rural Multilane Divided Roads

Variable		Total		KABC		KAB		PDO		Lane Departure	
		Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
Intercept		-5.0970	0.3454	-5.7277	0.4882	-4.8814	0.7038	-6.1478	0.4046	-2.8604	0.5622
ln(AADT)		0.7309	0.0371	0.6868	0.0522	0.4735	0.0754	0.7945	0.0432	0.2951	0.0607
k_1		0.4011		0.5100		0.4808		0.4466		0.3814	
R^2_{FT}		0.149		0.101		0.114		0.109		0.193	
R^2 Pseudo		0.215		0.212		0.144		0.250		0.082	
Observed Crashes		4642		1608		500		2977		716	
Predicted Crashes		4523.9		1587.6		499.1		2888.6		715.6	
Annual Factors	2004	1.013		0.992		*1.197*		1.026		*1.148*	
	2005	1.080		1.090		*0.972*		1.077		*1.089*	
	2006	1.084		1.044		*1.080*		1.103		*0.918*	
	2007	1.099		1.099		*1.061*		1.094		*1.033*	
	2008	0.865		0.849		*0.725*		0.865		*0.831*	
		Single vehicle		Multi vehicle		Wet		Night			
Intercept		-2.0613	0.4565	-7.0747	0.4347	-6.5216	0.6288	-6.1506	0.5415		
ln(AADT)		0.2512	0.0494	0.9105	0.0465	0.7019	0.0670	0.6981	0.0577		
k_1		0.2624		0.6094		0.6079		0.5042			
R^2_{FT}		0.276		0.018		0.056		0.120			
R^2 Pseudo		0.073		0.249		0.226		0.237			
Observed Crashes		1060		3582		837		1173			
Predicted Crashes		1059.5		3440.1		827.5		1157.5			
Annual Factors	2004	1.170		0.971		1.009		1.035			
	2005	0.990		1.114		1.200		1.065			
	2006	0.951		1.133		0.998		0.938			
	2007	1.057		1.121		0.928		1.159			
	2008	0.847		0.877		0.938		0.873			
AADT	Min	600									
	Average	11504									
	Max	40000									
Segment Length	Min	0.01									
	Average	0.17									
	Max	4.52									
Observations		5011									
Mile-years		829.1									

SPFs for Rural Multilane Undivided Roads

Variable		Total		KABC		KAB		PDO		Lane Departure	
		Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
Intercept		-6.5287	0.1410	-7.6112	0.2116	-7.5830	0.3331	-7.3781	0.1697	-4.4394	0.2192
ln(AADT)		0.8777	0.0158	0.8801	0.0235	0.7381	0.0370	0.9212	0.0189	0.4597	0.0248
k_1		0.3518		0.4219		0.3346		0.3939		0.3028	
R^2_{FT}		0.110		0.062		0.033		0.081		0.048	
R^2 Pseudo		0.349		0.372		0.360		0.380		0.187	
Observed Crashes		16105		5511		1554		10241		2988	
Predicted Crashes		15765.1		5459.3		1552.2		10014.1		2987.5	
Annual Factors	2004	1.128		1.081		1.041		1.147		1.162	
	2005	1.006		1.011		1.065		0.999		0.950	
	2006	0.997		1.022		0.977		0.983		0.924	
	2007	1.018		1.007		0.947		1.012		0.980	
	2008	0.963		0.931		0.977		0.977		0.988	
		Single vehicle		Multi vehicle		Wet		Night			
Intercept		-3.7731	0.2011	-8.3571	0.1759	-8.4141	0.2869	-6.7158	0.2231		
ln(AADT)		0.3982	0.0228	1.0512	0.0196	0.8872	0.0317	0.7440	0.0249		
k_1		0.2700		0.4945		0.4589		0.3672			
R^2_{FT}		0.062		0.075		0.033		0.058			
R^2 Pseudo		0.155		0.393		0.404		0.337			
Observed Crashes		3377		12728		2631		3921			
Predicted Crashes		3377.0		12341.9		2607.6		3896.9			
Annual Factors	2004	1.036		1.158		1.166		1.041			
	2005	0.906		1.037		1.059		0.907			
	2006	0.951		1.014		1.024		0.989			
	2007	1.004		1.025		0.724		1.053			
	2008	1.101		0.929		1.073		1.043			
AADT	Min	510									
	Average	7072									
	Max	30000									
Segment Length	Min	0.01									
	Average	0.12									
	Max	2.00									
Observations		40328									
Mile-years		4651.2									

SPFs for Urban Two Lane Roads

Variable		Total		KABC		KAB		PDO		Lane Departure	
		Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
Intercept		-9.3620	0.9387	-11.0663	1.2461	-7.0119	1.8929	-9.9272	1.0818	-3.9755	1.1625
ln(AADT)		1.0594	0.0859	1.1076	0.1133	0.6166	0.1727	1.0754	0.0989	0.4626	0.1068
k_1		0.2599		0.1889		0.2104		0.3128		0.3362	
R^2_{FT}		0.471		0.379		0.146		0.423		0.350	
R^2 Pseudo		0.293		0.423		0.177		0.286		0.079	
Observed Crashes		2354		728		188		1596		724	
Predicted Crashes		2348.4		727.7		188.1		1592.8		725.6	
Annual Factors	2004	1.387		*1.356*		*0.987*		1.395		*1.317*	
	2005	1.003		*1.038*		*1.207*		0.989		*1.220*	
	2006	1.032		*1.059*		*0.942*		1.014		*0.823*	
	2007	0.837		*0.787*		*0.968*		0.865		*0.845*	
	2008	0.803		*0.819*		*0.894*		0.795		*0.793*	
		Single vehicle		Multi vehicle		Wet		Night			
Intercept		-3.5645	1.0329	-16.2203	1.3404	-8.8571	1.5142	-6.9370	1.2752		
ln(AADT)		0.4433	0.0949	1.6358	0.1220	0.8682	0.1382	0.7167	0.1166		
k_1		0.2781		0.3613		0.3767		0.2967			
R^2_{FT}		0.373		0.413		0.278		0.268			
R^2 Pseudo		0.085		0.412		0.221		0.188			
Observed Crashes		885		1469		471		608			
Predicted Crashes		886.0		1477.1		471.7		608.2			
Annual Factors	2004	1.310		1.447		*1.312*		*1.395*			
	2005	1.206		0.875		*0.907*		*1.288*			
	2006	0.860		1.131		*1.098*		*0.867*			
	2007	0.829		0.825		*0.679*		*0.751*			
	2008	0.803		0.788		*1.044*		*0.728*			
AADT	Min	1200									
	Average	56135									
	Max	117000									
Segment Length	Min	0.01									
	Average	0.29									
	Max	1.76									
Observations		872									
Mile-years		252.5									

SPFs for 4 lane Urban Freeways outside the influence of interchanges

Variable		Total		KABC		KAB		PDO		Lane Departure	
		Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
Intercept		-9.7466	1.0364	-12.1747	1.4958	-5.7495	2.1880	-9.6485	1.1459	0.0000	
ln(AADT)		1.0896	0.0930	1.1974	0.1336	0.5189	0.1963	1.0487	0.1028	0.1273	0.0049
k_1		0.1807		0.2609		0.2649		0.1949		0.2768	
R^2_{FT}		0.507		0.395		0.248		0.465		0.389	
R^2 Pseudo		0.358		0.415		0.172		0.338		0.018	
Observed Crashes		1960		586		177		1360		712	
Predicted Crashes		1946.2		578.2		176.8		1354.4		712.4	
Annual Factors	2004	1.476		*1.652*		*1.754*		1.380		*1.467*	
	2005	1.144		*1.308*		*1.179*		1.081		*1.132*	
	2006	1.003		*0.887*		*0.785*		1.059		*1.092*	
	2007	0.862		*0.848*		*0.796*		0.868		*0.742*	
	2008	0.838		*0.780*		*0.886*		0.862		*0.845*	
		Single vehicle		Multi vehicle		Wet		Night			
Intercept		0.0000		-19.9118	1.2970	-4.2387	1.6655	-8.8697	1.5166		
ln(AADT)		0.1460	0.0046	1.9329	0.1153	0.4921	0.1499	0.9019	0.1359		
k_1		0.2473		0.1827		0.4084		0.2740			
R^2_{FT}		0.425		0.484		0.335		0.383			
R^2 Pseudo		0.008		0.647		0.069		0.264			
Observed Crashes		877		1083		594		569			
Predicted Crashes		877.1		1053.3		594.0		567.0			
Annual Factors	2004	1.444		1.542		*1.151*		*1.429*			
	2005	1.133		1.145		*1.112*		*1.216*			
	2006	1.060		0.991		*1.321*		*0.960*			
	2007	0.802		0.918		*0.697*		*0.937*			
	2008	0.825		0.857		*0.880*		*0.755*			
AADT	Min	22000									
	Average	69769									
	Max	153000									
Segment Length	Min	0.01									
	Average	0.37									
	Max	2.55									
Observations		474									
Mile-years		174.0									

SPFs for 6 lane Urban Freeways outside the influence of interchanges

Variable		Total		KABC		KAB		PDO		Lane Departure	
		Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
Intercept		-15.0814	2.6849	-16.6910	3.5614	-9.5988	5.2398	-16.3297	3.1336	-11.9200	4.0734
ln(AADT)		1.5760	0.2324	1.6082	0.3077	0.8854	0.4532	1.6526	0.2712	1.1900	0.3526
k_1		0.1267		0.1033		0.1172		0.1443		0.1915	
R^2_{FT}		0.624		0.541		0.338		0.580		0.467	
R^2 Pseudo		0.191		0.297		0.157		0.175		0.091	
Observed Crashes		2033		592		165		1413		548	
Predicted Crashes		2027.4		588.7		164.8		1413.7		548.7	
Annual Factors	2004	1.158		*1.204*		*1.221*		1.144		*1.319*	
	2005	1.116		*1.090*		*0.967*		1.125		*0.996*	
	2006	0.924		*1.078*		*1.231*		0.870		*0.879*	
	2007	0.988		*0.901*		*0.968*		1.010		*1.008*	
	2008	0.858		*0.812*		*0.704*		0.871		*0.894*	
		Single vehicle		Multi vehicle		Wet		Night			
Intercept		-7.7348	3.4144	-20.1093	3.2191	-13.9727	4.5834	-10.2526	3.4822		
ln(AADT)		0.8444	0.2957	1.9757	0.2784	1.3590	0.3967	1.0526	0.3013		
k_1		0.1430		0.1492		0.2583		0.1582			
R^2_{FT}		0.503		0.589		0.419		0.492			
R^2 Pseudo		0.061		0.249		0.091		0.117			
Observed Crashes		661		1372		497		594			
Predicted Crashes		661.5		1364.3		497.8		592.7			
Annual Factors	2004	*1.228*		1.129		*1.180*		*1.210*			
	2005	*0.994*		1.180		*1.001*		*1.044*			
	2006	*0.912*		0.930		*1.001*		*0.929*			
	2007	*0.979*		0.992		*0.887*		*1.000*			
	2008	*0.957*		0.811		*0.983*		*0.886*			
AADT	Min	32000									
	Average	103580									
	Max	160000									
Segment Length	Min	0.01									
	Average	0.29									
	Max	2.13									
Observations		307									
Mile-years		88.3									

SPFs for 8+ lane Urban Freeways outside the influence of interchanges

Variable		Total		KABC		KAB		PDO		Lane Departure	
		Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
Intercept		-11.1319	0.5362	-11.8788	0.6989	-8.8167	0.9819	-12.1334	0.6069	-6.3954	0.6406
ln(AADT)		1.2944	0.0493	1.2506	0.0640	0.8511	0.0899	1.3517	0.0557	0.7327	0.0589
k_1		0.2275		0.2074		0.2190		0.2527		0.2032	
R^2_{FT}		0.416		0.317		0.147		0.389		0.296	
R^2 Pseudo		0.293		0.312		0.206		0.305		0.157	
Observed Crashes		12816		3723		991		8851		3051	
Predicted Crashes		12769.0		3734.6		991.9		8814.3		3060.7	
Annual Factors	2004	1.388		1.384		1.320		1.392		1.275	
	2005	1.023		1.062		1.117		1.005		1.091	
	2006	0.982		0.940		0.950		0.997		0.953	
	2007	0.877		0.882		0.814		0.871		0.826	
	2008	0.786		0.751		0.819		0.795		0.858	
		Single vehicle		Multi vehicle		Wet		Night			
Intercept		-5.9043	0.5945	-15.0010	0.6762	-10.5826	0.7843	-11.1403	0.7322		
ln(AADT)		0.7012	0.0547	1.6157	0.0620	1.1033	0.0719	1.1646	0.0670		
k_1		0.1904		0.2896		0.2763		0.2237			
R^2_{FT}		0.330		0.369		0.254		0.303			
R^2 Pseudo		0.160		0.338		0.239		0.293			
Observed Crashes		3536		9280		2693		3028			
Predicted Crashes		3543.0		9245.7		2701.2		3032.0			
Annual Factors	2004	1.259		1.445		1.320		1.368			
	2005	1.058		1.007		0.957		1.058			
	2006	1.004		0.971		1.004		1.005			
	2007	0.841		0.886		0.783		0.826			
	2008	0.848		0.759		0.947		0.774			
AADT	Min	3500									
	Average	53239									
	Max	121000									
Segment Length	Min	0.01									
	Average	0.23									
	Max	1.00									
Observations		2847									
Mile-years		643.3									

SPFs for 4 lane Urban Freeways within the influence of interchanges

Variable		Total		KABC		KAB		PDO		Lane Departure	
		Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
Intercept		-11.9916	0.6310	-12.3658	0.8190	-10.3456	1.1836	-12.5595	0.6811	-5.5056	0.7617
ln(AADT)		1.3748	0.0557	1.2985	0.0720	0.9972	0.1038	1.3931	0.0601	0.6701	0.0672
k_1		0.1309		0.1502		0.1246		0.1400		0.1136	
R^2_{FT}		0.557		0.447		0.235		0.533		0.392	
R^2 Pseudo		0.341		0.346		0.292		0.345		0.162	
Observed Crashes		11702		3359		803		8197		2428	
Predicted Crashes		11787.4		3374.3		804.8		8250.4		2432.5	
Annual Factors	2004	1.076		1.047		0.946		1.087		1.208	
	2005	0.979		1.100		1.102		0.936		1.050	
	2006	1.025		1.043		1.045		1.023		0.961	
	2007	1.015		0.978		0.973		1.026		0.944	
	2008	0.881		0.839		0.934		0.901		0.884	
		Single vehicle		Multi vehicle		Wet		Night			
Intercept		-4.7701	0.7260	-15.7380	0.7465	-9.8027	0.8695	-11.8954	0.8037		
ln(AADT)		0.6196	0.0641	1.6762	0.0658	1.0540	0.0766	1.2368	0.0705		
k_1		0.1126		0.1674		0.1643		0.1201			
R^2_{FT}		0.429		0.521		0.387		0.462			
R^2 Pseudo		0.142		0.381		0.242		0.391			
Observed Crashes		2853		8849		2637		2664			
Predicted Crashes		2858.8		8945.1		2654.5		2661.8			
Annual Factors	2004	1.183		1.037		1.029		1.089			
	2005	1.005		0.964		0.952		1.033			
	2006	0.961		1.047		1.111		1.023			
	2007	0.979		1.028		0.840		0.980			
	2008	0.904		0.875		1.037		0.902			
AADT	Min	22000									
	Average	85414									
	Max	160000									
Segment Length	Min	0.01									
	Average	0.20									
	Max	1.00									
Observations		1526									
Mile-years		302.8									

SPFs for 6 lane Urban Freeways within the influence of interchanges

Variable		Total		KABC		KAB		PDO		Lane Departure	
		Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
Intercept		-16.0765	1.8331	-15.3488	2.3246	-10.0961	3.3748	-17.0107	1.9890	-7.8470	2.2343
ln(AADT)		1.7117	0.1570	1.5478	0.1989	0.9672	0.2887	1.7596	0.1703	0.8640	0.1913
k_1		0.1076		0.1112		0.0842		0.1087		0.0951	
R^2_{FT}		0.553		0.437		0.254		0.544		0.373	
R^2 Pseudo		0.193		0.176		0.094		0.201		0.083	
Observed Crashes		5830		1772		378		4017		1073	
Predicted Crashes		5844.6		1773.7		378.3		4025.8		1073.5	
Annual Factors	2004	1.215		1.195		*1.234*		1.221		1.175	
	2005	1.062		1.189		*1.283*		1.008		1.040	
	2006	0.996		0.968		*0.967*		1.010		1.031	
	2007	0.978		0.911		*0.850*		1.009		1.015	
	2008	0.834		0.850		*0.803*		0.829		0.806	
		Single vehicle		Multi vehicle		Wet		Night			
Intercept		-5.9137	2.0151	-21.0584	2.2058	-12.2384	2.6196	-13.9229	2.2935		
ln(AADT)		0.7144	0.1727	2.1154	0.1889	1.2477	0.2243	1.4042	0.1962		
k_1		0.0831		0.1282		0.1453		0.0853			
R^2_{FT}		0.427		0.528		0.404		0.482			
R^2 Pseudo		0.065		0.224		0.105		0.202			
Observed Crashes		1290		4540		1181		1371			
Predicted Crashes		1291.2		4559.8		1183.8		1371.9			
Annual Factors	2004	1.202		1.226		1.183		1.196			
	2005	1.030		1.074		0.849		1.027			
	2006	0.972		1.001		1.168		0.973			
	2007	0.974		0.975		0.865		0.998			
	2008	0.892		0.815		0.964		0.880			
AADT	Min	22000									
	Average	117452									
	Max	160000									
Segment Length	Min	0.01									
	Average	0.19									
	Max	1.00									
Observations		607									
Mile-years		114.7									

SPFs for 8+ lane Urban Freeways within the influence of interchanges

Variable		Total		KABC		KAB		PDO		Lane Departure	
		Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
Intercept		-4.1053	0.2118	-4.5964	0.2822	-4.8039	0.4411	-5.0670	0.2377	-5.4482	0.4145
ln(AADT)		0.6750	0.0211	0.6113	0.0281	0.4886	0.0438	0.7297	0.0237	0.5657	0.0410
k_1		0.2553		0.3003		0.3023		0.2798		0.3036	
R^2_{FT}		0.103		0.064		0.031		0.079		0.096	
R^2 Pseudo		0.104		0.101		0.079		0.122		0.134	
Observed Crashes		32830		10479		2455		21882		2821	
Predicted Crashes		32635.0		10461.7		2456.4		21743.0		2814.3	
Annual Factors	2004	1.109		1.131		1.139		1.099		1.155	
	2005	0.959		0.975		0.964		0.948		0.973	
	2006	0.990		0.983		1.034		0.990		1.020	
	2007	1.010		0.971		0.949		1.029		0.984	
	2008	0.966		0.953		0.918		0.970		0.888	
		Single vehicle		Multi vehicle		Wet		Night			
Intercept		-5.0573	0.3925	-4.4787	0.2302	-7.1957	0.3563	-5.6393	0.3232		
ln(AADT)		0.5358	0.0389	0.7021	0.0230	0.8060	0.0352	0.6711	0.0321		
k_1		0.2818		0.3013		0.3306		0.3201			
R^2_{FT}		0.129		0.050		0.038		0.055			
R^2 Pseudo		0.133		0.100		0.172		0.126			
Observed Crashes		3082		29748		5640		6784			
Predicted Crashes		3073.3		29577.9		5625.3		6766.4			
Annual Factors	2004	1.050		1.115		1.175		1.061			
	2005	0.930		0.961		0.941		0.954			
	2006	1.030		0.986		1.025		1.008			
	2007	1.024		1.008		0.778		1.009			
	2008	0.985		0.963		1.097		0.985			
AADT	Min	550									
	Average	24327									
	Max	70000									
Segment Length	Min	0.01									
	Average	0.11									
	Max	2.34									
Observations		20406									
Mile-years		2235.3									

SPFs for Urban Multilane Divided Roads

Variable		Total		KABC		KAB		PDO		Lane Departure	
		Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
Intercept		-5.2849	0.3632	-5.8160	0.4926	-6.9439	0.8164	-6.4065	0.4189	-5.8399	0.7854
ln(AADT)		0.8114	0.0376	0.7560	0.0508	0.7198	0.0837	0.8813	0.0432	0.6024	0.0807
k_1		0.2355		0.2708		0.2462		0.2644		0.1817	
R^2_{FT}		0.138		0.105		0.078		0.113		0.067	
R^2 Pseudo		0.155		0.156		0.197		0.184		0.168	
Observed Crashes		8726		2949		667		5651		640	
Predicted Crashes		8576.3		2931.5		666.3		5540.3		639.0	
Annual Factors	2004	1.049		1.060		*1.087*		1.040		*1.043*	
	2005	0.859		0.841		*0.855*		0.864		*0.792*	
	2006	1.065		1.079		*1.107*		1.048		*1.103*	
	2007	1.086		1.024		*1.006*		1.124		*1.089*	
	2008	1.033		1.031		*0.960*		1.031		*0.991*	
		Single vehicle		Multi vehicle		Wet		Night			
Intercept		-4.9262	0.7778	-5.6674	0.3842	-6.8115	0.6344	-6.9903	0.6180		
ln(AADT)		0.5038	0.0801	0.8430	0.0397	0.7876	0.0652	0.8196	0.0635		
k_1		0.1635		0.2635		0.3517		0.3602			
R^2_{FT}		0.080		0.117		0.067		0.037			
R^2 Pseudo		0.132		0.157		0.186		0.175			
Observed Crashes		611		8115		1491		1704			
Predicted Crashes		610.6		7967.4		1474.8		1688.5			
Annual Factors	2004	*0.996*		1.053		1.011		1.035			
	2005	*0.741*		0.868		0.949		0.826			
	2006	*1.110*		1.062		1.151		1.097			
	2007	*1.030*		1.090		0.787		1.094			
	2008	*1.130*		1.026		1.153		1.004			
AADT	Min	840									
	Average	16398									
	Max	49000									
Segment Length	Min	0.01									
	Average	0.08									
	Max	1.10									
Observations		8245									
Mile-years		657.3									

SPFs for Urban Multilane Undivided Roads

APPENDIX B. TYPE 2 SAFETY PERFORMANCE FUNCTIONS FOR RURAL TWO LANE ROADS IN NORTH CAROLINA

This Appendix shows the type 2 SPFs developed for 9 crash types for rural two lane roads in North Carolina. In type 2 SPFs, which include other variables in addition to AADT, the relationship between crash frequency and the independent variables was again log-linear (again, consistent with the state of the art):

$$Y = L * \exp\{\alpha + \beta_1 * f_1(AADT) + \beta_2 f_2(X_2) + \beta_3 f_3(X_3) + \beta_4 f_4(X_4) + ...\}$$

Where f_1, f_2, f_3, and f_4, represent functions of the independent variables AADT, X_2, X_3, and X_4. To predict the crash frequency for a particular year, this number is multiplied by the annual calibration factor for that year. For a log-linear model to be estimated using generalized linear modeling techniques, the functions are typically either identity or natural logarithm. Again, α, β_1, β_2, β_3, β_4, are parameters (coefficients) estimated as part of the negative binomial model.

For rural two lane roads, the following variables were included in the model:

- ln(AADT/10000)
- AADT/10000
- Terrain: This included three categories: Flat, Rolling, and Mountainous. When a model is estimated with 3 categories, one category is considered a 'reference' category (i.e, coefficient of zero). In these models, the *mountainous* category was used as a reference.
- Shoulder width (in feet): This was the average of the shoulder width from the two sides of the road
- Shoulder type: This included two categories: unpaved and paved (Sections with *curb* were removed because only a very small number of sites had this shoulder type). Paved shoulder indicates that at least part of the shoulder is paved. Paved shoulder was considered a reference category.

For example, for the total crash SPF, the coefficients for the different variables are as follows:

Variable		**Estimate**
Intercept		0.8727
ln(AADT/10000)		0.4414
AADT/10000		0.4293
Terrain	Flat	0.1264
	Rolling	0.1368
	Mountainous	0.0000
Shoulder Type	Unpaved	0.0354
	Paved	0.0000
Shoulder Width (in feet)		-0.0164

The average predicted total crashes in 2006 for a 1.5 mile rural two lane road segment in rolling terrain with a 2 foot unpaved shoulder and an AADT of 1500 will be the following:

Y = 1.5*0.973*exp{0.8727 + 0.4414**ln*(1500/10000) + 0.4293*(1500/10000) + 0.1368 + 0.0354 - 0.0164*2}=1.854

On the other hand, if the road was in mountainous terrain with a 2 foot paved shoulder, the average predicted total crashes will be:

Y = 1.5*0.973*exp{0.8727 + 0.4414**ln*(1500/10000) + 0.4293*(1500/10000) - 0.0164*2}=1.560

Overall, the increase in the shoulder width was associated with fewer crashes, and unpaved shoulders were associated with more crashes compared to paved shoulders. Shoulder width seems to make the largest difference for lane departure crashes. Type of shoulder was most important for lane departure and wet crashes. Regarding terrain, for some of the crash types (e.g. lane departure), mountainous terrain was associated with more crashes compared to flat and rolling terrain, whereas, for other crash types (e.g., night crashes), flat and rolling terrain were associated with more crashes compared to mountainous terrain.

Variable		Total		KABC		KAB		PDO		Lane Departure	
		Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
Intercept		0.8727	0.0269	0.1221	0.0379	-0.5681	0.0564	0.2325	0.0324	0.4526	0.0399
ln(AADT/10000)		0.4414	0.0085	0.4924	0.0127	0.4643	0.0181	0.4340	0.0102	0.4397	0.0124
AADT/10000		0.4293	0.0212	0.3723	0.0305	0.1782	0.0438	0.4578	0.0249	-0.1475	0.0325
Terrain	Flat	0.1264	0.0130			-0.1725	0.0263	0.2177	0.0162	-0.1047	0.0178
	Rolling	0.1368	0.0119			-0.1134	0.0234	0.2290	0.0149	-0.0754	0.0159
	Mountainous										
Shoulder Type	Unpaved	0.0354	0.0132	0.0479	0.0192	0.0586	0.0282	0.0250	0.0156	0.0872	0.0206
	Paved										
Shoulder Width (in feet)		-0.0164	0.0015	-0.0244	0.0021	-0.0277	0.0033	-0.0114	0.0019	-0.0461	0.0023
k_1		0.3006		0.4074		0.4040		0.3283		0.3623	
R^2_{FT}		0.291		0.159		0.093		0.225		0.188	
R^2 Pseudo		0.391		0.402		0.351		0.408		0.220	
Observed Crashes		150900		55937		23292		90312		56372	
Predicted Crashes		150896.6		55950.2		23294.4		90299.7		56369.3	
Annual Factors	2004	1.052		1.096		1.079		1.019		1.102	
	2005	0.960		0.987		0.989		0.941		0.984	
	2006	0.973		0.988		0.998		0.963		0.964	
	2007	1.012		1.005		1.012		1.020		0.976	
	2008	1.003		0.920		0.920		1.058		0.972	

AADT	Min	504
	Average	2971
	Max	20000
Segment Length	Min	0.01
	Average	0.37
	Max	18.98
Observations		250771
Mile-years		92907.2

SPFS for Rural two lane roads - Type 2 - page 1

Variable		Single Vehicle		Multi vehicle		Wet		Night	
		Estimate	S.E.	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
Intercept		0.5978	0.0322	0.2401	0.0376	-0.7180	0.0543	-0.2910	0.0387
ln(AADT/10000)		0.4414	0.0099	0.7188	0.0147	0.4929	0.0182	0.4259	0.0119
AADT/10000		-0.3227	0.0265	0.7152	0.0339	0.3383	0.0429	0.0426	0.0298
Terrain	Flat	0.1969	0.0152	0.0019	0.0212			0.5117	0.0201
	Rolling	0.1863	0.0139	0.0456	0.0192			0.4426	0.0188
	Mountainous								
Shoulder Type	Unpaved	0.0372	0.0158			0.0858	0.0276	0.0453	0.0183
	Paved								
Shoulder Width (in feet)		-0.0136	0.0018	-0.0216	0.0023	-0.0236	0.0030	-0.0071	0.0021
k_1		0.2384		0.7147		0.5739		0.2825	
R^2_{FT}		0.288		0.104		0.083		0.207	
R^2 Pseudo		0.217		0.554		0.372		0.315	
Observed Crashes		89804		61096		24697		58939	
Predicted Crashes		89805.0		61292.7		24700.4		58941.4	
Annual Factors	2004	1.010		1.110		1.086		0.993	
	2005	0.934		0.995		0.989		0.938	
	2006	0.963		0.984		1.006		0.970	
	2007	1.034		0.977		0.792		1.038	
	2008	1.060		0.916		1.120		1.064	

AADT	Min	504
	Average	2971
	Max	20000
Segment Length	Min	0.01
	Average	0.37
	Max	18.98
Observations		250771
Mile-years		92907.2

SPFS for Rural two lane roads - Type 2 - Page 2

APPENDIX C. SAMPLE SIZES BY FACILITY TYPES

The appendix shows the amount of sample (miles for segments or sites for intersections) that was used for each facility type by geographic area in the calibration process.

	Coast	Mountain	Piedmont	All Regions Combined
Segments	**(miles)**	**(miles)**	**(miles)**	**(miles)**
Rural 4 Lane Divided	18.59	21.31	9.87	49.77
Urban 2 Lane Undivided (2U)	11.47	18.33	29.59	59.39
Urban 2 Lane with TWLTL (3T)*	3.15	0.72	3.7	7.57*
Urban 4 Lane divided (4D)	2.94	2.73	9.83	15.5
Urban 4 Lane Undivided (4U)	3.52	4.3	7.47	15.29
Urban 4 Lane with TWLTL (5T)	4.16	3.88	4.42	12.46
Intersections	**(sites)**	**(sites)**	**(sites)**	**(sites)**
Rural 2 Lane, minor rd stop controlled 3-leg (3ST)	75	32	26	133
Rural 2 Lane, signalized 4-leg (4SG)	4	3	12	19
Rural 2 Lane, minor rd stop controlled 4-leg (4ST)	40	4	15	59
Rural 4 Lane, signalized 4-leg (4SG)	10	4	9	23
Urban arterial, signalized 3-leg (3SG)	12	9	10	31
Urban arterial, minor rd stop controlled 3-leg (3ST)	26	32	15	73
Urban arterial, signalized 4-leg (4SG)	47	35	40	122
Urban arterial, minor rd stop controlled 4-leg (4ST)	6	5	9	20

* There was very limited mileage of this facility type throughout the state, so all available mileage was used in the calibration sample.

APPENDIX D. COUNTY CLASSIFICATIONS INTO GEOGRAPHICAL AREA

This appendix provides the geographic area classifications that were assigned to each county.

County Name	Geographic Area
ALAMANCE	Piedmont
ALEXANDER	Piedmont
ALLEGHANY	Mountain
ANSON	Piedmont
ASHE	Mountain
AVERY	Mountain
BEAUFORT	Coast
BERTIE	Coast
BLADEN	Coast
BRUNSWICK	Coast
BUNCOMBE	Mountain
BURKE	Mountain
CABARRUS	Piedmont
CALDWELL	Mountain
CAMDEN	Coast
CARTERET	Coast
CASWELL	Piedmont
CATAWBA	Piedmont
CHATHAM	Piedmont
CHEROKEE	Mountain
CHOWAN	Coast
CLAY	Mountain
CLEVELAND	Piedmont
COLUMBUS	Coast
CRAVEN	Coast
CUMBERLAND	Coast
CURRITUCK	Coast
DARE	Coast
DAVIDSON	Piedmont
DAVIE	Piedmont
DUPLIN	Coast
DURHAM	Piedmont
EDGECOMBE	Coast
FORSYTH	Piedmont
FRANKLIN	Piedmont
GASTON	Piedmont
GATES	Coast

GRAHAM	Mountain
GRANVILLE	Piedmont
GREENE	Coast
GUILFORD	Piedmont
HALIFAX	Coast
HARNETT	Coast
HAYWOOD	Mountain
HENDERSON	Mountain
HERTFORD	Coast
HOKE	Piedmont
HYDE	Coast
IREDELL	Piedmont
JACKSON	Mountain
JOHNSTON	Coast
JONES	Coast
LEE	Piedmont
LENIOR	Coast
LINCOLN	Piedmont
MACON	Mountain
MADISON	Mountain
MARTIN	Coast
MCDOWELL	Mountain
MECKLENBERG	Piedmont
MITCHELL	Mountain
MONTGOMERY	Piedmont
MOORE	Piedmont
NASH	Coast
NEW HANOVER	Coast
NORTHAMPTON	Coast
ONSLOW	Coast
ORANGE	Piedmont
PAMLICO	Coast
PASQUOTANK	Coast
PENDER	Coast
PERQUIMANS	Coast
PERSON	Piedmont
PITT	Coast
POLK	Mountain
RANDOLPH	Piedmont
RICHMOND	Piedmont
ROBESON	Coast

ROCKINGHAM	Piedmont
ROWAN	Piedmont
RUTHERFORD	Mountain
SAMPSON	Coast
SCOTLAND	Piedmont
STANLY	Piedmont
STOKES	Piedmont
SURRY	Mountain
SWAIN	Mountain
TRANSYLVANIA	Mountain
TYRRELL	Coast
UNION	Piedmont
VANCE	Piedmont
WAKE	Piedmont
WARREN	Piedmont
WASHINGTON	Coast
WATAUGA	Mountain
WAYNE	Coast
WILKES	Mountain
WILSON	Coast
YADKIN	Mountain
YANCEY	Mountain

APPENDIX E. CALIBRATION FACTORS BY FACILITY TYPE BY GEOGRAPHIC AREA

This appendix presents the results of calibration factors developed separately for each geographic area of the state for each facility type of segment or intersection. For comparison, the tables below also show the calibration factor for the entire sample with all regions combined. It is clear from the tables below that calibration factors can vary widely across the geographic areas. Some of this may also be due to limited sample size (many cells are represented by a sample with fewer than 100 crashes) leading to more variability in the results. Factors that are based on the desired sample size of at least 100 observed crashes per year are indicated in ***bold italics***.

For each combination of geographic area and facility type, the tables below show the calibration factor along with the number of total observed crashes and total predicted crashes for that year.

	2007											
	Coast			Mountain			Piedmont			All Regions Combined		
Segments	Calibration Factor	Observed Crashes	Predicted Crashes	Calibration Factor	Observed Crashes	Predicted Crashes	Calibration Factor	Observed Crashes	Predicted Crashes	Calibration Factor	Observed Crashes	Predicted Crashes
Rural 4 Lane Divided	0.21	10	48.0	0.71	40	56.5	2.19	90	41.2	***0.96***	140	145.6
Urban 2 Lane Undivided (2U)	1.92	64	33.3	2.03	66	32.6	***1.34***	142	105.8	***1.58***	272	171.6
Urban 2 Lane with TWLTL (3T)	4.24	42	9.9	0.00	0	2.2	5.11	63	12.3	***4.30***	105	24.4
Urban 4 Lane divided (4D)	2.60	35	13.4	***11.50***	101	8.8	***2.91***	144	49.5	***3.90***	280	71.7
Urban 4 Lane Undivided (4U)	***4.77***	102	21.4	***2.99***	106	35.4	***4.58***	236	51.6	***4.10***	444	108.3
Urban 4 Lane with TWLTL (5T)	1.32	52	39.4	1.24	46	37.1	***2.52***	119	47.3	***1.75***	217	123.8
Intersections												
Rural 2 Lane, minor rd stop controlled 3-leg (3ST)	0.59	20	34.1	0.54	17	31.4	0.57	23	40.6	0.57	60	106.0
Rural 2 Lane, signalized 4-leg (4SG)	0.94	12	12.8	0.89	13	14.5	1.23	83	67.6	***1.14***	108	94.9
Rural 2 Lane, minor rd stop controlled 4-leg (4ST)	0.85	24	28.3	0.72	5	7.0	0.72	33	45.6	0.77	62	80.8
Rural 4 Lane, signalized 4-leg (4SG)	0.69	70	101.3	1.00	19	19.0	0.42	75	179.5	***0.55***	164	299.8
Urban arterial, signalized 3-leg (3SG)	2.60	52	20.0	2.83	46	16.2	3.24	50	15.4	***2.86***	148	51.7
Urban arterial, minor rd stop controlled 3-leg (3ST)	2.19	51	23.2	0.83	11	13.3	1.94	21	10.8	1.75	83	47.4
Urban arterial, signalized 4-leg (4SG)	***2.95***	378	128.2	***1.54***	157	101.6	***4.44***	481	108.2	***3.01***	1016	338.1
Urban arterial, minor rd stop controlled 4-leg (4ST)	0.97	9	9.2	1.49	9	6.1	2.31	22	9.5	1.61	40	24.8

	2008											
	Coast			**Mountain**			**Piedmont**			**All Regions Combined**		
Segments	Calibration Factor	Observed Crashes	Predicted Crashes	Calibration Factor	Observed Crashes	Predicted Crashes	Calibration Factor	Observed Crashes	Predicted Crashes	Calibration Factor	Observed Crashes	Predicted Crashes
Rural 4 Lane Divided	0.10	5	47.97	0.93	53	56.69	1.97	81	41.17	***0.95***	139	145.83
Urban 2 Lane Undivided (2U)	1.97	67	33.99	1.90	62	32.57	***1.50***	162	108.25	***1.66***	291	174.81
Urban 2 Lane with TWLTL (3T)	3.75	38	10.14	0.92	2	2.18	4.06	50	12.32	3.65	90	24.64
Urban 4 Lane divided (4D)	2.68	36	13.44	11.28	99	8.78	***3.44***	173	50.28	***4.25***	308	72.49
Urban 4 Lane Undivided (4U)	4.02	86	21.38	***3.08***	109	35.40	***5.33***	349	65.42	***4.45***	544	122.20
Urban 4 Lane with TWLTL (5T)	1.68	66	39.39	1.24	46	37.11	***2.14***	101	47.30	***1.72***	213	123.80
Intersections												
Rural 2 Lane, minor rd stop controlled 3-leg (3ST)	0.54	19	35.11	0.52	17	32.67	0.67	28	41.98	0.58	64	109.76
Rural 2 Lane, signalized 4-leg (4SG)	0.53	7	13.09	1.41	21	14.89	1.18	82	69.26	***1.13***	110	97.24
Rural 2 Lane, minor rd stop controlled 4-leg (4ST)	1.06	31	29.11	0.41	3	7.26	0.40	19	47.00	0.64	53	83.37
Rural 4 Lane, signalized 4-leg (4SG)	0.49	51	104.50	0.76	15	19.62	0.42	78	185.17	***0.47***	144	309.29
Urban arterial, signalized 3-leg (3SG)	1.82	38	20.83	2.37	40	16.88	3.37	54	16.04	***2.46***	132	53.74
Urban arterial, minor rd stop controlled 3-leg (3ST)	1.82	44	24.20	1.52	21	13.82	3.10	35	11.29	2.03	100	49.31
Urban arterial, signalized 4-leg (4SG)	***2.74***	365	133.17	***1.10***	116	105.54	***4.62***	519	112.38	***2.85***	1000	351.08
Urban arterial, minor rd stop controlled 4-leg (4ST)	0.42	4	9.52	0.48	3	6.23	2.24	22	9.81	1.13	29	25.56

	2009											
	Coast			Mountain			Piedmont			All Regions Combined		
Segments	Calibration Factor	Observed Crashes	Predicted Crashes	Calibration Factor	Observed Crashes	Predicted Crashes	Calibration Factor	Observed Crashes	Predicted Crashes	Calibration Factor	Observed Crashes	Predicted Crashes
Rural 4 Lane Divided	0.29	14	48.97	1.05	61	57.88	1.74	73	42.03	***0.99***	148	148.89
Urban 2 Lane Undivided (2U)	2.33	81	34.81	1.38	46	33.30	***1.04***	116	111.11	***1.36***	243	179.22
Urban 2 Lane with TWLTL (3T)	3.76	39	10.38	0.90	2	2.23	2.54	32	12.61	2.90	73	25.21
Urban 4 Lane divided (4D)	3.05	42	13.76	9.57	86	8.99	***2.48***	128	51.51	***3.45***	256	74.26
Urban 4 Lane Undivided (4U)	***5.07***	111	21.89	***2.90***	105	36.26	***3.45***	231	67.01	***3.57***	447	125.16
Urban 4 Lane with TWLTL (5T)	2.06	83	40.24	0.69	26	37.91	***2.13***	103	48.32	***1.68***	212	126.46
Intersections												
Rural 2 Lane, minor rd stop controlled 3-leg (3ST)	0.63	23	36.24	0.59	20	34.00	0.51	22	43.52	0.57	65	113.77
Rural 2 Lane, signalized 4-leg (4SG)	0.97	13	13.41	1.05	16	15.24	0.78	55	70.92	0.84	84	99.57
Rural 2 Lane, minor rd stop controlled 4-leg (4ST)	0.94	29	30.81	0.66	5	7.55	0.45	21	46.53	0.65	55	84.89
Rural 4 Lane, signalized 4-leg (4SG)	0.51	55	107.82	0.94	19	20.24	0.38	73	191.06	***0.46***	147	319.13
Urban arterial, signalized 3-leg (3SG)	2.03	44	21.66	2.28	40	17.56	1.98	33	16.67	***2.09***	117	55.89
Urban arterial, minor rd stop controlled 3-leg (3ST)	1.90	48	25.20	0.42	6	14.35	1.45	17	11.75	1.38	71	51.30
Urban arterial, signalized 4-leg (4SG)	***2.94***	406	138.29	***1.19***	130	109.60	***3.24***	380	117.26	***2.51***	916	365.16
Urban arterial, minor rd stop controlled 4-leg (4ST)	1.22	12	9.80	0.47	3	6.42	1.68	17	10.10	1.22	32	26.32

APPENDIX F. SAS CODE FOR ESTIMATING SPFS

This Appendix provides the SAS code that could be used by NCDOT to estimate SPFs. This program will estimate a type 1 SPF using negative binomial regression for total crashes in rural two lane roads. Two common modules are used to estimate negative binomial models in SAS: PROC GENMOD and PROC GLIMMIX. Both modules provide similar results. The code shown in the Appendix uses PROC GLIMMIX to estimate the model. The second part of the code is used for calculating the different goodness of fit measures and for the cumulative residual plot.

```
/* Program for total crashes on rural 2 lane rural roads - only AADT */
/* Developed by Raghavan Srinivasan, UNC Highway Safety Research Center for
NCDOT Project 2010-09 */

libname NC 'C:\Documents and Settings\rsrini\My Documents\NCDOT-SPF\rural-
2lane';

data input;

set NC.rural_2lane;

observed = total_crashes; /* the right hand side will change depending on the
crash type or severity */

/* Calculate F - this is for calculating the Freeman-Tukey R-square */

F = sqrt(observed) + sqrt(observed + 1);

run;
/***** MODEL where k is for 1 mile ***/

/* model with just intercept terms for calculating pseudo R-square based on
Miaou */

proc glimmix data = input;
weight seg_lng; /* this statement ensures that k is for 1 mile.  If k is to
be a constant, this statement should be removed */
model observed =    /
              dist = negbin link = log solution offset = lg_length;
/* lg_length is the natural logarithm of segment length */
run;
/***** Model with just AADT - same format as SafetyAnalyst */

/* lg_length is the natural log of the length.  This is included as offset to
ensure that the predicted crash frequency
is per mile */
/* lg_aadt is the natural log of AADT */

proc glimmix data = input;
weight seg_lng; /* this statement ensures that k is for 1 mile.  If k is to
be a constant, this statement should be removed */
model observed =   lg_aadt  / dist = negbin link = log solution offset =
lg_length ;
                    Output out = work.temp_out
                    residual(ilink)=r predicted(ilink)=p;
run;

data input out work.temp_out (keep = year AADT aadt_length seg_lng
observed r p F E resid_sq E_sq);
set work.temp_out;
resid_sq = r*r;
E = F - sqrt(4*p + 1);
E_sq = E*E;
run;

/* Calculating the traditional R-square and Freeman-Tukey R-square */
```

```
proc sql ;
select 1-(sum(resid_sq)/((count(*)-1)*var(observed))) into:r_square
from work.temp_out;
select 1-(sum(E_sq)/((count(*)-1)*var(F))) into:r_square_ft
from work.temp_out;
select 1-((count(*))/((count(*)-1)*var(F))) into:p_square_ft
from work.temp_out;
quit;
run;

/* Sort the data by Year to calculate total observed and total predicted
crashes */
proc sort data = work.temp_out out = out1;
by Year;
run;
/* Calculate total observed and total predicted */
proc means data = out1 SUM MIN MEAN MAX;
Var observed p aadt aadt_length seg_lng;
run;

/* Calculate total observed and total predicted by year */
/* Output is stored in a file to calculate annual factors later */
proc means data = out1 SUM;
Var observed p;
By Year;
output out = NC.tot_2lane_AADT_AF (keep = year observed_sum p_sum)
                  SUM=observed_sum p_sum;
run;
/* CURE PLOTS by AADT */

/* Sort the dataset you output from PROC GENMOD BY AADT */
proc sort data = work.temp_out out = out1;
by AADT;
run;

/* Assign the sum of all residuals to a macro variable SUM. This will
be used later to develop limits of CURE plot */

proc sql;
select sum(r*r) into: sum
from out1;
quit;
run;

/* To compute the Cummulative Residuals and the CURE Plot limits. I try
to replicate the equations suggested by Ezra in one of his papers */

data dat1;
cure = 0;
crsq = 0;
set out1 nobs=n;/* Assign number of observations in the output dataset to
                           a variable n */
do i = 1 to n; /* Go from first obs to last observation */
      set out1 point = i;
      cure = r + cure;
      rsq = r * r;
      crsq = rsq + crsq ;
```

```
      lim = (sqrt(crsq))*sqrt(1 - (crsq/&sum));
      uplim = 2*lim;
      lowlim = -2*lim;
      output;
end;
stop;
run;

/* To generate the CURE PLOT */
proc gplot data = dat1 ;
plot (cure uplim lowlim) * AADT / overlay;
run;
```