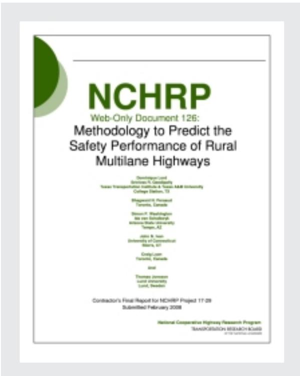
THE NATIONAL ACADEMIES PRESS



Methodology to Predict the Safety Performance of Rural Multilane Highways

DETAILS

0 pages | null | PAPERBACK
ISBN 978-0-309-43601-4 | DOI 10.17226/23097

AUTHORS

NCHRP

Web-Only Document 126:

Methodology to Predict the Safety Performance of Rural Multilane Highways

Dominique Lord
Srinivas R. Geedipally
Texas Transportation Institute & Texas A&M University
College Station, TX

Bhagwant N. Persaud
Toronto, Canada

Simon P. Washington
Ida van Schalkwyk
Arizona State University
Tempe, AZ

John N. Ivan
University of Connecticut
Storrs, CT

Craig Lyon
Toronto, Canada

And

Thomas Jonsson
Lund University
Lund, Sweden

Contractor's Final Report for NCHRP Project 17-29
Submitted February 2008

National Cooperative Highway Research Program
TRANSPORTATION RESEARCH BOARD
OF THE NATIONAL ACADEMIES

ACKNOWLEDGMENT

This work was sponsored by the American Association of State Highway and Transportation Officials (AASHTO), in cooperation with the Federal Highway Administration, and was conducted in the National Cooperative Highway Research Program (NCHRP), which is administered by the Transportation Research Board (TRB) of the National Academies.

COPYRIGHT PERMISSION

DISCLAIMER

THE NATIONAL ACADEMIES

Advisers to the Nation on Science, Engineering, and Medicine

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. On the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Ralph J. Cicerone is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. Charles M. Vest is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, on its own initiative, to identify issues of medical care, research, and education. Dr. Harvey V. Fineberg is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both the Academies and the Institute of Medicine. Dr. Ralph J. Cicerone and Dr. Charles M. Vest are chair and vice chair, respectively, of the National Research Council.

The **Transportation Research Board** is one of six major divisions of the National Research Council. The mission of the Transportation Research Board is to provide leadership in transportation innovation and progress through research and information exchange, conducted within a setting that is objective, interdisciplinary, and multimodal. The Board's varied activities annually engage about 7,000 engineers, scientists, and other transportation researchers and practitioners from the public and private sectors and academia, all of whom contribute their expertise in the public interest. The program is supported by state transportation departments, federal agencies including the component administrations of the U.S. Department of Transportation, and other organizations and individuals interested in the development of transportation. **www.TRB.org**

www.national-academies.org

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER II

CHAPTER III

CHAPTER IV

CHAPTER V

CHAPTER VI

CHAPTER VII

LIST OF FIGURES

CHAPTER V

ACKNOWLEDGMENTS

The research reported herein was performed under NCHRP Project 17-29 by the Texas Transportation Institute (TTI). The Texas A&M Research Foundation was the contractor for this study.

Dominique Lord of the Texas Transportation Institute was the principal investigator (PI). Primary authors of this report are: Dominique Lord (Assistant Professor, Texas A&M University), Bhagwant N. Persaud (Persaud and Lyon Inc.), Simon P. Washington (Professor, Arizona State University), John N. Ivan (Associate Professor, University of Connecticut), Ida van Schalkwyk (Research Associate, Arizona State University), Craig Lyon (Persaud and Lyon Inc.), Thomas Jonsson (Lund University), and Srinivas R. Geedipally (Texas Transportation Institute). The work was performed under the general supervision of Dr. Lord.

Several other researchers contributed to this research project. Dr. Shaw-Pin Miaou acted as co-PI between July 2004 and August 2005. His input and guidance while he was working on the project are greatly appreciated. Dr. Eun Sug Park was the lead researcher for the development of multivariate Poisson-lognormal models. Her contribution was extremely valuable for this project and provided a new approach for estimating multivariate models. Dr. Sudeshna Mitra developed several preliminary models for rural segments while she completed her doctoral degree at the University of Arizona and Arizona State University. Dr. Chen Zhang estimated several of the crash type models while she was working toward her Ph.D. at the University of Connecticut.

The authors would also like to recognize the following TTI staff for assistance with the research:

- Collecting and assembling documents for the literature review: Ms. Susie Catalina (TTI Librarian) and Ms. Sandy Tucker (former Main TTI Librarian);
- Reducing field data: Mr. Jeff Whitacre (Assistant Researcher), Mr. Danny Morris (Senior Research Associate);
- Interim Report preparation: Ms. Mel Campbell (Student Worker);
- Administrative Assistants: Ms. Nora Martinez, Ms. Christie Havemann, and Ms. Maria Medrano; and,
- The people at ITEC: Ms. Rhonda Brinkmann, Ms. Dolores Hott, and Mr. Richard Cole.

The authors wish to acknowledge the many other individuals who contributed to this research by participating in the on-line survey; by providing documents, papers, reports, or other types of information relevant for this research project; by providing crash and other related data; and by participating in the Joint NCHRP 17-25/17-29 Expert Panel meeting.

Participating in the on-line survey were Mr. Bob Euler and Mr. Barry Warhoftig (Western Virginia DOT), Mr. Michael Shea, Brian Chandler, and Mr. John Schaefer (Montana DOT), Mr. Ron Lipps (Maryland DOT), Mr. David L. Lippert (Illinois DOT), Ms. Debra Vermillion (Texas DOT), Mr. Tim McDowell (Wyoming DOT), Mr. Collin Franco (Rhode Island DOT), Dr. Mirinmay Biswas (North Carolina DOT), and Mr. Barry Partridge (Indiana DOT).

The following individuals provided documents, papers, reports, or other types of information relevant for this research project: Mr. Aaron Roozenburg (Beca Engineering, Christchurch, New Zealand) and Dr. Shane Turner (Beca Engineering, Christchurch, New Zeeland), Margaret Parkhill (iTRANS Consulting, Inc., Richmond, Ontario), Geni Bahar (iTRANS Consulting, Inc., Richmond, Ontario), Mr. Doug Harwood (Mid-West Research Insititute, Kansas City, Kansas), Dr. Sudeshna Mitra (California Polytechnic Institute, San Luis Obispo, California), and Dr. Graham Wood (Macquarie University, Sydney, Australia).

Crash and other related data were provided by Dr. David Harkey (Highway Safety Research Center, Charlotte, North Carolina), Dr. Forest Council (Highway Safety Research Center, Charlotte, North Carolina), Dr. Yusef Mohamedshah (Federal Highway Administration, Washington, D.C.), all the staff at the Transportation Data Office of the Washington State Department of Transportation (Tumwater, Washington), and Mr. Donald A. Terry (NYSDOT).

The following colleagues accepted the invitation to attend the Joint NCHRP 17-25/17-29 Expert Panel meeting held in Raleigh, North Carolina and provide very useful feedback: Ms. Kimberly A. Eccles (Vanesse Hagen Brustin Inc.), Dr. David Harkey (HSRC), Dr. Ezra Hauer (Consultant), Dr. Forrest Council (HSRC), Mr. Thomas M. Welch (Iowa DOT), Dr. Raghavan Srinivasan (HSRC), Mr. Loren Hill (Minnesota DOT), Dr. James A. Bonneson (TTI), Ms. Kim Eccles (HSRC), and Mr. Brian Mayhew (North Carolina DOT).

Last, but not least, we thank the Texas A&M University Research Foundation for providing their services for managing the contractual obligations associated with this project. Ms. Charlene Miller, Ms. Pam Allen, Ms. Sharon Moore, Ms. Joanie Birdwell, Ms. Pattie Sims, Ms. Lesa Feldhousen, Ms. Marcia Walker, and Ms. Phyllis Bonifazi were all very helpful.

SUMMARY

There is a significant need to improve the explicit consideration of highway safety in making decisions on roadway planning, design, and operations. To receive appropriate consideration, safety needs to be dealt with quantitatively within the transportation planning and highway design processes. The lack of available tools is a deterrent to quantifying the safety of a transportation facility during the planning or highway design process. Recognizing this problem, a group of Transportation Research Board (TRB) committees has identified the need for more explicit and quantitative consideration of safety within the above-mentioned processes. This important need eventually led to the development of the forthcoming Highway Safety Manual (HSM). The Manual will serve as a tool to help practitioners make planning, design, and operations decisions based on safety. It will serve the same role for safety analysis that the Highway Capacity Manual (HCM) serves for traffic-operations analyses. The product of this research will provide the necessary tools for estimating the safety performance of multilane rural highways and will be incorporated into Chapter 9 of the HSM (note: At the time this report was prepared, Chapter 11 had been tentatively designated as the chapter covering multilane rural highways; to be consistent with previous work on this project, the research team will keep referring to Chapter 9 for the methodology to estimate the safety performance of multilane rural highways).

The objective of this research is to develop a methodology that predicts the safety performance of various elements considered in the planning, design, and operation of non-limited-access rural multilane highways.

The literature review illustrated that the safety performance of multilane rural highways was seldom investigated, both for segments and intersections located on these facilities. Despite the limited availability of studies on this topic, previous research has found that multilane rural highway segments experience, on average, less crashes than two-lane rural highways for the same level of exposure. For intersections, only one study specifically focused on estimating the safety performance for these types of facilities. In short, the review indicated that there is a need to develop models and a methodology for estimating the safety performance of multilane rural highways.

To accomplish the study objective, the research team divided the work into several tasks. The first task consisted of conducting a limited survey of state transportation agencies. The objectives of this survey were to: a) determine whether the selected DOTs are currently using or developing statistical models to predict the safety performance of multilane rural highways; b) find out candidate input variables of interest to the survey participants; c) determine the availability of, and accessibility to, various databases, such as crash data, geometric design information for segments and intersections, traffic flows for segments, major and minor approaches.

The survey results showed that only two state agencies currently have a methodology for estimating the safety performance of multilane rural highways. The results also showed that crash data and segment files could be made available by all study participants. However, other databases, such as intersection databases, or access to georeferenced data, were not always available.

The second task consisted of collecting data using five state databases: California, Minnesota, New York, Texas, and Washington. The data were used for developing statistical models and AMFs for intersections and segments, as well as for a cross-validation study to evaluate the recalibration procedure for jurisdictions other than those for which the models were estimated. The data collected in Texas, California, Minnesota, and Washington were utilized for developing the models and AMFs. The New York data were used for the cross-validation. The data included detailed information about geometric design characteristics, traffic flow, and motor vehicle crashes.

In the third task, the research team developed the proposed accident prediction methodology. The methodology proposed in this research separated rural multilane road networks into segments and intersections. Specific models were developed for each transportation element. Three classes of models were proposed: models with covariates, baseline models applicable for specific values of covariates, and general Average Daily Traffic (ADT) models. For the first model class (models with covariates), the relationship between crashes and geometric design features was captured via the covariates inside the statistical model. It was proposed that models be estimated for undivided and divided segments, as well as for most types of intersections, and where possible by injury severity and crash type. The models with covariates produced for divided segments were used to estimate baseline models by substituting variables meeting baseline conditions for Chapter 9 of the HSM. For the second model class, statistical models were developed using a given set of baseline conditions. The baseline conditions usually reflect the nominal conditions agencies most often used for designing segments and intersections. Most models proposed for Chapter 9 were calibrated using this approach. For the third model class, general ADT models were developed for the following transportation elements: 4-lane undivided segments, 3- and 4-legged signalized intersections, as well as 3- and 4-legged unsignalized intersections. These models reflect the average conditions found in the data for each transportation element and can be used for cases where the user has limited information about the geometric design features for the particular project under study. For 4-legged signalized intersections, the general ADT models were used as baseline models in Chapter 9. Due to the small sample size, the general ADT models for 3-legged signalized intersections were not utilized as baseline models for the chapter.

The fourth task focused on describing the framework for developing the models. This framework, which is very important for developing sound and statistically valid predictive models, included four steps: determine the modeling objective matrix; establish the appropriate processes to develop the models; determine the inferential goals; and select the computation techniques and tools. All the models were developed using this modeling framework. The coefficients of the models were estimated using negative binomial (NB) regression methods, with the exception of models of crash counts by severity. In this task, various methods used for estimating AMFs were also described.

The last task consisted of summarizing the modeling results. Models were proposed for the three types of intersections, undivided and divided highway segments, by crash type, and by crash severity. More than 80 models were estimated in this research or derived from recent relevant research. The models were assessed using various goodness-of-fit measures, and most provided good statistical properties. From the models developed or assembled, a few were recommended to be included in Chapter 9 of the HSM. Table 1 summarizes the elements for which baseline models are recommended and the approach used to derive them. This table includes a reference to the table in the report in which the actual model is presented.

Table 1 Recommended Baseline Models for Chapter 9 of the HSM

Transportation Elements	**Approach Used**
Segments	
4-Lane Undivided	Models estimated from data meeting baseline conditions (Table 6.23). General ADT crash type models also estimated (Tables 6.43 – 6.48).
4-Lane Divided	Models estimated from models with covariates by substituting variables meeting baseline conditions (Table 6.34).
Intersections	
3-Legged Unsignalized	Models estimated from data meeting baseline conditions (Table 6.2).
4-Legged Unsignalized	Models estimated from data meeting baseline conditions (Table 6.1).
3-Legged Signalized	No models recommended due to small sample size.
4-Legged Signalized	General ADT models estimated (Table 6.4).

Several AMFs were produced from this work. Some of these AMFs were estimated using the models with covariates, while others were vetted by the Joint NCHRP 17-25/17-29 Expert Panel committee. The AMFs assembled from this work are summarized in Table 2.

Table 2 List of AMFs for Rural Multilane Highways

Intersections	**Segments**
● Sight distance (unsignalized 3- and 4-legged) ● Left-turning lane on major approach (unsignalized 3- and 4-legged) ● Right turning lane on major approach (unsignalized 3- and 4-legged) ● Intersection angle (unsignalized 3- and 4-legged)	● Lane width (divided and undivided) ● Shoulder width (divided and undivided) ● Sideslope ● Horizontal curve density ● Median width ● Median barrier

Finally, the cross-validation study showed that some models transferred very well, as long as they performed well in the jurisdiction where they were estimated.

CHAPTER I

Introduction

Background

There is a significant need to improve the explicit consideration of highway safety in making decisions on roadway planning, design, and operations. To receive appropriate consideration, safety needs to be dealt with quantitatively within the transportation planning and highway design processes (De Leur and Sayed, 2002; Lord and Persaud, 2004; Ladron de Guevara et al., 2004; Bonneson et al., 2005). For instance, many other factors, such as environmental impacts, capacity, and vehicle delays, among others, are already estimated in quantitative terms. Unfortunately, safety is often incorporated into these processes in a qualitative sense and, more often than not, the final outcome of the analysis is left to the judgment of the engineer or planner.

The lack of available tools is a deterrent to quantifying the safety of a transportation facility during the planning or highway design process. Recognizing this problem, a group of Transportation Research Board (TRB) committees has identified the need for more explicit and quantitative consideration of safety within the above-mentioned processes and proposed the development of a Highway Safety Manual (HSM). Portions of the Manual are being developed under several NCHRP projects, and it is expected that the HSM will be recommended for adoption by AASHTO. The HSM will serve as a tool to help practitioners make planning, design, and operations decisions based on safety. It will serve the same role for safety analysis that the Highway Capacity Manual (HCM) serves for traffic-operations analyses. A TRB Task Force for the Development of Highway Safety Manual was formed to guide the development of the HSM. The purpose of the HSM, as identified by the HSM Task Force (HSM TF), will be to provide the best factual information and tools in a useful and widely accepted form to facilitate specific consideration of safety in the decision-making process.

The TRB committees interested in the HSM have formed a Task Force to guide its development. Several NCHRP studies are currently under way or completed to develop key aspects of the HSM. NCHRP Project 17-18(4) recommended the content and outline of the first edition of the HSM and developed a plan for a research program needed to develop that first edition within five years. Recommendations developed in NCHRP Project 17-18(4) and approved by the HSM TF indicate that the first edition of the HSM should include models for making quantitative estimates of the expected safety performance of rural multilane highways, rural two-lane highways, and urban and suburban arterials. Research to address the latter two of these facility types has already been initiated and these projects were near completion at the time this report was prepared. Predictive models for rural two-lane highways have been developed by FHWA, and NCHRP Project 17-18(4) has developed a HSM chapter (i.e., Original Chapter 8 – now, tentatively labeled as Chapter 10) incorporating those models. NCHRP Project 17-26 is developing models for predicting the safety performance of urban and suburban arterials and the results will be incorporated into Chapter 10 (or 12). Therefore, the only facility type to be included in the first edition of the HSM for which further modeling work is needed is rural multilane highways.

Research Objectives and Scope

The objective of this research is to develop a methodology that will be used to predict the safety performance of various transportation elements considered in the planning, design, and operation of non-limited-access rural multilane highways. This methodology will be incorporated as Chapter 9 in Part III of the HSM (see Appendix G) (note: At the time this report was prepared, Chapter 11 has been tentatively designated as the chapter covering multilane rural highways; to be consistent with previous work on this project, the research team will keep referring to Chapter 9 for the methodology to estimate the safety performance of multilane rural highways).

The rural network to be considered in the research includes all highway facilities located in rural areas, other than freeways, that serve the primary function of providing movements among medium to large urban areas. The term “rural” is defined using FHWA and AASHTO guidelines, since these are commonly used by most transportation and highway agencies for consistency with Federal and national usage. These guidelines classify rural areas as places outside the boundaries of urban places where the population is less than 5,000 inhabitants. Consequently, any highway located outside the city limits of an urban agglomeration above 5,000 inhabitants would be considered rural (AASHTO, 2004).

The term “multilane” refers to facilities with four or more through lanes, that is, at least two lanes in each direction, as defined by the HSM TF. These facilities may be divided with a rigid or flexible barrier, paved or landscaped median, or a two-way left-turn (TWLT) lane, but should not have access and egress limited to grade-separated interchanges (i.e., not freeways). Facilities may have occasional grade-separated interchanges, but these should not be the primary form of access and egress, and the methodology proposed in this project would not apply to the sections passing through the interchanges.

Organization of The Report

A brief description of the content included in the remainder of the chapters is provided below:

Chapter II documents the literature review on existing statistical models, Accident Modification Factors (AMFs), and other predictive methodologies relevant to this study.

Chapter III summarizes the survey results conducted in this study. This chapter also describes the characteristics of the survey instrument, as well as the candidate input variables of interest by the study participants.

Chapter IV describes the characteristics of the data collected in this project. The data were assembled for developing statistical models, estimating AMFs, and evaluating the re-calibration of models. Data collected from the states of Texas, California, Minnesota, New York, and Washington were used in this project.

Chapter V describes the model development and accident prediction methodology. The chapter provides details about the model classes and functions, the modeling framework, and methods for estimating AMFs.

Chapter VI summarizes the modeling results for estimating the safety performance of rural multilane highways. The chapter provides the modeling results for intersection and segment models, as well as models developed by crash type and accident modifications factors recommended for application in the prediction methodology. The chapter also includes the results of the cross-validation study.

Chapter VII provides a summary of the work accomplished in this project and proposes recommendations for further work.

The appendices document additional models from related research for consideration, HSM Chapter 9, the survey instrument, Cumulative Residuals (CURE plots), crash severity models, a methodology for estimating the variance of the product between baseline models and AMFs, and the Joint NCHRP 17-25/17-29 Expert Panel assessment.

CHAPTER II

Literature Review

This chapter documents the literature review relevant for this project. The chapter summarizes key outcomes of published and unpublished documents as well as existing research activities in the United States (U.S.). The chapter is divided into two sections. The first section summarizes key studies on the estimation of the safety performance of multilane highways in North America and around the world. The second section describes studies related to the safety performance of rural intersections located on multilane highways.

Safety Performance of Multilane Rural Highways

There are more than 4 million miles of roadway in the United States (BTS, 2005). About 77% of these roadways, roughly 3.1 million miles, are located in or near rural areas (based on a population below 5,000). Wang et al. (1998) reported that about 35,000 miles of arterial highways are classified as non-freeway rural highways. This immense rural network results in a widespread dispersal of fatal and non-fatal injuries, and property damage only (PDO) crashes across the U.S. Although fewer vehicles travel in rural areas, a driver is at greater risk to be involved in a collision than in urban areas.

According to the United States General Accounting Office (U.S. GAO, 2004), the National Highway Traffic Safety Administration (NHTSA) estimates that crashes occurring on rural highways account for more than 60% of all fatal crashes occurring on the United States highway network. The significance of these deaths is even more striking considering that the rural network carries only 40% of the nation’s traffic in any given year.

As reported by the US DOT (2002), non-freeway rural highway facilities (i.e., two-lane and multilane highways combined) experience between 1.2 and 2 times more fatal crashes than urban highways grouped under the same functional highway classification. When the number of vehicle-miles traveled is included in the comparison, the gap between rural and urban highways becomes even larger, as seen in Table 2.1.

Table 2.1 Fatal Crashes and Fatal Crash Rate by Highway Classification (US DOT, 2002)

Functional Classification	**1998**		**1999**		**2000**	
Rural	**Number**	**Rate**	**Number**	**Rate**	**Number**	**Rate**
Principal Arterial	5,485	2.3	5,385	2.2	5,236	2.1
Minor Arterial	4,300	2.6	4,300	2.6	4,352	2.5
Major Collector	5,956	2.9	5,933	2.9	5,783	2.8
Urban						
Principal Arterial	5,322	1.4	5,107	1.3	5,157	1.3
Minor Arterial	3,359	1.1	3,227	1.0	3,335	1.0
Major Collector	1,044	0.8	1,039	0.8	1,036	0.8

There has been a significant amount of research conducted on the safety performance of rural highways. Most of this research has been concentrated on rural two-lane highways (Vogt and Bared, 1998; Harwood et al., 2000; Qin et al., 2004). This focus is not surprising given the fact that the rural network is composed mainly of two-lane highways. Consequently, there has been very little research conducted on the safety performance of multilane highways in rural areas.

The research carried out on multilane rural highways can usually be grouped under two categories. The first category regroups studies that specifically looked at the safety performance of multilane rural highway facilities. The second category encompasses research that focused on the safety effects of converting rural two-lane highways to either undivided or divided multilane facilities. Studies categorized under both categories are described in the next two sections.

Multilane Highways

There have been few studies that examined the safety performance of multilane rural highways in North America. In a 1991 study, Persaud (1991) investigated the safety performance of rural and urban multilane highway segments in Ontario, Canada. To accomplish this task, Persaud developed several statistical models and used the empirical Bayes (EB) method to refine (or improve) the estimated long-term safety performance of these facilities. He found that statistical models predicted more collisions for rural multilane highways than urban segments for the same level of exposure and number of lanes. As expected, divided rural segments performed better than undivided segments.

Fitzpatrick and Balke (1995) examined the safety performance of TWLT lanes on four-lane rural highways in Texas. This type of design is often used on rural highways located adjacent to urban areas, where the adjacent land is in the process of being fully urbanized for commercial activity. Based on a cross-sectional study, their results show no statistical differences in the number of crashes between highways with TWLT lanes and highways with flush medians with low driveway densities. The authors also found no difference in the way these two median treatments operated in rural areas.

In a third study, Wang et al. (1998) evaluated the safety performance of non-freeway multilane rural highways in five states using data from the Highway Safety Information System (HSIS), a database managed by the University of North Carolina under contract from FHWA. These authors collected supplemental information at the sites selected as part of their study. They developed a statistical model using explanatory variables, such as access control, shoulder width, area type (e.g., rural municipal or non-municipal), and functional classification (e.g., principal and minor arterials). The outcome of the analysis showed that functional class and area type explained most of the variation for the number of crashes on multilane highway facilities. Several highway geometric design elements influenced the safety performance, but their impacts were less significant.

Council and Stewart (1999) developed statistical models for rural four-lane undivided and divided highways. The models were developed as part of a cross-sectional study for comparing the safety performance between rural two-lane and four-lane highways. Council and Stewart used data collected in California, Minnesota, North Carolina, and Washington. The models were developed using traffic flow and shoulder width as input variables. Larger shoulder widths were associated with fewer crashes for divided four-lane highways.

Persaud and Bahar (2000) investigated the use of statistical models for screening high-risk sections of rural highways in Ontario, Canada. They developed a Potential for Safety Improvement Index (PSI) for identifying these sites. Several statistical models were produced, including models for rural divided and undivided four+-lane highways. Similar to the study conducted by Persaud (1991), the models predicted more crashes for undivided than divided rural multilane highways.

The *SafetyAnalyst* project has developed a suite of models for predicting the safety performance of multilane highways that are detailed in an unpublished 2004 technical memorandum. These models will be used in the *SafetyAnalyst* interim tools and will subsequently be improved for use in the final tools if required. Crash models for both all crash severities (i.e., total) and fatal and non-fatal injuries, including possible injury, crashes were calibrated using data from Minnesota, Ohio, Washington, and North Carolina. Separate models were developed for rural multilane undivided and rural multilane divided roadways. All statistical models used input flows as the only covariate.

At the international level, a more substantial amount of research has been conducted on the safety performance of multilane highways. However, given the fact that the applicability of this research to the U.S. is limited, only the most relevant study outcome is summarized herein. Mountain et al. (1996) developed statistical models for predicting the safety performance of four-lane rural highways in the U.K. The number of minor intersections on multilane highways was positively associated with the number of crashes, although in a non-linear fashion. Amoros et al. (2003) compared the crash experience for different types of highway and severity levels in several counties in France. They found that rural non-freeway facilities usually experienced higher crash frequency and severity than freeway facilities for the same level of exposure. Abbas (2004) developed a total of 200 statistical models for predicting the safety of rural highways in Egypt. Although two-lane and multilane rural highways were mixed together, Abbas reported that rural desert roads were more dangerous than any other type of rural highways.

In Australia, McLean (1996) evaluated the safety performance of cross-section design elements of multilane rural highways. He found that shoulder widths and roadside design influenced the safety of multilane highway facilities. Prinsloo and Goudanas (2003) produced descriptive models (in a table format) for determining the safety effects of cross-section design elements for four-lane rural highways. Roadway curvature, the presence of a median, and shoulder widths influenced the safety performance of rural highways. Turner et al. (2003) developed several statistical models predicting the safety performance of multilane rural highways by crash type in New Zealand. The proposed models were developed as part of a safety management system that will eventually be implemented by various transportation authorities in New Zealand.

In a recent study conducted in Italy, Caliendo et al. (2007) developed a series of statistical models to estimate the effects of geometric design features on the safety performance of multilane rural highways; tangents and curves were modeled separately. Three model types were estimated: the Poisson model, the NB model, and the NB multinomial, which is a model where the dispersion parameter is modeled as a function of the segment length. A similar approach was used in this project, as detailed in Chapter V. The models showed that the number of intersections on the segments was associated with an increase in the number of collisions. For curves, it was found that a smaller radius is positively associated with crashes.

Converting Rural Two-Lane to Multilane Highways

Three studies have been identified on the evaluation of converting two-lane to four-lane rural highways in the literature. In the first study, Rogness et al. (1982) evaluated the safety effects of converting rural two-lane with full-paved shoulders to four-lane undivided highways. Data were collected at 60 sites in Texas for highways converted between 1969 and 1976. During this period, about 394 miles of rural two-lane highways were upgraded. The analysis was performed using a simple before-after study; hence, the results were not adjusted for the regression-to-the-mean (RTM) and site selection biases. The authors reported that highways experienced less crashes per unit of exposure after they were converted. However, the magnitude of the difference was not consistent across crash types.

As described above, Council and Stewart (1999) developed statistical models for estimating the safety of rural two-lane and four-lane highways. The goal of the study consisted of comparing both models and determining the change in safety if a two-lane highway were to be converted to a four-lane divided or undivided highway. The model comparison has shown that converting two-lane rural highways would reduce the number of crashes for the same level of exposure.

The third study was performed by the Kentucky Transportation Center (Agent and Pigman, 2001). This study evaluated the safety effects of converting rural two-lane to four-lane highways, by adding an additional lane on each side of the traveled-way. The authors examined 25 highway segments. Using a simple before-after study, they found that four-lane highways experienced up to 45% fewer crashes after the conversion. Similar to the study by Rogness et al. (1982), the site selection and RTM biases were not corrected for in this analysis.

In summary, the literature on the safety performance of multilane highways indicates that multilane highways perform on average better than two-lane highways for the same level of exposure. The outcome was similar for both cross-sectional and before-after studies. The next section summarizes the safety performance of rural signalized and unsignalized intersections on multilane highways.

Safety Performance of Intersections on Multilane Rural Highways

This section summarizes work performed on the safety performance of intersections located on multilane rural highways. Because of their relevance, models developed in two recent FHWA sponsored research projects are covered below in relatively intense detail. This is preceded by a brief description of the only other known set of models of possible relevance. These models were developed for the *SafetyAnalyst* project.

For the *SafetyAnalyst* project, a suite of models detailed in an unpublished 2004 technical memorandum have been developed for use in the *SafetyAnalyst* interim tools for network screening, and development and evaluation of treatments. The models of relevance were calibrated from intersections on both multilane roadways and two-lane roadways. At present, the only independent variables are major and minor road entering Annual Average Daily Traffic (AADT). Calibration of models with covariates (i.e., sometimes they are referred to as full models), possibly with separate ones for intersections on multilane roads, and incorporating additional variables available in HSIS data, will be done within the next year with a view to including these in the final tools.

Currently, total crash (AADT only) models are provided for the following rural intersection types using Minnesota HSIS data for the years 1995-1999:

- Rural 3-legged intersections with minor-road stop control
- Rural 3-legged intersections with signal control
- Rural 4-legged intersections with minor-road stop control
- Rural 4-legged intersections with all-way stop control
- Rural 4-legged intersections with signal control

The functional form for intersection statistical models for the interim tools is:

$$\mu = \beta_0 F_1^{\beta_1} F_2^{\beta_2} \tag{2.1}$$

Where,

μ = mean number of crashes per intersection per year;

F_1 = entering AADT on the major road (veh/day);

F_2 = entering AADT on the minor road (veh/day); and

$\beta_0, \beta_1, \beta_2$ = coefficients estimated from data.

That these current models only contain AADT variables, and that they are calibrated from combined two-lane and multilane data, would make them only marginally relevant to NCHRP 17-29.

In the FHWA projects, there were separate models for rural 3-legged (Type III) and 4-legged (Type IV) stop-controlled intersections, both with four lanes on the major and two lanes on the minor. In the first project, Vogt (1999) used three years of data from sites located in California and Michigan. Models for all crashes occurring within 250 ft of the intersection center are shown in Table 2.3 below, which is preceded by Table 2.2 that describes the variables used. The model form was as follows:

$$\mu = \beta_0 F_1^{\beta_1} F_2^{\beta_2} \, e^{(\sum_{j=3}^{n} \beta_j x_j)} \tag{2.2}$$

Where,

μ = mean number of crashes per year;

F_1 = entering AADT on the major road (veh/day);

F_2 = entering AADT on the minor road (veh/day);

$x_3, x_4, \ldots, x_n$ = the values of the non-traffic highway variables (e.g., sight distance, skew angle, etc.); and

$\beta_0, \beta_1, \beta_2, \ldots, \beta_n$ = coefficients estimated from data.

The second FHWA project was reported by Lyon et al. (2003) and Oh et al. (2003) based on the project report by Washington et al. (2005). These documents present the results of an effort to validate and to subsequently recalibrate the accident prediction models for rural intersections detailed in two previous FHWA reports. The first report is the aforementioned one by Vogt (1999), which developed the models for 3- and 4-legged stop-controlled intersections of four-lane by two-lane rural roads and also models for signalized intersections of two-lane roads. The other report (Vogt and Bared, 1998) documents the development of models for 3- and 4-legged stop-controlled intersections of two-lane rural roads. These models are not of direct relevance to NCHRP 17-29, but they are worth noting in that they do form the basis for the models currently in the prototype HSM chapter (now Chapter 8).

In the recalibration effort, the data used to calibrate the original models were combined with the additional years' data for the same sites as well as the additional sites from other States. Table 2.2 provides the definition of variables used in the recalibrated model for total accidents, for which the coefficients are shown in Table 2.3. These recalibrated models generally had better goodness-of-fit (GOF) measures than comparable models from the earlier FHWA research as measured by the value of the dispersion parameter. They also had more variables that were significant at the 10% level. The models' coefficients associated with the major and minor road AADTs were also of a similar magnitude to the earlier models. For all models, the major road AADT exerts a larger influence on accident predictions than the minor road AADT.

Table 2.2 Variable Definitions for Models in Table 2.3 (Washington et al., 2005)

Variable Name	Variable Description
F_1	Average daily traffic on major road (vehicles per day)
F_2	Average daily traffic on minor road (vehicles per day)
DRWY1	Number of driveways within 250 ft. of intersection on major road
COMDRWY1	Number of commercial driveways on major road within 250 ft of the intersection center
HAU	Intersection angle variable defined where the angle between the major and minor roads is measured from the far side of the minor road: *3-Legged Intersections* Angle minus 90 if minor road is to the right of the major road in the increasing direction 90 minus angle if minor road is to the left of the major road in the increasing direction *4-Legged Intersections* (right angle - left angle)/2
MEDTYPE1	Median type on major road (0 = no median, 1 = painted, 2 = curbed, 3 = others)
MEDWDTH1	Median width on major road (ft)
LTLN1S	Presence of left-turn lane on major road (0=no, 1=yes)
PKLEFT	Peak left-turn percentage (%)
PKTHRU2	Peak through percentage on minor road (%)
PKTRUCK	Peak truck percentage passing through the intersection (%)
SDR2	Right-side sight distance on minor road (ft)
VEI1	Sum of absolute change of grade in percent per hundred ft for each curve on major road any portion of which is within 800 ft of the intersection center, divided by the number of such curves

Table 2.3 Models from Two FHWA Projects for Total Accidents at Stop-Controlled Intersections on Rural Multilane Roads (Washington et al., 2005)
Note: Model form is given by Eq. 2.2

Variables	Model coefficients and p-values for 3-legged intersections		Model coefficients and p-values for 4-legged intersections	
	Original (Vogt & Bared, 1998)	Recalibrated by (Washington et al. 2005)	Original (Vogt & Bared, 1998)	Recalibrated (Washington et al. 2005)
Intercept [$\ln(\beta_0)$]	-12.2196 (0.0001)*	-10.1914 (0.0000)	-9.4631 (0.0003)	-7.4713 (0.0001)
F_1 (β_1)	1.1479 (0.0001)	0.8877 (0.0000)	0.8503 (0.0022)	0.7350 (0.0001)
F_2 (β_2)	0.2624 (0.0024)	0.3228 (0.0000)	0.3294 (0.0087)	0.2390 (0.0099)
LTNIS			-0.4841 (0.0362)	
DRWY1	0.0391 (0.1023)			
COMDRWY1		0.0681 (0.0154)		
VEI1		0.1081 (0.0519)		
HAU		0.0101 (0.0861)		
MEDWDTH1	-0.0546 (0.0285)	-0.0106 (0.0760)		
MEDTYPE1		-0.3209 (0.0700)		
SDR2				-0.0003 (0.0403)
PKTRUCK				-0.0479 (0.0000)
PKTHRU				0.0249 (0.0034)
PKLEFT			0.1100 (0.0076)	0.0229 (0.0525)
Dispersion parameter	0.389	0.423	0.458	0.400

*Standard Error

The authors of the two papers concluded that except for AADT counts, there was little consistency in the variables found to be significant predictors of intersection crashes across the various models. Given that fact, it was concluded that the results of the recalibration of models with covariates provided strong support for the proposed Interactive Highway Safety Design Model (IHSDM) modeling approach (and that are currently in the HSM Chapter 8, now HSM

Chapter 10) of applying independently derived AMFs to baseline models rather than models with covariates to estimate the safety of an intersection. To that end, that research also developed baseline models and used these in a special procedure to derive initial AMF estimates for some variables. Those results, which are documented in the full project report, are of relevance in that they are recommended for consideration for the accident prediction algorithm for stop-controlled intersections on rural multilane roads and are presented in detail later in this report (Appendix A).

Chapter Summary

This chapter has summarized the review of the literature of documents related to the safety performance of rural multilane highway segments as well as rural intersections located on rural multilane facilities. The chapter showed that the available literature on safety performance of multilane rural highways is not extensive, and with a few exceptions, not solid enough to form the basis of a Highway Safety Manual accident prediction methodology. The next chapter summarizes the survey results conducted in this study.

CHAPTER III

Survey Design and Results

This chapter describes the characteristics of the survey design and results. The chapter is divided into three sections. The first section describes the survey instrument used in this study. The second section summarizes the candidate input variables that could be used for estimating the safety performance of multilane rural highways. The third section describes the survey results.

Survey Instrument

This section describes key elements of the survey conducted in this study. The first part provides details on the selection process for identifying state DOTs that were surveyed. The second part summarizes the characteristics of the survey instrument.

Selection Process

At the request of the NCHRP project panel, the research team conducted a limited survey of state DOTs rather than performing the full-fledged survey originally planned in the work plan. In order to avoid duplicating the survey effort of other projects (e.g., contacting the same survey participants within a year interval), the panel has asked the research team to determine whether the information contained in these surveys could be used for NCHRP 17-29. A brief assessment of the survey results carried out in this project is presented at the end of this chapter.

The research team requested the original survey responses from the original NCHRP 17-26 project. The intent of the request was two-fold. First, the request was used to determine potential state agencies that have shown interest in the HSM effort and the development and application of crash predictive models. Second, the team wanted to obtain information about accessing potential databases that could be used for this study. The limited survey allowed the team to focus the survey effort to a selected number of agencies.

After analyzing study results of NCHRP 17-26, the research team has identified 10 state agencies that have shown an interest in the development and the application of the HSM. Table 3.1 shows the state agencies selected to be part of this survey. This table also shows whether or not the agencies could provide crash data, and information on segments and intersections.

Table 3.1 State Transportation Agencies Selected for Limited Survey

State Agencies	Predictive Methodology	Crash Database	Segment Database	Intersection Database
Illinois	No	Yes	Yes	Yes
Indiana	Yes	Yes	Yes	No
Maryland	No	Yes	Yes	No
Missouri	No	Yes	No	No
Montana	No	Yes	Yes	Yes
North Carolina	No	Yes	Yes	No
Rhode Island	No	Yes	No	No
Texas	No	Yes	Yes	No*
West Virginia	No	Yes	Yes	No
Wyoming	No	Yes	No	No
* Partial data are available.				

Survey Characteristics

The survey was sent by e-mail to the contact person identified by AASHTO and TRB for each state DOT identified above. The contact person was usually the head of the research division of the state agency. If the contact representative did not send back the survey within three weeks after the survey was e-mailed, the research team contacted the contact person by phone to ensure he or she received the survey and to answer any potential questions.

The survey was divided into three parts. The first part sought to determine if the selected DOTs are currently using or developing statistical models to predict the safety performance of multilane rural highways. The second part focused on input variables. The list shown in Tables 3.2 and 3.3 was submitted to the survey participants. The survey participants were asked to rate these input variables on a scale from one to five, with one representing the lowest priority and five representing the highest priority. The survey participants were able to add input variables not listed. The last part covered questions on data availability and sought information on access to, and availability of crash data; geometric design data for segments and intersections; traffic flow data for segments, and major and minor approaches of intersections; and georeferenced data. The survey also sought to determine if the state agency can provide data for before-after studies for the aim of developing AMFs. A copy of the survey instrument is provided in Appendix B.

Identification of Candidate Model Inputs

This section describes a list of input variables that were identified by the research team as variables showing promise in terms of predicting the variation in crashes on multilane rural roads. The list was submitted to the survey participants.

Tables 3.2 and 3.3 summarize candidate input variables for the predictive methodology of multilane rural highway segments and intersections. The list contains a mix of geometric design, traffic operations, and human factors variables. The list of input variables was determined based on their potential effects on the safety performance of rural highways, as discussed in Chapter II, and the likelihood that the variables can be realistically collected in the field.

Table 3.2 Candidate Input Variables for Segments

Variables	Type
Delineation	Traffic Operations
Design or posted speed	Traffic Operations
Grades	Geometric Design
Horizontal curves	Geometric Design
Illumination	Traffic Operations
Land-use adjacent to traveled-way	Traffic Operations
Lane widths	Geometric Design
Median type	Geometric Design
Median width	Geometric Design
Number and type of median openings	Geometric Design
Number and type of driveways	Geometric Design
Number of through lanes	Geometric Design
Pavement friction	Traffic Operations
Raised pavement markers	Geometric Design
Roadside design/clear zones/roadside objects	Geometric Design
Roadside distractions (e.g., billboards, signage, etc.)	Geometric Design
Shoulder width/curb type	Geometric Design
Shoulder rumble strips	Geometric Design
Spacing between driveways	Geometric Design
Speed variance of vehicular traffic	Traffic Operations
Traffic volume (AADT) (veh/day)	Traffic Operations
Traffic volume in peak period (veh/hr)	Traffic Operations
Traffic volume for different time periods (average veh/hr)	Traffic Operations
Traffic volumes for individual driveways	Traffic Operations
Vehicle mix (e.g., percent trucks)	Traffic Operations
Vehicle speed (average for different time periods)	Traffic Operations
Vertical curvature	Geometric Design
Wet pavement	Traffic Operations
Ice on pavement	Traffic Operations
Snow on pavement	Traffic Operations
Visibility restrictions (e.g., fog, glare, etc.)	Human Factors

Table 3.3 Candidate Input Variables for Intersections

Variables	Type
Approach speed (observed)	Traffic Operations
Approach speed (posted or design speed)	Traffic Operations
Horizontal alignment of intersection approaches	Geometric Design
Illumination	Traffic Operations
Intersection sight distance	Geometric Design
Intersection skew angle	Geometric Design
Lane widths on intersection approaches	Geometric Design
Level of service (LOS) (only at signalized intersections)	Traffic Operations
Median type/presence of median	Geometric Design
Number of intersection legs	Geometric Design
Number of through lanes on intersection approaches	Geometric Design
Number and length of added through lanes at intersections	Geometric Design
Presence/number of left-turn lanes	Geometric Design
Presence of right-turn lanes	Geometric Design
Shoulder/curb type on intersection approaches	Geometric Design
Shoulder/curb width on intersection approaches	Geometric Design
Signal phasing (e.g., left-turn phasing)	Traffic Operations
Signal timing	Traffic Operations
Signal visibility	Human Factors
Spacing between intersection and nearby driveways	Geometric Design
Type of traffic control	Traffic Operations
Traffic volumes (AADTs) for major- and minor-road legs (AADTs)	Traffic Operations
Type of left-turn channelization (painted vs. raised curb)	Geometric Design
Vehicle mix (e.g., percent trucks)	Traffic Operations
Weather variables	Traffic Operations

In the process of identifying candidate input variables, other variables specifically related to human factors issues were also identified. For instance, age and experience of older drivers have an influence on the risk of crashes. However, many of these variables cannot be collected easily. Nonetheless, they were included in the survey instrument.

As discussed in Chapter II, only a few predictive models currently exist for estimating the safety performance of multilane rural highways. The characteristics of these models are summarized in Tables 3.4A and 3.4B. These tables are used to show input variables that have been used from previous research done on multilane rural highways. Tables 3.4A and B do not provide any inferences, i.e. standard errors, on the coefficients since they were not available from the original source. The functional form for the models presented in these tables is the following:

$$\mu = \beta_0 L F^{\beta_1} \, e^{\left(\sum_{i=2}^{n} \beta_i x_i\right)} \tag{3.1}$$

Where,

μ = mean number of crashes per year;

L = the length of the segment;

F = flow in vehicles per day (AADT);

x_i = a series of covariates (access control, etc.); and

$\beta_0, \beta_1, \beta_2, \ldots, \beta_n$ = coefficients estimated from data.

Table 3.4A Summary of Existing Statistical Models for Estimating the Safety Performance of Multilane Highway Segments

Variables	Study									
Author(s)	**Persaud (1991)**		**Wang et al. (1998)**	**Council and Stewart (1999)**					**Persaud and Bahar (2000)**	
Highway type	**Divided**	**Undivided**	**Divided/ Undivided**	**Divided (NC)[a]**	**Divided (WA)**	**Divided (MN)**	**Divided (CA)**	**Undivided (CA)**	**Divided**	**Undivided**
Number of lanes	4	4	4+	4	4	4	4	4	4+	4+
Severity	All	All	All	All	All	All	All	All	All	All
Intercept	0.460	0.653	0.0002	0.0092	0.0107	0.00071	0.000125	0.000164	0.00272	0.000175
Flow	1.304	1.304	1.073	0.761	0.635	1.064	1.071	1.121	0.730	1.031
Average roadside rating	--	--	0.131	--	--	--	--	--	--	--
Access control (1=partial control, 0=no control)	--	--	-0.151	--	--	--	--	--	--	--
Driveway density	--	--	0.034	--	--	--	--	--	--	--
Intersections with left turn lanes/mile	--	--	0.163	--	--	--	--	--	--	--
Intersections without turn lanes/mile	--	--	0.052	--	--	--	--	--	--	--
Functional class (1=rural municipal, 2=minor collector)	--	--	-0.572	--	--	--	--	--	--	--
Shoulder width (ft)	--	--	-0.094	-0.288	--	-0.234	--	--	--	--
Median width (ft)	--	--	-0.003	--	--	--	--	--	--	--
Area location type (1=rural municipal, 2=rural non-municipal)	--	--	0.429	--	--	--	--	--	--	--
ϕ (inverse of dispersion parameter)	2.90		Poisson	2.19	1.61	Poisson	1.46	1.81	5.56	5.70

[a] NC=North Carolina, WA=Washington, MN=Minnesota, CA=California

Table 3.4B Interim *SafetyAnalyst* Statistical Models for Estimating the Safety Performance of Multilane Highway Segments

Variables													
Highway type	**Divided (OH)** [a]	**Divided (NC)**	**Divided (MN)**	**Divided (OH)**	**Divided (NC)**	**Divided (MN)**	**Undivided (OH)**	**Undivided (NC)**	**Undivided (MN)**	**Undivided (WA)**	**Undivided (OH)**	**Undivided (NC)**	**Undivided (MN)**
Number of lanes	4	4	4	4	4	4	4	4	4	4	4	4	4
Severity	Injury	Injury	Injury	All	All	All	Injury	Injury	Injury	All	All	All	All
Intercept	2.20E-4	0.042	2.23E-4	0.017	1.68E-5	0.015	4.45E-5	0.0064	0.002	0.014	5.76E-4	1.14E-3	4.18E-4
Flow	1.07	.49	1.03	0.60	1.21	0.50	1.06	0.66	0.77	0.60	0.72	0.74	0.81
ϕ (inverse of dispersion parameter)	4.54	1.88	2.22	9.09	3.45	1.89	3.03	3.12	1.79	5.26	11.11	3.22	5.00

[a] OH=Ohio, WA=Washington, MN=Minnesota, NC=North Carolina

Table 3.4 shows that most available models use traffic flow as the only input variables. The model proposed by Wang et al. (1998) includes several input variables, including the number of intersections on the selected segment. This table also shows that the exponent for the traffic flow variable is, in several cases, larger than one. As suggested by Persaud (1991), multilane facilities may have an increasing rate of accident occurrence as traffic flow increases because more lane changes occur at high flows.

Survey Results

Six of the ten states surveyed returned the survey instrument. They included Illinois, Washington, Maryland, Missouri, Texas, and West Virginia.

The survey results show that the States of Texas and Missouri employed a methodology for predicting the safety performance of multilane highways. The State of Missouri uses the methodology for identifying highly hazardous locations on the rural network.

The input variables that were considered of highest priorities by the survey participants are summarized in Table 3.5. These variables shown in this table have been rated with an overall average of four or above.

Table 3.5 Highest Priority Candidate Input Variables

Segments	Intersections
Delineation	Type of Traffic Control
Roadside Design	Intersection Sight Distance
Horizontal Curves	Signal Visibility
Median Width	Presence/Number of Left-Turn Lanes
Shoulder Width	Horizontal Alignment (approaches)
Traffic Volumes	Approach Speed (observed)
	Approach Speed (design)
	Traffic Volumes
	Signal Phasing
	Signal Visibility
	Spacing between intersections and nearby driveways

The survey results show that all of the states surveyed can provide electronic crash databases and roadway segment inventory data. In addition, the databases can be linked electronically. On the other hand, access to intersection inventory data is currently not available for the State of Texas. Finally, access to georeferenced data is limited to the States of Washington and Texas.

Table 3.6 summarizes the length of multilane rural network, as reported by the survey participants. This table shows that the States of Illinois, Texas, and West Virginia have, as expected, a vast rural network.

Table 3.6 Multilane Rural Highway Network

State	Miles
Illinois	11,474
Maryland	770
Missouri	1,019
Texas	7,667
West Virginia	36,000

Chapter Summary

This chapter has summarized the characteristics of the survey carried out for this project. A limited survey was conducted to: a) determine whether the selected DOTs are currently using or developing statistical models to predict the safety performance of multilane rural highways; b) find out candidate input variables of interest by the survey participants; and c) determine the accessibility to various databases, such as crash data, geometric design information for segments and intersections, and traffic flows for segments, and major and minor approaches.

The survey results have shown that only two of the state agencies surveyed currently have a methodology for estimating the safety performance of multilane rural highways. Crash data and segment files could be made available by all study participants. However, other databases, such as intersection databases or access to georeferenced data, were not always available. The next chapter documents the data collection activities carried out in this study.

CHAPTER IV

Characteristics of Data Collected

This chapter describes the summary statistics for the data that have been used for developing the statistical models and AMFs as well as for the cross-validation study. The chapter is divided into five sections and each covers a state database that was assembled for this project: Texas, California, Minnesota, New York, and Washington.

Texas Data

The Texas data were used for developing statistical models and AMFs for rural multilane segments. For this dataset, five years of data were collected for this project. The data were obtained from the Department of Public Safety (DPS) and the Texas Department of Transportation (TxDOT). Three databases were merged together into one common database using SAS (SAS, 2002): Accident Database (DPS), Road Inventory Database (TxDOT), and the Texas Reference Marker (TRM) (TxDOT). These databases did not have a common linking variable and the research team had to create specific codes to link the databases together.

Table 4.1 shows the summary statistics of the rural multilane segments. The database only included segments that have not been modified during the study period and are greater than or equal to 0.1 mile in length. The network included 1,733 divided and 1,522 undivided segments for a total length equal to 1,749.53 and 848.29 miles, respectively. The minimum and maximum AADT values ranged from about 200 to 90,000 vehicles per day. Among the variables collected, they included lane and shoulder widths, median width, number of intersections located along the segment, and the number of horizontal curves on the segment. It should be pointed out that in this table and all subsequent tables, the values in parentheses represent the standard deviation (SD).

Table 4.1 Summary Statistics for the Rural Segments in Texas

Variables	Category	Min.	Max.	Mean (SD)	Total
Segment Length (Miles)	Undivided	0.1	6.275	0.55 (0.67)	830.5
	Divided	0.1	11.21	1.01 (1.21)	1746
AADT (Vehicles/day)	Undivided	402	24,800	6,613.61 (4010.01)	
	Divided	158	89,264	10,403.63 (7809.61)	
Lane Width (ft)	Undivided	9.75	16.5	12.57 (1.59)	
	Divided	9.5	16	12.06 (0.39)	
Shoulder Width (Right + Left) (ft)[a]	Undivided	0	40	9.96 (8.02)	
Right Shoulder Width (ft)[b]	Divided	0	24	13.65 (3.65)	
Median Width (ft)	Divided	1	240	47.71 (28.87)	
Number of Intersections	Undivided	0	47	2.33 (2.62)	3,493
	Divided	0	11	1.22 (1.64)	2,093
Number of Horizontal Curves	Undivided	0	16	0.70 (1.32)	1,052
	Divided	0	13	0.78 (1.20)	1,342

[a] Sum of right and left shoulders in ft; [b] sum of right shoulder widths in ft

Table 4.2 summarizes important crash data characteristics. During the study period, a total of 15,753 crashes occurred on the rural network, with about 28% occurring on undivided highways. In addition, single-vehicle crashes accounted for about two-thirds of all crashes.

Table 4.2 Summary Crash Statistics for Rural Segments in Texas (5 years)

Crash Type	Severity	Undivided				Divided				Together			
		Min.	Max.	Mean (SD)	Total	Min.	Max.	Mean (SD)	Total	Min.	Max.	Mean (SD)	Total
Single-vehicle	Injury	0	19	0.63 (1.60)*	942	0	20	1.59 (2.48)	2,729	0	20	1.14 (2.17)	3,671
	Total	0	56	1.71 (3.80)	2,559	0	54	4.57 (6.71)	7,872	0	56	3.24 (5.73)	10,431
Multi-vehicle	Injury	0	22	0.40 (1.11)	605	0	22	0.72 (1.57)	1,246	0	22	0.57 (1.38)	1,851
	Total	0	41	1.13 (2.52)	1,694	0	72	2.11 (4.55)	3,628	0	72	1.65 (3.77)	5,322
All	Injury	0	39	1.03 (2.45)	1,547	0	35	2.31 (3.52)	3,975	0	39	1.71 (3.13)	5,522
	Total	0	97	2.84 (5.69)	4,253	0	108	6.68 (9.93)	11,500	0	108	4.89 (8.45)	15,753

California Data

The California data were used for developing statistical models and AMFs for rural segments and intersections. Depending upon the observation, between three and ten years of data were collected. The data are obtained from the FHWA's HSIS maintained by the University of North Carolina. Similar to the Texas database, the data were assembled into one common database using SAS (SAS, 2002).

Table 4.3 summarizes the key variable statistics for rural multilane segments. The segment network included 1,087 divided and 356 undivided segments for a total length equal to 518.9 and 150.5 miles, respectively. The minimum and maximum AADT values ranged from about 1,300 to 61,000 vehicles per day. The variables collected include lane and shoulder widths, median width, and the number of intersections located along the segment.

Table 4.3 Summary Statistics for Rural Segments in California

Variables	Category	Min.	Max.	Mean (SD)	Total
Segment Length (Miles)	Undivided	0.1	3.858	0.42 (0.43)	150.5
	Divided	0.1	3.895	0.48 (0.51)	518.9
AADT (Vehicles/day)	Undivided	1,372	609,512	9,312.41 (8378.25)	
	Divided	3,044.4	39,744	12,280.76 (8545.70)	
Lane Width (ft)	Undivided	6	15	11.99 (0.61)	
	Divided	11	15	12.02 (0.27)	
Shoulder Width (Right + Left) (ft)[a]	Undivided	0	31	10.94 (5.59)	
Right Shoulder Width (ft)[b]	Divided	0	19	8.64 (2.18)	
Median Width (ft)	Divided	5	99	45.06 (30.67)	
Number of Intersections	Undivided	0	8	0.34 (0.87)	120
	Divided	0	9	0.46 (0.89)	501

[a] Sum of right and left shoulders in ft; [b] sum of right shoulder widths in ft

Tables 4.4 to 4.7 summarize the key statistics for the variables in the data for multilane rural intersections. The data contained 37 4-legged and 13 3-legged signalized intersections as well as 267 4-legged two-way stop-controlled (TWSC) and 403 3-legged stop-controlled intersections, respectively.

Table 4.4 Summary Statistics for Rural 4-Legged Signalized Intersections in California

Variable	Level	Frequency	Min.	Max.	Mean
Traffic Control	Semi-actuated multi-phase	1			
	Fully-actuated 2 phase	1			
	Fully-actuated multi-phase	35			
Light Condition	Lighted	37			
	No lighting	0			
Mainline Signal Mastarm	Mastarm	36			
	No mastarm	1			
Mainline Left-Turn	None	1			
	Curbed median	13			
	Painted	23			
Mainline Right-Turn	Channelization provided	17			
	None	20			
Permitted Left-Turn	Permitted	37			
	Restricted	0			
	Prohibited	0			
Crossroad Signal Mastarm	Mastarm	31			
	No mastarm	6			
Crossroad Left-Turn	None	19			
	Curbed median	3			
	Painted	15			
Crossroad Right-Turn	Channelization provided	16			
	None	21			
Crossroad Number of Lanes	2	30			
	>2	7			
Mainline Median	Divided	32			
	Undivided	5			
Major Road AADT			5,923	43,500	18,478
Minor Road AADT			101	18,000	3,689

Table 4.5 Summary Statistics for Rural 3-Legged Signalized Intersections in California

Variable	Level	Frequency	Min.	Max.	Mean
Traffic Control	Semi-actuated multi-phase	3			
	Fully-actuated multi-phase	10			
Intersection Light Type	Lighted	13			
	No lighting	0			
Mainline Signal Mastarm	Mastarm	13			
	No mastarm	0			
Mainline Left-Turn	None	0			
	Curbed median	3			
	Painted	10			
Mainline Right-Turn	Channelization provided	8			
	None	5			
Permitted Left-Turn	Permitted	13			
	Restricted	0			
	Prohibited	0			
Crossroad Signal Mastarm	Mastarm	7			
	No mastarm	6			
Crossroad Left-Turn	None	4			
	Curbed median	1			
	Painted	8			
Crossroad Right-Turn	Channelization provided	7			
	None	6			
Crossroad Number of Lanes	2	11			
	>2	2			
Mainline Median	Divided	13			
	Undivided	0			
Major Road AADT			2,792	31,550	18,270
Minor Road AADT			101	14,000	3,810

Table 4.6 Summary Statistics for Rural 4-Legged TWSC Intersections in California

Variable	Level	Frequency	Min.	Max.	Mean
Intersection Light Type	Lighted	85			
	No lighting	182			
Mainline Left-Turn	None	80			
	Curbed median	56			
	Painted	130			
	Raised bar	1			
Mainline Right-Turn	Channelization provided	86			
	None	181			
Permitted Left-Turn	Permitted	263			
	Restricted	0			
	Prohibited	4			
Crossroad Left-Turn	None	264			
	Curbed median	0			
	Painted	3			
Crossroad Right-Turn	Channelization provided	36			
	None	231			
Crossroad Number of Lanes	2	267			
	>2	0			
Mainline Median	Divided	239			
	Undivided	28			
Major Road AADT			2,295	78,300	15,058
Minor Road AADT			10	7,400	429

Table 4.7 Summary Statistics for Rural 3-Legged SC Intersections in California

Variable	Level	Frequency	Min.	Max.	Mean
Intersection Light Type	Lighted	89			
	No lighting	314			
Mainline Left-Turn	None	160			
	Curbed median	56			
	Painted	186			
	Raised bar	1			
Mainline Right-Turn	Channelization provided	76			
	None	327			
Permitted Left-Turn	Permitted	387			
	Restricted	0			
	Prohibited	16			
Crossroad Left-Turn	None	401			
	Curbed median	1			
	Painted	1			
Crossroad Right-Turn	Channelization provided	27			
	None	376			
Crossroad Number of Lanes	2	400			
	>2	3			
Mainline Median	Divided	345			
	Undivided	58			
Major Road AADT			2,430	78,300	17,339
Minor Road AADT			10	23,111	447

Table 4.8 summarizes important data characteristics for crashes occurring on segments in California. During the study period, a total of 2,267 crashes occurred on rural multilane highway segments, with about 17% occurring on undivided highways. In addition, single-vehicle crashes accounted for about two-thirds of all crashes happening on the network included in the database.

Table 4.8 Summary Statistics for the Crash Data (per Year) on Segments in California (Segment Models)

Crash Type	Severity	Undivided				Divided				Together			
		Min.	Max.	Mean (SD)	Total	Min.	Max.	Mean (SD)	Total	Min.	Max.	Mean (SD)	Total
Single-vehicle	Injury	0	3.6	0.20 (0.36)	695	0	11	0.29 (0.62)	3,100	0	11	0.27 (0.57)	3,795
	Total	0	10.9	0.61 (1.09)	2,175	0	53.7	0.93 (2.35)	9,999	0	53.7	0.85 (2.11)	12,174
Multi-vehicle	Injury	0	3.1	0.10 (0.26)	364	0	3.8	0.14 (0.30)	1,487	0	3.8	0.13 (0.29)	1,851
	Total	0	14.2	0.48 (1.28)	1,718	0	65.5	0.80 (2.74)	8,615	0	65.5	0.72 (2.46)	10,333
All	Injury	0	4.5	0.30 (0.55)	1,059	0	14.8	0.43 (0.86)	4687	0	14.8	0.39 (0.80)	5,646
	Total	0	21.7	1.09 (2.21)	3,893	0	119.2	1.73 (4.95)	18,614	0	119	1.57 (4.44)	22,507

Table 4.9 summarizes important crash data characteristics for intersections for the California data. During the study period, 1,298 and 379 crashes occurred at 4-legged and 3-legged signalized intersections. For unsignalized intersections, 2,881 and 4,484 crashes happened at 4-legged and 3-legged locations, respectively.

Table 4.9 Summary Statistics for Crash Data (per Year) at Intersections in California (Intersection Models)

Crash Type	Severity	Crashes			
		Min.	**Max.**	**Mean (SD)**	**Total**
Rural 4-Legged Signalized Intersections	Injury	0.0	2	0.88 (0.62)	240
	Total	0.2	15.3	5.03 (3.53)	1,298
Rural 3-Legged Signalized Intersections	Injury	0.0	0.9	0.30 (0.28)	37
	Total	0.4	7.9	3.22 (2.21)	379
Rural 4-Legged TWSC Intersections	Injury	0.0	2.2	0.32 (0.44)	839
	Total	0.0	12.8	1.10 (1.47)	2,881
Rural 3-Legged SC Intersections	Injury	0.0	13.9	1.14 (0.41)	994
	Total	0.0	2.7	0.26 (1.83)	4,484

Tables 4.10 and 4.11 summarize the input crash data for the models by crash type. The data described in Table 4.10 were used for the models estimated using intersection data, while the data described in Table 4.11 were employed for the model development for segments.

Table 4.10 Summary Statistics for Crash Data (per Year) for Intersections in California (Crash Type Models)

Crash Type	Severity	3-Legged Stop-Controlled				4-Legged Stop-Controlled				Together			
		Min.	Max.	Mean (SD)	Total	Min.	Max.	Mean (SD)	Total	Min.	Max.	Mean (SD)	Total
Same direction	Injury	0	0.9	0.06 (0.12)	272	0	0.8	0.05 (0.10)	418	0	0.9	0.05 (0.11)	690
	Total	0	7.8	0.44 (0.90)	1,641	0	5.0	0.30 (0.59)	796	0	7.8	0.39 (0.79)	2,437
Opposite direction	Injury	0	1.5	0.05 (0.14)	178	0	1.1	0.06 (0.14)	164	0	1.5	0.05 (0.14)	342
	Total	0	2.9	0.14 (0.34)	530	0	2.3	0.16 (0.33)	410	0	2.9	0.15 (0.33)	940
Intersecting direction	Injury	0	2.0	0.08 (0.19)	272	0	1.5	0.16 (0.28)	418	0	2.0	0.11 (0.24)	690
	Total	0	3.9	0.24 (0.47)	875	0	3.7	0.43 (0.66)	1,085	0	3.9	0.32 (0.56)	1,960
Single vehicle	Injury	0	0.9	0.05 (0.11)	201	0	0.6	0.05 (0.09)	128	0	0.9	0.05 (0.11)	329
	Total	0	4.7	0.24 (0.43)	869	0	2.3	0.21 (0.29)	547	0	4.7	0.23 (0.38)	1,416

Table 4.11 Summary Statistics for Crash Data (per Year) on Segments in California (Crash Type Models)

Crash Type	Severity	Undivided				Divided				Together			
		Min.	**Max.**	**Mean (SD)**	**Total**	**Min.**	**Max.**	**Mean (SD)**	**Total**	**Min.**	**Max.**	**Mean (SD)**	**Total**
Crashes at minor intersections	Injury	0	0.7	0.02 (0.06)	59	0	0.5	0.02 (0.06)	168	0	0.7	0.02 (0.06)	227
	Total	0	3.5	0.09 (0.24)	300	0	2	0.09 (0.20)	634	0	3.5	0.09 (0.21)	934
Single-vehicle and Opposite direction without turning	Injury	0	3.3	0.16 (0.32)	503	0	5.6	0.25 (0.45)	1,875	0	5.6	0.22 (0.42)	2,378
	Total	0	8.1	0.43 (0.75)	1,392	0	14	0.72 (1.24)	5,395	0	14	0.64 (1.13)	6,787
Same direction without turn	Injury	0	0.5	0.03 (0.06)	81	0	2.6	0.06 (0.17)	471	0	2.6	0.05 (0.15)	552
	Total	0	3.8	0.14 (0.31)	448	0	10.6	0.38 (0.92)	2,786	0	10.6	0.30 (0.79)	3,234

Minnesota Data

The Minnesota data were used for developing statistical models and AMFs for rural intersections. Up to ten years of data were collected for each site. The data were obtained from HSIS. Similar to the previous state databases, the data were assembled into one common database using SAS (SAS, 2002).

Tables 4.12 to 4.15 summarize the statistics of the data used for intersection models. The data contained 43 4-legged intersections, eight 3-legged signalized intersections, 224 unsignalized 4-legged intersections, and 171 3-legged unsignalized intersections. There were insufficient sites to calibrate a reliable model for 3-legged signalized intersections. However, data for this intersection type are shown in Table 4.13 for completeness.

Table 4.12 Summary Statistics for Rural 4-Legged Signalized Intersections in Minnesota

Variable	Level	Frequency	Min.	Max.	Mean
Intersection Light Type	Lighted	39			
	No lighting	4			
Crossroad Number of Lanes	2	35			
	4	8			
	Unknown	0			
Major Posted Speed			30	65	58
Minor Posted Speed			30	55	49
Major Road AADT			6,445	37,800	21,351
Minor Road AADT			337	18,489	5,137

Table 4.13 Summary Statistics for Rural 3-Legged Signalized Intersections in Minnesota

Variable	Level	Frequency	Min.	Max.	Mean
Intersection Light Type	Lighted	8			
	No lighting	0			
Crossroad Number of Lanes	2	7			
	4	1			
	Unknown	0			
Major Posted Speed			55	60	57.5
Minor Posted Speed			25	55	45.0
Major Road AADT			10,770	38,963	23,591
Minor Road AADT			695	26,128	9,181

Table 4.14 Summary Statistics for Rural 4-Legged TWSC Intersections in Minnesota

Variable	Level	Frequency	Min.	Max.	Mean
Intersection Light Type	Lighted	33			
	No lighting	191			
Crossroad Number of Lanes	2	218			
	4	3			
	Unknown	3			
Major Posted Speed			30	65	59
Minor Posted Speed			30	55	50
Major Road AADT			903	34,082	11,379
Minor Road AADT			14	5,209	743

Table 4.15 Summary Statistics for Rural 3-Legged SC Intersections in Minnesota

Variable	Level	Frequency	Min.	Max.	Mean
Intersection Light Type	Lighted	31			
	No lighting	140			
Crossroad Number of Lanes	2	159			
	4	5			
	Unknown	7			
Major Posted Speed			40	65	59
Minor Posted Speed			15	65	48
Major Road AADT			1,503	41,013	13,070
Minor Road AADT			11	5,287	795

Table 4.16 summarizes important crash data characteristics for intersections for the Minnesota data. During the study period, 2,024 and 475 crashes occurred at 4-legged and 3-legged intersections, respectively. For unsignalized intersections, 3,184 and 1,190 crashes happened at 4-legged and 3-legged locations, respectively.

Table 4.16 Summary Statistics for Crash Data (per Year) at Intersections in Minnesota

Crash Type	Severity	Crashes			
		Min.	**Max.**	**Mean (SD)**	**Total**
Rural 4-Legged Signalized Intersections	Injury	0.0	8.0	2.2 (2.01)	314
	Total	0.0	30.0	13.2 (8.14)	2,024
Rural 3-Legged Signalized Intersections	Injury	0.0	3.5	0.9 (1.24)	65
	Total	1.3	13.8	8.3 (4.12)	475
Rural 4-Legged TWSC Intersections	Injury	0.0	7.0	0.7 (1.05)	974
	Total	0.0	19.6	2.4 (3.32)	3,184
Rural 3-Legged SC Intersections	Injury	0.0	2.4	0.3 (0.53)	299
	Total	0.0	5.4	1.2 (1.34)	1,190

New York Data

The New York data were used for evaluating the re-calibration procedure for rural segments and intersections. Seven years of data were collected. The data were obtained from the NY State DOT. The data were assembled into one common database using SAS (SAS, 2002). Table 4.17 summarizes the statistics of the data used for segment models.

Table 4.17 Summary Statistics for Rural Segments in New York

Variables	Category	Min.	Max.	Mean (SD)	Total
Segment Length (Miles)	Undivided	0.1	3.84	0.43 (0.54)	85.38
	Divided	0.1	4.96	0.87 (0.93)	138.79
ADT (Vehicles/day)	Undivided	271.3	18,889.7	7,477.92 (4,738.57)	
	Divided	1,082	46,717.4	10,287.86 (7,034.76)	
Lane Width (ft)	Undivided	9	17.5	11.73 (1.47)	
	Divided	10	19	12.25 (1.40)	
Shoulder Width (ft)	Undivided	0	12	4.45 (3.87)	
	Divided	0	14	7.06 (4.00)	
Median Width (ft)	Divided	0	9	3.04 (2.12)	
Number of Intersections	Undivided	0	14	1.72 (1.83)	339
	Divided	0	11	1.86 (1.89)	296

The data contained 159 undivided and 197 divided highway segments, respectively. The data collected included lane and shoulder widths, median width, and the number of intersections located on segments.

Table 4.18 shows the important crash data characteristics for rural highways in New York. A total of 2,031 and 2,800 crashes occurred on undivided and divided segments, respectively.

Table 4.18 Summary Statistics for Crash Data on Rural Segments in New York

Crash Type	Severity	Undivided				Divided				Together			
		Min.	Max.	Mean (SD)	Total	Min.	Max.	Mean (SD)	Total	Min.	Max.	Mean (SD)	Total
Single-vehicle	Injury	0	28	1.72 (3.35)	339	0	32	3.01 (3.84)	479	0	32	2.30 (3.63)	818
	Total	0	87	4.40 (8.58)	867	0	72	10.49 (11.86)	1,668	0	87	7.12 (10.61)	2,535
Multi-vehicle	Injury	0	32	3.21 (5.28)	633	0	35	4.14 (6.18)	658	0	35	3.63 (5.71)	1,291
	Total	0	55	5.91 (9.71)	1,164	0	71	7.12 (10.25)	1,132	0	71	6.45 (9.96)	2,296
All	Injury	0	60	4.93 (7.62)	972	0	66	7.15 (8.30)	1,137	0	66	5.92 (8.00)	2,109
	Total	0	142	10.31 (15.77)	2,031	0	116	17.61 (18.05)	2,800	0	142	13.57 (17.19)	4,831

Tables 4.19 and 4.20 summarize the statistics of the data used for intersection models. Only sufficient data for unsignalized intersections was available. The data contained 71 4-legged signalized intersections and 282 3-legged signalized intersections.

Table 4.19 Summary Statistics for Rural 4-Legged Signalized Intersections in New York

Variable	Level	Frequency	Min.	Max.	Mean
Presence of left-turn lane	Yes	3			
	No	68			
Terrain	Flat	47			
	Rolling	21			
	Hilly	3			
Major Road AADT			1,552	23,116	8,597
Minor Road AADT			155	3,778	911

Table 4.20 Summary Statistics for Rural 3-Legged Signalized Intersections in New York

Variable	Level	Frequency	Min.	Max.	Mean
Presence of left-turn lane	Yes	9			
	No	272			
Terrain	Flat	176			
	Rolling	95			
	Hilly	10			
Major Road AADT			271	20,320	8,566
Minor Road AADT			14	1,889	477

Table 4.21 summarizes important crash data characteristics for intersections for the Minnesota data. During the study period, 472 and 673 crashes occurred at 4-legged and 3-legged intersections, respectively.

Table 4.21 Summary Statistics for Crash Data (per Year) at Intersections in New York

Crash Type	Severity	Crashes			
		Min.	**Max.**	**Mean (SD)**	**Total**
Rural 4-Legged Signalized Intersections	Injury	0.00	1.33	0.18 (0.24)	117
	Total	0.00	3.44	0.74 (0.82)	472
Rural 3-Legged Signalized Intersections	Injury	0.00	0.56	0.05 (0.10)	127
	Total	0.00	3.11	0.27 (0.41)	673

Washington Data

The Washington data were used for developing statistical models for rural segments. Four years of data were collected for this dataset. The data were obtained from the HSIS. The data were assembled into one common database using SAS (SAS, 2002).

Table 4.22 summarizes the statistics of the data used for segment models. The data contained 476 and 35 divided and undivided segments, respectively. The ADT varied from about 4,000 vehicles per day to 61,947 vehicles per day. The variables collected included lane and shoulder widths, median width, and the number of horizontal curves.

Table 4.22 Summary Statistics for Rural Segments in Washington

Variables	Category	Min.	Max.	Mean (SD)	Total
Segment Length (Miles)	Undivided	0.1	0.49	0.19 (0.11)	6.67
	Divided	0.1	2.76	0.41 (0.42)	195.55
ADT (Vehicles/day)	Undivided	4,014	33,118	17,539.23 (8161.07)	
	Divided	3,187	61,947	15,625.86 (10,271.42)	
Lane Width (ft)	Undivided	11	17	12.97 (1.32)	
	Divided	11	17	12.13 (0.68)	
Shoulder Width (Right + Left) (ft)[a]	Undivided	0	27	6.80 (8.51)	
Right Shoulder Width (ft)[b]	Divided	0	25	9.53 (1.89)	
Median Width (ft)	Divided	4	620	67.06 (73.46)	
Number of Intersections	Undivided	0	2	0.49 (0.56)	17
	Divided	0	2	0.22 (0.42)	107
Number of Horizontal Curves	Undivided	0	2	0.23 (0.55)	8
	Divided	0	13	0.83 (1.04)	393

[a] Sum of right and left shoulders in ft; [b] sum of right shoulder widths in ft

Table 4.23 summarizes important crash data characteristics about segments for the Washington data. During the four-year study period, a total of 2,416 crashes occurred on rural segments, with only 5% occurring on undivided highways. In addition, single-vehicle crashes accounted for about two-thirds of all crashes happening on the network.

Table 4.23 Summary Statistics for Crash Data on Segments in Washington (Segment Models)

Crash Type	Severity	Undivided				Divided				Together			
		Min.	Max.	Mean (SD)	Total	Min.	Max.	Mean (SD)	Total	Min.	Max.	Mean (SD)	Total
Single-vehicle	Injury	0	3	0.20 (0.63)	7	0	9	0.87 (1.48)	415	0	9	0.83 (1.45)	422
	Total	0	10	1.09 (2.03)	38	0	46	3.61 (4.95)	1,716	0	46	3.43 (4.85)	1,754
Multi-vehicle	Injury	0	3	0.29 (0.62)	10	0	5	0.28 (0.65)	131	0	5	0.28 (0.65)	141
	Total	0	35	2.74 (5.84)	96	0	19	1.19 (2.21)	566	0	35	1.30 (2.65)	662
All	Injury	0	3	0.49 (0.82)	17	0	12	1.15 (1.74)	546	0	12	1.10 (1.70)	563
	Total	0	36	3.83 (6.35)	134	0	62	4.79 (6.06)	2,282	0	62	4.73 (6.08)	2,416

Chapter Summary

This chapter described the summary statistics for the data collected for developing statistical models and AMFs for intersections and segments as well as for evaluating the re-calibration procedure. The models estimated in this research used five state databases: Texas, California, Minnesota, New York, and Washington. The data included detailed information about geometric design characteristics, traffic flow, and motor vehicle crashes. The next chapter describes the methodology used for developing the models and AMFs.

CHAPTER V

Modeling and Accident Prediction Methodologies

This chapter describes the characteristics of the model development and modeling methodologies. The chapter is divided into five sections. The first section describes the accident prediction methodology. The second section provides details about the model classes and functions. The third section covers the modeling framework for developing statistical models. The fourth explains the characteristics of the models developed in this work. The last section describes the methods used for estimating AMFs.

Accident Prediction Methodology

The models presented in this document estimate the safety performance of existing and proposed multilane rural highways operating under current and projected traffic demand. The models apply to four-lane undivided and divided rural highways. The safety performance measure is the expected annual crash frequency by severity, which can be calculated for a particular segment, intersection, or an entire project when both are combined.

As described in Chapter I, the term **rural** in this project is defined using the guidelines proposed by AASHTO (2004). These guidelines classify rural areas as places outside the boundaries of urban places where the population is less than 5,000 inhabitants. Consequently, any highway located outside the city limits of an urban agglomeration above 5,000 inhabitants is considered rural. The boundary delimitating rural and urban areas can at times be difficult to determine, especially since most multilane rural highways are located on the outskirts of urban agglomerations. In any case, these procedures may be used for any multilane (defined in the next paragraph) road in which the general design features and land use setting are rural rather than urban or suburban in nature. In other words, if the road is designed according to the rural road design standards in the AASHTO (2004) "Green Book," and development along the road is relatively sparse, these procedures apply.

The term **multilane** refers to facilities with four through lanes. These facilities may be divided with a rigid (concrete) or flexible barrier (cable), paved or landscaped median, but should not have access and egress limited by grade-separated interchanges (i.e., not freeways). Facilities may have occasional grade-separated interchanges, but these should not be the primary form of access and egress, and these procedures do not apply to the sections passing through the interchanges.

Separate models were estimated for intersections and segments and for sections that comprise both segments and intersections. The model classes are described in more detail in the next section. For segments, models were estimated for different types of multilane highways classified according to median type (i.e., divided and undivided facilities). For intersections, there are different models for 3- and 4-legged intersections, and for stop- and signal-controlled intersections. Crashes that have already been defined as intersection or intersection-related in the accident report and that occurred within 250 ft (76 m) of the intersection center were assigned to the intersection. For cases where no such definitions are available, all crashes occurring within 250 ft from the middle of the intersection were assigned to that intersection. The models estimate expected crash frequencies by crash type and severity to aid in the identification of safety concerns at existing locations or for a planned scenario.

In this work, major intersections are intersections between the highway segment being analyzed and other primary roads, such as major and minor arterials, or major collectors, and where traffic volumes (ADT) are available on all approaches. The latter requirement is extremely important, as the application of the intersection procedures requires ADT on all intersection approaches.

All other intersections are then referred to as minor intersections. These are generally intersections between the facility being analyzed and minor collectors, local roads, access driveways, or any intersection for which traffic volumes (ADT) are not available on approaches intersecting the facility being analyzed (it is assumed that ADT is available for the facility being analyzed).

Finally, segments are defined as portions of the facility delimited by major intersections or significant changes in the roadway cross-section, geometric characteristics of the facility, or the surrounding land uses. Roadway segments can be either undivided or divided.

Crashes are assigned to either a major or minor intersection, or a segment. Crashes occurring within or near the intersection are assigned to the intersection, and all other crashes are assigned to the respective segment. It is recommended that the following two criteria be used to define intersection and intersection-related crashes. In the first criterion, crashes that have already been defined as intersection or intersection-related in the accident report and that occurred within 250 ft (76 m) of the intersection center are assigned to the intersection. For cases where crashes are not identified as intersection or intersection-related, all crashes occurring within 250 ft from the middle of the intersection are assigned to that intersection. Figure 5.1 illustrates the key transportation elements of the predictive methodology.

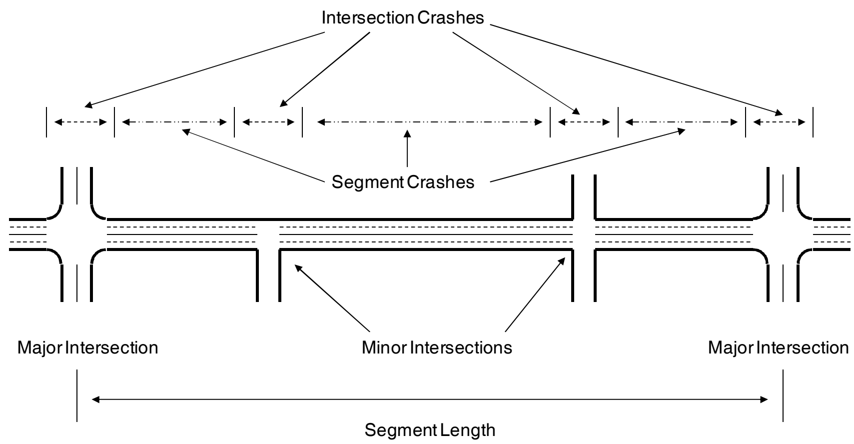


Figure 5.1 Definition of Segments and Intersections

Table 5.1 provides an overview of the key transportation elements of multilane rural highways for which specific procedures are used for estimating the models.

Table 5.1 Transportation Elements Included for the Model Development

Segment Type	Intersection Type
Undivided segments	Signalized 4-legged Signalized 3-legged Unsignalized 4-legged Unsignalized 3-legged
Divided segments	Signalized 4-legged Signalized 3-legged Unsignalized 4-legged Unsignalized 3-legged

Model Classes and Functions

This section describes the three classes of models that were estimated in this work. The three model classes are baseline models, general Average Daily Traffic (ADT) models, and models with covariates. For this study, AADT values are used as estimates of the ADT to be consistent with other sections of the HSM.

Baseline Models

For the first model class, statistical models are developed using a given set of baseline conditions. The baseline conditions usually reflect the nominal conditions agencies most often used for designing segments and intersections. For instance, nominal baseline conditions for segments may include paved 12-ft lane and 8-ft shoulder widths. For intersections, nominal conditions may include 12-ft lanes and no turning lanes on all approaches. Detailed characteristics for the nominal conditions for each model are described in the next chapter. Baseline models are currently part of the methodology for estimating the safety performance of rural two-lane highways (HSM Chapter 8) and urban and suburban arterial (HSM Chapter 10).

With these types of models, changes in geometric design characteristics are estimated using AMFs. The output of the models is adjusted by multiplying it with the modification factors. Additional details about AMFs can be found later in the chapter.

The functional forms used for estimating the baseline models for intersections are the following:

$$\mu_{it} = \beta_0 F_{1it}^{\beta_1} F_{2it}^{\beta_2} \tag{5.1a}$$

$$\mu_{it} = \beta_0 F_{Tit}^{\beta_1} \tag{5.1b}$$

Where,

μ_{it} = mean number of crashes for intersection i and year t;

F_{1it} = entering traffic flows in vehicles per day (ADT) on the major approaches for intersection i and year t;

F_{2it} = entering traffic flows in vehicles per day (ADT) on the minor approaches for intersection i and year t;

F_{Tit} = entering traffic flows in vehicles per day (ADT) on the major and minor approaches ($F_T = F_1 + F_2$) for intersection i and year t; and

β_0, β_1, β_2 = coefficients estimated from data.

The functional form used for estimating the baseline models for segments is the following:

$$\mu_{it} = \beta_0 L_i F_{Sit}^{\beta_1} \tag{5.2}$$

Where,

μ_{it} = mean number of crashes per year for segment i and year t;

L_i = segment length in miles for site i;

F_{Sit} = traffic flow in vehicles per day (ADT) for segment i and year t; and

β_0, β_1 = coefficients estimated from data.

It is important to note that an important limitation associated with this model class is related to the sample size. When baseline conditions are precisely defined, the sample size can be significantly reduced. In return, a small sample size can affect the robustness and statistical power of the model. In some circumstances, the model may become biased, particularly when the sample mean value is very low (Lord, 2006).

General ADT Models

For the second model class, general ADT models are developed for the following transportation elements: 4-lane undivided segments, 3- and 4-legged signalized intersections, as well as 3- and 4-legged unsignalized intersections. These models therefore reflect the average conditions found in the data for each transportation element. The data were described in Chapter IV. These models can be used for cases where the user has limited information about the geometric design features for the particular project under study. They can still be useful and provide an average value for the safety performance of multilane rural segments and rural intersections.

For this model class, AMFs can be used to adjust for changes in geometric design features. However, the AMFs need to be re-calibrated or adjusted to reflect the average conditions found in the data.

Similar to baseline models, the functional forms used for estimating the general ADT models for intersections are the following:

$$\mu_{it} = \beta_0 F_{1it}^{\beta_1} F_{2it}^{\beta_2} \tag{5.3a}$$

$$\mu_{it} = \beta_0 F_{Tit}^{\beta_1} \tag{5.3b}$$

Where,

μ_{it} = mean number of crashes for intersection i and year t;

F_{1it} = entering traffic flows in vehicles per day (ADT) on the major approaches for intersection i and year t;

F_{2it} = entering traffic flows in vehicles per day (ADT) on the minor approaches for intersection i and year t;

F_{Tit} = entering traffic flows in vehicles per day (ADT) on the major and minor approaches ($F_T = F_1 + F_2$) for intersection i and year t; and

β_0, β_1, β_2 = coefficients estimated from data.

The functional form used for estimating the general ADT models for segments is the following:

$$\mu_{it} = \beta_0 L_i F_{Sit}^{\beta_1} \tag{5.4}$$

Where,

μ_{it} = mean number of crashes per year for segment i and year t;

L_i = segment length in miles for site i;

F_{Sit} = traffic flow in vehicles per day (ADT) for segment i and year t; and

β_0, β_1 = coefficients estimated from data.

Models with Covariates

The third model class is the model with covariates. For this model class, the relationship between crashes and geometric design features is captured via the covariates in the statistical model. The selection of the covariates to be included into the model can be governed by various statistical criteria, such as the statistical significance of each variable and Akaike Information Criterion (AIC). This model class does not require the use of AMFs.

The functional form used for estimating the models with covariates for intersections is the following:

$$\mu_{it} = \beta_0 F_{1it}^{\beta_1} F_{2it}^{\beta_2} e^{\sum_{k=3}^{n} \beta_k x_{itk}} \tag{5.5}$$

Where,

μ_{it} = mean number of crashes for site i and year t;

F_{1it} = entering traffic flows in vehicles per day (ADT) on the major approaches for site i and year t;

F_{2it} = entering traffic flows in vehicles per day (ADT) on the minor approaches for site i and year t;

x_{itk} = model covariates (e.g., right-turning lanes, lighting, etc.) for site i and year t; and

β_0, β_1, β_2, β_k = coefficients estimated from data.

The functional form used for estimating the models with covariates for segments is the following:

$$\mu_{it} = \beta_0 L_i F_{Lit}^{\beta_1} e^{\sum_{k=2}^{n} \beta_k x_{itk}} \tag{5.6}$$

Where,

μ_{it} = mean number of crashes per year for site i and year t;

L_i = segment length in miles for site i;

F_{Lit} = traffic flow in vehicles per day (ADT) for site i and year t;

x_{itk} = model covariates (e.g., lane width, shoulder width, etc.) for site i and year t; and

β_0, β_1, β_k = coefficients estimated from data.

In this work, the research team has explored the development of multivariate crash count models by severity (Song et al., 2006; Miaou and Song, 2005; Ma and Kockelman, 2006; Park and Lord, 2007). Since these models are currently in the exploratory stages, they will not be part of the proposed HSM predictive methodology for Chapter 9. The multivariate model was applied to only one dataset (i.e., 3-legged signalized intersections) and the modeling results are summarized in Appendix E.

This new type of model uses the same functional form as the one depicted in Equation 5.6. However, the models include multivariate variables or covariates (i.e., each variable is a vector) instead of univariate variables, as traditionally used in highway safety analysis (e.g., Eqs.

5.5 and 5.6). In multivariate models, the different levels of the variables are assumed to be dependent. Ignoring this dependence may lead to inclusion or exclusion of key covariates that have detrimental effects on the model (Park and Lord, 2007). An example of the functional form for this model type is the following:

$$\mu_{ijt} = \beta_{j0} F_{1ijt}^{\beta_{j1}} F_{2ijt}^{\beta_{j2}} e^{\sum_{k=3}^{n} \beta_{jk} x_{ijtk}} \tag{5.7}$$

Where,

μ_{ijt} = mean number of crashes for site i, severity j, and year t;

F_{1it} = entering traffic flows in vehicles per day (ADT) on the major approaches for site i, severity j, and year t;

F_{2it} = entering traffic flows in vehicles per day (ADT) on the minor approaches for site i, severity j, and year t;

x_{ijtk} = model covariates (e.g., right-turning lanes, lighting, etc.) site i and severity j, and and year t; and

β_{j0}, β_{j1}, β_{j2}, β_{jk} = coefficients estimated from data.

Modeling Framework

This section describes the characteristics associated with the development of statistical models performed in this work. The development of statistical models was divided into four major steps: (1) determine the modeling objective matrix; (2) establish the appropriate processes to develop the models; (3) determine the inferential goals; and (4) select the computation techniques and tools.

Determine Modeling Objective Matrix

The first step in developing predictive or statistical models is to lay out the objectives of the modeling effort. The main considerations in this step include application needs, project requirements, data availability, logical scales—both spatial and temporal scales—of modeling units and their definitions, and range, definition, and unit of key input and output variables.

Table 5.2 shows a modeling objective matrix that was used for this project. The first column provides a list of physical elements on the network that are typically used in modeling and possible types and levels of aggregations in the applications. The first row lists crash outcome parameters from the predictive models that are potentially of interest to the users.

Table 5.2 Modeling Objective Matrix

Network Elements	Crash Frequency	Crash Rate	Crash Injury Severity (KABCO)	Crash Type[a]	AMFs[b]	Crash Cost
• Intersections	Y[c]	Y	T	Y	Y	N
• Segments	Y	Y	N	Y	Y	N

[a] Single vehicle: Single vehicle run-off-road (SVROR) and single vehicle main lanes; Multiple vehicles: Head-on, rear-end, sideswipe, angled, etc.
[b] AMFs can be developed for different crash types and severities.
[c] Y = was performed for this project; T = tried for this project; N = was not be attempted for this project; in some cases, it was initially proposed, but was eventually not used in this study.
Note: This table represents a simplification of the modeling objective.

It is critically important in this step to determine logical scales of modeling units and their definitions, as well as range, unit, and definition of key input and output variables. For instance, it is important to have a spatial and physical definition of intersections and segments and the exact types of traffic crashes (e.g., intersection, intersection-related, pedestrian-involved, or animal-involved crashes) to be assigned to each observational unit or observation. The range of traffic flows can be used as another example. There is a need to determine the range of flows and geometric characteristics of interest to this study (e.g., ADT = 200 to 20,000) and make sure commensurable data can be obtained. The time unit of analysis (i.e., number of crashes per unit of time) is another critical element when developing statistical models. Whether one uses crashes per month, per year, per 3-year, etc., will have serious effects on modeling assumptions and consequently on model interpretation and applicability. Not properly defining the modeling scales could lead to datasets with too many zeros or too much variation. This may cause problems with the development of statistical models, as documented by Lord et al. (2005b) and Lord (2006).

Establish Appropriate Process to Develop Models

This step is to make sure that the best possible statistical models are developed to achieve the modeling objectives identified in the last step. This includes ensuring that: a) data sources and limitations, sampling design, and statistical, functional, and logical assumptions are clearly spelled out; b) supporting theories are properly defended and/or cited; c) models are systematically developed and tested; and d) modeling results are properly interpreted.

Typical modeling procedures employed in developing statistical models can be grouped into five major processes: (1) establish a sampling model; (2) choose an observational model; (3) develop a process/state/system model; (4) develop a parameter model (for the Bayesian approach only); and (5) construct model choice and interrogation methods, including model comparison, sensitivity or robustness analysis, and specification test among others.

Process #1: Sampling Model

The first process in developing a predictive model is to establish a sampling model, through which the sampling design used in collecting sampling units is statistically characterized. Typical sampling methods include simple random sampling, systematic sampling, cluster sampling, stratified sampling, and other variants or combinations of these sampling methods. This sampling model allows the analyst to recognize: (1) the population or subpopulations that the model is intended to represent, (2) potential sampling or selection bias that may exist in the collected data which needs to be corrected in the model, and (3) sampling properties (e.g., correlation due to geographical clustering) that need to be considered in the subsequent modeling processes.

Process #2: Observational Model

This process determines the probability model for an individual site if repeated observations can be made under the same conditions, i.e., with conditions frozen in space and time. For modeling crash frequency, the Poisson model and its variants (i.e., Poisson-gamma and Poisson-lognormal, etc.) are the time-honored observational model for modeling rare and discrete events that often exhibit highly skewed distribution in their occurrence. The mixed-Poisson model is well accepted in the traffic safety community, as well as many other scientific and engineering disciplines. When modeling injury severity distribution, logistic, multinomial probit, ordered multinomial probit, multinomial logit, and ordered multinomial logit models are commonly employed, each of which has an underlying distributional assumption regarding the distribution of crashes by severity.

Another important issue to consider is to determine whether to choose a series of conditionally independent univariate models or a multivariate model to predict crashes by, e.g., crash types or crash injury severity level, as described below. Quite a few models have been employed in the literature and to the best of the research team's knowledge, no systematic evaluation of their underlying assumptions and model performance has been reported at this point. For example, in modeling crashes by severity level, some used separate Poisson-based models for crashes of different severity levels, e.g., one for total crashes or all crash severities, one for KAB (i.e., Fatal Injury, Injury Type A, and Injury Type B) or KABC (all injuries, including Type C) crashes, and one for property damage only (PDO or O) crashes. Others used the mixed-Poisson model for total crashes and then separately developed a (possibly ordered) logit-based model for characterizing severity distributions. Some, in recent years, have begun to model crashes by severity simultaneously with multivariate mixed-Poisson models (Song et al., 2006; Miaou and Song, 2005). The first two types of models require conditional independence assumption for crashes of different severities and use only partial information at a time. This is still an open research area that needs to be addressed, both in terms of the statistical and logical strengths and limitations of each approach. The research team has evaluated this model type and the results are presented in Appendix E.

Process #3: Process/State/System Model

In this process, functional and statistical relationships are established and possibly logical constraints are applied to characterize the effect of inter- and intra-sites-times heterogeneities on outcome variables based on logical, empirical, and other supporting theories. For example, in a mixed-Poisson model, a ‘log link-function” is usually adopted to link the mean of Poisson with covariates and other unobserved effects. A large body of literature has been published in the highway safety community in this area. Some of the decisions to be made during the process include which types of fixed and random effects to consider, including covariate effects, temporal effects, spatial effects, and exchangeable effects. Variance (as well as other higher order) structures change when different random effects are assumed. So far, most of the predictive models have been formulated parametrically; some semi-parametric or even non-parametric models are possible to relax some of the assumptions made about random effects and functional form for the Poisson mean.

Process #4: Parameter Model

This process is relevant only if the analyst decides to take a full-Bayes modeling approach where the distributions of "priors" (and "hyperpriors") need to be specified. So far, non-informative priors are typically employed in the highway safety literature. However, in some instances, adopting partially informative priors could be beneficial. For instance, Miranda-Moreno et al. (2008) have discussed how to incorporate information collected from previous studies into Bayesian models. Models that incorporate prior information usually performed better than models with vague or non-informative priors.

Process #5: Model Choice and Interrogation Methods

This process selects a series of statistical tests and logical arguments to decide which models may have the best predictive power. This includes developing statistical criteria to compare model performance, developing sensitivity or robustness analysis by allowing competitive yet plausible assumptions to be evaluated, including some specification tests.

Determine Inferential Goals

Without encroaching onto the decision theory area, where some sort of "loss function" (such as benefit-cost ratios or some sorts of utility functions) will need to be determined by the users of the predictive model for their individual applications, this process determines whether a point prediction coupled with a simple estimate of its standard error, an interval prediction (e.g., 2.5th and 97.5th percentile "credible" intervals in Bayesian statistics), or a full probability distribution for the prediction is needed. As will be discussed in the next section, more detailed inferential goals will require more sophisticated computational methods to fully capture the sampling variations in producing estimates and predictions.

Select Computation Techniques and Tools

This is the process where frequentists (analysts who use the likelihood-based method) (McCullagh and Nelder, 1989), empirical Bayesian (EB) (Hauer, 1997), and Bayesian (Carlin and Louis, 2000; Gelman et al., 2004) analysts are likely to differ in their estimating approaches and use of different "stochastic approximations" to reduce computational burden, such as asymptotic normality, "plug-in" estimator, linearization method, Laplace approximation, Gaussian quadrature approximation, etc. Many statistical tools are now available for modeling nonlinear mixed-effects models, using both Bayesian and frequentist methods, under the so-called exponential family of probability distributions, where the Poisson model is a special case. More difficult inferential goals will require more sophisticated computational methods to fully capture the sampling variations in producing estimates and predictions. By being able to take advantage of the unprecedented computing power available today, simulation-based methods, including various bootstrap, cross-validation, and Markov Chain Monte Carlo (MCMC) methods (Gilks et al., 1996), have been particularly popular in the statistical community in the last 15 years, regardless of whether the likelihood-based, EB, or Bayesian approach is taken.

Regardless of the statistical approach, effective prediction procedures attempt to reduce the mean squared error (MSE) of predictions, which strikes a balance between reducing predictive variance and reducing predictive bias. In addition, recent statistical developments in the area of mixed-effect models (i.e., models with both fixed and random effect components), including the mixed-effect Poisson, logit, and probit models, appear to be a unifying force among various statistical approaches, including the frequentist and Bayesian approaches.

Model Development

This section describes the approach used for developing the different classes of statistical models. The approach was based on the four-step modeling framework described in the previous section.

For all model classes, the modeling objective consists of estimating the crash frequency of motor vehicle crashes for each transportation element listed in Table 5.1. The units for these models were in number of crashes per year. For each class, models are estimated for two severity levels: (1) all crashes (defined as Total) or (2) KABCO and KAB. The crash severity levels were defined above.

The statistical models developed in this body of work are defined as observational models. Given the nature of the data, a mixed-Poisson modeling framework was used for developing these models. This modeling framework is the preferred method for modeling non-negative and independent discrete events, such as motor vehicle crashes. The selected mixed-Poisson model was the Poisson-gamma or negative binomial model. Although this model has some limitations (see Lord, 2006), the negative binomial offers a straightforward way to accommodate the over-dispersion commonly found in crash data (Lord et al., 2005b); over-dispersion refers to the amount of variation in the data that cannot be entirely captured by a Poisson model, where the variance must be equal to the mean. In addition, the mixed-distribution of the Poisson-gamma function has a closed form and the mathematics to manipulate the relationship between the mean and the variance structures is relatively simple (Hauer, 1997).

Poisson-gamma models in highway safety applications have been shown to have the following probabilistic structure: the number of crashes at the i-th entity (road section, intersections, etc.) and t-th time period, Y_{it}, when conditional on its mean μ_{it}, is assumed to be Poisson distributed and independent over all entities and time periods as:

$$Y_{it} \mid \mu_{it} \sim Po(\mu_{it}) \qquad \text{i = 1, 2, …, I and t = 1, 2, …, T} \tag{5.8}$$

The mean of the Poisson is structured as:

$$m_{it} = f(\mathbf{X};\boldsymbol{\beta})\exp(e_{it}) \tag{5.9}$$

Where,

$f(.)$ is a function of the covariates (X);

$\boldsymbol{\beta}$ = a vector of unknown coefficients; and

e_{it} = the model error independent of all the covariates.

It is usually assumed that $\exp(e_{it})$ is independent and gamma distributed with a mean equal to 1 and a variance $1/\phi$ for all i and t (with $\phi > 0$). With this characteristic, it can be shown that Y_{it}, conditional on $f(.)$ and ϕ, is distributed as a Poisson-gamma random variable with a mean $f(.)$ and a variance $f(.)(1+f(.)/\phi)$ respectively. The term ϕ is usually defined as the "inverse dispersion parameter" of the Poisson-gamma distribution (note: in the statistical and econometric literature, $\alpha = 1/\phi$ is usually defined as the dispersion parameter; in some published documents, the variable α has also been defined as the "over-dispersion parameter"). Usually, the dispersion parameter or its inverse is assumed to be fixed, but recent research in highway safety has shown that the variance structure can potentially be dependent on the covariates (Hauer, 2001; Heydecker and Wu, 2001; Miaou and Lord, 2003; Lord et al., 2005a). In this work, models with both fixed and varying dispersion parameters were estimated. For the varying dispersion parameter, the following functional form was used:

$$\alpha_{it} = {}^{1}\!/_{\phi_{it}} = \exp\left(Z_{it} \times \delta_t\right) \tag{5.10}$$

Where,

Z_{it} = a vector of secondary covariates (not necessarily the same as the covariates in estimating the mean function m_{it}); and

d_t = regression coefficients corresponding to covariates Z_{it}.

The inferential goals for this project were to provide point estimates and associated standard errors. In other words, the modeling output consists of providing the estimated mean value for each transportation element analyzed. Given these goals, the statistical software SAS (SAS, 2002) was used to estimate the coefficients for all the models, with the exception of the multivariate models. Given the complexity of these models (and discussed in the previous section), the coefficients were estimated using the Bayesian function in MathLAB (MathWorks, Inc., 2006). It should be pointed out that a Poisson-lognormal modeling framework was used for the multivariate model (see Appendix E). This modeling framework offers more flexibility than the negative binomial model, because the correlation matrix can be fully generalized as opposed to the Poisson or the negative binomial multivariate model.

AMF Estimation

This section describes the different approaches that were evaluated or used for selecting and estimating the AMFs for rural multilane highways and intersections located on this type of facility. The first part describes the expert panel process that was employed for this work. The second part covers the methods used for estimating AMFs based on regression models and data.

Joint NCHRP 17-25/17-29 Expert Panel

Since literature on AMFs for multilane rural roadways is limited, it was hoped that expert panels might be able to combine the limited past evaluations specific to this roadway class with modified versions of two-lane AMFs, already developed through expert panels, to develop the needed estimates. It was also hoped that this option might produce multiple AMFs at a lower cost than would be required by new analyses.

A critical requirement was that the panel be analysis-driven. The AMFs derived by the panel were to be based on critical reviews of the existing research literature and on a consensus decision that the results from the research literature were robust enough to allow development of an AMF with at least a medium-high level of predictive certainty. At times, the AMF was based not only on the findings from the original research study, but on additional limited analyses of the data from one or more studies. Given this orientation, the composition of panel membership needed to include both expert researchers knowledgeable about the AMFs of interest and about the strengths and weaknesses of study methods, and a group of expert state and local AMF users (i.e., safety engineers) with knowledge of both the specifics of the AMFs needed and knowledge of real-world conditions under which those evaluated treatments were probably implemented.

The decisions on panel membership were made jointly by the Principal Investigators of the two involved NCHRP projects. Members of the panels for each project are shown in Figure 5.2.

- Dr. Dominique Lord, Texas A&M University (Principal Investigator, NCHRP 17-29)
- David Harkey, UNC Highway Safety Research Center (Principal Investigator, NCHRP 17-25)
- Dr. James Bonneson, Texas Transportation Institute
- Dr. Forrest Council, VHB
- Ms. Kim Eccles, VHB
- Dr. Ezra Hauer , University of Toronto (Retired)
- Mr. Loren Hill, State Highway Safety Engineer, Minnesota DOT
- Mr. Brian Mayhew, North Carolina DOT
- Dr. Bhagwant Persaud, Ryerson University (representing NCHRP 17-29)
- Dr. Raghavan Srinivasan, UNC Highway Safety Research Center
- Dr. Simon Washington, Arizona State University
- Mr. Tom Welch, State Transportation Safety Engineer, Iowa DOT

Figure 5.2 Members of the Analysis-Driven Expert Panel

The panel met for a three-day period. Because the number of potential AMFs was greater than could be studied and discussed in this time frame, the AMFs were prioritized before the meeting. Here, the 17-25 project team and the Principal Investigator for NCHRP 17-29 developed two lists of candidate variables to consider for AMF development: one for roadway segment variables and one for intersection variables. These lists were based on the results of an AMF Knowledge Matrix developed by the NCHRP 17-25 team and on the specific AMF needs of the companion project. The two lists were sent to each member of the expert panel before the meeting. The panel members reviewed each list separately (intersections and segments) and ranked each variable with respect to level of importance as either primary (P) or secondary (S). Primary variables were those believed to be the most important predictors of safety on the road type in question and in definite need of discussion at the meeting. Secondary variables were those believed to be of less importance with respect to predicting safety and should only be considered for discussion at the meeting if time permitted. The primary variables were also ranked from most important (first) to least important (last). These rankings were based on each member's assessment of: (1) the perceived magnitude of the effect of the variable on safety and (2) the quality and extent of reliable information in the literature on which an AMF could be based. The panel inputs were then compiled by the NCHRP 17-25 team to develop a final list for discussion.

The project team then developed and distributed to each panel member a resource notebook. This notebook included the results of the prioritization task, contact information for all panel members, resource materials for each variable/treatment, and pre-meeting assignments for the panel members. The resource materials included the following for each AMF under consideration:

- The AMF summary material developed earlier in NCHRP 17-25. This material included the draft Research Results Digest for those AMFs considered to have high or medium-high levels of predictive certainty and summary pages from the interim report for those of lower quality.
- The AMF summary from NCHRP 17-27 (Parts I and II of the Highway Safety Manual). This draft summary had been developed within the past two years, and included a second recent assessment of AMFs along with a discussion of the studies from which they were taken or derived; a discussion of materials reviewed without the recommendation of an AMF; or a listing of possible resources that may be reviewed for future AMF development consideration.
- Copies of five cross-sectional studies that included a number of the high-priority elements. A description of the models developed and the variables included was

provided for each study. These models could possibly provide additional insight about the direction and magnitude of the effect of the variables found to be significant.

- A draft procedure for adjusting AMF estimates and standard errors. This procedure had been developed and applied within the NCHRP 17-27 project.

In order to make sure all treatments or variables were adequately addressed, the panel members were given pre-meeting assignments. The panel was divided into three groups, and each group was assigned a subset of the variables to review prior to the meeting and asked to help lead the discussion on those topics. The panel was asked to focus on three questions:

1) Do the materials presented include enough quantitative information to potentially develop an AMF for urban and suburban arterials (or rural multilane highways)? The materials provided to the panel included a wide spectrum of study types (e.g., rigorous before-after studies, simple before-after studies, cross-sectional studies, or less vigorous data assessments). In many cases, the materials and existing AMFs were related to rural two-lane roads. The panel member was asked to assess whether or not the material was sufficient for the specification of an AMF for a rural multilane highway.
2) If an AMF can be developed from the material provided, what is the magnitude of the effect and to what types and severities of crashes does it apply?
3) Are there other studies that are not included in the existing set of materials that should be discussed at the meeting?

The panel members prepared accordingly for the meeting, and the meeting was held at the UNC Highway Safety Research Center facilities. The NCHRP 17-25 project team recorded detailed notes of the ensuing discussions and continually displayed both the notes and possible findings. Final decisions were then made by panel consensus. Four AMFs were used from this process. A report on that review is included in Appendix C.

AMFs from Data and Models

AMFs can be estimated using various statistical methods. Three approaches were considered for estimating the AMFs. Each one is briefly described below.

The first method is based on the before-after study framework. This method consists of estimating the safety effects of changes in geometric design features, traffic operations, or other characteristics by examining the increase or reduction in crash counts between the before and after periods. Three techniques have been proposed for this kind of study: 1) the simple or naïve before-after study, 2) the before-after study with a control group, and 3) the before-after study using the EB method. These techniques, including their limitations, have been well documented by others and are not described here (Hauer, 1997; Persaud et al., 2001 and 2003; Ye and Lord, 2007). Given the scope and the type of data collected in this study, the before-after method was not used for this research.

The second method consists of estimating AMFs using the coefficients of statistical models. This method has been used by Lord and Bonneson (2007) and Washington et al. (2005) for estimating AMFs for rural frontage roads in Texas and rural intersections in various states, respectively. The AMFs are estimated the following way:

$$AMF_j = e^{\left(\beta_j \times \left[x_j - y_j\right]\right)} \tag{5.10}$$

Where,

x_j = range of values or a specific value investigated (e.g., lane width, shoulder width, etc.) for AMF j;

y_j = baseline conditions or average conditions for the variable x_j (when needed or available); and

β_j = regression coefficient associated for the variable j (estimated from data).

This method provides a simple way to estimate the effects of changes in geometric design features. However, although the variables are assumed to be independent, they may be correlated, which could affect the coefficients of the model. The Variance Inflation Factor (VIF) can be used for detecting correlated variables, but this procedure usually flags only extreme cases of correlation (Myers, 2000). This method was used for estimating AMFs in this work, and the results are presented in the next chapter. Two AMFs could be produced for this method.

For the third method, AMFs are estimated using baseline models and applying them to data that do not meet the nominal conditions. The characteristics of baseline models are described above. This method has been proposed by Washington et al. (2005), who have re-calibrated models for estimating the safety performance of rural signalized and unsignalized intersections. For this method, the baseline model is first applied to sites not meeting all of the baseline conditions; then, the predicted and observed values per year are compared, and a linear relationship between these two values is estimated via a regression model to determine whether or not AMFs could be produced from its coefficients. The linear equation is given by the following:

$$Y_i - \mu_i = \gamma_1 X_1 + \ldots + \gamma_m X_m \qquad (5.11)$$

Where,

μ_i = mean number of crashes for site i per year estimated by the baseline model;

Y_i = observed number of crashes for site i per year;

X_m = a vector of the baseline variables (each site not meeting one or more of these variables); and

γ_m = coefficients estimated from data.

The AMFs are estimated using the following relationship when the coefficients are found to be statistically significant (e.g., 5%- or 10%-level):

$$AMF_m = \frac{\sum_{i=1}^{n} Y_i / n}{\sum_{i=1}^{n} Y_i / n - \gamma_m} \tag{5.12}$$

Where,

AMF_m = AMF for coefficient m; and

n = the number of observations in the sample.

This method was applied and evaluated in this work, but no AMFs could be produced.

Chapter Summary

This chapter described the modeling methodologies performed in this work. The first section covered the accident prediction methodology. The proposed methodology separates the multilane rural networks for each state into segments and intersections. Specific models would be developed for each element. The second section provided information about the model classes and functions. Three types of models were proposed: baseline models, general ADT models, and models with covariates. The third section presented the modeling framework for developing statistical models. The modeling framework was divided into four steps: determine the modeling objective matrix; establish the appropriate processes to develop the models; determine the inferential goals; and select the computation techniques and tools. The fourth section described the characteristics of the models developed in this work. All the models were developed using the Poisson-gamma or NB regression framework. The Poisson-lognormal was used for developing models by crash severity. The last section described the methods used for estimating AMFs. The next chapter summarizes the modeling results.

CHAPTER VI

Modeling Results

This chapter describes the results of modeling undertaken for this project. The chapter is divided into three sections. The first section describes the modeling results for intersections, segments, and by crash type. The second section provides a discussion about the recommended AMFs. The third section describes the results of the cross-validation study.

Modeling Output

This section provides details about the results for the statistical models developed for this project. The first sub-section covers intersection models. The second sub-section describes the segment models. The third sub-section presents the characteristics of the models by crash type.

Intersection Models

This section summarizes the modeling results for intersection models. Also relevant are models recently estimated by Washington et al. (2005) for 3- and 4-legged stop-controlled intersections on multilane rural roads as part of an effort that re-estimated models for two-lane rural roads to be presented in HSM Chapter 8. Those relevant models are presented in Appendix A. The section is primarily divided into four parts. The first part describes the two methods used for assessing the statistical fit of models. The second part covers baseline models. The third part describes general ADT models. The fourth part summarizes models with covariates. The coefficients for all the models were estimated using SAS (SAS, 2002).

Assessment of the Goodness-of-Fit of Models

The assessment of the models (i.e., goodness-of-fit) was performed using the statistical tools provided by the software program (i.e., deviance, log likelihood, etc.) and the two methods proposed by Oh et al. (2003). The two methods are described below.

The Mean Absolute Deviance (MAD) provides a measure of the average mis-prediction of the model (Oh et al., 2003). It is computed using the following equation:

$$MAD = \frac{1}{n}\sum_{i=1}^{n}\left|\hat{y}_i - y\right| \tag{6.1}$$

The Mean Squared Predictive Error (MSPE) is typically used to assess the error associated with a validation or external data set (Oh et al., 2003). It can be computed using Equation (6.2):

$$MSPE = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2 \tag{6.2}$$

Baseline Models

It was possible to develop baseline models for total (KABCO) and injury (KAB) crashes for 3- and 4-legged stop-controlled intersections using the California data. These models are shown below.

4-Legged Two-Way Stop-Controlled (TWSC)

Table 6.1 summarizes the modeling results for the baseline models for 4-legged TWSC intersections. The models were estimated for the following baseline conditions: no turn lanes, no illumination, median on major, adequate sight distance, and angle between -5° and +5°. The functional form for the models is the following (in crashes per year):

$$\mu = \beta_0 F_1^{\beta_1} F_2^{\beta_2} \tag{6.3}$$

Table 6.1 Baseline Models for 4-Legged TWSC Intersections (California Data)

	Total	Injury
Parameter	**Estimate (Std Err)**	**Estimate (Std Err)**
Intercept ($\ln \beta_0$)	-10.7137 (1.6507)	-11.4399 (2.2568)
F_1 (β_1)	0.8482 (0.1703)	0.8281 (0.2302)
F_2 (β_2)	0.4481 (0.0827)	0.4122 (0.1108)
Dispersion Parameter (α)	0.4935 (0.1286)	0.6551 (0.2008)
Goodness-of-fit Statistics		
Number of Observations	62	62
Deviance	75.5453	65.8148
Pearson Chi-Square	6.9760	60.4798
Log Likelihood	1092.6824	111.4977
MAD[a]	5.557	2.3864
MSPE[b]	69.537	14.1147

[a]MAD = Mean absolute deviation (see description above)
[b]MSPE = Mean square prediction errors (see description above)

3-Legged Two-Way Stop-Controlled (SC)

Table 6.2 summarizes the modeling results for the baseline models for 3-legged SC intersections. Similar to the unsignalized 4-legged models described above, the baseline conditions were as follows: no turn lanes, no illumination, median on major, adequate sight distance, and angle between -5° and +5°. The functional form for the models is the following (in crashes per year):

$$\mu = \beta_0 F_1^{\beta_1} F_2^{\beta_2} \tag{6.4}$$

Table 6.2 Baseline Models for 3-Legged SC (California Data)

	Total	Injury
Parameter	**Estimate (Std Err)**	**Estimate (Std Err)**
Intercept ($\ln \beta_0$)	-13.0982 (0.9497)	-12.5606 (1.3239)
F_1 (β_1)	1.2040 (0.0973)	1.0130 (0.1321)
F_2 (β_2)	0.2357 (0.0515)	0.2280 (0.0673)
Dispersion Parameter (α)	0.4602 (0.0755)	0.5661 (0.1290)
Goodness-of-fit Statistics		
Number of Observations	133	133
Deviance	144.4375	135.0602
Pearson Chi-Square	171.5762	148.4170
Log Likelihood	3820.6699	202.2043
MAD	7.415	2.077
MSPE	229.902	13.888

General ADT Models

General ADT models were calibrated for consideration for use as baseline models where these are not suitable or where information about nominal conditions is not available. Separate general ADT models were estimated for California and Minnesota, and are presented below.

<u>4-Legged Signalized Intersections</u>

Tables 6.3 and 6.4 summarize the modeling results for 4-legged signalized intersections based on the California and Minnesota data, respectively. It was not possible to calibrate an injury model for California. The functional forms for the models are the following (in crashes per year):

$$\mu = \beta_0 F_1^{\beta_1} F_2^{\beta_2} \tag{6.5a}$$

$$\mu = \beta_0 \left(F_1 + F_2\right)^{\beta_3} = \beta_0 F_T^{\beta_3} \tag{6.5b}$$

Table 6.3 General ADT Models for 4-Legged Signalized Intersections (California Data)

	Total	Injury
Parameter	**Estimate (Std Err)**	**Estimate (Std Err)**
Intercept ($\ln \beta_0$)	-8.5482 (2.5110)	N/A
F_1 (β_1)	0.9575 (0.2605)	
F_2 (β_2)	0.0958 (0.0914)	
Dispersion Parameter (α)	0.3238 (0.0823)	
Goodness-of-fit Statistics		
Number of Observations	37	N/A
Deviance	38.9363	
Pearson Chi-Square	33.6301	
Log Likelihood	3636.9062	
MAD	16.979	
MSPE	453.146	

Table 6.4 General ADT Models for 4-Legged Signalized Intersections (Minnesota Data)

	Total	Injury
Parameter	**Estimate (Std Err)**	**Estimate (Std Err)**
Intercept ($\ln \beta_0$)	-7.4234 (1.9916)	-12.2515 (3.6628)
F_1 (β_1)	0.7224 (0.1865)	--
F_2 (β_2)	0.3369 (0.1012)	--
F_T (β_3)	--	1.2787 (0.3602)
Dispersion Parameter (α)	0.2767 (0.0709)	0.5658 (0.1873)
Goodness-of-fit Statistics		
Number of Observations	43	43
Deviance	49.1457	52.2327
Pearson Chi-Square	52.2477	35.1398
Log Likelihood	6448.0434	410.9344
MAD	15.411	4.193
MSPE	480.955	31.061

3-Legged Signalized Intersections

Table 6.5 summarizes the modeling results for 3-legged signalized intersections. The models were estimated using the California data. Due to the small sample size, the models presented in Table 6.5 were not recommended for application in the HSM Chapter 9. The functional form for the models is the following:

$$\mu = \beta_0 \left(F_1 + F_2 \right)^{\beta_1} = \beta_0 F_T^{\beta_1} \tag{6.6}$$

Table 6.5 General ADT Models for 3-Legged Signalized Intersections (California Data)

	Total	**Injury**
Parameter	**Estimate (Std Err)**	**Estimate (Std Err)**
Intercept ($\ln \beta_0$)	-6.4369 (2.6583)	-14.0930 (6.5936)
F_T (β_1)	0.7610 (0.2667)	1.2881 (0.6503)
Dispersion Parameter (α)	0.2835 (0.1288)	0.1660 (0.1942)
Goodness-of-fit Statistics		
Number of Observations	13	13
Deviance	13.4935	12.4248
Pearson Chi-Square	9.8716	11.1029
Log Likelihood	998.2771	11.0262
MAD	12.522	1.5772
MSPE	256.169	4.9500

4-Legged TWSC Intersections

Tables 6.6 and 6.7 summarize the modeling results for 4-legged TWSC intersections based on the California and Minnesota data. The functional form for the models is the following (in crashes per year):

$$\mu = \beta_0 F_1^{\beta_1} F_2^{\beta_2} \tag{6.7}$$

Table 6.6 General ADT Models for 4-Legged TWSC Intersections (California Data)

	Total	Injury
Parameter	**Estimate (Std Err)**	**Estimate (Std Err)**
Intercept ($\ln\beta_0$)	-8.3366 (0.7983)	-8.7935 (1.0551)
F_1 (β_1)	0.6697 (0.0827)	0.6016 (0.1089)
F_2 (β_2)	0.3659 (0.0351)	0.3517 (0.0462)
Dispersion Parameter (α)	0.6762 (0.0759)	0.8409 (0.1241)
Goodness-of-fit Statistics		
Number of Observations	267	267
Deviance	306.9926	291.3532
Pearson Chi-Square	371.7419	347.0326
Log Likelihood	5501.2040	497.6786
MAD	6.090	2.402
MSPE	90.142	13.432

Table 6.7 General ADT Models for 4-Legged TWSC Intersections (Minnesota Data)

	Total	Injury
Parameter	**Estimate (Std Err)**	**Estimate (Std Err)**
Intercept ($\ln\beta_0$)	-9.3818 (0.8863)	-10.1463 (1.2815)
F_1 (β_1)	0.6862 (0.0964)	0.5631 (0.1402)
F_2 (β_2)	0.5984 (0.0561)	0.7001 (0.0892)
Dispersion Parameter (α)	0.6813 (0.0906)	1.2384 (0.2080)
Goodness-of-fit Statistics		
Number of Observations	224	224
Deviance	265.4875	224.3631
Pearson Chi-Square	202.5639	208.7450
Log Likelihood	7224.2526	1016.3425
MAD	7.621	2.902
MSPE	239.267	31.510

3-Legged SC Intersections

Tables 6.8 and 6.9 summarize the modeling results for 3-legged SC intersections. The models were estimated from the California and Minnesota data. The functional form for the models is the following (in crashes per year):

$$\mu = \beta_0 F_1^{\beta_1} F_2^{\beta_2} \tag{6.8}$$

Table 6.8 General ADT Models for 3-Legged SC Intersections (California Data)

	Total	Injury
Parameter	**Estimate (Std Err)**	**Estimate (Std Err)**
Intercept ($\ln\beta_0$)	-11.7868 (0.5926)	-12.4338 (0.7674)
F_1 (β_1)	1.0457 (0.0603)	0.9824 (0.0771)
F_2 (β_2)	0.3149 (0.0302)	0.2781 (0.0358)
Dispersion Parameter (α)	0.6659 (0.0607)	0.6302 (0.0894)
Goodness-of-fit Statistics		
Number of Observations	403	403
Deviance	451.9755	421.0516
Pearson Chi-Square	532.3831	473.6664
Log Likelihood	9405.2866	430.2737
MAD	6.974	1.864
MSPE	181.859	10.069

Table 6.9 General ADT Models for 3-Legged SC Intersections (Minnesota Data)

	Total	Injury
Parameter	**Estimate (Std Err)**	**Estimate (Std Err)**
Intercept ($\ln\beta_0$)	-8.1815 (1.1867)	-8.5187 (1.6377)
F_1 (β_1)	0.5798 (0.1151)	0.4074 (0.1644)
F_2 (β_2)	0.4591 (0.0598)	0.5498 (0.0913)
Dispersion Parameter (α)	0.6702 (0.1122)	0.9486 (0.2568)
Goodness-of-fit Statistics		
Number of Observations	172	172
Deviance	206.6663	179.8830
Pearson Chi-Square	155.4650	203.0010
Log Likelihood	1542.0855	-13.5549
MAD	4.207	1.423
MSPE	39.283	3.843

Models with Covariates

Models with covariates were calibrated for certain HSM applications (e.g., network screening, countermeasure development, or before-after safety evaluations). The models described below have one of the following two forms (in crashes per year):

$$\mu = \beta_0 F_1^{\beta_1} F_2^{\beta_2} e^{\sum_{k=3}^{n} \beta_k x_k} \qquad (6.9a)$$

$$\mu = \beta_0 \left(F_1 + F_2\right)^{\beta_1} = \beta_0 F_T^{\beta_1} e^{\sum_{k=2}^{n} \beta_k x_k} \tag{6.9b}$$

The modeling results are shown in Tables 6.10 to 6.22. For several intersection categories or crash severities (marked "N/A"), it was not possible to calibrate or estimate these models with covariates. For some intersection types or severities, a number of alternative models were calibrated.

Table 6.10 Models with Covariates for 3-Legged Signalized Intersections (California Data)

	Total	**Injury**
Parameter	**Estimate (Std Err)**	**Estimate (Std Err)**
Intercept ($\ln \beta_0$)	-4.8630 (2.6385)	N/A
F_T (β_1)	0.5458 (0.2741)	
MinLT (β_2)* 0=Yes, 1=No	0.7334 (0.3575)	
Dispersion Parameter (α)	0.2066 (0.1013)	
Goodness-of-fit Statistics		
Number of Observations	13	N/A
Deviance	13.4478	
Pearson Chi-Square	11.4871	
Log Likelihood	999.9934	
MAD	11.621	
MSPE	212.922	

*MinLT = Left-turn lane on minor approach exists

Table 6.11 Models with Covariates for 4-Legged TWSC Intersections (California Data) (Model 1)

	Total	Injury
Parameter	**Estimate (Std Err)**	**Estimate (Std Err)**
Intercept ($\ln \beta_0$)	-8.4508 (0.7897)	N/A
F_1 (β_1)	0.7013 (0.0825)	
F_2 (β_2)	0.3835 (0.0352)	
MajLT (β_3)* 0=Yes, 1= No	-0.4175 (0.1209)	
Dispersion Parameter (α)	0.6365 (0.0726)	
Goodness-of-fit Statistics		
Number of Observations	267	N/A
Deviance	307.0439	
Pearson Chi-Square	365.9891	
Log Likelihood	5507.2510	
MAD	6.022	
MSPE	90.775	

*MajLT = Left-turn lane on major approaches exist

Table 6.12 Models with Covariates for 4-Legged TWSC Intersections (California Data) (Model 2)

	Total	Injury
Parameter	**Estimate (Std Err)**	**Estimate (Std Err)**
Intercept ($\ln \beta_0$)	-8.5999 (0.7986)	-8.9643 (1.0595)
F_1 (β_1)	0.7455 (0.0857)	0.6551 (0.1135)
F_2 (β_2)	0.3753 (0.0349)	0.3550 (0.0461)
Median (β_3)* 1=Yes, 0=No	-0.5742 (0.1860)	-0.3991 (0.2386)
Dispersion Parameter (α)	0.6441 (0.0732)	0.8277 (0.1226)
Goodness-of-fit Statistics		
Number of Observations	267	267
Deviance	306.5520	290.7433
Pearson Chi-Square	389.7000	360.3213
Log Likelihood	5506.3188	499.1261
MAD	5.948	2.381
MSPE	85.876	13.373

*Median = Divided/undivided median

Table 6.13 Models with Covariates for 4-Legged TWSC Intersections (California Data) (Model 3)

	Total	Injury
Parameter	**Estimate (Std Err)**	**Estimate (Std Err)**
Intercept ($\ln \beta_0$)	-8.5884 (0.7899)	N/A
F_1 (β_1)	0.7492 (0.0847)	
F_2 (β_2)	0.3866 (0.0350)	
Median (β_3)* 1=Yes, 0=No	-0.4403 (0.1894)	
MajLT (β_4)* 0=Yes, 1=No	-0.3365 (0.1234)	
Dispersion Parameter (α)	0.6205 (0.0713)	
Goodness-of-fit Statistics		
Number of Observations	267	N/A
Deviance	306.4200	
Pearson Chi-Square	376.2774	
Log Likelihood	5510.0936	
MAD	5.921	
MSPE	86.575	

*Median = Divided/undivided median
*MajLT = Left-turn lane on major approaches exist

Table 6.14 Models with Covariates for 4-Legged TWSC Intersections (California Data) (Model 4)

	KABCO	KAB
Parameter	**Estimate (Std Err)**	**Estimate (Std Err)**
Intercept ($\ln \beta_0$)	-7.5515 (0.9041)	N/A
F_1 (β_1)	0.6921 (0.0833)	
F_2 (β_2)	0.3559 (0.0349)	
MinLT (β_3)* 0=Yes, 1=No	-0.9634 (0.4911)	
Dispersion Parameter (α)	0.6587 (0.0745)	
Goodness-of-fit Statistics		
Number of Observations	267	N/A
Deviance	307.3116	
Pearson Chi-Square	362.5429	
Log Likelihood	5503.6971	
MAD	6.104	
MSPE	92.334	

*MinLT = Left-turn lane on minor approaches exist

Table 6.15 Models with Covariates for 4-Legged TWSC Intersections (California Data) (Model 5)

	Total	Injury
Parameter	**Estimate (Std Err)**	**Estimate (Std Err)**
Intercept ($\ln \beta_0$)	-8.4599 (0.7929)	N/A
F_1 (β_1)	0.6637 (0.0820)	
F_2 (β_2)	0.3698 (0.0350)	
MajRT (β_3)* 1=Yes, 0=No	0.2239 (0.1223)	
Dispersion Parameter (α)	0.6646 (0.0750)	
Goodness-of-fit Statistics		
Number of Observations	267	N/A
Deviance	307.2130	
Pearson Chi-Square	379.7275	
Log Likelihood	5502.8411	
MAD	6.166	
MSPE	91.104	

*MajRT = Right-turn lane on major approaches

Table 6.16 Models with Covariates for 4-Legged TWSC Intersections (California Data) (Model 6)

	Total	Injury
Parameter	**Estimate (Std Err)**	**Estimate (Std Err)**
Intercept ($\ln \beta_0$)	-8.2757 (0.8007)	N/A
F_1 (β_1)	0.6794 (0.0831)	
F_2 (β_2)	0.3624 (0.0352)	
Light (β_3)* 1=No, 0=Yes	-0.2064 (0.1187)	
Dispersion Parameter (α)	0.6675 (0.0750)	
Goodness-of-fit Statistics		
Number of Observations	267	N/A
Deviance	306.5502	
Pearson Chi-Square	370.2564	
Log Likelihood	5502.7320	
MAD	6.088	
MSPE	91.019	

*Light = Presence of lighting at the intersection

Table 6.17 Models with Covariates for 4-Legged TWSC Intersections (California Data) (Model 7)

	Total	Injury
Parameter	**Estimate (Std Err)**	**Estimate (Std Err)**
Intercept ($\ln \beta_0$)	-8.0123 (0.9030)	N/A
F_1 (β_1)	0.7432 (0.0846)	
F_2 (β_2)	0.3690 (0.0347)	
MajRT (β_3)* 1=Yes, 0=No	0.2266 (0.1213)	
MinLT (β_4)* 0=Yes, 1=No	-0.8066 (0.4923)	
Median (β_4)* 1=Yes, 0=No	-0.4644 (0.1900)	
Dispersion Parameter (α)	0.6238 (0.0717)	
Goodness-of-fit Statistics		
Number of Observations	267	N/A
Deviance	307.0927	
Pearson Chi-Square	382.3304	
Log Likelihood	5509.2353	
MAD	6.003	
MSPE	86.669	

*MajRT = Right-turn lane on major approaches
*MinLT = Left-turn lane on minor approaches
*Median = Divided/undivided median

Table 6.18 Models with Covariates for 4-Legged TWSC Intersections (Minnesota Data)

	Total	Injury
Parameter	**Estimate (Std Err)**	**Estimate (Std Err)**
Intercept ($\ln \beta_0$)	9.3028 (0.8896)	N/A
F_1 (β_1)	0.7265 (0.0991)	
F_2 (β_2)	0.5707 (0.0569)	
Light (β_3)* 0=No, 1=Yes	-0.3501 (0.1665)	
Dispersion Parameter (α)	0.6611 (0.0885)	
Goodness-of-fit Statistics		
Number of Observations	224	N/A
Deviance	265.6219	
Pearson Chi-Square	200.5806	
Log Likelihood	7226.5513	
MAD	7.397	
MSPE	238.132	

*Light = Presence of lighting at the intersection

Table 6.19 Models with Covariates for 3-Legged SC Intersections (California Data) (Model 1)

	Total	Injury
Parameter	**Estimate (Std Err)**	**Estimate (Std Err)**
Intercept ($\ln \beta_0$)	-11.7156 (0.5768)	-12.3996 (0.7652)
F_1 (β_1)	1.0962 (0.0600)	1.0107 (0.0782)
F_2 (β_2)	0.2737 (0.0312)	0.2564 (0.0371)
Light (β_3)* 1=No, 0=Yes	-0.4841 (0.1117)	-0.2698 (0.1341)
Dispersion Parameter (α)	0.6278 (0.0578)	0.6198 (0.0881)
Goodness-of-fit Statistics		
Number of Observations	403	403
Deviance	449.9605	419.8631
Pearson Chi-Square	485.5578	460.0294
Log Likelihood	9414.9390	432.3090
MAD	7.049	1.859
MSPE	181.571	9.925

*Light = Presence of lighting at the intersection

Table 6.20 Models with Covariates for 3-Legged SC Intersections (California Data) (Model 2)

	Total	Injury
Parameter	**Estimate (Std Err)**	**Estimate (Std Err)**
Intercept ($\ln \beta_0$)	-11.9778 (0.5987)	N/A
F_1 (β_1)	1.0730 (0.0615)	
F_2 (β_2)	0.3269 (0.0306)	
MajLT (β_3)* 0=Yes, 1=No	-0.2169 (0.0990)	
Dispersion Parameter (α)	0.6575 (0.0600)	
Goodness-of-fit Statistics		
Number of Observations	403	N/A
Deviance	450.8811	
Pearson Chi-Square	523.5833	
Log Likelihood	9407.7106	
MAD	6.900	
MSPE	181.042	

*MajLT = Left-turn lane on major approaches exist

Table 6.21 Models with Covariates for 3-Legged SC Intersections (California Data) (Model 3)

	Total	Injury
Parameter	Estimate (Std Err)	Estimate (Std Err)
Intercept ($\ln \beta_0$)	-11.9236 (0.5817)	N/A
F_1 (β_1)	1.1107 (0.0606)	
F_2 (β_2)	0.3128 (0.0294)	
Median (β_3)* 1=Yes, 0=No	-0.5792 (0.1295)	
Dispersion Parameter (α)	0.6221 (0.0576)	
Goodness-of-fit Statistics		
Number of Observations	403	N/A
Deviance	450.7708	
Pearson Chi-Square	511.1065	
Log Likelihood	9415.8518	
MAD	6.883	
MSPE	180.382	

*Median = Divided/undivided median

Table 6.22 Models with Covariates for 3-Legged SC Intersections (California Data) (Model 4)

	Total	Injury
Parameter	Estimate (Std Err)	Estimate (Std Err)
Intercept ($\ln \beta_0$)	-11.8354 (0.5730)	N/A
F_1 (β_1)	1.1345 (0.0603)	
F_2 (β_2)	0.2808 (0.0309)	
Median (β_3)* 1=Yes, 0=No	-0.4535 (0.1340)	
Light (β_4)* 1=No, 0=Yes	-0.3610 (0.1152)	
Dispersion Parameter (α)	0.6023 (0.0561)	
Goodness-of-fit Statistics		
Number of Observations	403	N/A
Deviance	450.1355	
Pearson Chi-Square	484.6914	
Log Likelihood	9420.8527	
MAD	6.784	
MSPE	174.409	

*Median = Divided/undivided median
*Light = Presence of lighting at the intersection

Segment Models

This sub-section summarizes the modeling results for rural multilane segments. The dependent variable is the number of crashes per year per mile. The coefficients of the models were estimated using SAS (SAS, 2002). The assessment of the models was performed using the statistical tools provided in SAS and the two methods proposed by Oh et al. (2003). The MAD and MSPE were described in the previous section.

Undivided Highways

This section provides the modeling results for undivided highways. The section is divided into three parts and covers baseline models, general ADT models, and models with covariates.

<u>Baseline Models</u>

This section describes the characteristics of the baseline models. To increase the sample size, the data from the states of Texas and Washington were merged together. The data could not be merged with the California data due to the lack of information about horizontal alignment. By merging the data, one can develop more robust models and minimize the biases associated with small sample size and low sample mean (Lord, 2006). The coefficients of the models were estimated using the following baseline conditions: lane width, 11-12 ft; shoulder width, 7-8 ft; no horizontal curves.

Models with segments that did not contain any intersections, in addition to the conditions listed above, were estimated, but were found unreliable. The sample size was too small, which affected the stability of the models.

Table 6.23 summarizes the modeling results for the baseline models. The functional form for the model with a varying dispersion parameter is the following (in crashes per year):

$$\mu = \beta_0 L F^{\beta_1} \tag{6.10a}$$

$$\alpha = e^{\gamma_0} L \tag{6.10b}$$

Table 6.23 Baseline Models for Segments (Texas and Washington Data)

	Total	**Injury**
Parameter	**Estimate (Std Err)**	**Estimate (Std Err)**
Model		
Intercept (ln β_0)	-11.4448 (1.307)	-10.4414 (1.710)
F (β_1)	1.2870 (0.149)	1.0642 (0.194)
Dispersion Parameter		
Intercept (ln γ_0)[a]	-0.6743 (0.561)	-3.5973 (5.531)
Goodness-of-fit Statistics		
Number of Observations	183	183
-2 Log Likelihood	447.9	278.3
AIC	453.9	284.3
BIC	463.6	293.9
MAD	0.958	1.073
MSPE	2.596	5.699

[a] Note: $\alpha = e^{\gamma_0} L = e^{-0.6743} L$

General ADT Models

This section describes the characteristics of the general ADT models. To increase the sample size, the data from the states of Texas, California, and Washington were merged into one single database.

Table 6.24 summarizes the modeling results for the general ADT models. The functional form for the model with a fixed dispersion parameter is the following (in crashes per year):

$$\mu = \beta_0 L F^{\beta_1} \tag{6.11}$$

Table 6.24 General ADT Models for Segments (Texas, California, and Washington Data)

	Total	**Injury**
Parameter	**Estimate (Std Err)**	**Estimate (Std Err)**
Model		
Intercept (ln β_0)	-10.2351 (0.284)	-9.8320 (0.379)
F (β_1)	1.1901 (0.031)	1.0083 (0.042)
Dispersion Parameter		
Intercept (ln γ_0)	0.1330 (0.066)	-0.4058 (0.108)
Goodness-of-fit Statistics		
Number of Observations	2083	2083
-2 Log Likelihood	8800.0	5143.8
AIC	8806.0	5149.8
BIC	8822.9	5166.7
MAD	3.020	1.080
MSPE	46.469	6.084

Models with Covariates

This section describes the modeling results for models with covariates. Distinct models were developed for California, Texas, Minnesota, and Washington.

California Models

Tables 6.25 and 6.26 summarize the model results for the California data. Table 6.25 shows the results for the models with a fixed dispersion parameter, whereas Table 6.26 summarizes the results for the models with a varying dispersion parameter. The tables illustrate that lane width, shoulder width, and the number of intersections located on the segment is associated with motor vehicle crashes.

The functional form for the model with a fixed dispersion parameter is the following (in crashes per year):

$$\mu = \beta_0 L F^{\beta_1} e^{\{\beta_2 LW + \beta_3 SW + \beta_4 INT\}} \tag{6.12}$$

Table 6.25 Segment Models for Models with Covariates and Fixed Dispersion Parameter (California Data)

Parameter	Total	Injury
	Estimate (Std Err)	**Estimate (Std Err)**
Intercept (ln β_0)	-6.7469 (0.927)	-7.0615 (1.326)
F (β_1)	1.1298 (0.057)	1.0967 (0.078)
LW (β_2)*	-0.1905 (0.070)	-0.2414 (0.096)
SW (β_3)*	-0.0370 (0.008)	-0.0381 (0.010)
INT (β_4)*	0.1005 (0.049)	--
Dispersion Parameter (α)	0.4746 (0.050)	0.5103 (0.078)
Goodness-of-fit Statistics		
Number of Observations	357	357
Deviance	380.647	353.5314
Pearson Chi-Square	431.136	394.0243
Log Likelihood	8730.513	862.0125
MAD	6.537	2.037
MSPE	198.073	14.760

*LW = Lane width in ft; *SW = Total shoulder width (both sides) in ft;
*INT = Number of intersections located on the segment

The functional form for the model with a varying dispersion parameter is the following (in crashes per year):

$$\mu = \beta_0 L F^{\beta_1} e^{\{\beta_2 LW + \beta_3 SW + \beta_4 INT\}} \qquad (6.13a)$$

$$\alpha = e^{\gamma_0} L \qquad (6.13b)$$

Table 6.26 Segment Models for Models with Covariates and Varying Dispersion Parameter (California Data)

	Total	Injury
Parameter	**Estimate (Std Err)**	**Estimate (Std Err)**
	Model	
Intercept (ln β_0)	-7.1636 (0.801)	-7.7696 (1.0878)
F (β_1)	1.1591 (0.051)	1.1043 (0.07135)
LW (β_2)*	-0.1807 (0.056)	-0.1927 (0.07485)
SW (β_3)*	-0.03418 (0.0072)	-0.03351(0.009720)
INT (β_4)*	0.1306 (0.042)	--
	Dispersion Parameter	
Intercept (ln γ_0)	0.05690 (0.124)	-0.08261 (0.1808)
	Goodness-of-fit Statistics	
Number of Observations	357	357
-2 Log Likelihood	2015.5	1281.1
AIC	2027.5	1291.1
BIC	2050.8	1310.4
MAD	6.515	2.030
MSPE	195.267	14.835

*LW = Lane width in ft; *SW = Total shoulder width (both sides) in ft;
*INT = Number of intersections located on the segment

Texas Models

Tables 6.27 and 6.28 summarize the model results for the Texas data. Table 6.27 shows the results for the models with a fixed dispersion parameter. On the other hand, Table 6.28 presents the results for the models with a varying dispersion parameter. The tables indicate that lane width, shoulder width, and the number of horizontal curves per mile (horizontal curve density) located on the segment are associated with motor vehicle crashes.

The functional form for the model with a fixed dispersion parameter is the following (in crashes per year):

$$\mu = \beta_0 LF^{\beta_1} e^{\{\beta_2 LW + \beta_3 SW + \beta_4 CURVE\_DEN\}} \quad (6.14)$$

Table 6.27 Segment Models for Models with Covariates and Fixed Dispersion Parameter (Texas Data)

	KABCO	KAB
Parameter	**Estimate (Std Err)**	**Estimate (Std Err)**
Intercept (ln β_0)	-7.9488 (0.406)	-6.8242 (0.547)
F (β_1)	0.9749 (0.044)	0.7768 (0.058)
LW (β_2)*	-0.0533 (0.017)	-0.0844 (0.023)
SW (β_3)*	-0.0100 (0.003)	-0.0114 (0.005)
CURVE_DEN (β_4)*	0.0675 (0.012)	0.0635 (0.016)
Dispersion Parameter (α)	0.3906 (0.036)	0.3793 (0.057)
Goodness-of-fit Statistics		
Number of Observations	1499	1499
Deviance	1577.7	1304.7
Pearson Chi-Square	1799.7	1524.7
Log Likelihood	3429.9	-222.8
MAD	1.702	0.826
MSPE	11.236	2.727

*LW = Lane width in ft; *SW = Total shoulder width (both sides) in ft;
*CURVE_DEN = Number of horizontal curves per mile located on the segment

The functional form for the model with a varying dispersion parameter is the following (in crashes per year):

$$\mu = \beta_0 L F^{\beta_1} e^{\{\beta_2 LW + \beta_3 SW + \beta_4 CURVE\_DEN\}} \quad (6.15a)$$

$$\alpha = e^{\gamma_0} L \quad (6.15b)$$

Table 6.28 Segment Models for Models with Covariates and Varying Dispersion Parameter (Texas Data)

	KABCO	KAB
Parameter	**Estimate (Std Err)**	**Estimate (Std Err)**
Model		
Intercept (ln β_0)	-8.4815 (0.387)	-6.9887 (0.524)
F (β_1)	1.0184 (0.042)	0.7889 (0.056)
LW (β_2)*	-0.04292 (0.015)	-0.08185 (0.021)
SW (β_3)*	-0.00859 (0.003)	-0.01014 (0.004)
CURVE_DEN (β_4)*	0.06202 (0.011)	0.06083 (0.016)
Dispersion Parameter		
Intercept (ln γ_0)	-1.0586 (0.134)	-1.6155 (0.2003)
Goodness-of-fit Statistics		
Number of Observations	1499	1499
-2 Log Likelihood	5268.1	3212.0
AIC	5280.1	3224.0
BIC	5311.9	3255.8
MAD	1.698	0.823
MSPE	11.437	2.765

*LW = Lane width in ft; *SW = Total shoulder width (both sides) in ft;
*CURVE_DEN = Number of intersections per mile located on the segment

Minnesota Models

Tables 6.29 and 6.30 summarize the model results for the Minnesota data. Table 6.29 displays the results for models with a fixed dispersion parameter. Alternatively, Table 6.30 shows the results for the models with a varying dispersion parameter. The tables indicate that lane width and shoulder widths are associated with motor vehicle crashes. However, the coefficient for the lane width variable is counterintuitive and is possibly caused by the small sample size.

The functional form for the model with a fixed dispersion parameter is the following:

$$\mu = \beta_0 LF^{\beta_1} e^{\{\beta_2 LW + \beta_3 SW\}} \tag{6.16}$$

Table 6.29 Segment Models for Models with Covariates and Fixed Dispersion Parameter (Minnesota Data)

	Total	Injury
Parameter	**Estimate (Std Err)**	**Estimate (Std Err)**
Intercept (ln β_0)	-7.8894 (0.917)	-9.7699 (1.952)
F (β_1)	0.9650 (0.109)	0.9546 (0.225)
LW (β_2)*	0.0449 (0.031)	0.0131 (0.058)
SW (β_3)*	-0.0469 (0.012)	-0.0204 (0.020)
Dispersion Parameter (α)	0.3443 (0.097)	0.0117 (0.255)
Goodness-of-fit Statistics		
Number of Observations	114	114
Deviance	124.818	93.178
Pearson Chi-Square	114.925	118.549
Log Likelihood	257.097	-78.571
MAD	2.359	0.460
MSPE	11.180	0.425

*LW = lane width in ft; *SW = Total shoulder width (both sides) in ft

The functional form for the model with a varying dispersion parameter is the following:

$$\mu = \beta_0 L F^{\beta_1} e^{\{\beta_2 LW + \beta_3 SW\}} \quad (6.17a)$$

$$\alpha = e^{\gamma_0} L \quad (6.17b)$$

Table 6.30 Segment Models for Models with Covariates and Varying Dispersion Parameter (Minnesota Data)

	Total	Injury
Parameter	**Estimate (Std Err)**	**Estimate (Std Err)**
	Model	
Intercept (ln β_0)	-8.2848 (0.925)	-9.7665 (1.947)
F (β_1)	0.9993 (0.102)	0.9540 (0.224)
LW (β_2)*	0.05533 (0.029)	0.01323 (0.057)
SW (β_3)*	-0.04327 (0.012)	-0.02057 (0.019)
	Dispersion Parameter	
Intercept (ln γ_0)	-0.3125 (0.341)	-7.8167 (.)
	Goodness-of-fit Statistics	
Number of Observations	114	114
-2 Log Likelihood	474.5	174.0
AIC	484.5	184.0
BIC	498.2	197.7
MAD	2.447	0.460
MSPE	12.003	0.424

*LW = Lane width in ft; *SW = Total shoulder width (both sides) in ft

Washington Models

Tables 6.31 and 6.32 summarize the model results for the Washington data. Table 6.31 exhibits the results for models with a fixed dispersion parameter. Table 6.32 presents the results for models with a dispersion parameter that varies as a function of the segment length. The tables illustrate that lane width, shoulder width, the number of intersections, and horizontal curves per mile located on the segment are associated with motor vehicle crashes. However, it should be noted that the effects for some of the variables, such as lane width, horizontal curve density, and intersections are very large, compared to previous models documented above and in the literature. For instance, the change in the total number of crashes per ft is -40.0% ($e^{-0.5076}-1$). Consequently, the authors caution about using these models for predicting motor vehicle crashes.

The functional form for the model with a fixed dispersion parameter is the following:

$$\mu = \beta_0 LF^{\beta_1} e^{\{\beta_2 LW + \beta_3 SW + \beta_4 INT + \beta_5 CURVE\_DEN\}} \quad (6.18)$$

Table 6.31 Segment Models for Models with Covariates and Fixed Dispersion Parameter (Washington Data)

	KABCO	KAB
Parameter	**Estimate (Std Err)**	**Estimate (Std Err)**
Intercept (ln β_0)	-5.7332 (3.268)	N/A
F (β_1)	1.4289 (0.320)	
LW (β_2)*	-0.5354 (0.144)	
SW (β_3)*	-0.0362 (0.027)	
INT (β_4)*	0.1144 (0.056)	
CURVE_DEN (β_5)*	0.1414 (0.068)	
Dispersion Parameter (α)	0.5742 (0.217)	
Goodness-of-fit Statistics		
Number of Observations	37	N/A
Deviance	35.608	
Pearson Chi-Square	43.518	
Log Likelihood	112.660	
MAD	2.692	
MSPE	27.696	

*LW = Lane width in ft; *SW = Total shoulder width (both sides) in ft
*INT = Number of intersections located on the segment
*CURVE_DEN = Number of horizontal curves per mile located on the segment

The functional form for the model with a varying dispersion parameter is the following:

$$\mu = \beta_0 L F^{\beta_1} e^{\{\beta_2 LW + \beta_3 SW + \beta_4 INT\}} \tag{6.19a}$$

$$\alpha = e^{\gamma_0} L \tag{6.19b}$$

Table 6.32 Segment Models for Models with Covariates and Varying Dispersion Parameter (Washington Data)

	KABCO	KAB
Parameter	**Estimate (Std Err)**	**Estimate (Std Err)**
Model		
Intercept (ln β_0)	-4.9817 (3.443)	N/A
F (β_1)	1.4298 (0.330)	
LW (β_2)*	-0.5862 (0.144)	
SW (β_3)*	-0.04193 (0.026)	
INT (β_4)*	0.1134 (0.054)	
CURVE_DEN (β_5)*	0.1238 (0.065)	
Dispersion Parameter		
Intercept (ln γ_0)	1.2105 (0.403)	
Goodness-of-fit Statistics		
Number of Observations	37	N/A
-2 Log Likelihood	158.4	
AIC	172.4	
BIC	183.7	
MAD	2.806	
MSPE	27.002	

*LW = Lane width in ft; *SW = Total shoulder width (both sides) in ft;
*INT = Number of intersections located on the segment
*CURVE_DEN = Number of horizontal curves per mile located on the segment

Divided Highways

This section summarizes the modeling results for divided highways. Statistical models were estimated for California and Texas, and only included the ones with covariates. Given the characteristics of the data, there were not enough observations for establishing baseline and general ADT models. In order to estimate models with baseline conditions, only segments with 12-ft lane width were used for developing the models. About 90% of all segments contained 12-ft lane width. For the models with a fixed dispersion parameter, a single intercept was estimated for each state. Due to the kind of modeling framework used for this part of the work, it was not possible to estimate a single intercept for each state for the model with a varying dispersion parameter. It should be pointed out that models for divided highways (median width equal to 30 ft) predict fewer crashes than models for undivided highways for the same level of exposure.

Table 6.33 summarizes the model results for the model with a fixed dispersion parameter. The table indicates that right shoulder width and median width are associated with motor vehicle crashes.

The functional form for the model with a fixed dispersion parameter is the following:

$$\mu = \beta_0 LF^{\beta_1} e^{\{\beta_2 MW + \beta_3 RSW\}} \tag{6.20}$$

Table 6.33 Segment Models for Models with Covariates and Fixed Dispersion Parameter (Texas and California Data)

	KABCO	KAB
Parameter	**Estimate (Std Err)**	**Estimate (Std Err)**
State1:Intercept (ln β_0)[†]	-8.1080 (0.236)	-7.9600 (0.279)
State2:Intercept (ln β_0)[†]	-8.4588 (0.220)	-8.1396 (0.261)
F (β_1)	1.0258 (0.022)	0.8584 (0.026
MW (β_2)*	-0.0023 (0.0005)	-0.0009 (0.001
RSW (β_3)*	-0.0740 (0.009)	-0.0612 (0.010
Dispersion Parameter (α)	0.3791 (0.016)	0.3008 (0.020)
Goodness-of-fit Statistics		
Number of Observations	2587	2587
Deviance	2826.0072	2717.9407
Pearson Chi-Square	3841.1498	3129.3143
Log Likelihood	68370.7398	6747.3372
MAD	5.290655	1.647256
MSPE	670.2857	16.48959

*MW = median + left shoulder width in ft; *RSW = average right shoulder width (both sides) in ft
[†] State 1: Texas, State 2: California

Table 6.34 summarizes the model results for the model with a varying dispersion parameter. Similar to the previous results, the table shows that right shoulder width and median width are associated with motor vehicle crashes. The models listed in Table 6.34 were used to produce baseline models for the HSM Chapter 9 by incorporating variables that describe typical baseline conditions (i.e., median width = 30 ft and right shoulder width = 8 ft).

The functional form for the model with a varying dispersion parameter is the following:

$$\mu = \beta_0 L F^{\beta_1} e^{\{\beta_2 MW + \beta_3 RSW\}} \tag{6.21a}$$

$$\alpha = e^{\gamma_0} L \tag{6.21b}$$

Table 6.34 Segment Models for Models with Covariates and Varying Dispersion Parameter (Texas and California Data)

	KABCO	KAB
Parameter	**Estimate (Std Err)**	**Estimate (Std Err)**
Model		
Intercept (ln β_0)	-9.7776 (0.215)	-8.7721 (0.242)
F (β_1)	1.1714 (0.0207)	0.9394 (0.0233)
MW (β_2)*	-0.00390 (0.0006)	-0.00181 (0.001)
RSW (β_3)*	-0.04210 (0.007)	-0.06008 (0.008)
Dispersion Parameter		
Intercept (ln γ_0)	-0.3715 (0.046)	-1.2824 (0.081)
Goodness-of-fit Statistics		
Number of Observations	2587	2587
-2 Log Likelihood	15165	9772.2
AIC	15175	9782.2
BIC	15204	9811.5
MAD	5.689934	1.677349
MSPE	683.4206	16.46669

*MW = median + left shoulder width in ft; *RSW = average right shoulder width (both sides) in ft

Crash Type Models

This section summarizes the models by collision types. Models were estimated for segments and intersections.

Models for Stop-Controlled Intersections

This section presents general ADT models for intersections. Four collision type models were estimated for: 1) single-vehicle, 2) intersecting (angle), 3) same-direction, and 4) opposing direction.

Single-Vehicle Crashes

Tables 6.35 and 6.36 summarize the modeling results for single-vehicle crashes for 3-legged and 4-legged stop-controlled intersections. The functional form for the model is the following:

$$\mu = \beta_0 F_T^{\beta_1} \qquad (6.22)$$

Table 6.35 Models for Single-Vehicle Crashes at 3-Legged SC Intersections (California Data)

	Total	**Injury**
Parameter	**Estimate (Std Err)**	**Estimate (Std Err)**
Intercept (ln β_0)	-11.5575 (0.7673)	-12.1798 (1.2312)
F_T (β_1)	1.0349 (0.0785)	0.9519 (0.1248)
Dispersion Parameter (α)	0.6407 (0.0887)	0.8378 (0.2279)
Goodness-of-fit Statistics		
Number of Observations	378	378
Deviance	389.0444	291.9866
Pearson Chi-Square	416.5742	386.6110
MAD	1.8104	0.6245
MSPE	12.2547	1.0262

Table 6.36 Models for Single-Vehicle Crashes at 4-Legged TWSC Intersections (California Data)

	Total	Injury
Parameter	**Estimate (Std Err)**	**Estimate (Std Err)**
Intercept (ln β_0)	-10.7050 (0.9420)	-10.6699 (1.6752)
F_T (β_1)	0.9501 (0.0972)	0.7995 (0.1726)
Dispersion Parameter (α)	0.4522 (0.0954)	1.0102 (0.3582)
Goodness-of-fit Statistics		
Number of Observations	264	264
Deviance	286.8640	202.7639
Pearson Chi-Square	256.7380	253.8210
MAD	1.5273	0.6171
MSPE	5.106	0.7531

Intersecting Crashes

Tables 6.37 and 6.38 summarize the models for intersecting crashes for 3-legged and 4-legged stop-controlled intersections. The functional form for the model is the following:

$$\mu = \beta_0 F_1^{\beta_1} F_2^{\beta_2} \qquad (6.23)$$

Table 6.37 Models for Intersecting Crashes at 3-Legged SC Intersections (California Data)

	Total	Injury
Parameter	**Estimate (Std Err)**	**Estimate (Std Err)**
Intercept (ln β_0)	-10.7594 (0.9599)	-12.6563 (1.3603)
F_1 (β_1)	0.6708 (0.0966)	0.7960 (0.1348)
F_2 (β_2)	0.5288 (0.0508)	0.4419 (0.0658)
Dispersion Parameter (α)	1.1842 (0.1497)	1.5375 (0.3011)
Goodness-of-fit Statistics		
Number of Observations	378	378
Deviance	365.9121	281.6272
Pearson Chi-Square	540.8687	562.8762
MAD	2.093	0.8237
MSPE	14.7569	2.1318

Table 6.38 Models for Intersecting Crashes at 4-Legged TWSC Intersections (California Data)

	Total	Injury
Parameter	**Estimate (Std Err)**	**Estimate (Std Err)**
Intercept (ln β_0)	-7.8013 (1.2808)	-8.2442 (1.5012)
F_1 (β_1)	0.4580 (0.1342)	0.4198 (0.1559)
F_2 (β_2)	0.4617 (0.0555)	0.4411 (0.0654)
Dispersion Parameter (α)	1.5195 (0.1964)	1.5059 (0.2558)
Goodness-of-fit Statistics		
Number of Observations	264	264
Deviance	278.4834	244.0316
Pearson Chi-Square	542.6259	502.1700
MAD	3.2612	1.546
MSPE	28.3496	6.0261

Opposing Direction Crashes

Tables 6.39 and 6.40 summarize the modeling results for opposing direction crashes for 3-legged and 4-legged stop-controlled intersections. The functional form for the model is the following:

$$\mu = \beta_0 F_1^{\beta_1} F_2^{\beta_2} \tag{6.24}$$

Table 6.39 Models for Opposing Direction Crashes at 3-Legged SC Intersections (California Data)

	Total	Injury
Parameter	**Estimate (Std Err)**	**Estimate (Std Err)**
Intercept (ln β_0)	-14.3796 (1.1890)	-16.0467 (1.6730)
F_1 (β_1)	1.0431 (0.1157)	1.1045 (0.1584)
F_2 (β_2)	0.4245 (0.0560)	0.4171 (0.0790)
Dispersion Parameter (α)	1.5713 (0.2363)	1.9430 (0.4454)
Goodness-of-fit Statistics		
Number of Observations	378	378
Deviance	312.5243	219.4241
Pearson Chi-Square	446.9706	413.5255
MAD	1.4319	0.5977
MSPE	8.51	1.5423

Table 6.40 Models for Opposing Direction Crashes at 4-Legged TWSC Intersections (California Data)

	Total	Injury
Parameter	**Estimate (Std Err)**	**Estimate (Std Err)**
Intercept (ln β_0)	-9.2451 (1.3211)	-10.7639 (1.7688)
F_1 (β_1)	0.4363 (0.1338)	0.5469 (0.1791)
F_2 (β_2)	0.5696 (0.0622)	0.4972 (0.0842)
Dispersion Parameter (α)	1.0679 (0.1918)	1.4260 (0.3654)
Goodness-of-fit Statistics		
Number of Observations	264	264
Deviance	241.0278	189.7150
Pearson Chi-Square	361.0404	354.5546
MAD	1.4129	0.7198
MSPE	7.3987	1.6864

Same Direction Crashes

Tables 6.41 and 6.42 summarize the modeling results for same direction crashes for 3-legged and 4-legged stop-controlled intersections. The functional form for the model is the following:

$$\mu = \beta_0 F_1^{\beta_1} F_2^{\beta_2} \tag{6.25}$$

Table 6.41 Models for Same Direction Crashes at 3-Legged SC Intersections (California Data)

	Total	Injury
Parameter	**Estimate (Std Err)**	**Estimate (Std Err)**
Intercept (ln β_0)	-16.0291 (0.8343)	-15.3078 (1.2842)
F_1 (β_1)	1.3807 (0.0831)	1.1985 (0.1277)
F_2 (β_2)	0.3063 (0.0405)	0.1471 (0.0590)
Dispersion Parameter (α)	0.8293 (0.1030)	0.6540 (0.1959)
Goodness-of-fit Statistics		
Number of Observations	378	378
Deviance	392.8642	279.4014
Pearson Chi-Square	686.4972	397.1157
MAD	2.9207	0.5849
MSPE	36.4117	0.985

Table 6.42 Models for Same Direction Crashes at 4-Legged SC Intersections (California Data)

	Total	Injury
Parameter	**Estimate (Std Err)**	**Estimate (Std Err)**
Intercept (ln β_0)	-12.1658 (1.0188)	-13.9293 (1.9287)
F_1 (β_1)	0.9707 (0.1043)	1.0320 (0.1964)
F_2 (β_2)	0.2911 (0.0445)	0.1835 (0.0793)
Dispersion Parameter (α)	0.8034 (0.1245)	1.2827 (0.3999)
Goodness-of-fit Statistics		
Number of Observations	264	264
Deviance	275.0518	178.9540
Pearson Chi-Square	388.6292	277.8590
MAD	2.1794	0.5801
MSPE	16.1487	0.9068

Models for Segments

This section presents general ADT models for segments. Three collision type models were estimated: 1) intersecting direction and turning, 2) single-vehicle and opposing direction, and 3) same direction. The models were estimated for undivided and divided segments.

Intersecting Direction and Turning Crashes—Undivided Segments

Table 6.43 summarizes the modeling results for multi-vehicle crashes occurring at minor intersections located on undivided segments (3-legged and 4-legged). The functional form for the model is the following:

$$\mu = \beta_0 F_1^{\beta_1} \tag{6.26}$$

Table 6.43 Models for Intersecting Direction and Turning Crashes on Undivided Segments (California Data)

	Total	Injury
Parameter	**Estimate (Std Err)**	**Estimate (Std Err)**
Intercept (ln β_0)	-10.8862 (1.2331)	-14.8938 (2.3708)
F_1 (β_1)	1.0463 (0.1364)	1.3017 (0.2580)
Dispersion Parameter (α)	0.8456 (0.1815)	1.5090 (0.7053)
Goodness-of-fit Statistics		
Number of Observations	321	321
Deviance	271.9292	133.5401
Pearson Chi-Square	432.8899	403.1290
MAD	0.9435	0.2764
MSPE	4.5183	0.3341

Single-Vehicle and Opposing Direction Crashes—Undivided Segments

Table 6.44 summarizes the models for single-vehicle and opposing crashes on undivided segments. The functional form for the model is the following:

$$\mu = \beta_0 F_1^{\beta_1} \tag{6.27}$$

Table 6.44 Models for Single-Vehicle and Opposing Direction Crashes on Undivided Segments (California Data)

	Total	Injury
Parameter	**Estimate (Std Err)**	**Estimate (Std Err)**
Intercept (ln β_0)	-6.1959 (0.7936)	-8.0949 (1.0468)
F_1 (β_1)	0.6955 (0.0894)	0.7895 (0.1171)
Dispersion Parameter (α)	0.7769 (0.0939)	0.9623 (0.1673)
Goodness-of-fit Statistics		
Number of Observations	321	321
Deviance	343.9405	296.2000
Pearson Chi-Square	403.5599	358.1533
MAD	3.1362	1.3684
MSPE	39.6649	7.1471

Same Direction Crashes—Undivided Segments

Table 6.45 summarizes the models for same direction crashes on undivided segments.

The functional form for the model is the following:

$$\mu = \beta_0 F_1^{\beta_1} \tag{6.28}$$

Table 6.45 Models for Same Direction Crashes on Undivided Segments (California Data)

	Total	Injury
Parameter	**Estimate (Std Err)**	**Estimate (Std Err)**
Intercept (ln β_0)	-15.8125 (1.0344)	-15.8310 (1.7839)
F_1 (β_1)	1.6206 (0.1123)	1.4415 (0.1922)
Dispersion Parameter (α)	0.5251 (0.1233)	0.5141 (0.3573)
Goodness-of-fit Statistics		
Number of Observations	321	321
Deviance	290.1979	179.6563
Pearson Chi-Square	398.9027	355.8598
MAD	1.1179	0.3317
MSPE	3.9842	0.3102

Intersecting Direction and Turning Crashes—Divided Segments

Table 6.46 summarizes the models for crashes related to minor intersections located on divided segments (3-legged and 4-legged). The functional form for the model is:

$$\mu = \beta_0 F_1^{\beta_1} \tag{6.29}$$

Table 6.46 Models for Intersecting Direction and Turning Crashes on Divided Segments (California Data)

	Total	Injury
Parameter	**Estimate (Std Err)**	**Estimate (Std Err)**
Intercept (ln β_0)	-10.5393 (0.9755)	-10.1174 (1.3242)
F_1 (β_1)	0.9079 (0.1006)	0.7229 (0.1356)
Dispersion Parameter (α)	1.2441 (0.1858)	0.4014 (0.2540)
Goodness-of-fit Statistics		
Number of Observations	755	755
Deviance	611.5160	431.2496
Pearson Chi-Square	1128.8914	1109.8216
MAD	0.8331	0.289
MSPE	2.2768	0.259

Single-Vehicle and Opposing Direction Crashes—Divided Segments

Table 6.47 summarizes the models for single-vehicle and opposing crashes on divided segments. The functional form for the model is the following:

$$\mu = \beta_0 F_1^{\beta_1} \tag{6.30}$$

Table 6.47 Models for Single-Vehicle and Opposing Direction Crashes on Divided Segments (California Data)

	Total	Injury
Parameter	**Estimate (Std Err)**	**Estimate (Std Err)**
Intercept (ln β_0)	-6.7476 (0.4205)	-7.2107 (0.5186)
F_1 (β_1)	0.7341 (0.0438)	0.6722 (0.0537)
Dispersion Parameter (α)	0.3573 (0.0317)	0.3059 (0.0425)
Goodness-of-fit Statistics		
Number of Observations	755	755
Deviance	852.0748	790.9872
Pearson Chi-Square	873.7402	773.2152
MAD	3.5566	1.5633
MSPE	42.0779	7.3658

Same Direction Crashes – Divided Segments

Table 6.48 summarizes the models for same direction crashes on divided segments. The functional form for the model is the following:

$$\mu = \beta_0 F_1^{\beta_1} \tag{6.31}$$

Table 6.48 Models for Same Direction Crashes on Divided Segments (California Data)

	Total	Injury
Parameter	**Estimate (Std Err)**	**Estimate (Std Err)**
Intercept (ln β_0)	-16.6290 (0.6800)	-15.0526 (0.9976)
F_1 (β_1)	1.6681 (0.0694)	1.3229 (0.1008)
Dispersion Parameter (α)	0.6112 (0.0592)	0.3486 (0.1121)
Goodness-of-fit Statistics		
Number of Observations	755	755
Deviance	771.9307	567.5936
Pearson Chi-Square	1815.3283	1003.7265
MAD	2.2515	0.5423
MSPE	29.891	1.0422

In this work, the predicted values from general ADT models by crash type were compared with the values estimated from general ADT models used for intersections and segments. The comparison analysis results show that models by crash type should be used whenever crash types are of interest. They produced more accurate results than multiplying crash type ratios with the output of general ADT models for intersections and segments. On the other hand, if the purpose is to estimate the number of crashes for an entire intersection, segment, or both, general ADT models should be used rather than summing up all the values estimated from general ADT models by crash type.

Recommended AMFs

This section describes the recommended AMFs that can be applied with the baseline models described above. These AMFs were taken from the literature and vetted through the Joint NCHRP 17-25/17-29 Expert Panel associated with this project, estimated from the data collected in this work, or estimated from previous work conducted by the research team. The AMFs are grouped under intersections and segments. It is important to point out that very few AMFs exist for multilane highways, as reported by the Joint NCHRP 17-25/17-29 Expert Panel (Harkey et al., 2008). Thus, further work will be needed to estimate new AMFs and possibly recalibrate existing AMFs. Table 6.49 summarizes the geometric and operational features that were examined by the Joint Expert Panel.

Table 6.49 Geometric and Operational Features Investigated by the Joint Expert Panel

Intersections	Segments
• Illumination • Turning Lanes • Signalization (i.e., phasing) • Approach Speed • Channelization • Traffic Control	• Median Width and Type • Shoulder Type and Width • Shoulder Rumble Strips • Roadside Hazards, Clear Zone, and Slideslope • Posted and Operating Speed • Horizontal and Vertical Alignment • Access, Driveways, and Median Openings

Intersection AMFs

The Joint NCHRP 17-25/17-29 Expert Panel convened for multilane rural roads did not develop any AMFs for intersections. Neither was it possible to use the limited data collected for this project to develop these. Some AMFs were developed for stop-controlled intersections by Washington et al. (2005) and these are documented below. The models from which the AMFs were derived are summarized in Appendix A.

The AMFs were derived using the third of the three methods listed in Section 5.5 and are given in Tables 6.50 and 6.51, respectively. These AMFs apply to total and injury crashes.

Table 6.50 AMFs for 3-Legged Unsignalized Intersections

AMF	Total	Injury
Left-turning lane on major road	0.71 (One approach)	1 (One approach)
Right-turning lane on major road	1 (One approach)	1 (One approach)
Sight Distance[d]	1	1
Intersection Angle	1+(0.016*SKEW[a])/ (0.98[b]+0.016*SKEW)	1+(0.017*SKEW[a])/ (0.52[c]+0.017*SKEW)

[a] SKEW = intersection skew angle (degrees), expressed as the absolute value of the difference between 90 degrees and the actual intersection angle.
[b] 0.98 = mean of the observed TOTACC accidents per year of the sites meeting no angle, no right lane, and presence of left lane.
[c] 0.52 = mean of the observed INJACC accidents per year of the sites meeting no angle, no right lane, and presence of left lane.
[d] Sight Distance = if limited in any quadrant.

Table 6.51 AMFs for 4-Legged Unsignalized Intersections

AMF	Total		Injury	
	One Approach	Both Approaches	One Approach	Both Approaches
Left-turn lane on major road	1	1	0.86	0.74
Right-turn lane on major road	1	1	1	1
Sight Distance[d]	1		1	
Intersection Angle	1+(0.053*SKEW[a])/ (1.43[b]+0.053*SKEW)		1+(0.048*SKEW[a])/ (0.72[c]+0.048*SKEW	

[a] SKEW = intersection skew angle (degrees), expressed as the absolute value of the difference between 90 degrees and the actual intersection angle

[b] 0.43 = mean of the observed TOTACC accidents per year of the sites meeting no angle, no right lane, and presence of left lane

[c] 0.72 = mean of the observed INJACC accidents per year of the sites meeting no angle, no right lane, and presence of left lane

[d] Sight Distance = if limited in any quadrant.

Segment AMFs

The AMFs for estimating changes in safety for geometric design and traffic control features of roadway segments are presented below. All the AMFs affect changes in the total number of crashes occurring on the segment. Some AMFs have been adjusted based on the percentage of targeted number of crashes.

AMFs for Lane Width

An AMF for lane width has been adopted from the work of Harwood et al. (2003) and Harkey et al. (2008) via the joint NCHRP 17-25/17-29 Meeting. The AMF is described in Table 6.52. The nominal condition for the lane width variable is 12 ft; for segments with traffic volumes equal to or greater than 2,000 vehicles per day; and the percentage of targeted crashes equal to 35%. For segments with smaller traffic volumes or a different percentage of targeted crashes, the reader is referred to Harkey et al. (2008) for obtaining additional information about estimating a different AMF for lane width. This table shows that as the lane width narrows, the expected number of crashes increases.

Table 6.52 AMF for Lane Width

	Lane Width (ft)			
	9	**10**	**11**	**12**
Four-lane undivided	1.13	1.08	1.02	1.00
Four-lane divided	1.09	1.05	1.01	1.00

AMFs for Paved Shoulder Width

An AMF for shoulder width on undivided segments has been adopted from the work of Harkey et al. (2008) via the Joint NCHRP 17-25/17-29 Meeting. The AMF for divided segments has been developed in this work and is based on the model documented in Table 6.34. Both AMFs are described in Table 6.53. For undivided segments, the nominal condition for the paved shoulder width variable is 6 ft; for segments with traffic volumes equal to or greater than 2,000 vehicles per day; and the percentage of targeted crashes equal to 35%. For segments with smaller traffic volumes or a different percentage of targeted crashes, the reader is referred to Harkey et al. (2008) for obtaining additional information about estimating a different AMF for shoulder width. For divided segments, the nominal conditions are 12-ft lane width and 8-ft shoulder widths. There are no values beyond 8.0 ft, since the exact safety effects are unknown.

Table 6.53 AMF for Paved Shoulder Width

	Average Shoulder Width (ft)(both sides)				
	0	**2**	**4**	**6**	**8**
Four-lane undivided	1.18	1.11	1.05	1.00	0.95
Four-lane divided (right shoulders)	1.18	1.13	1.09	1.04	1.00

AMFs for Median Width

Table 6.54 summarizes AMFs for median width. This AMF has been adopted from the work of Miaou et al. (2005). The nominal baseline conditions are set for a median width equal to 30 ft because existing guidelines usually recommend the installation of a median barrier for median widths below 30 ft. Baseline models for nominal conditions could be developed using the models described in Tables 6.33 and 6.34. The values below account for the total number of crashes occurring on the segment, but do reflect the fact that median width mainly affects median-related (20% of all crashes) and cross-median crashes (2% of all crashes).

Table 6.54 AMF for Medians with Barriers

Median Width (ft)	AMF
20	1.006
30	1.000
40	0.994
50	0.988
60	0.983
70	0.978
80	0.973
90	0.968
100	0.963

AMFs for Median Barrier

Table 6.55 describes AMFs for the presence of a concrete median barrier. This AMF has been adopted from the work of Miaou et al. (2005). The nominal baseline conditions are for a median width equal to 30 ft, with the concrete barrier located in the middle of the median. Baseline models for nominal conditions could be developed using the models described in Tables 6.33 and 6.34. The AMFs apply to the total number of crashes occurring on the segment but do reflect the fact that this AMF mainly affects median-related crashes (20% of all crashes). Note that this AMF should be used only when the segment has a median barrier, but the width of the median is changed.

Table 6.55 AMF for Median Barriers

Median Width (ft)	AMF
20	1.012
30	1.000
40	0.988
50	0.977
60	0.967
70	0.953
80	0.944
90	0.935
100	0.957

AMFs for Sideslopes

An AMF for the sideslope has been adopted from the work of Zegeer et al. (1988) and from the work of Harkey et al. (2008) via the Joint NCHRP 17-25/17-29 Meeting. The AMF is described in Table 6.56. The nominal conditions are for a sideslope 1:7 (vertical : horizontal) or flatter.

Table 6.56 AMF for Sideslope

1:2 or Steeper	1:4	1:5	1:6	1:7 or Flatter
1.18	1.12	1.09	1.05	1.00

AMFs for Horizontal Curves

Table 6.57 summarizes the AMF for estimating the safety effects of the number of horizontal curves per mile (curve density) on the segment. This AMF was produced from the models with covariates for the State of Texas (Table 6.27). The nominal conditions are for segments with no horizontal curves. This AMF does not address the exact design characteristics of horizontal curves.

Table 6.57 AMF for the Number of Horizontal Curves per Mile

Number of Horizontal Curves per Mile					
0	**1**	**2**	**3**	**4**	**5**
1.00	1.07	1.14	1.22	1.31	1.40

Cross-Validation Study

This section presents the results of the cross-validation study for intersections and segments. The purpose of the cross-validation study was to examine how a model that was developed from data collected in one jurisdiction predicts crashes when it is applied in another jurisdiction. In order to evaluate the application of models to a new jurisdiction, data collected from the State of New York were used for this part of the study. For this cross-validation exercise, only general ADT models, described above, were used; the variables included in various models were not available in all datasets. Previous work has shown that ADT models provide better flexibility to be transferred from one jurisdiction to another (Washington et al., 2005). The baseline models are not validated against other datasets because the required information for the various base conditions is not always available and can therefore not be used in the cross-validation study.

The analysis was carried out in two steps. The first step consisted of re-calibrating the models estimated from the original datasets using the data collected from the second dataset (can be referred to as the data collected in the new jurisdiction). The models were recalibrated using the existing method proposed for the HSM (Hughes et al., 2005). In this method, a multiplicative factor is calculated by dividing the sum of observed crashes in the new jurisdiction by the sum of predicted crashes estimated from the original model. This factor is then added to the original model. The second step consisted of applying the newly recalibrated model to the second dataset and running various statistical tests for assessing their performance.

The effectiveness of the cross-validation study was performed using several goodness-of-fit evaluation criteria. These criteria have been proposed by Oh et al. (2003) and Hauer and Bamfo (1997). They have been used by several researchers in highway safety (Lord and Persaud; 2000; Wang and Abdel-Aty, 2007). The evaluation criteria include the following:

- pearson product-moment linear correlation coefficients (PPMC);
- mean absolute deviation (MAD);
- mean square and mean square prediction errors (MSE and MSPE); and,
- cumulative residuals (CURE) plots for major and minor road ADTs.

The CURE plots are provided in Appendix D.

Table 6.58 shows the results for 3-legged signalized intersections. For this table, only the California models were used. With only fourteen 3-legged signalized sites, the sample size is probably not adequate for evaluating the California model applied to New York data. The results shown in Table 6.58 and in the cumulative residual graphs indicate that the recalibrated model is performing satisfactorily for 3-legged signalized intersections.

Table 6.58 Summary Results for 3-Legged Signalized Intersections

New York Data 14 Sites 65 Total Crashes 35 Injury Crashes					
Statistical Model	**Pearson's Correlation Coefficient**	**Mean Absolute Deviation**	**Mean Squared Prediction Error**	**Maximum Deviation From CURE Plot – Major AADT**	**Maximum Deviation From CURE Plot – Minor AADT**
MN-Total	-	-	-	-	-
CA-Total	0.53	2.59	11.20	8.53	8.99
MN-Injury	-	-	-	-	-
CA-Injury	0.48	0.56	0.49	1.56	1.66

Note: No model for Minnesota was successfully calibrated.

Table 6.59 shows the results for 3-legged stop-controlled intersections. This table shows that the recalibrated models from Minnesota and California perform very similarly although the Minnesota models have an advantage according to the CURE plots. The cumulative residual graphs are substantially within the standard deviation boundaries (see Appendix D). The maximum absolute deviations fall within approximately 10% of the total sum of crashes. These results indicate that both the recalibrated models are performing satisfactorily.

Table 6.59 Summary Results for 3-Legged Unsignalized Intersections

New York Data 281 Sites 673 Total Crashes 396 Injury Crashes					
Statistical Model	**Pearson's Correlation Coefficient**	**Mean Absolute Deviation**	**Mean Squared Prediction Error**	**Maximum Deviation From Cure Plot – Major AADT**	**Maximum Deviation From Cure Plot – Minor AADT**
MN-Total	0.46	2.02	10.63	36.85	60.53
CA-Total	0.46	2.04	10.70	67.97	49.84
MN-Injury	0.24	0.66	0.87	12.80	13.15
CA-Injury	0.33	0.67	0.91	19.01	18.30

Table 6.60 summarizes the results for 4-legged signalized intersections. With only eighteen 4-legged signalized sites, the sample size is perhaps not adequate for evaluating the Minnesota and California models applied to the New York data. Nonetheless, a comparison of goodness-of-fit measures shows that the recalibrated models from Minnesota and California perform very similarly. The cumulative residual graphs show significant deviations outside of the standard deviation limits, indicating that the recalibrated models are not performing as well as might be expected (see Appendix D).

Table 6.60 Summary Results for 4-Legged Signalized Intersections

New York Data 18 Sites 222 Total Crashes 107 Injury Crashes					
Statistical Model	**Pearson's Correlation Coefficient**	**Mean Absolute Deviation**	**Mean Squared Prediction Error**	**Maximum Deviation From Cure Plot – Major AADT**	**Maximum Deviation From Cure Plot – Minor AADT**
MN-Total	0.27	6.28	67.32	51.44	45.83
CA-Total	0.24	6.30	69.52	52.93	43.14
MN-Injury	-0.01	1.71	4.00	12.45	11.47
CA-Injury	-	-	-	-	-

Table 6.61 summarizes the results for 4-legged stop-controlled intersections. A comparison of goodness-of-fit measures shows that the recalibrated models from Minnesota and California perform very similarly. However, an examination of the cumulative residual graphs show that the California models perform better when applied to the New York data. While the California models largely stay within the two standard deviation limits, the Minnesota models deviate further outside these limits (see Appendix D). The results indicate that the California model performs satisfactorily while the Minnesota model is less so, although perhaps still reasonable.

Table 6.61 Summary Results for 4-Legged Unsignalized Intersections

New York Data 71 Sites 472 Total Crashes 299 Injury Crashes					
Statistical Model	**Pearson's Correlation Coefficient**	**Mean Absolute Deviation**	**Mean Squared Prediction Error**	**Maximum Deviation From Cure Plot – Major AADT**	**Maximum Deviation From Cure Plot – Minor AADT**
MN-Total	0.47	4.47	45.19	72.26	78.04
CA-Total	0.49	4.41	41.21	46.59	53.31
MN-Injury	0.01	1.82	6.46	23.76	23.27
CA-Injury	0.00	1.73	5.75	15.47	14.94

Table 6.62 shows the results for 4-lane undivided segments. This table shows that the recalibrated general ADT model and the model developed using the Minnesota data performed better than the California model. This is reflected in the cumulative residual graphs, which show that the curve oscillates within the standard deviation limits for most of the ADT ranges (see Appendix D). These results indicate that both the recalibrated general ADT and the model from models are performing satisfactorily.

Table 6.62 Summary Results for 4-Lane Undivided Segments

New York Data 199 Sites 2048 Total Crashes (7 years)				
Statistical Model	**Pearson's Correlation Coefficient**	**Mean Absolute Deviation**	**Mean Squared Prediction Error**	**Maximum Deviation From Cure Plot**
General ADT – Total (all states)	0.874	0.711	1.564	18.529
California – Total	0.757	0.936	12.308	49.310
Minnesota – Total	0.865	0.687	1.326	11.947

This section has shown that the recalibration procedure works relatively well if the statistical model to be recalibrated is a good candidate for transferring to the second jurisdiction. For this to work, it is important to test and ensure that the recalibrated statistical model, in fact, performs well for a given jurisdiction, from which it was developed. It is also believed that the present exercise was applied in under extreme conditions where, for example, the definition of an intersection-related crash is different in each of the three jurisdictions.

Chapter Summary

This chapter described the modeling results for this study. The results were separated into three sections. The first section described the results for the statistical models produced for estimating the safety performance of rural multilane highways and intersections. Models were produced for intersections, undivided and divided highway segments, by crash type, and by crash severity. Above 80 models were estimated in this body of work. The models were assessed using various GOF and other statistical measures. The second section covered the AMFs produced from this work as well as the ones taken from various sources in the literature and the Joint NCHRP 17-25/17-29 Expert Panel members. Some of these AMFs have been vetted by the Joint NCHRP 17-25/17-29 Expert Panel assembled for the benefits of this project. The last section described the results of the cross-validation study. The study was carried out with data collected from the State of New York. The cross-validation study shows that some general ADT models transferred very well, as long as they work well in the jurisdiction where they were estimated. The next chapter summarizes the main activities of this research project and discusses avenues for further work.

CHAPTER VII

Summary, Conclusions, and Recommendations

The primary objective of this research was to develop a methodology for estimating the safety performance of various transportation elements considered in the planning, design, and operation of non-limited-access rural multilane highways. The first section of this Chapter describes the summary of the work performed in this research. The second section summarizes the proposed models for Chapter 9 of the HSM. The third section provides recommendations for further work.

Summary of Project

There is a significant need to improve the explicit consideration of highway safety in making decisions on roadway planning, design, and operations. To receive appropriate consideration, safety needs to be dealt with quantitatively within the transportation planning and highway design processes. The lack of available tools is a deterrent to quantifying the safety of a transportation facility during the planning or highway design process. Recognizing this problem, a group of TRB committees has identified the need for more explicit and quantitative consideration of safety within the above-mentioned processes. This important need eventually led to the development of the forthcoming HSM. The Manual will serve as a tool to help practitioners make planning, design, and operations decisions based on safety. It will serve the same role for safety analysis that the HCM serves for traffic-operations analyses. The product of this research will provide the necessary tools for estimating the safety performance of multilane rural highways and will be incorporated into Chapter 9 of the HSM.

The literature review illustrated that the safety performance of multilane rural highways was seldom investigated, both for segments and intersections located on these facilities. Researchers have found though that multilane rural highway segments experience, on average, less crashes than two-lane rural highways for the same level of exposure. For intersections, only one study specifically focused on estimating the safety performance for these types of facilities. In short, the review indicated that there is a need to develop models and a methodology for estimating the safety performance of multilane rural highways.

Chapter III summarized the survey of selected DOTs. The objectives of this survey were to: a) determine whether the selected DOTs are currently using or developing statistical models to predict the safety performance of multilane rural highways; b) find out candidate input variables of interest to the survey participants; c) determine the availability of, and accessibility to various databases, such as crash data, geometric design information for segments and intersections, traffic flows for segments, and major and minor approaches.

The survey results showed that only two state agencies currently have a methodology for estimating the safety performance of multilane rural highways. The results also showed that crash data and segment files could be made available by all study participants. However, other databases, such as intersection databases, or access to georeferenced data, were not always available.

The summary statistics for the data collected were presented in Chapter IV. The data were used for developing statistical models and AMFs for intersections and segments as well as for a cross-validation study to evaluate the recalibration procedure for jurisdictions other than those for which the models were estimated. The models and AMFs were estimated using four state databases: Texas, California, Minnesota, and Washington. New York data were used for the cross-validation. The data included detailed information about geometric design characteristics, traffic flow, and motor vehicle crashes.

Chapter V described the modeling methodology proposed in this research. The proposed accident prediction methodology separated rural multilane road networks into segments and intersections. Specific models were developed for each transportation element. Three classes of models were proposed: models with covariates, baseline models applicable for specific values of covariates, and general ADT models. For the first model class (models with covariates), the relationship between crashes and geometric design features was captured via the covariates inside the statistical model. It was proposed that models be estimated for undivided and divided segments, as well as for most types of intersections, and where possible by injury severity and crash type. The models with covariates for divided segments were used to estimate baseline models for the HSM Chapter 9 by substituting baseline values for the variables in the model. For the second model class, baseline models were directly estimated using data meeting baseline conditions that typically reflect the nominal conditions agencies most often used for designing segments and intersections. Several models proposed for the HSM Chapter 9 were calibrated using this approach. For the third model class, general ADT models were developed for the following transportation elements: 4-lane undivided segments, 3- and 4-legged signalized intersections, as well as 3- and 4-legged unsignalized intersections. These models reflect the average conditions found in the data for each transportation element. These models can be used for cases where the user has limited information about the geometric design features for the particular project under study. For 4-legged signalized intersections, the general ADT models were used as baseline models in HSM Chapter 9 as neither of the other approaches was feasible for developing baseline models for this entity type. Due to the small sample size, the general

ADT models for 3-legged signalized intersections were not recommended as baseline models for the chapter.

Chapter V also described the framework for developing the models. This framework, which is very important for developing sound and statistically valid predictive models, included four steps: (1) determine the modeling objective matrix; (2) establish the appropriate processes to develop the models; (3) determine the inferential goals, and (4) select the computation techniques and tools. All the models were developed using this modeling framework. The coefficients of the models were estimated using negative binomial (NB) regression methods, with the exception of models of crash counts by severity. The last section of the chapter described the various methods to be employed for estimating AMFs.

The modeling results for this study were presented in Chapter VI. The results were separated into three sections. The first section described the results for the statistical models produced for estimating the safety performance of rural multilane highways and intersections. Models were produced for the four types of intersections, undivided and divided highway segments, by crash type, and by crash severity. More than 80 models were estimated in this research or derived from recent relevant research. The models were assessed using various goodness-of-fit and other statistical measures. The second section covered the AMFs produced from this work as well as the ones taken from various sources in the literature and proposed by the Joint NCHRP 17-25/17-29 Expert Panel. The last section described the results of the cross-validation study. The study was performed with the data collected from the State of New York. The cross-validation study showed that some models transferred very well, as long as they performed well in the jurisdiction where they were estimated.

Recommended Models for HSM

As described above, the series of models developed in this research will be incorporated into Chapter 9 of the HSM. The current predictive methodology used for the other chapters for estimating the safety performance of rural two-lane highways as well as urban and suburban arterials consists of using baseline models combined with the use of AMFs to account for safety effects of non-baseline design and traffic operations conditions. Distinct models were estimated for intersections and segments.

The proposed methodology for estimating the safety performance of multilane rural highways is very similar to the ones used for the other two "predictive" chapters of the HSM. However, the research team provided additional tools for cases when baseline conditions are not available or data are limited. General ADT models and models with covariates can be used in cases where baseline conditions are not available. The proposed methodology also includes the use of general ADT models by crash type. This type of model could be used when the safety for specific crash types is of interest. Table 7.1 summarizes the recommended models for Chapter 9 of the HSM.

Table 7.1 Recommended Baseline Models for Chapter 9 of the HSM

Transportation Elements	Approach Used
Segments	
4-Lane Undivided	Models estimated from data meeting baseline conditions (Table 6.23). General ADT crash type models also estimated (Tables 6.43 – 6.45).
4-Lane Divided	Models estimated from models with covariates by substituting variables meeting baseline conditions (Table 6.34).
Intersections	
3-Legged Unsignalized	Models estimated from data meeting baseline conditions (Table 6.2).
4-Legged Unsignalized	Models estimated from data meeting baseline conditions (Table 6.1).
3-Legged Signalized	No models recommended due to small sample size.
4-Legged Signalized	General ADT models estimated (Table 6.4).

Further Work

There are a few recommendations for further work. Firstly, there is a general need to develop more robust baseline models that are applicable to a wider range of variables and AMFs that are of interest to designers. To this end, the commitment of resources to special field data collection efforts should be considered.

Further work should also be performed on the application of predictive models by crash severity. These models have shown to provide better estimates of crash counts for different severity levels than the models currently used for the HSM. Preliminary work has been done as part of this project (documented in Appendix E and in Park and Lord, 2007) and by others (e.g., Miaou and Song, 2005; Ma and Kockelman, 2006), but more research is needed in this area.

Another important topic is the difficulty created by considering intersections and segments separately in developing models for accident prediction. This approach has been proposed here and in other chapters of the HSM. By using this approach, each intersection or segment is considered independent of each other. In reality, the transportation network should be viewed as an interactive system and should not be defined using artificial boundaries delimiting segments from intersections. Estimation problems are certain to occur at the boundaries. The research team initially proposed research activities on this topic. However, this topic was eventually removed from this research due to the unavailability of data for a large connected network. (Urban networks would be more suitable for this kind of research activities.)

The recalibration of the models developed for application in another jurisdiction was not specifically addressed in this research. It is understood that a parallel effort is underway as part of the HSM production contract (NCHRP 17-36) to develop and document a recalibration procedure that would be common to all three HSM Part III Chapters.

Finally, as documented in previous chapters, very few AMFs are available for multilane highways. Thus, further research should be done to fill this important gap. At the time this report was written, several research projects were already underway to improve this situation.

References

Abbas, K.A., "Traffic Safety Assessment and Development of Predictive Models for Accidents on Rural Roads in Egypt." *Accident Analysis & Prevention*, Vol. 36, No. 2, 2004, pp.149-163.

Agent, K.R., and J.G. Pigman, Safety Impacts of Rural Road Construction. Research Report KTC-01-01/SPR219-00-1I, Kentucky Transportation Center, Lexington, KY, 2001.

AASHTO, A Policy on Geometric Design of Highways and Streets, 5th Edition, American Association of State Highway and Transportation Officials Washington, D.C. 2004.

Amoros, E., J. L. Martin, and B. Laumon, "Comparison of Road Crashes Incidence and Severity between some French Counties." *Accident Analysis & Prevention*, Vol. 35, No. 4, 2003, pp. 537-547.

Bonneson, J.A., K. Fitzpatrick, M.D. Wooldridge, and D. Lord, A Plan for Incorporating Safety into the Highway Design Process. TTI Report 0-4703-1, Texas Transportation Institute, College Station, TX, 2005.

Bureau of Transportation Statistics, Pocket Guide to Transportation 2005. The Transportation Network: 2003. U.S. Department of Transportation, Washington, D.C. 2005. http://www.bts.gov/publications/pocket_guide_to_transportation/2005/pdf/entire.pdf (accessed on July 3rd, 2007)

Caliendo, C., M. Guida, and A. Parisi, "A Crash-Prediction Models for Multilane Roads." *Accident Analysis & Prevention*, Vol. 39, No. 4, 2007, pp. 657-670.

Carlin, B.P., and T.A. Louis, *Bayes and Empirical Bayes Methods for Data Analysis*, 2nd edition, Chapman and Hall/CRC, London, U.K., 2000.

Council, F.M., and J.R. Stewart, "Safety Effects of the Conversion of Rural Two-Lane to Four-Lane Roadways Based on Cross-Sectional Models." *Transportation Research Record 1665*, 1999, pp. 35-43.

De Leur, P., and T. Sayed, "Developing a Systematic Framework for Proactive Road Safety Planning." *Transportation Research Record 1784*, 2002, pp. 34-42.

Fitzpatrick, K., and K. Balke, "Evaluation of Flush Medians and Two-Way, Left-Turn Lanes on Four-Lane Rural Highways." *Transportation Research Record 1500*, 1995, pp. 146-152.

Gelman, A., J.B. Carlin, H.S. Stern, and D.B. Rubin, *Bayesian Data Analysis*, 2nd edition, Chapman & Hall/CRC Press, 2004.

Gilks, W.R., S. Richardson, and D.J. Spiegelhalter, *Markov Chain Monte Carlo in Practice*. London: Chapman and Hall, 1996.

Harkey, D.L., S. Raghavan, B. Jongdea, F.M. Council, K. Eccles, N. Lefler, F. Gross, B. Persaud, C. Lyon, E. Hauer, and J. Bonneson, Crash Reduction Factors for Traffic Engineering and ITS Improvement. Draft Final Report, Highway Safety Research Centre, University of North Carolina, Raleigh, NC, 2008.

Harwood, D.W., E.R.K. Rabbani, K.R. Richard, H.W. McGee, and G.L. Gittings, Systemwide Impact of Safety and Traffic Operations Design Decisions for 3R Projects. NCHRP Report No. 486, Transportation Research Board, Washington, D.C., 2003.

Harwood, D.W., F.M. Council, E. Hauer, W.E. Hughes, and A. Vogt. Prediction of the Expected Safety Performance of Rural Two-Lane Highways. Report FHWA-RD-99-207, FHWA, U.S. Department of Transportation, Washington, D.C., 2000.

Hauer, E., *Observational Before-After Studies in Road Safety: Estimating the Effect of Highway and Traffic Engineering Measures on Road Safety*. Pergamon, Elsevier Science Ltd, Oxford, U.K., 1997.

Hauer, E., “Overdispersion in Modelling Accidents on Road Sections and in Empirical Bayes Estimation.” *Accident Analysis & Prevention*, Vol. 33, No. 6, 2001, pp. 799-808.

Hauer, E. and J. Bamfo, Two tools for finding what function links the dependent variable to the explanatory variables. In Proceedings of the ICTCT 1997 Conference, Lund, Sweden, 1997.

Heydecker, B.G., and J. Wu, "Identification of Sites for Road Accident Remedial Work by Bayesien Statistical Methods: An Example of Uncertain Inference." *Advances in Engineering Software*, Vol. 32, 2001, pp. 859-869.

Hughes, W., K. Eccles, D. Harwood, I. Potts, and E. Hauer, Development of a Highway Safety Manual. Appendix C: Highway Safety Manual Prototype Chapter: Two- Lane Highways. NCHRP Web Document 62 (Project 17-18(4)). Washington, D.C., 2005. (http://onlinepubs.trb.org/onlinepubs/nchrp/nchrp_w62.pdf, accessed October 2007)

Ladron de Guevara, F., S.P. Washington, and J. Oh, "Forecasting Crashes at the Planning Level: Simultaneous Negative Binomial Crash Model Applied in Tucson, Arizona." *Transportation Research Record 1897*, 2004, pp. 191-199.

Lord, D., "Modeling Motor Vehicle Crashes using Poisson-gamma Models: Examining the Effects of Low Sample Mean Values and Small Sample Size on the Estimation of the Fixed Dispersion Parameter." *Accident Analysis & Prevention*, Vol. 38, No. 4, 2006, pp. 751-766.

Lord, D., and J.A. Bonneson, "Development of Accident Modification Factors for Rural Frontage Road Segments in Texas." *Transportation Research Record 2023*, 2007, pp. 20-27.

Lord, D., A. Manar, and A. Vizioli, "Modeling Crash-Flow-Density and Crash-Flow-V/C Ratio for Rural and Urban Freeway Segments." *Accident Analysis & Prevention*. Vol. 37, 2005a, pp. 185-199.

Lord, D., and B.N. Persaud, "Accident prediction models with and without trend: Application of the Generalized Estimating Equations (GEE) procedure." *Transportation Research Record 1717*, 2000, pp. 102-108.

Lord, D., and B. Persaud, "Estimating the Safety Performance of Urban Transportation Networks." *Accident Analysis & Prevention*, Vol. 36, 2004, pp. 609-620.

Lord, D., S.P. Washington, and J.N. Ivan, "Poisson, Poisson-Gamma and Zero Inflated Regression Models of Motor Vehicle Crashes: Balancing Statistical Fit and Theory." *Accident Analysis & Prevention*. Vol. 37, No. 1, 2005b, pp. 35-46.

Lyon, C., J. Oh, B.N. Persaud, S.P. Washington, and J. Bared, "Empirical Investigation of the IHSDM Accident Prediction Algorithm for Rural Intersections." *Transportation Research Record 1840*, 2003, pp. 78-86.

Ma, J, and K.M. Kockelman, "Bayesian Multivariate Poisson Regression for Models of Injury Count by Severity." *Transportation Research Record 1950*, 2006, pp. 24-34.

MathWorks, Inc., MATLAB Neural Network Toolbox 5. Natick, MA, 2006.

McCullagh P., and J.A. Nelder, *Generalized Linear Models*. Chapman and Hall, London, U.K., 1989.

McLean, J., Review of Accidents and Rural Cross Section Elements Including Roadsides. Research Report ARR 297, Australian Road Research Board, Vermont South, Vic., 1996.

Miaou, S.P., R. Bligh, and D. Lord, "Developing Median Barrier Installation Guidelines: A Benefit/Cost Analysis using Texas Data." *Transportation Research Record 1904*, 2005, pp. 3-19.

Miaou, S.-P., and D. Lord, "Modeling Traffic-Flow Relationships at Signalized Intersections: Dispersion Parameter, Functional Form and Bayes vs Empirical Bayes." *Transportation Research Record 1840*, 2003, pp. 31-40.

Miaou, S.-P., and J.J. Song, "Bayesian Ranking of Sites for Engineering Safety Improvements: Decision parameter, Treatability Concept, Statistical Criterion and Spatial Dependence." *Accident Analysis & Prevention*, Vol. 37, No. 4, 2005, pp. 699-720.

Miranda-Moreno, L., D. Lord, and L. Fu, "Evaluation of Alternative Hyper-Priors for Bayesian Road Safety Analysis." Paper presented at the 87th Annual Meeting of the Transportation Research Board, Washington D.C., 2008.

Mountain, L., B. Fawaz, and D. Jarrett," Accident Prediction Models for Roads with Minor Junctions." *Accident Analysis & Prevention*, Vol. 28, No. 6, 1996, pp. 695-707.

Myers, R.H., *Classical and Modern Regression with Applications*. 3rd Edition, Duxbury Press, Belmont, CA, 2000.

Oh, J., C. Lyon, S.P. Washington, B.N. Persaud, and J. Bared, Validation of the FHWA Crash Models for Rural Intersections: Lessons Learned. *Transportation Research Record 1840*, 2003, pp. 41-49.

Park, E.S., and D. Lord, "Multivariate Poisson-Lognormal Models for Jointly Modeling Crash Frequency by Severity." *Transportation Research Record 2019*, 2007, pp. 1-6.

Persaud, B.N.. "Estimating Accident Potential of Ontario Road Sections." *Transportation Research Record 1327*, pp.47-54, 1991.

Persaud, B.N., and G. Bahar, "Applications of Safety Performance Functions in the Management of Highway Safety." Paper presented at the 2000 CSCE Annual Meeting, June 7-10, London, Canada, 2000.

Persaud, B.N., H. McGee, C. Lyon, and D. Lord, "Development of a Procedure for Estimating the Safety Effects for a Contemplated Traffic Signal Installation." *Transportation Research Record 1840*, 2003, pp. 96-103.

Persaud, B.N., R. Retting, P. Garder, and D. Lord, "Crash and Injury Reduction Following Installation of Roundabouts in the United States." *American Journal of Public Health*, Vol. 91 No. 4, 2001.

Prinsloo B., and C. Goudanas, Development of a Crash Prediction Model for Rural Roads in NSW. Proceedings of the 21st ARRB and 11th REAAA Conference. ARRB Transport Research, Road Engineering Association of Asia and Australasia, Cairns, Australia, 2003.

Qin, X., J.N. Ivan, and N. Ravishankar, "Selecting Exposure Measures in Crash Rate Prediction for Two-Lane Highway Segments." *Accident Analysis & Prevention*, Vol. 36, No. 1, 2004, pp. 183-191.

Rogness, R.O., D.B. Fambro, and D.S. Turner, "Before-After Accident Analysis for Two Shoulder Upgrading Alternatives." *Transportation Research Record 855*, TRB, 1982, pp. 41-47.

SAS Institute Inc., Version 9 of the SAS System for Windows. Cary, N.C., 2002.

Schluter, P. J., J. J. Deely, and A. J. Nicholson, "Ranking and Selecting Motor Vehicle Accident Sites by Using a Hierarchical Bayesian Model." *The Statistician*, Vol. 46, No. 3, 1997, pp. 293-316.

Song, J.J., M. Gosh, S.-P. Miaou, and B. Malik, "Bayesian Multivariate Spatial Models for Roadway Traffic Crash Mapping." *Journal of Multivariate Analysis*, Vol. 97, No. 1, 2006, pp. 246-273.

Turner, S., I. Bone, and M. Jackett, New Zealand Accident Prediction Models and Their Applications. Proceedings of the 21st ARRB and 11th REAAA Conference. ARRB Transport Research, Road Engineering Association of Asia and Australasia, Cairns, Australia, 2003.

U.S. DOT, Status of the Nation's Highways, Bridges, and Transit: Conditions and Performance, Chapter 5: Safety Performance. Federal Highway Administration, Washington, D.C., 2002.

U.S. General Accounting Office, Highway Safety: Federal and State Efforts to Address Rural Road Safety Challenges. GAO-04-663, 2004. (http://www.gao.gov/new.items/d04663.pdf, accessed on July 14th, 2007)

Vogt A., and J. Bared, Accident Models for Rural Two-Lane Roads: Segments and Intersections. FHWA-RD-98-133, FHWA, U.S. Department of Transportation, Washington, D.C., 1999.

Vogt, A., Crash Models for Rural Intersections: Four-Lane by Two-Lane Stop-Controlled and Two-Lane by Two-Lane Signalized. FHWA-RD-99-128, FHWA, U.S. Department of Transportation, Washington, D.C., 1999.

Wang, X., and M. Abdel-Aty. "Investigation of Signalized Intersection Right-Angle Crash Occurrence at Intersection, Roadway, and Approach Levels." Presented at the 84th Annual Meeting of the TRB, Washington, D.C., 2007.

Wang, J., W.E. Hughes, and R. Stewart, Safety Effects of Cross-Section Design on Rural, Four-Lane, Non-Freeway Highways. FHWA-RD-98-071, U.S. Department of Transportation, Washington, D.C., 1998.

Washington, S., B. Persaud, C. Lyon, and J. Oh, Validation of Accident Models for Intersections. FHWA-RD-03-037, Federal Highway Administration, Washington, D.C., 2005.

Ye, Z., and D. Lord, "Estimating the Variance in Before-After Studies." Presented at the 86th Annual Meeting of the TRB, Washington, D.C., 2007.

Zegeer, C.V., D.W. Reinfurt, W.W. Hunter, J. Hummer, R. Stewart, and L. Herf, "Accident Effects of Sideslope and Other Roadside Features on Two-Lane Roads." *Transportation Research Record 1195*, 1988, pp. 33-47.

APPENDIX A

Prediction Models and AMFs for Stop-Controlled Intersections from FHWA Study (Washington et al., 2005)

NOTE: Tables and some text are taken verbatim from the source report.

INTRODUCTION

Two types of intersection models were addressed in the FHWA research (Washington et al., 2005):

- Type III: Three-legged stop controlled intersections with two lanes on minor and four lanes on major roads.
- Type IV: Four-legged stop controlled intersections with two lanes on minor and four lanes on major roads.

The models were originally calibrated by Vogt (1999). The FHWA model recalibration effort was focused on improving the Vogt crash models through use of an improved and expanded database and through lessons learned in the validation and recalibration activities in that effort.

The data in support of that research were derived from three sources:

1. The original data used for the calibration of the main models for total accidents were obtained from the researchers who developed those models.
2. Highway Safety Information Systems (HSIS) data were obtained for additional years for the same intersections used in the calibration and for injury accidents for the original and additional years.
3. An independent validation data set of intersections and their relevant crash, traffic, and geometric data in Georgia was specially assembled for this project.

For each intersection type, the FHWA effort developed and/or refined three different sets of models. The first type is Annual Average Daily Traffic (AADT) Models, which represent base models for predicting crashes as a function of major and minor road AADT. The second type of model is Full Models. These statistical models forecast crashes as a function of a relatively large set of independent variables. The third type of model is AMF. These models, better described as countermeasure correction factors, represent the researchers' best efforts to estimate the effect of geometric countermeasures on safety relative to base model predictions.

VARIABLE ABBREVIATIONS

TOTACC: Total number of accidents within 76.25 m (250 ft) of the intersection.
INJACC: Total number of injury crashes within 76.25 m (250 ft) of the intersection.
F_1: Average daily traffic on major road (vehicles per day).
F_2: Average daily traffic on minor roads (vehicles per day).
COMDRWY1: Commercial driveways on major roads within 76.25 m (250 ft) of the intersection center.
COMDRWY2: Commercial driveways on minor roads within 76.25 m (250 ft) of the intersection center.
DRWY1: Driveways on major roads within 76.25 m (250 ft) of the intersection.
DRWY2: Driveways on minor roads within 76.25 m (250 ft) of the intersection.
HAU: Intersection angle variable defined where the angle between the major and minor roads is measured from the far side of the minor road:

- Three-legged intersections: Angle minus 90 if minor road is to the right of the major road in the increasing direction; 90 minus angle if minor road is to the left of the major road in the increasing direction.

- Four-legged intersections: (right angle – left angle)/2.

HAZRAT1: Roadside hazard rating on major road within 76.25 m (250 ft) of the intersection center (from 1, least hazardous case, to 7, most hazardous case).
HAZRAT2: Roadside hazard rating on minor road within 76.25 m (250 ft) of the intersection center (from 1, least hazardous case, to 7, most hazardous case).
LTLN1S: Left-turn lane on major roads (0 = no, 1 = yes).
MEDTYPE: Median type (0 = no median, 1 = painted, 2 = curbed, 3 = others).
MEDWDTH1: Median width on major roads (feet).
MEDWDTH2: Median width on minor roads (feet).
PKLEFT: Peak left-turn percentage (percent).
PKLEFT1: Peak left-turn percentage on major roads (percent).
PKLEFT2: Peak left-turn percentage on minor roads (percent).
PKTHRU1: Peak through percentage on major roads (percent).
PKTHRU2: Peak through percentage on minor roads (percent).
PKTRUCK: Peak truck percentage passing through the intersection (percent).
PKTURN: Peak turning percentage (percent).
PROT_LT: Protected left lane (0 = no, 1 = yes).
SDR2: Right-side sight distance on minor road (feet).
SPD1: The average posted speed on major roads in vicinity of the intersection (mph).
SPD2: The average posted speed on minor roads in vicinity of the intersection (mph).

SUMMARY STATISTICS OF RECALIBRATION DATA

Table A1 Sources of Data

State	Years of Data Available	No. of Sites		No. of Total (Injury Accidents)	
		Type III	Type IV	Type III	Type IV
California HSIS	1991-98	294	222	2136 (847)	1956 (899)
California (Original)	1993-98	60	54	427 (196)	478 (268)
Michigan	1993-97	24	18	248 (63)	277 (92)
Georgia	1996-97	52	52	124 (56)	222 (104)
Total		430	346	2935 (1162)	2933 (1363)

Table A2 Summary Statistics for Type III Sites

Variables	Frequency	Mean	Median	Minimum	Maximum
TOTACC per year	136	1.35	0.80	0.00	10.60
INJACC per year	136	0.55	0.33	0.00	4.00
F_1	136	13011	12100	2360	33333
F_2	136	709	430	15	9490
MEDTYPE1 Total	136				
No Median	69(50.7%)				
Painted	45(33.1%)				
Curbed	14(10.3%)				
Other	8(5.9%)	N/A			
MEDWIDTH1	136	12.6	6	0	63
HAU	136	1.3	0	-65	90
HAZRAT1 Total	136				
1	16(11.8%)				
2	58(42.6%)				
3	26(19.1%)				
4	25(18.4%)				
5	8(5.9%)				
6	2(1.5%)				
7	1(0.7%)	N/A			
HAZRAT2 Total	52				
1	0(0%)				
2	2(4.0%)				
3	20(40.0%)				
4	16(32.0%)				
5	6(12.0%)				
6	6(12.0%)				
7	2(4.0%)	N/A			
COMDRWY1	136	1.5	0	0	14
DRWY1	136	2.5	1.0	0.0	15.0
NoCOMDRWY2	52	0.4	0	0	3
DRWY2	52	1.0	1.0	0.0	6.0
SPD1	136	52.5	55	30	65
SPD2	136	33.7	35	15	55
PKTRUCK	84	9.15	7.79	1.18	28.16
PKTURN	84	6.68	4.28	0.27	53.09
PKLEFT	84	3.28	2.16	0.13	25.97
PKLEFT1	84	1.47	0.69	0.00	21.29
PKLEFT2	84	55.31	60.29	0.00	100.00
SD1	136	1515	2000	500	2000
SDL2	136	1418	1510	40	2000
SDR2	136	1428	1555	80	2000

Table A3 Summary Statistics for Type IV Sites

Variables	Frequency	Mean	Median	Minimum	Maximum
TOTACC per YEAR	124	2.0	1.4	0.0	10.8
INJACC per YEAR	124	0.9	0.5	0.0	5.7
F_1	124	12881	11496	3150	73799
F_2	124	621	430	21	2990
MEDTYPE on major Total 0: No Median 1: Painted 2: Curbed 3: Other	124 70(56.5%) 27(21.8%) 22(17.7%) 5(4.0%)	N/A			
MEDTYPE on minor Total 0: No Median	52 52(100%)	N/A			
MEDWDTH1	124	16.1	6.5	0	60
MEDWDTH2	52	0.0	0	0	1
HAZRAT1 1 2 3 4 5 6 7	124 24(19.4%) 43(34.7%) 32(25.8%) 21(16.9%) 2(1.6%) 2(1.6%) 0(0%)	N/A			
HAZRAT2 1 2 3 4 5 6 7	52 0(0%) 7(13.5%) 15(28.8%) 16(30.8%) 12(23.1%) 2(3.8%) 0(0%)	N/A			
COMDRWY1	124	0.6	0	0	12
DRWY1	124	1.3	0	0	15
COMDRWY2	52	0.4	0	0	4
DRWY2	52	0.8	0	0	6
HAU	124	1.5	0	-50	55
SPD1	124	55.6	55	25	65
SPD2	124	34.7	35	25	55
PKTRUCK	72	10.95	8.36	1.75	37.25
PKTHRU1	72	94.41	96.95	67.77	100.00
PKTURN	72	9.47	6.56	0.00	48.52
PKLEFT	72	4.80	3.08	0.00	25.26
PKLEFT1	72	2.78	1.51	0.00	13.96
PKTHRU2	72	15.69	10.82	0.00	68.09
PKLEFT2	72	38.89	36.66	0.00	100.00

AADT MODEL ESTIMATION RESULTS

Recommended models were calibrated using all available data from the HSIS California database, the original sites from Minnesota and Michigan, and the Georgia validation data. Models were calibrated for data meeting specified base conditions listed in Table A5.

Table A5 Base Conditions for AADT Models

Variable	Type III and IV Base Condition	Type III Frequency (Percent)	Type IV Frequency (Percent)
Right turn on major	No	253 (86.05)	164 (73.87)
Right turn on minor	No	268 (91.16)	176 (79.28)
Left turn on major	Yes	174 (59.18)	145 (65.32)
Left turn on minor	No	292 (99.32)	219 (98.65)
Median on major	Yes	212 (73.87)	148 (66.67)
Terrain on major	Flat	164 (55.78)	148 (66.67)
Total sites meeting all base conditions		62 (21.09)	34 (15.32)
Total sites		294 (100.00)	222 (100.00)

Table A6 Number of Sites Used for Type III AADT Models

Dataset	All Sites	Group B Base Condition Sites	Percent (Base/All Sites)
Michigan 1993–97	24	0	0.0
Georgia 1996–97	52	14	27.0
California 1991–98	218	48	22.0

Table A7 Number of Sites Used for Type IV AADT Models

Dataset	All Sites	Base Condition Sites	Percent (Base/All Sites)
Michigan 1993–97	18	0	0.0
Georgia 1996–97	52	1	1.9
California 1991–98	152	33	21.7

Table A8 Summary Statistics for Type III Sites for AADT Models

	All Sites				Base Condition Sites			
Variable	Mean	Median	Minimum	Maximum	Mean	Median	Minimum	Maximum
TOTACC/yr	1.13	0.50	0.00	15.13	1.05	0.63	0	6.88
INJACC/yr	0.45	0.25	0.00	5.13	0.57	0.38	0	4.13
AADT1	17002	12909	1902	74500	18933	15433	6500	57731
AADT2	449	206	1	9490	466	325	10	2500

Table A9 Summary Statistics for Type IV Sites for AADT Models

Variable	All Sites				Base Condition Sites			
	Mean	Median	Minimum	Maximum	Mean	Median	Minimum	Maximum
TOTACC/yr	1.6	1.0	0.0	10.8	1.3	0.6	0	5.4
INJACC/yr	0.7	0.4	0.0	4.5	0.7	0.4	0	3.3
F_1	15477	12950	2192	69521	18385	13865	2367	43167
F_2	552	420	10	7400	345	186	11	2625

Table A10 Parameter Estimates for Type III AADT Models
(model form given by Eq. 2.2)

Variable	TOTACC (s.e., *p*-value)	INJACC (s.e., *p*-value)
Intercept ($\ln \beta_0$)	-12.1332 (1.9357, 0.0000)	-15.2817 (2.2629, 0.0000)
F_1 (β_1)	1.0941 (0.1762, 0.0000)	1.3316 (0.2081, 0.0000)
F_2 (β_2)	0.2544 (0.0636, 0.0001)	0.2648 (0.0717, 0.0002)
K	0.3125	0.3074
Pearson product-moment correlation coefficient	0.67	0.63
MPB/yr	0.02	0.01
MAD/yr	0.60	0.32

Table A11 Parameter Estimates for Type IV AADT Models
(model form given by Eq. 2.2)

Variable	TOTACC (s.e., *p*-value)	INJACC (s.e., *p*-value)
Intercept ($\ln \beta_0$)	-14.9469 (1.5082, 0.0000)	-15.1858 (1.7442, 0.0000)
F_1 (β_1)	1.2826 (0.1398, 0.0000)	1.2513 (0.1688, 0.0000)
F_2 (β_2)	0.4671 (0.0779, 0.0000)	0.4535 (0.0811, 0.0000)
K	0.2070	0.1486
Pearson product-moment correlation coefficient	0.90	0.89
MPB/yr	0.00	-0.01
MAD/yr	0.48	0.29

MODELS WITH COVARIATES

For these the main model is judged best from the goodness-of-fit measures. Variant 1 is the next best alternative.

Table A12 Parameter Estimates for TOTACC Full Model: Type III

(model form given by Eq. 2.2)

Variables	Main Model Coeff. (s.e., *p*-value)	Variant 1 Coeff. (s.e., *p*-value)
Intercept (ln β_0)	-10.1914 (1.5232,0.0000)	-9.9214 (1.5100,0.0000)
F_1 (β_1)	0.8877 (0.1666,0.0000)	0.8509 (0.1665,0.0000)
F_2 (β_2)	0.3228 (0.0585,0.0000)	0.2972 (0.0590,0.0000)
COMDRWY1	0.0681 (0.0281,0.0154)	0.0912 (0.0276,0.0010)
VEI1	0.1081 (0.0556,0.0519)	0.1044 (0.0523,0.0461)
HAU	0.0101 (0.0059,0.0861)	0.0088 (0.0054,0.1014)
MEDWDTH1	-0.0106 (0.0060,0.0760)	
MEDTYPE1 (painted on main road)	-0.3209 (0.1771,0.0700)	
DRWY1	N/A[3]	
K	0.4229 (0.1064,0.0001)	0.4552 (0.1109,0.0000)
Pearson product-moment correlation coefficients	0.70	0.70
MPB/yr	0.09	-0.02
MAD/yr	0.84	0.88

Table A13 Parameter Estimates for INJACC Full Models: Type III

(model form given by Eq. 2.2)

Variables	Main Model Coeff. (s.e., *p*-value)	Variant 1 Coeff. (s.e., *p*-value)
Intercept (ln β_0)	-10.6443 (2.0474,0.0000)	-10.4453 (2.0845,0.0000)
F_1 (β_1)	0.8498 (0.2097,0.0001)	0.8260 (0.2146,0.0001)
F_2 (β_2)	0.2188 (0.0949,0.0212)	0.2460 (0.0901,0.0063)
COMDRWY1	0.0627 (0.0353,0.0756)	0.0607 (0.0346,0.0797)
HAZRAT1	0.1889 (0.0923,0.0407)	0.1897 (0.0930,0.0412)
HAU	0.0163 (0.0053,0.0021)	0.0168 (0.0054,0.0019)
PKTRUCK	-0.0253 (0.0135,0.0605)	-0.0331 (0.0186,0.0762)
PKTURN	0.0254 (0.0135,0.0592)	
PKLEFT		0.0333 (0.0188,0.0758)
K	0.5102 (0.1426,0.0003)	0.5178 (0.1437,0.0003)
Pearson product-moment correlation coefficients	0.66	0.64
MPB/yr	-0.05	-0.14
MAD/yr	0.43	0.47

Table A14 Parameter Estimates for TOTACC Full Models: Type IV

Variables	Main Model Coeff. (s.e., *p*-value)	Variant 1 Coeff. (s.e., *p*-value)
Intercept (ln β_0)	-7.4713 (1.8930,0.0001)	-7.4350 (1.6933,0.0000)
F_1 (β_1)	0.7350 (0.1849,0.0001)	0.7193 (0.1722,0.0000)
F_2 (β_2)	0.2390 (0.0926,0.0099)	0.2586 (0.0975,0.0080)
SDR2	-0.0003 (0.0001,0.0403)	-0.0005 (0.0001,0.0018)
PKTRUCK	-0.0479 (0.0110,0.0000)	
PKTHRU2	0.0249 (0.0085,0.0034)	0.0154 (0.0082,0.0591)
PKLEFT	0.0229 (0.0118,0.0525)	
PKLEFT1		-0.0158 (0.0083,0.0565)
MEDTYPE1 (painted on major roads)		-0.4027 (0.2084,0.0533)
MICHIGAN INDICATOR		0.4823 (0.2645,0.0683)
LTLN1S (0 or 1)		
K	0.4001 (0.0958,0.0000)	0.4382 (0.0965,0.0000)
Pearson product-moment correlation coefficients	0.77	0.75
MPB/yr	0.12	0.28
MAD/yr	1.16	1.20

Table A15 Parameter Estimates for INJACC Type IV: Full Models

Variables	Main Model Coeff. (s.e., *p*-value)	Variant 1 Coeff. (s.e., *p*-value)
Intercept (ln β_0)	-7.3927 (2.1279,0.0005)	-7.9801 (2.0870,0.0001)
F_1 (β_1)	0.5008 (0.2186,0.0220)	0.5670 (0.2145,0.0082)
F_2 (β_2)	0.3027 (0.1341,0.0240)	0.3452 (0.1213,0.0044)
SPD2	0.0289 (0.0145,0.0465)	0.0262 (0.0149,0.0795)
PKTRUCK	-0.0520 (0.0127,0.0000)	
PKLEFT1	0.0523 (0.0128,0.0000)	
SDR2		-0.0003 (0.0002,0.0420)
MEDTYPE1 (painted on major roads)		-0.5299 (0.2560,0.0385)
K	0.4671 (0.1296,0.0003)	0.5400 (0.1345,0.0001)
Pearson product-moment correlation coefficients	0.71	0.70
MPB/yr	0.05	-0.05
MAD/yr	0.65	0.67

ESTIMATION OF ACCIDENT MODIFICATION FACTORS

AMFs Derived from Full Models

One approach to deriving AMFs was to apply a model using the estimated parameter values from only statistically significant variables in accident prediction models. This approach suffers from correlation between geometric variables and traffic, and the difference in accident experience between sites is possibly due to the substantial unexplained variation resulting from omitted factors. Nevertheless, AMFs derived in this manner from the full models in Tables A12 to A15 are listed in Table A16.

Table A16 AMFs Derived from Type III, IV, and V Full Models

<table>
<tr><th rowspan="2">AMF</th><th colspan="2">Type III</th><th colspan="2">Type IV</th></tr>
<tr><th>Total</th><th>Injury</th><th>Total</th><th>Injury</th></tr>
<tr><td>COMDRWY1</td><td>Exp (0.0681COMDRWY1)</td><td>exp (0.0627COMDRWY1)</td><td rowspan="6">Not calibrated</td><td rowspan="6">Not calibrated[2]</td></tr>
<tr><td>VEI1</td><td>exp(0.1081VEI1)</td><td>Not calibrated</td></tr>
<tr><td>HAU</td><td>exp(0.0101HAU)</td><td>exp(0.0163HAU)</td></tr>
<tr><td>MEDWIDTH1</td><td>exp (-0.0106MEDWDTH1)</td><td rowspan="2">Not calibrated</td></tr>
<tr><td>MEDTYPE1[a]</td><td>0.73</td></tr>
<tr><td>HAZRAT1</td><td rowspan="10">Not calibrated</td><td>exp(0.1889HAZRAT1)</td></tr>
<tr><td>PKTRUCK</td><td>exp(-0.0253PKTRUCK)</td><td>exp(-0.0479PKTRUCK)</td><td>exp(-0.0520PKTRUCK)</td></tr>
<tr><td>PKTURN</td><td>exp(0.0254PKTURN)</td><td>Not calibrated</td><td rowspan="3">Not calibrated</td></tr>
<tr><td>PKTHRU2</td><td rowspan="7">Not calibrated</td><td>exp(0.0249PKTHRU2)</td></tr>
<tr><td>PKLEFT</td><td>exp(0.0229PKLEFT)</td></tr>
<tr><td>PKLEFT1</td><td>Not calibrated</td><td>exp(0.0523PKLEFT1)</td></tr>
<tr><td>SDR2</td><td>exp(-0.0003SDR2)</td><td rowspan="4">Not calibrated</td></tr>
<tr><td>LIGHT</td><td rowspan="3">Not calibrated</td></tr>
<tr><td>HEICOM</td></tr>
<tr><td>HEI2</td></tr>
</table>

[a] Medtype1: Painted

198

AMFs Derived from Regression Models

The AMFs were derived using the third of the three methods listed in Section 5.5 and are given in Tables A17 and A18. These AMFs apply to total accidents and total injury accidents.

Table A17 AMFs for Type III Sites

AMFs	Recalibrated (TOTACC)	Recalibrated (INJACC)
Left lane on major road	0.71 (One approach)	1 (One approach)
Right lane on major road	1 (One approach)	1 (One approach)
Sight distance	1	1
Intersection angle	1+(0.016*SKEW[a])/ (0.98[b]+0.016*SKEW)	1+(0.017*SKEW[a])/ (0.52[c]+0.017*SKEW)

[a] SKEW = intersection skew angle (degrees), expressed as the absolute value of the difference between 90 degrees and the actual intersection angle

[b] 0.98 = mean of the observed TOTACC accidents per year of the sites meeting no angle, no right lane, and presence of left lane

[c] 0.52 = mean of the observed INJACC accidents per year of the sites meeting no angle, no right lane, and presence of left lane

Table A18 AMFs for Type IV Sites

AMFs	Recalibrated (TOTACC)		Recalibrated (INJACC)	
	One Approach	Both Approaches	One Approach	Both Approaches
Left-turn lane on major road	1	1	0.86	0.74
Right-turn lane on major road	1	1	1	1
Sight Distance	1		1	
Intersection Angle	1+(0.053*SKEW[a])/ (1.43[b]+0.053*SKEW)		1+(0.048*SKEW[a])/ (0.72[c]+0.048*SKEW	

[a] SKEW = intersection skew angle (degrees), expressed as the absolute value of the difference between 90 degrees and the actual intersection angle

[b] 0.43 = mean of the observed TOTACC accidents per year of the sites meeting no angle, no right lane, and presence of left lane

[c] 0.72 = mean of the observed INJACC accidents per year of the sites meeting no angle, no right lane, and presence of left lane

Comparison of AMFs

A comparison of AMFs from the two methods is shown in Tables 19 and 20. For Type III and IV intersections, intersection angle (SKEW) was estimated as significant in the regression models. Right-turn lanes on major roads provided significant AMFs for Type IV intersections.

Table A19 Comparison of Type III and IV AMFs for TOTACC

AMF	AMFs Derived From Full Models		AMFs Derived From Regression Models	
	Type III	Type IV	Type III	Type IV
SKEW	exp(0.010SKEW)	*Not calibrated*	1+(0.016*SKEW)/ (0.98+0.016*SKEW)	1+(0.053*SKEW)/ (1.43+0.053*SKEW)
RT MAJ	*Not calibrated*		1	1
LT MAJ			0.71	1
SIGHT DISTANCE			1	1
COMDRWY1	exp (0.0681COMDRWY1)		*Not calibrated*	*Not calibrated*
VEI1	exp(0.1081VEI1)			
MEDWIDTH1	exp (-0.0106MEDWDTH1)			
MEDTYPE1	0.73			
PKTRUCK	*Not calibrated*	exp (-0.0479PKTRUCK)		
PKTHRU2		exp (0.0249PKTHRU2)		
PKLEFT		exp (0.0229PKLEFT)		
SDR2		exp(-0.0003SDR2)		

Table A20 Comparison of AMFs for INJACC

<table>
<tr><th rowspan="2">AMF</th><th colspan="2">AMFs Derived From Full Models</th><th colspan="2">AMFs Derived From Regression Models</th></tr>
<tr><th>Type III</th><th>Type IV</th><th>Type III</th><th>Type IV</th></tr>
<tr><td>SKEW</td><td>exp(0.0163SKEW)</td><td rowspan="10">Not calibrated</td><td>1+(0.017SKEW)/ (0.52+0.017SKEW)</td><td>1+(0.048SKEW)/ (0.72+0.048SKEW)</td></tr>
<tr><td>RT MAJ</td><td rowspan="7">Not calibrated</td><td>1</td><td>0.86 one approach, 0.74 both approaches</td></tr>
<tr><td>LT MAJ</td><td>1</td><td>1</td></tr>
<tr><td>RT MIN</td><td>1</td><td>1</td></tr>
<tr><td>SIGHT DISTANCE</td><td>1</td><td>1</td></tr>
<tr><td>HI</td><td rowspan="11">Not calibrated</td><td rowspan="11">Not calibrated</td></tr>
<tr><td>DRWY1</td></tr>
<tr><td>MEDIAN</td></tr>
<tr><td>COMDRWY1</td><td>exp (0.0627COMDRWY1)</td></tr>
<tr><td>HAZRAT1</td><td>exp (0.1889HAZRAT1)</td></tr>
<tr><td>PKTRUCK</td><td>exp (-0.0253PKTRUCK)</td><td>exp (-0.0520PKTRUCK)</td></tr>
<tr><td>PKTURN</td><td>exp(0.0254PKTURN)</td><td>Not calibrated</td></tr>
<tr><td>PKLEFT1</td><td rowspan="4">Not calibrated</td><td>exp (0.0523PKLEF)</td></tr>
<tr><td>SDR2</td><td rowspan="3">Not calibrated</td></tr>
<tr><td>LIGHT</td></tr>
<tr><td>HEI2</td></tr>
</table>

REFERENCES

Washington, S., B. Persaud, C. Lyon, and J. Oh, Validation of Accident Models for Intersections. FHWA-RD-03-037, Federal Highway Administration, Washington, D.C., 2005.

Vogt A. Crash Models For Rural Intersections: Four-Lane by Two-Lane Stop-Controlled and Two-Lane by Two-Lane Signalized. FHWA-RD-99-128, FHWA, U.S. Department of Transportation, Washington, D.C., 1999.

APPENDIX B

Survey Instrument

SURVEY ON THE PREDICTION OF THE SAFETY PERFORMANCE FOR MULTILANE RURAL HIGHWAYS
NCHRP Project 17-29

You are being asked to participate in this short survey because your agency has shown, through a previous survey conducted for NCHRP 17-26 (Predicting the Safety Performance of Urban and Suburban Arterials) an interest in the development and application of the *Highway Safety Manual* (HSM). In addition, your agency has indicated that relevant databases are available for research purposes. In NCHRP Project 17-29, the Texas Transportation Institute (TTI) is leading an effort to develop a methodology to predict the safety performance of multilane rural highways.

The intent of the survey is to determine a more accurate evaluation of input variables available in your agency as well as the potential input variables that would be of interest for predicting the safety performance of rural multilane highways located within your jurisdiction. A definition of a rural multilane highway is provided below.

The Transportation Research Board (TRB) and the National Cooperative Highway Research Program (NCHRP), which is managed by TRB and cosponsored by the Federal Highway Administration (FHWA) and the American Association of State Highway and Transportation Officials (AASHTO), are working to create the *Highway Safety Manual* (HSM). The HSM will organize knowledge about highway safety for application by highway agencies and will include procedures to predict the quantitative safety performance of highways and streets, such as the *Highway Capacity Manual* (HCM) predicts the quantitative operational performance of highways and streets. A prototype HSM Chapter, for the prediction of the safety performance of two-lane rural roads, is available in Appendix C of http://gulliver.trb.org/publications/nchrp/nchrp_w62.pdf.

The results you provide to this survey are critical in the development of the HSM. Thus, your responses to the following questions are greatly appreciated.

CURRENT SAFETY PREDICTION METHODS

The following definition of rural multilane highways guides this research and should also help to identify the specific roadway segments of interest in this survey:

The research team has decided to use the guidelines proposed by FHWA and AASHTO. These guidelines define rural areas as places outside the boundaries of urban places where the population is less than 5,000 inhabitants. Consequently, a highway for the NCHRP 17-29 project will be classified as rural when it is located outside the city limits of the urban agglomeration above 5,000 inhabitants. Given this definition, we greatly appreciate if you could answer the questions below.

1. Does your agency use, are you developing, any statistical methods or other methods to predict or estimate the safety performance of multilane rural highways?

   ___YES___NO
   If YES, please
   (a) Indicate what the safety performance estimates or predictions are used for

   (b) describe or attach a copy of your agency's models or predictive methodology:

INPUT VARIABLES OF INTEREST TO PREDICT THE SAFETY PERFORMANCE OF MULTILANE RURAL HIGHWAYS

One of the main purposes of the proposed methodology is to enable designers and planners to quantify the effects of relevant design and operational variables on safety performance in order to assess the safety implications of design decisions or planned improvements. In developing the predictive methodology for estimating the safety of multilane rural highways, candidate input variables will be selected based on known relationships of specific variables to safety, relationships developed as part of this research, and the priorities of potential HSM users. We would appreciate your assessment of which design, weather, and operational variables should be included in the safety performance prediction process. Please rate the following variables in terms of their potential value for your agency for inclusion in the predictive methodology [5=high priority; 1=low priority]. <u>Please use the full range of ratings from 1 to 5 so that your ratings are useful for setting overall priorities; if your assessment is that a particular input variable is not needed in the first edition of the HSM, please use a rating of 1</u>. In assigning your ratings, please DO NOT consider the availability of data in your agency's databases.

Candidate Input Variables for Rural Multilane Roadway Segments

______ delineation
______ design or posted speed
______ grades
______ horizontal curves
______ illumination
______ land-use adjacent to traveled-way
______ lane widths
______ median type
______ median width
______ number and type of median openings
______ number and type of driveways

- ______ number of through lanes
- ______ pavement friction
- ______ raised pavement markers
- ______ roadside design/clear zones/roadside objects
- ______ roadside distractions (e.g. billboards, signage, etc.)
- ______ shoulder width/curb type
- ______ shoulder rumble strips
- ______ spacing between driveways
- ______ speed variance of vehicular traffic
- ______ traffic volume (AADT) (veh/day)
- ______ traffic volume in peak period (veh/hr)
- ______ traffic volume for different time periods (average veh/hr)
- ______ traffic volumes for individual driveways
- ______ vehicle mix (e.g., percent trucks)
- ______ vehicle speed (average for different time periods)
- ______ vertical curvature
- ______ wet pavement
- ______ ice on pavement
- ______ snow on pavement
- ______ visibility restrictions (e.g. fog, glare, etc.)

Candidate Input Variables for At-Grade Intersections

- ______ approach speed (observed)
- ______ approach speed (posted or design speed)
- ______ horizontal alignment of intersection approaches
- ______ illumination
- ______ intersection sight distance
- ______ intersection skew angle
- ______ lane widths on intersection approaches
- ______ level of service (LOS) (only at signalized intersections)
- ______ median type/presence of median
- ______ number of intersection legs
- ______ number of through lanes on intersection approaches
- ______ number and length of added through lanes at intersections
- ______ presence/number of left-turn lanes
- ______ presence of right-turn lanes
- ______ shoulder/curb type on intersection approaches
- ______ shoulder/curb width on intersection approaches
- ______ signal phasing (e.g., left-turn phasing)
- ______ signal timing
- ______ signal visibility
- ______ spacing between intersection and nearby driveways
- ______ type of traffic control
- ______ traffic volumes (AADTs) for major- and minor-road legs (AADTs)
- ______ type of left-turn channelization (painted vs. raised curb)

_____ vehicle mix (e.g., percent trucks)
_____ weather variables as previous

2. Are there other potential input variables, not listed above, that you think should have a high priority for inclusion in the safety prediction methodology for multilane rural highways?

DATA AVAILABILITY

The data availability questions are directed specifically to representatives of public agencies that operate and maintain roadways.

3. For multilane rural highways under your agency's jurisdiction, does your agency have computerized files of:

Crash data (i.e., records for each individual crash) ___YES___NO

Roadway segment inventory data (i.e., geometrics and traffic control for roadway segments between intersections) ___YES___NO

If YES, approximately how many miles of multilane highways is under your organization's jurisdiction? _______ miles

Intersection inventory (i.e., geometrics and and traffic control for each individual intersection) ___YES___NO

If YES, could you provide approximately how many signalized and unsignalized intersections are part of the multilane rural highways? _______

4. In your agency's computerized accident data, can driveway-related crashes be distinguished from other non-intersection crashes? ___YES___NO

5. Do your agency's computerized data files use a common location reference system to allow direct linking of:

Individual crash records to the inventory data for the roadway segment on which the accident occurred ___YES___NO

Individual crash records to the inventory data for the intersection at which the crash or the intersection to which the crash is related ___YES___NO

6. Can your agency's computerized databases provided georeferenced data (i.e., the longitude and latitude coordinates based on the Geographical Information System)?

   Crash data (i.e., records for each individual crash) ___YES___NO
   Road inventory (i.e., location of intersections, bridges, land-use characteristics, etc.) ___YES___NO

7. Can your agency provide vehicular traffic information for rural signalized and unsignalized intersections?

   Major approaches (i.e., approaches located on the main highway under your jurisdiction) ___YES___NO

   Minor approaches (i.e., approaches connected to the main highway under your jurisdiction) ___YES___NO

8. Can your agency provide before-after crash and traffic data for a significant number of rural multilane segments or intersections sites in which any of the candidate variables were changed (e.g. a left turn lane was added or shoulders were widened)? If so, please identify those variables and provide approximate details (if possible) on number of segments/miles or intersection sites and number of before/after years of data.

9. Does your agency have access to daily weather related information, such as rain, ice, and snow events?

CONTACT PERSON

10. May we have the name of an individual in your agency that we can contact for further information, if necessary?

Name: ____

Title: ____

Agency: ____

Address: ____

____

____

Phone: ______________________ Fax: _______________

E-mail: ____________________

Thank you very much for your assistance. Please send the completed survey to:

Dominique Lord
Associate Research Scientist
Texas Transportation Institute
3135 TAMU
Texas A&M University
College Station, TX
77843-3135

E-mail: d-lord@tamu.edu

APPENDIX C

Expert Panel Review: Critique and Assessment of the Expert Panel Process

This appendix was originally prepared by Dr. Simon P. Washington.

Purpose of Critique and Assessment

Expert panels have been used extensively in the development of the Highway Safety Manual (HSM). These panels have been used to extract useful research information from highway safety experts, research that is often inconsistent in the literature, that is relatively scant, or that for one reason or another was not conducted under ideal conditions. While the panels have been used to recommend agendas for new and continuing research, their primary role has been in the development of Accident Modification Factors (AMFs)—quantitative relationships between highway safety and various highway safety treatments. Because the expert panels derive quantitative relationships and because these relationships have potentially enormous impact on future highway safety investment decisions, the expert panel process and its viability needs to be assessed. The need for this review should not come as a surprise, as scientific evaluation of analytical and quantitative methods employed in highway safety is a routine activity employed in the profession.

The remainder of this report first provides a background of the role that expert panels serve in the development of the HSM, how they have been used, and some background on expert panels used in other transportation applications. The next section describes the review process in detail, so that the interested reader can understand how these expert panel reviews are conducted and learn how decisions are derived within these expert panels. Then, important questions surrounding the accuracy and precision of expert panel findings are discussed, serving as a critique of the expert panel process. Conclusions and Recommendations identify areas of potential improvement to be considered in future expert panels and highlight strengths of the existing process.

Background

The NCHRP 17-29 research program focused on the scientific evaluation of an array of crash prediction methodologies and analytical methodologies currently being integrated into mainstream products that will be used to evaluate and improve highway safety. Due to the extreme importance of these methods on highway safety and ultimately human lives, it is vital to scrutinize and evaluate methodologies employed in the profession. One procedure that has received little attention in the literature that plays a fairly significant role in highway safety is the expert panel. Expert panels have been and are being used extensively in various research programs related to and in support of the HSM development, and deserve the careful scrutiny that other methods receive.

The AMF is a quantitative measure of safety that is integral to the HSM and to the Interactive Highway Safety Design Model (IHSDM). The AMF is a safety performance factor or function that relates the safety of a highway with a specific countermeasure or treatment. The general analytical approach that incorporates the AMF makes use of a ‘baseline’ or ‘base’ model which predicts crashes for sites of interest (e.g., rural roads, intersections of rural roads, multi-lane highways, etc.) based on exposure (AADT or VMT) and perhaps one or two other factors (e.g., whether the facility is in an urban or rural area). A calibration factor is applied to correct for differences across regions (e.g., cities, states) so as to make predictions applicable to the local jurisdiction. Crash modification factors or functions are then applied to obtain estimates of the reduction in crashes after installation of a particular countermeasure or treatment. In simplified form, the AMF approach is given as:

$$\text{Crashes}_{\text{After}} = \text{Crashes}_{\text{Baseline}} \text{ x } \text{AMF}_{\text{Countermeasure i}} \text{ x Regional Calibration Factor,}$$

where $\text{Crashes}_{\text{After}}$ are the predicted crashes at a site (after application of countermeasure i) given the product of baseline predicted crashes, the expected reduction in crashes from countermeasure *i*, and a regional calibration factor. The AMF can be a simple factor (e.g., 0.80) or be a function of other variables such as AADT. Of course, different countermeasures will have different effects on crashes, and may have effects that are limited to certain crash types such as rear-end crashes, angle crashes, etc. Thus, the AMF can range from a simple factor to a function by crash type. For example, the expected reduction in crashes as a result of installation of a cable median barrier device may be a function of average vehicle speeds and may apply only to cross-median crashes.

There is solid support for the AMF and AMF approach, as described in Lyon et al. (2003):

> "After detailed examination of the data obtained across several states and time periods, the approach proposed in the IHSDM appears to be a sound and defensible approach for forecasting crashes. The approach offers two considerable technical advantages over conventional 'full model' approaches for forecasting crashes. First, the often high intercorrelation of explanatory variables with traffic volumes renders isolation of the safety effects of individual variables difficult at best, leading to inconsistent predictions. The algorithm approach skirts this issue by allowing only traffic volumes to be statistically associated with crashes, and by using AMFs derived independently of the prediction model. Second, corrections for driver populations, weather, environmental, and other factors, which are often hard to capture and are inter-correlated as well as correlated with traffic volume, are treated with a correction factor."

The conclusions by Lyon et al. arose from a detailed and extensive analysis of the safety effects of various geometric and traffic factors for rural intersections across several states. With additional details provided in a companion study (Oh et al., 2003), the authors identified omitted known variables, omitted unknown variables, site-selection bias, countermeasure-selection bias, poorly measured and surrogate variables, and model functional forms as factors that contribute to the difficulty in estimating suitable 'full' regression models that related traffic and geometric features to safety, further bolstering support for base models with AMFs.

The current practice involves the estimation of AMFs through the use of expert panels or through expert opinions. The use of expert opinions in transportation safety applications is not new. Dissayanake and Lu (1999) relied on expert opinions to assess the safety needs of special populations. European experts were convened to determine the most effective and reliable air traffic management system for improving safety, operational, and environment performance (Zografos and Giannouli, 2001). The required characteristics of cockpit weather information systems for NASA's Aviation Safety Program, including some intelligent transportation system technologies, were derived in part from the collection of expert opinions (Sireli et al., 2002). Fukuoka (2002) developed a unified reliability and analysis environment for assessing European railway network safety using expert opinions and information about failures. Recently and perhaps of most relevance to this research, Harwood et al. (2000) developed algorithms intended for use in the Interactive Highway Safety Design Model which include AMFs that are derived from expert opinions. In their approach, point estimates of AMFs were derived from a collection of expert opinions and were used to adjust baseline model predictions to estimate the impact of various countermeasures—as described previously. Little has been written on the exact procedure used to derive AMFs, and their derivation has not been as of yet scientifically scrutinized (like for example the evaluation of the AMF and base model approach cited previously).

The expert panel approach—step by step

Prior to critiquing and assessing the expert panel approach used in many HSM applications it is important to document and describe the process. The NCHRP 17-29 team attended and evaluated an expert panel meeting (June 28th through 30th, 2006) focused on the development of AMFs for multi-lane rural highways. The expert panel process was the same as that used to assess AMFs in prior and subsequent HSM-related research. While three team members attended the meeting and focused on the substantive matters, one team member focused on the expert panel review process itself, making notes of interactions, procedures, and deliberations. The following steps describe in detail the expert panel review process as observed during the expert panel meeting. The discussion is generalized so as to make the process generic.

Step 1: Identify expert panelists

The expert panel consists of nationally recognized experts in the subject matter of interest. It is extremely important that a substantial number of leading researchers be assembled to conduct the expert panel review. While there is no 'magic' number of experts, a panel that is too small may not represent the collective set of views in the profession, while a panel too large might be unwieldy to manage and reach consensus on AMF factors and functions. A number between 10 and 15 experts appears to be an appropriate range to satisfy the need to be representative and manage the tasks charged to the panel in a reasonable amount of time. A number of specific panel member attributes are needed:

1. Representation from experts in analytical methods and experimental design as applied to transportation safety, and preferably to the substantive area of interest. Preferably, the experts should have diverse backgrounds, with a mix of people coming from the academia, the consulting business, and public transportation agencies.

2. Geographical representation, so that the collective experience of the experts present can speak to the safety needs of various national stakeholders groups, including rural, urban, eastern, western, and mountainous regions of the U.S.
3. Specific subject-matter experience is needed in the substantive area being studied. It should not be surprising that selected experts also have authored a disproportionate number of the research studies that are discussed during the expert panel review. While this may create a potential conflict of interest, it cannot be avoided and efforts can be made during the panel proceedings if such a conflict creates a problem.

Step 2: Set panel meeting date and prepare supporting panel materials

A panel meeting date must be set when all expert panelists can attend. The expert panel can take between 2 and 4 full days of deliberation and so a comfortable meeting room with snacks, amenities (restrooms, phones, and internet access) is essential to support a quality meeting. Meeting minutes are needed and either transcribed from recordings or recorded by a meeting secretary. Also, a computer projector and flip charts are needed to support the decision-making and consensus building process.

A critical and significant undertaking at this stage is the preparation of materials used to support the expert panel review. Typically, this task is undertaken by the funded group or team conducting the expert panel review. The essential product of this task is to compile copies and/or summaries of all the completed and relevant research related to the countermeasures to be discussed by the expert panel. This compilation consists of all relevant and available peer-reviewed research and research summaries conducted nationally and internationally (if relevant) by countermeasure (e.g., all peer-reviewed research on replacing yield with stop signs in rural areas). In many cases, surprisingly little peer-reviewed research is available relevant to the objectives of the expert panel. A typical expert panel review may require between 15 and 25 countermeasures to be evaluated. This countermeasure list is circulated to the panel experts prior to the compilation of the materials, to make sure that important countermeasures have not been omitted.

Ideally, the binder of relevant research is assembled substantially in advance of the expert panel meeting and is distributed to all panelists for their review prior to the meeting. This binder becomes a pivotal tool in the expert panel review process, and also serves as an important reference prior to, during, and after the expert panel meeting.

Finally, experts are assigned a specific section of the binder to read in detail and asked to be prepared to discuss the material during the expert panel meeting. If there is a large number of topics, an expert may be assigned a set of countermeasures to review by his or herself, whereas a small number of countermeasures may result in overlap among experts. These experts are expected to summarize the research objectively during the expert panel meeting.

Step 3: Conduct expert panel meeting

By the time the expert panel is convened, a comprehensive list of countermeasures has been endorsed by the panel and a binder of all relevant and available research has been compiled, summarized, and distributed to the experts for prior review. In addition to their subject-matter experience, all experts should arrive at the expert panel meeting having reviewed the materials.

An agenda is determined and distributed that follows a logical sequence for discussing the countermeasures by group. For example, roadside countermeasures may make one group, whereas signing and striping may constitute another. The grouping of course depends on the subject matter. Finally, usually there is some hierarchy assigned to the groups, with 'first order' countermeasures being discussed first, 'second order' being discussed second, etc. The hierarchy may be determined by the availability of literature on the subject (assumed to be proportional to its importance), with high importance discussed first; or the speculated magnitude of the countermeasure effect (often correlated with the importance), with high-magnitude effects discussed first; or by the controversiality of the countermeasures, with less controversial countermeasures discussed first. Often these three hierarchies are related to one another (i.e., less importance is associated with less research which is associated with more controversy) and so the decision on which countermeasures to discuss first usually is not a difficult one.

A relatively unstructured open discussion technique is applied, with a designated moderator leading the general discussion. Countermeasures are discussed along with the research results for each countermeasure. The expert review panel's goal is to derive a 'weighted average' AMF factor or function through interactive and open discussion of relevant research and by assigning relevance weights to all of the relevant research. Although the weights are not explicitly (no numbers assigned) or even objectively determined through this process (e.g., ballots), discussion continues until consensus is reached. For one countermeasure, this process may take anywhere between 20 minutes and 3 hours, and is generally not time-constrained in any way—the panel deliberates until an AMF factor or function is agreed upon. It is important to note that an AMF of 1.0 (no effect whatsoever) and the lack of a suitable AMF are outcomes on which the panel may reach consensus. The analytical equivalent of the expert panel is a meta-analysis, although a meta-analysis is significantly more formalized and based on step-by-step procedures whereas the expert panel is based more on consensus building.

A number of important issues are discussed with respect to summarizing the available peer-reviewed research, typically quite systematically:

1. *Relevance of the research to the application being discussed*. For example, was the research conducted in an urban environment when a rural countermeasure is being sought? Was the research conducted on mountainous terrain when flat terrain is the setting of interest? Typically these questions of relevance surround issues of traffic exposure, driving population (e.g., country in which research was

conducted), range of conditions examined, and similarity of 'non-countermeasure' traffic controls.

2. *Timeliness of the research*. The age of the research and its affect on changes in relevance as regards road users, analysis methods, vehicle safety, and injury reporting thresholds is often relevant for discounting the relevance and weighting of research.
3. *Non-ideal conditions of the research design*. The research conditions that may lead to incorrect or weak conclusions such as omitted important variables, included irrelevant variables, endogeneity of variables, inappropriate analysis methods, or sampling procedure are discussed, with research studies conducted under non-ideal conditions typically down weighted.
4. *Sample size and sample representativeness*. Studies with large samples typically are given greater weight than studies using small samples, all else being equal. In addition, studies with greater sampling representativeness (heterogeneity) of the population are given greater weight than studies conducted on more limited or biased samples.
5. *Findings and conclusions of the research*. The conclusions of research are often viewed to make sure the expert panel arrives at the same conclusions as the study authors. While some of the previously listed issues may attract greater attention, studies where authors over- or mis-stated the conclusions are scrutinized.

The expert panel systematically discusses these various aspects of relevant research and goes through the pre-organized list of countermeasures one by one. The session recorder takes notes, records, and otherwise keeps track of conclusions that are drawn regarding all of the AMFs and AMF functions. All of the details necessary to derive an AMF or AMF function are decided in this meeting, such as the limits of the function, the shape of the function, and any non-linearities, spikes, or humps. In the majority of cases, a computer and computer projector are used so the AMFs can be shown during the meeting and revised to reflect consensus.

Step 4: Disseminate Results

The results of the session are distributed to panel members for review and comment. This final step is conducted to make sure all events and decisions made were captured and are reflected accurately in the AMF factors and functions. After panel members have provided comment, AMFs are described and detailed in a document intended for broader dissemination.

Critique of Expert Panel Process

With theoretical support for the AMF analytical approach and an established history of appropriate uses of expert opinions and panels, the use of expert panels is likely to continue into the near future.

There remain, however, some important questions that need to be addressed regarding the derivation of AMFs via expert panels.

1. Are the results derived from expert panels precise and/or accurate?

2. Can expert panels be used to derive estimates of uncertainty?
3. Do results across expert panels differ, and if so, how?
4. What guidance can be provided to expert panels to ensure repeatable and accurate results?
5. Should expert panels follow informal procedures (as they have been) or more formal expert panel procedures, such as the Delphi method?

These questions (and perhaps others), which raise issues regarding the scientific credibility and use of expert derived AMFs are now addressed in turn. An attempt is made to identify how deficiencies might be tested and/or addressed.

Are the results derived from expert panels accurate and precise?

Precision and accuracy of AMFs via expert panels are difficult to assess. If statistical accuracy and precision criteria are applied, then expert panels would need to be repeated numerous times in order to compute the relevant statistics. It is unlikely that this kind of controlled expert panel evaluation will be conducted given the enormous resources that would be required.

The answer to this question hinges upon the repeatability of experts in deriving an AMF factor or function given the relevant literature of information. Research by Melcher et al. (2001) showed that experts agreed with one another in evaluating safety countermeasure effectiveness. The researchers derived the AMFs in this study in a different manner than those derived in HSM expert panels, the differences of which are worthy to note. First, in the Melcher study, experts derived AMF factors independent of each other. Second, experts were given a random sample of crashes in order to examine the effectiveness of a set of countermeasures rather than a summary of literature findings. Third, multiple observations across experts were tallied to derive means and variances of the AMFs. Finally, only AMF factors were considered (no AMF functions were derived).

The most obvious practical differences between the expert-panel derived AMFs and the Melcher derived AMFs are independence and the explicit estimation of precision. The verbal and non-verbal interaction that occurs within an expert panel is likely to influence the opinions of some experts. Because consensus is one of the aims of the expert panel, precision is under-estimated. In other words, independent experts are likely to disagree more than experts in a consensus building exercise. It is quite reasonable to speculate, however, that the accuracy of the two approaches will be similar—and will be bounded by findings in the literature and represent some notion of a mean, median, or mode of AMFs as represented in the literature. Moreover, precision of AMFs is not explicitly estimated in the expert panel approach, and for all intensive purposes would reflect both the degree to which consensus was established and the uncertainty in the literature. As a result, it is unlikely that the scale of precision would be reliable across countermeasures. For example, countermeasures given ample discussion time may have smaller precision, whereas countermeasures discussed prior to when the panel is fatigued may have larger variance estimates. Again, the current HSM expert panel practice does not explicitly produce variance estimates and so this discussion is hypothetical.

To synthesize these results, it is quite likely that the accuracy of expert panel derived AMFs is quite acceptable and that experts will produce an AMF factor or function that is useful in practice and represents close to a mean, median, or mode AMF factor or function on the subject. For estimating precision of AMFs, in contrast, expert panels are not as reliable as methods that poll or query experts independently.

Can expert panels be used to derive estimates of uncertainty?

As currently practiced in the HSM development, uncertainty in an expert panel represents disagreement among experts. Since disagreement may become contentious and/or confrontational, especially when experts may be authors of contrasting research studies being studied (and conflict of interest is possible), it is not perhaps the most objective way of producing uncertainty estimates. Expert panels could arise, for example, where most countermeasure discussions are non-confrontational but several are, threatening objectivity and comparability of the derived AMFs.

It is possible, however, to modify the existing expert panel process to poll experts prior to the consensus building process to derive estimates of uncertainty. This slight modification, if applied consistently and in a structured way, could be used to develop reliable and objective precision estimates.

Do results across expert panels differ, and if so, how?

It is possible that different expert panels would produce different AMF factors or functions. As discussed previously, however, these expert panels are not likely to produce AMFs that are substantially different. As in all research endeavors, a robust expert panel process should not conclude with one expert panel, but will be improved with future expert panels refining and updating AMFs from previous panels. Thus, the expert panel process and the AMF factors and functions that result should be continually refined and improved with future expert panels.

What guidance can be provided to expert panels to ensure repeatable and accurate results?

Structure and formality of the expert panel procedures will yield repeatability. There is considerable structure already included in the expert panel process as described previously, yielding what are considered to be accurate AMFs. To improve the process, however, there is room for increased formality, particularly when it comes to developing estimates of AMF precision and to address potential problems that result from group dynamics.

Should expert panels follow informal procedures (as they have been) or more formal expert panel procedures such as the Delphi method?

The Delphi method for polling experts has been shown to produce forecasts that are more accurate than unstructured groups of experts (Green et al., 2007; Rowe and Wright, 1999; Rowe and Wright, 2001). The Delphi method makes use of questionnaires in two or more rounds of independent polling of panel experts. A facilitator is used to help reach consensus on a forecast (e.g., an AMF factor or function) so that a group of experts may converge on an accurate answer. The Delphi method rests on the following principals:

Structured information flow: Unstructured expert panels suffer from the inclusion of irrelevant information and problems associated with group dynamics. In the Delphi method, the initial contributions from the experts are collected via questionnaires, along with open-ended comments to their answers. The panel facilitator controls the interactions among the participants by summarizing the information anonymously and filtering out irrelevant content. This procedure purportedly avoids many of the negative effects of face-to-face panel discussions and solves the usual problems of group dynamics (Rowe and Wright, 1999, 2001).

Regular feedback: In the Delphi process, participants comment on their forecasts, the responses of other experts, and on the progress of the panel as a whole. There are various opportunities to revise their earlier statements, and these revisions are done anonymously. These revisions are in contrast to unstructured and interactive group meetings, whereby participants tend to stick to previously stated opinions and often conform too much to the group leader. It is for these reasons that the Delphi method is believed to lead to more accurate and objective forecasts.

Anonymity of the participants: In the Delphi method, all expert panelists maintain anonymity throughout the expert panel review process. Their identity is not revealed even after completion of a final report or product. The anonymity purportedly prevents expert panelists from dominating others in the consensus building process by using their authority or personality, frees panelists (to some extent) from their personal biases, and minimizes the “bandwagon” or “halo effect” as discussed previously. The method allows experts to freely express their opinions, and encourages open critique and the revision of prior judgments given the current group consensus.

Whether or not the current expert panel process used in support of HSM development morphs into a process akin to the Delphi method depends upon the goal of future panels, the willingness of HSM organizers to revisit the expert panel process, and the professional communities’ acceptance of the current HSM expert panel process. It is clear, however, that the Delphi method has been shown to produce more objective forecasts than unstructured panels. Future expert panels used to illicit highway safety AMFs—after the first edition of the HSM is produced—are well-advised to consider the positive attributes of the Delphi method when considering possible modifications to the process.

Conclusions

The NCHRP 17-29 research team attended an expert panel meeting on June 28th through 30th, 2006 with the explicit objective to evaluate the expert panel process. Upon attending this meeting and through experiences conducting numerous other expert panel evaluations, the following conclusions are drawn:

1. The current HSM expert panel process, with all its strengths and weaknesses, is being consistently applied. Consistency is one hallmark of a credible scientific process. In addition, breaking consistency is detrimental to any scientific method. Thus, any

changes and/or enhancements made to the current expert panel process should be considered after completion of the first edition of the Highway Safety Manual.
2. The HSM expert panels do not currently derive precision estimates of AMF factors and functions, and there is a need to have such information.
3. The expert panels clearly agree on the mission of the panel—to derive 'the best' and most reliable estimates of AMFs and AMF functions. Persons are selected with this mission in mind and with a track record of conducting scientific research in the subject areas. Thus, although estimates are subjectively derived, all of the participants are intimate with objective procedures for deriving estimates.
4. It is quite likely that the accuracy of expert panel derived AMFs is quite acceptable and that experts will produce an AMF factor or function that is useful in practice and represents close to a mean, median, or mode AMF factor or function on the subject.
5. For estimating precision of AMFs, in contrast, expert panels are not as reliable as methods that poll or query experts independently. It is possible to modify the existing expert panel process (after the first edition of the HSM) to poll experts prior to the consensus building process to derive estimates of uncertainty, or to develop a hybrid Delphi process. This slight modification to current practice, if applied consistently and in a structured way, could be used to develop reliable and objective precision estimates.
6. The expert panel process and the AMF factors and functions that result should be continually refined and improved with future expert panels.
7. Whether or not the current expert panel process used in support of HSM development morphs into a process akin to the Delphi method depends upon the goal of future panels and the professional communities' acceptance of the current process. It is clear, however, that the Delphi method has been shown to produce more objective forecasts than unstructured panels. Future expert panels used to illicit highway safety AMFs—after the first edition of the HSM is produced—are well-advised to consider the positive attributes of the Delphi method when considering possible modifications to the process.
8. The use of the Delphi process would enable the expert panel to avoid a physical meeting which reduces the logistics and cost burden of expert panel meetings considerably.

There are two overall recommendations as a result of the HSM expert panel review. First, the current HSM expert panel process should be revisited upon completion of the first edition of the HSM. No changes should be made to the expert panel process prior to completion of the current HSM edition, and it is believed that the current expert panel process will produce reliable and quite reasonable AMF factors and functions. Existing shortcomings are lack of reliable precision estimates of the AMFs, possible complications arising from interactions and group dynamics, and possible forecasting bias as a result. It may be possible to develop a hybrid expert panel process that utilizes the strengths of the existing HSM expert panel process and the Delphi method.

Second, a comparison of the existing HSM expert panel process and the Delphi method should be conducted. To accomplish this, a panel of say 16 experts could be selected and randomly assigned either to the Delphi or HSM expert panel. The two expert panel approaches should be conducted on a very limited set of countermeasures to produce AMF factors and functions. The results obtained from these two approaches should be compared, documented, and reported. Specific recommendations on how to conduct expert panels for the update of the HSM (or other highway safety effort that involves expert panels) should be provided upon completion of this review.

References

Dissayanake, S., J.J. Lu, X. Chu, and P. Turner (1999) Evaluation of Critical Highway Safety Issues of Special Population Groups. *ITE Journal*, Vol. 69, No. 9, pp. 28-30.

Fukuoka, H. (2002) Reliability Evaluation Method for the Railway System: A Model for Complicated Dependency. *Quarterly Report of Railway Technical Research Institute*. Tokyo, Japan.

Green, Armstrong, and Graefe (2007) Methods to Elicit Forecasts from Groups: Delphi and Prediction Markets Compared. Forthcoming in *Foresight: The International Journal of Applied Forecasting*, Fall issue.

Harwood, D.W., F.M. Council, E. Hauer, W.E. Hughes, and A. Vogt (2000) Prediction of the Expected Safety Performance of Rural Two Lane Highways. Federal Highway Administration Report, FHWA-RD-99-207. Washington, D.C.

Lyon, Craig, J. Oh, B. Persaud, S. Washington, and J. Bared (2003). Empirical Investigation of the IHSDM Accident Prediction Algorithm: Rural Intersections. *Journal of the Transportation Research Board*. National Academy Press, Washington D.C., No 1840. pp. 78-86.

Melcher, Daniel, K. Dixon, S. Washington, and E. Wu (2001). Feasibility of 'Subjective' Engineering Assessments of Road Safety Improvements: Bayesian Development. TRB, National Research Council, Washington, D.C. *Journal of the Transportation Research Board*, No. 1758. pp. 36-43.

Oh, Jutaek, C. Lyon, S. Washington, B. Persaud, and J. Bared. (2003). Validation of the FHWA Crash Models for Rural Intersections: Lessons Learned. *Journal of the Transportation Research Board*. National Academy Press, Washington D.C., No. 1840. pp. 41-49.

Rowe, G., and G. Wright (1999) The Delphi technique as a forecasting tool: issues and analysis. *International Journal of Forecasting*, Vol. 15, Issue 4, October 1999.

Rowe G., and G. Wright (2001) Expert Opinions in Forecasting. Role of the Delphi Technique. In: Armstrong (Ed.): *Principles of Forecasting: A Handbook of Researchers and Practitioners*, Boston: Kluwer Academic Publishers.

Sireli, Y., E. Ozan, P. Kauffmann, S. Gupta, and P. Kachroo (2002) A QFD-Based Decision Model for Application of ITS Technologies to Cockpit Weather Information

Systems for General Aviation. *Proceedings of 9th World Congress of Intelligent Transportation Systems*. Chicago, Ill. ITS American, Washington D.C.

Zografos, K.G. and I.M. Giannouli (2001) Evaluating Cost-Effectiveness of an Air Traffic Management System for Europe: Development and Application of Methodological Framework. *Transportation Research Record 1744*. Transportation Research Board, Washington D.C.

APPENDIX D

Cumulative Residuals (CURE) Plots for the Cross-Validation Study

3-LEGGED SIGNALIZED INTERSECTIONS

Table D1 Summary Results for 3-Legged Signalized Intersections

New York Data 14 Sites 65 Total Crashes 35 Injury Crashes					
SPF	**Pearson's Correlation Coefficient**	**Mean Absolute Deviation**	**Mean Squared Prediction Error**	**Maximum Deviation From Cure Plot – Major AADT**	**Maximum Deviation From Cure Plot – Minor AADT**
MN-Total-1	-	-	-	-	-
CA-Total-1	0.53	2.59	11.20	8.53	8.99
MN-Injury-1	-	-	-	-	-
CA-Injury-1	0.48u	0.56	0.49	1.56	1.66

*No model for Minnesota was successfully calibrated

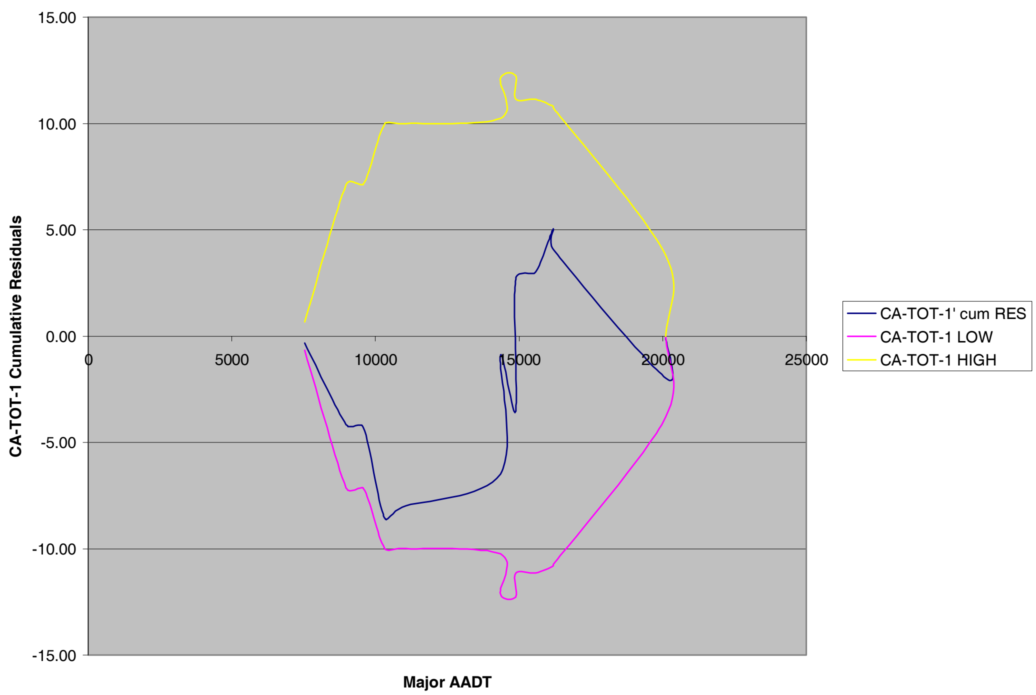


Figure D1 CURE Plot for the California Model and Total Crashes (Major AADT)

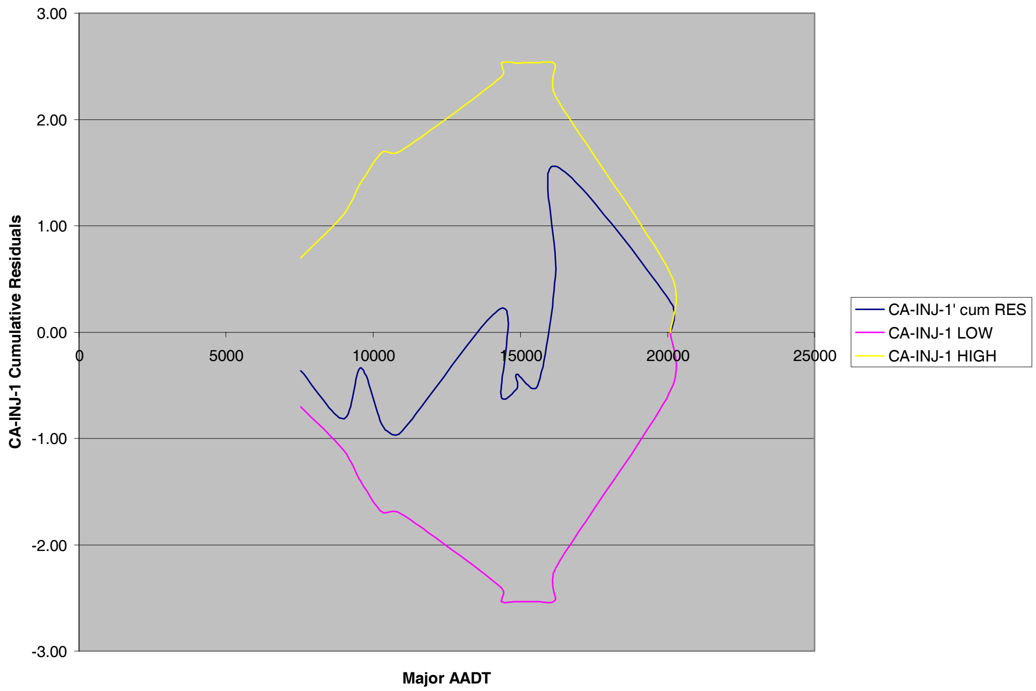


Figure D2 CURE Plot for the California Model and Injury Crashes (Major AADT)

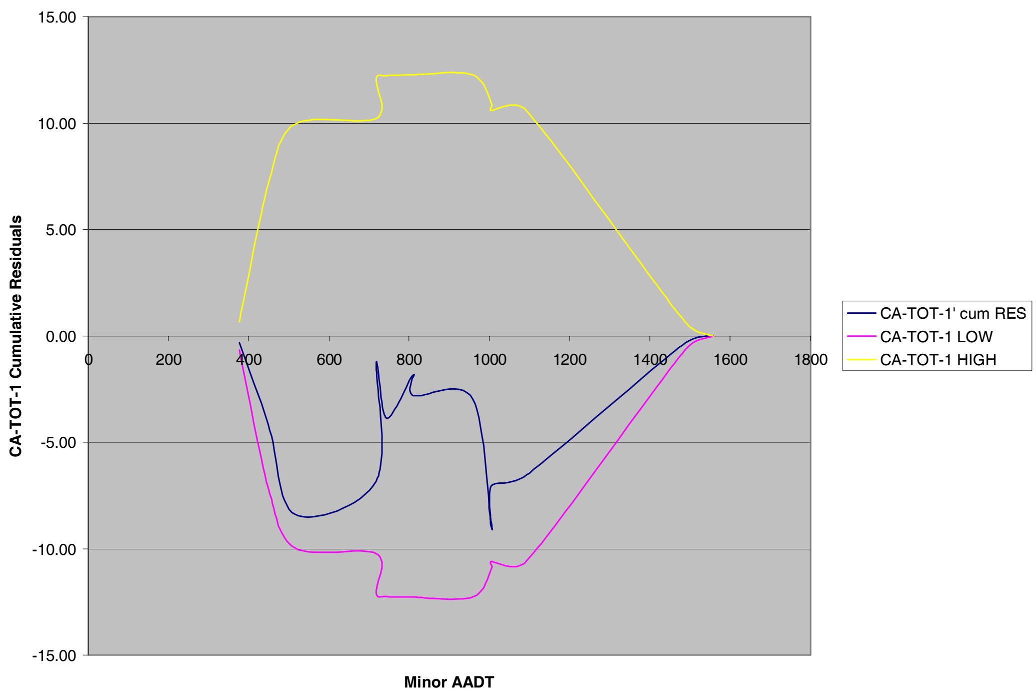


Figure D3 CURE Plot for the California Model and Total Crashes (Minor AADT)

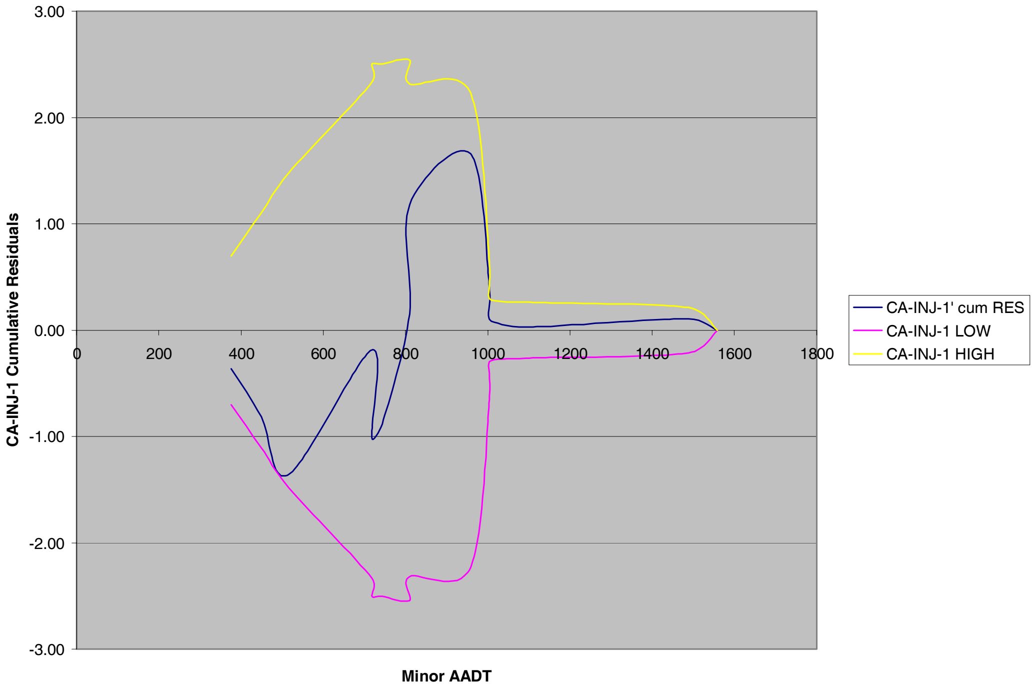


Figure D4 CURE Plot for the California Model and Injury Crashes (Minor AADT)

3-LEGGED STOP-CONTROLLED INTERSECTIONS

Table D2 Summary Results for 3-Legged Stop-Controlled Intersections

New York Data 281 Sites 673 Total Crashes 396 Injury Crashes					
SPF	**Pearson's Correlation Coefficient**	**Mean Absolute Deviation**	**Mean Squared Prediction Error**	**Maximum Deviation From Cure Plot – Major AADT**	**Maximum Deviation From Cure Plot – Minor AADT**
MN-TOT-1	0.46	2.02	10.63	36.85	60.53
CA-TOT-1	0.46	2.04	10.70	67.97	49.84
MN-INJ-1	0.24	0.66	0.87	12.80	13.15
CA-INJ-1	0.33	0.67	0.91	19.01	18.30

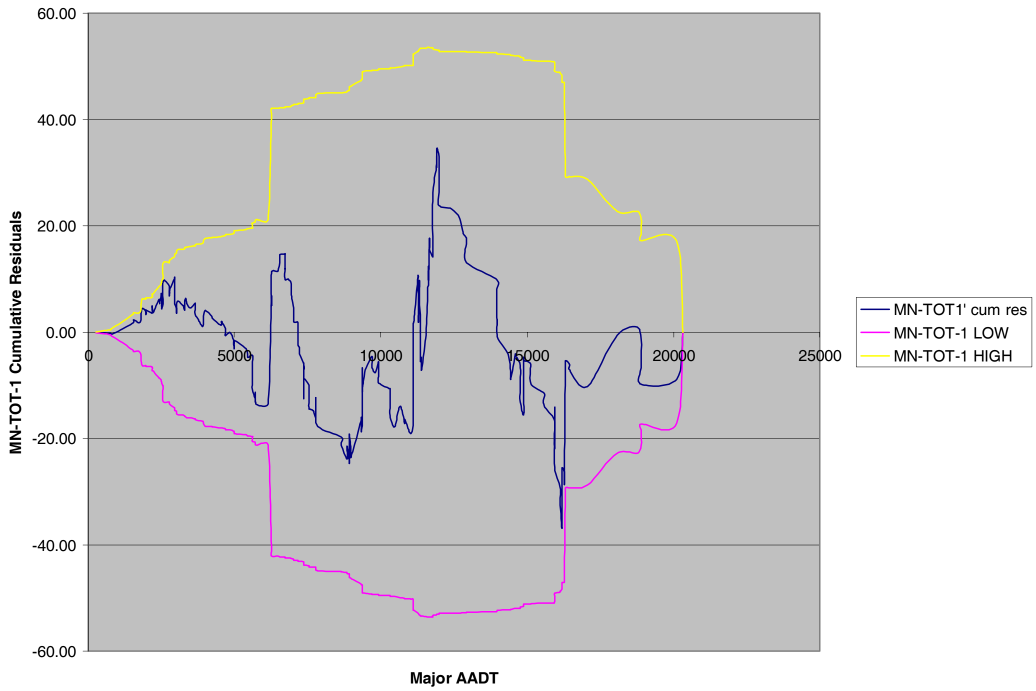


Figure D5 CURE Plot for the Minnesota Model and Total Crashes (Major AADT)

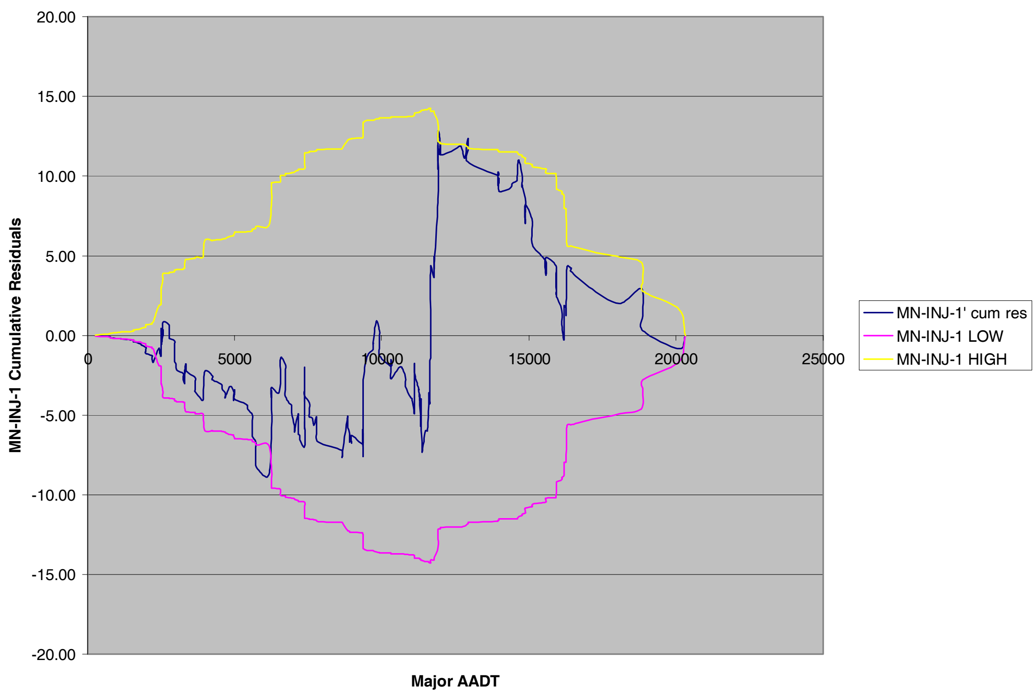


Figure D6 CURE Plot for the Minnesota Model and Injury Crashes (Major AADT)

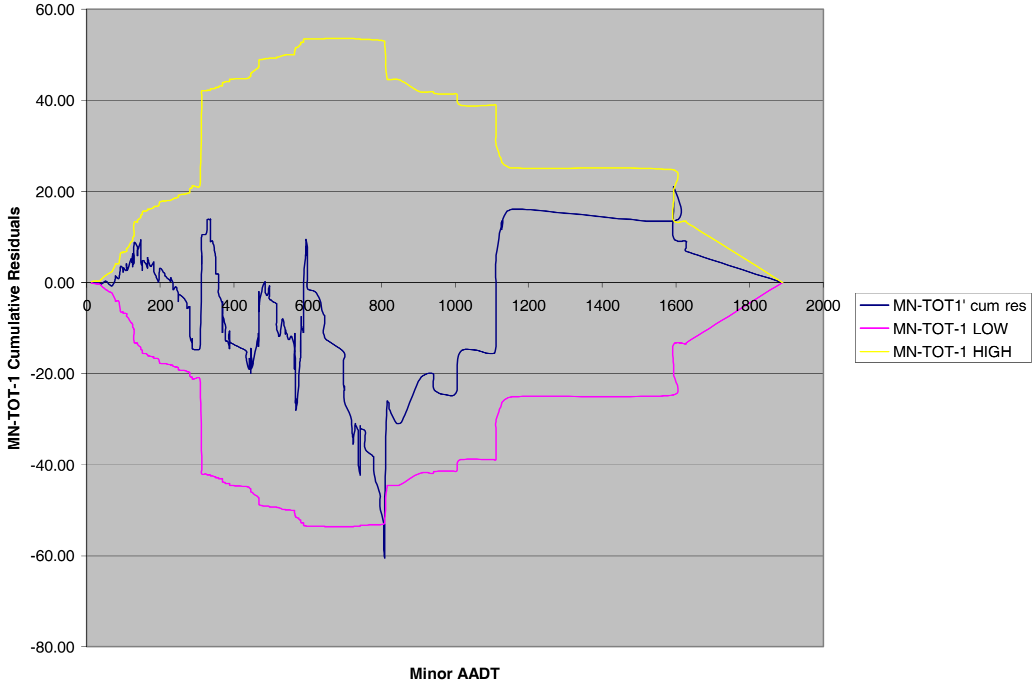


Figure D7 CURE Plot for the Minnesota Model and Total Crashes (Minor AADT)

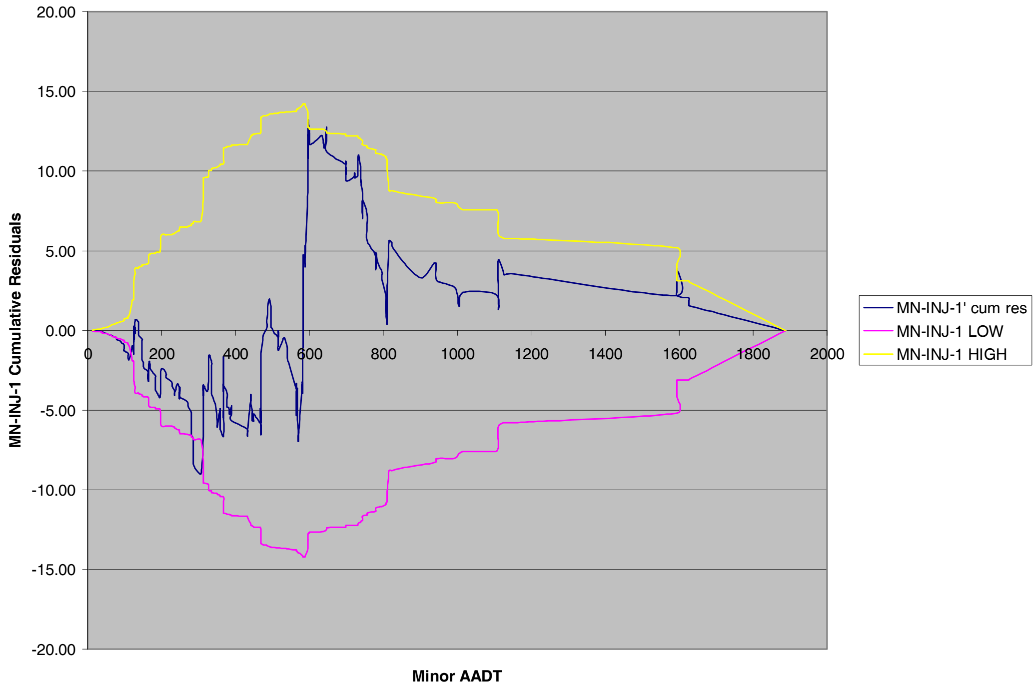


Figure D8 Cure Plot for the Minnesota Model and Injury Crashes (Minor AADT)

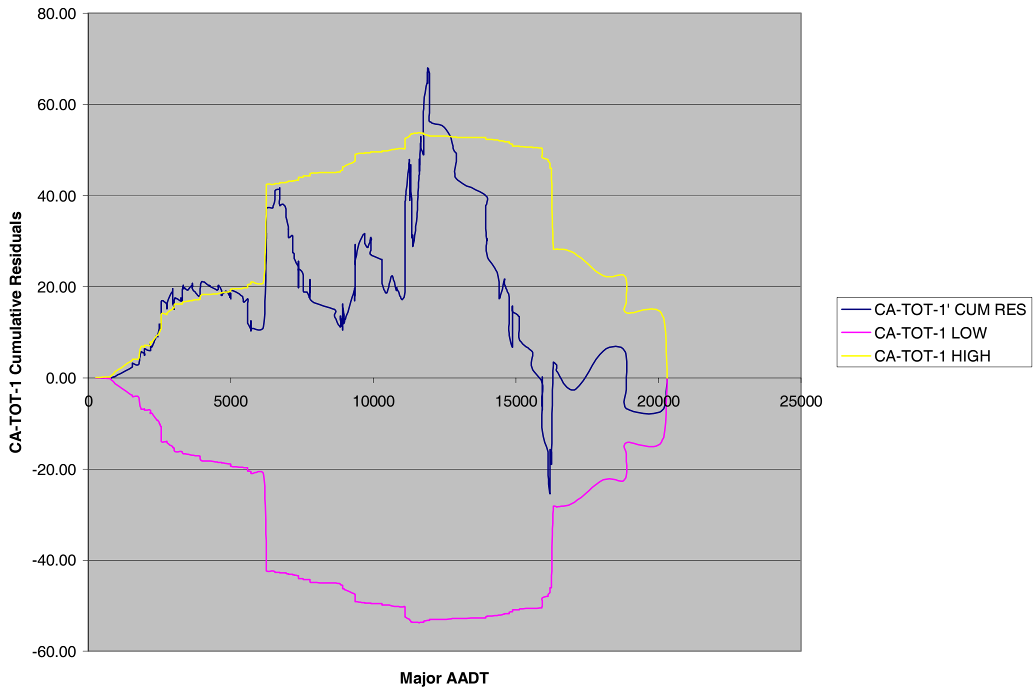


Figure D9 CURE Plot for the California Model and Total Crashes (Major AADT)

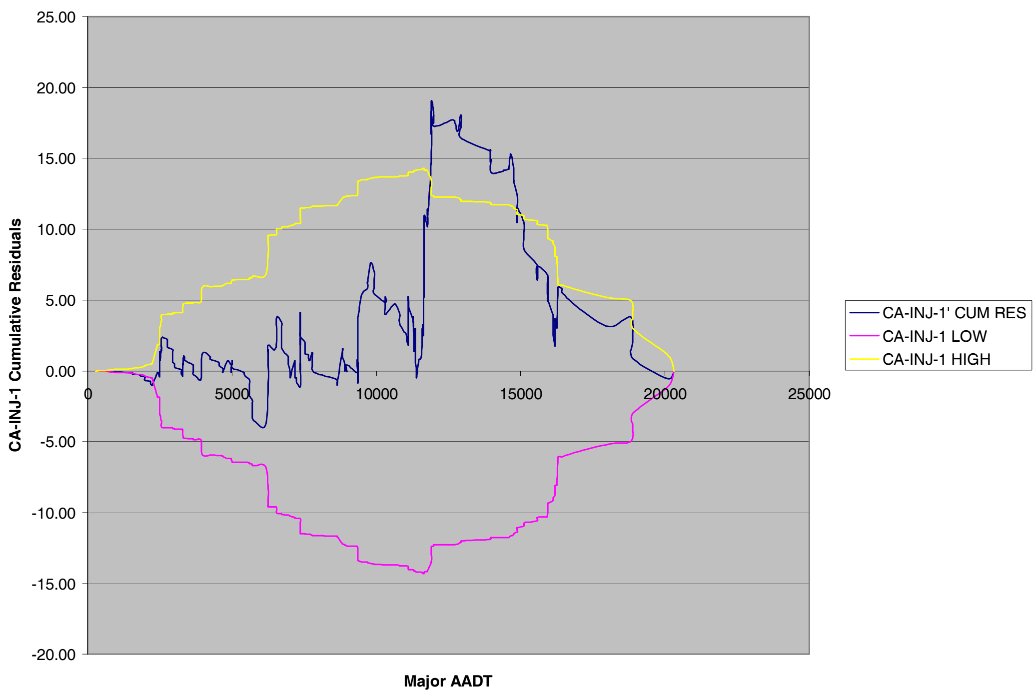


Figure D10 CURE Plot for the California Model and Injury Crashes (Major AADT)

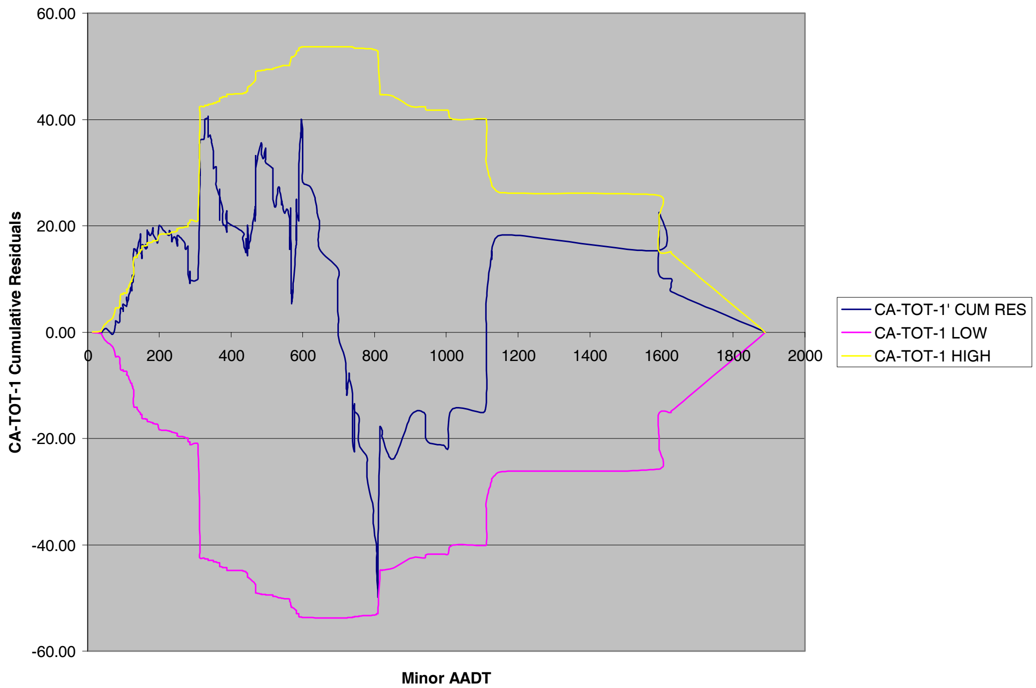


Figure D11 CURE Plot for the California Model and Total Crashes (Minor AADT)

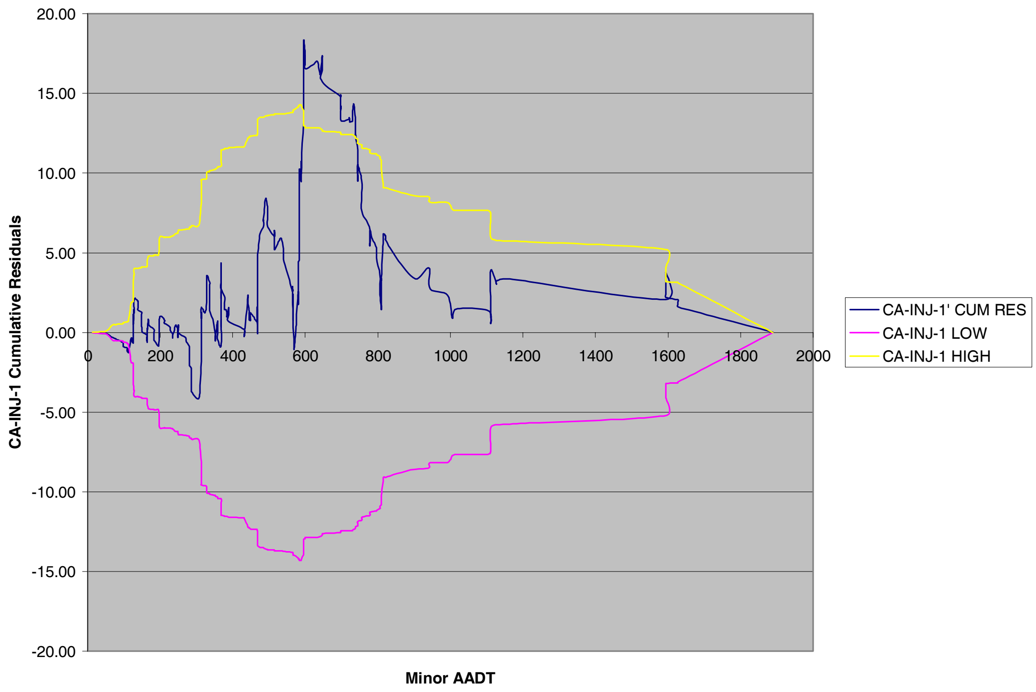


Figure D12 CURE Plot for the California Model and Injury Crashes (Minor AADT)

4-LEGGED SIGNALIZED INTERSECTIONS

Table D3 Summary Results for 4-Legged Signalized Intersections

New York Data 18 Sites 222 Total Crashes 107 Injury Crashes					
SPF	Pearson's Correlation Coefficient	Mean Absolute Deviation	Mean Squared Prediction Error	Maximum Deviation From Cure Plot – Major AADT	Maximum Deviation From Cure Plot – Minor AADT
MN-Total-1	0.27	6.28	67.32	51.44	45.83
CA-Total-1	0.24	6.30	69.52	52.93	43.14
MN-Injury-1	-0.01	1.71	4.00	12.45	11.47
CA-Injury-1	-	-	-	-	-

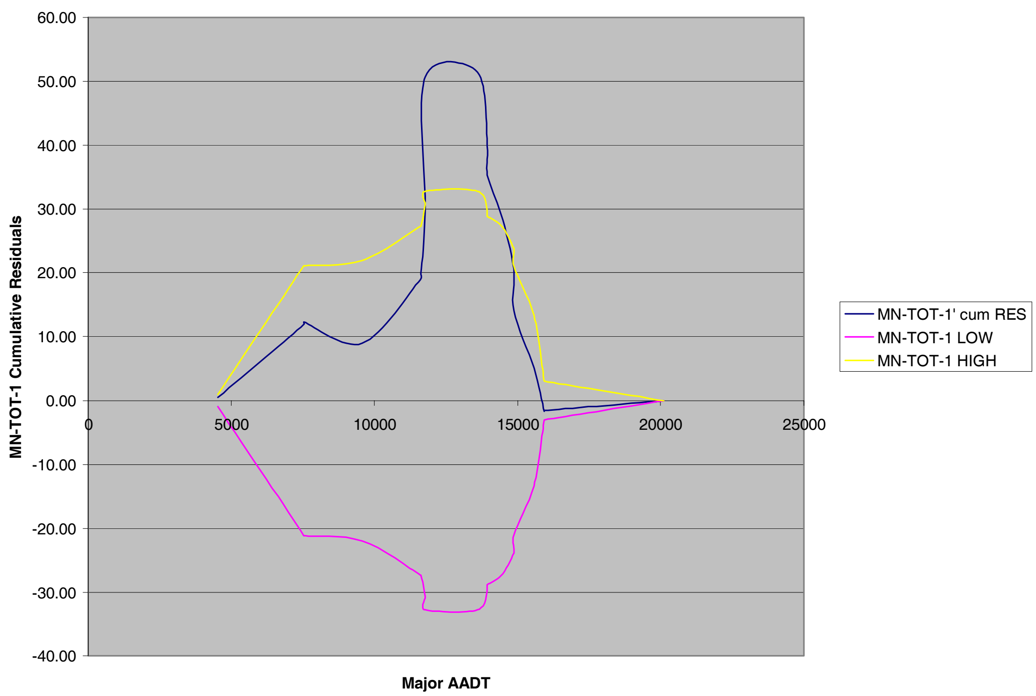


Figure D13 CURE Plot for the Minnesota Model and Total Crashes (Major AADT)

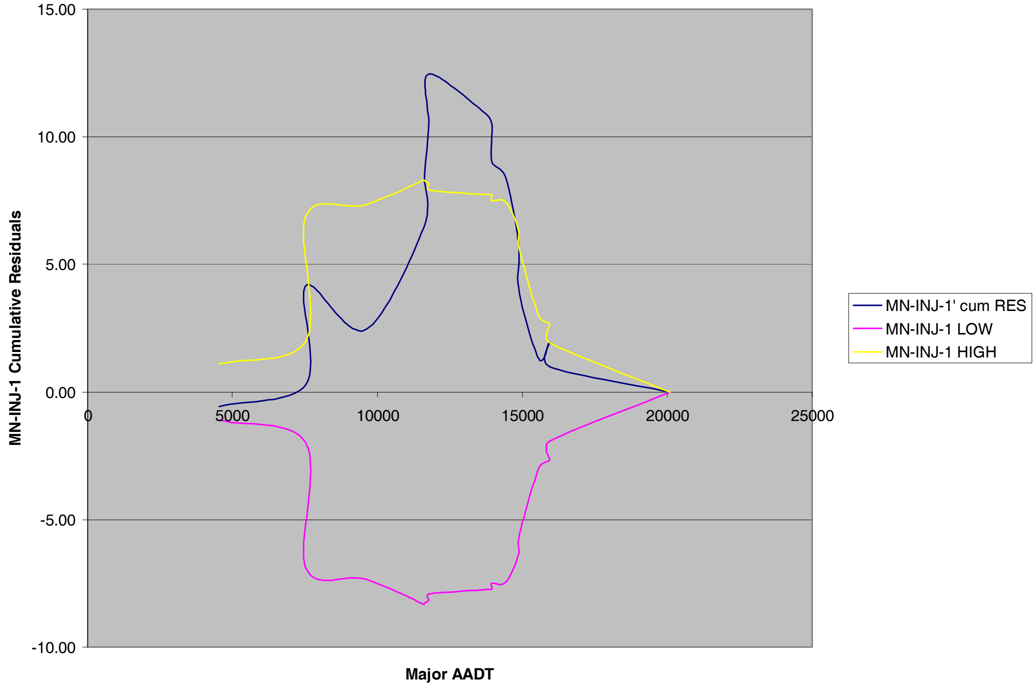


Figure D14 CURE Plot for Minnesota Model and Injury Crashes (Major AADT)

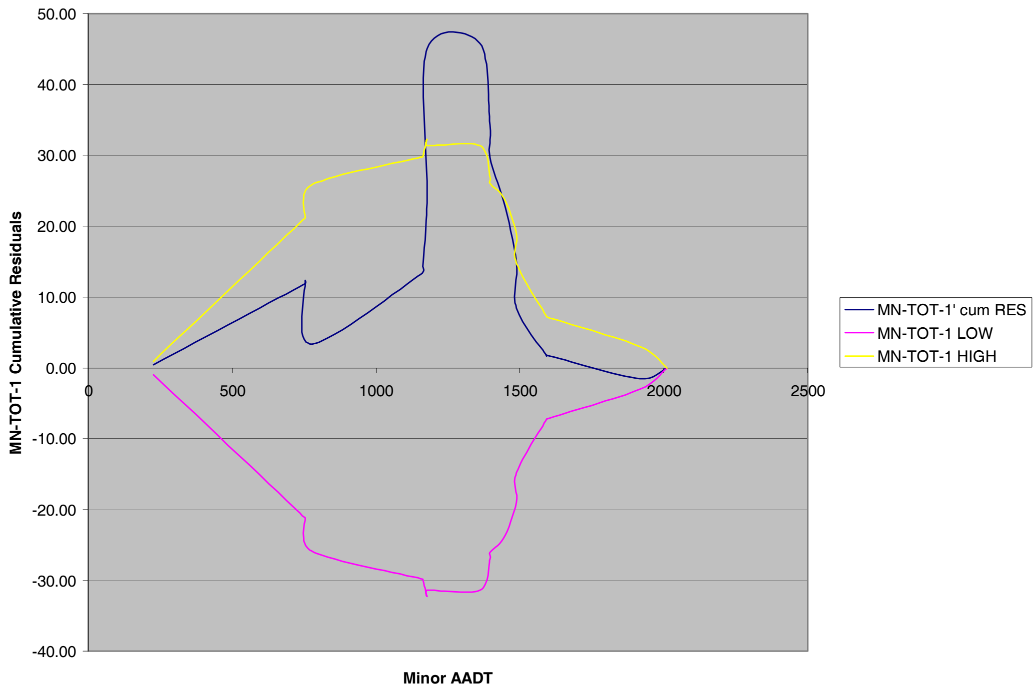


Figure D15 CURE Plot for the Minnesota Model and Total Crashes (Minor AADT)

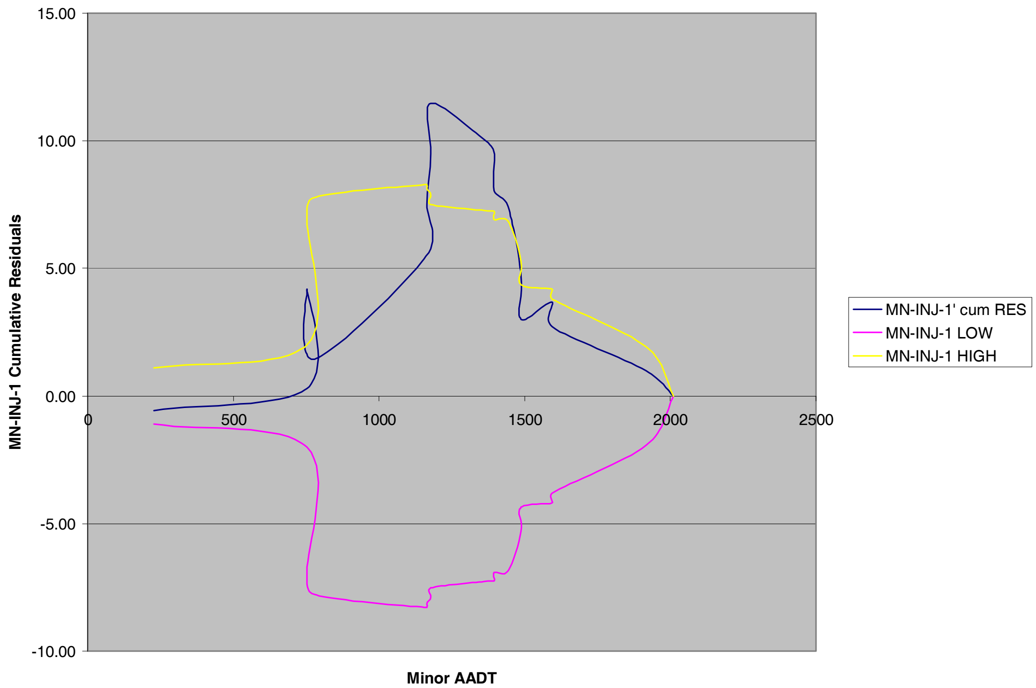


Figure D16 CURE Plot for the Minnesota Model and Injury Crashes (Minor AADT)

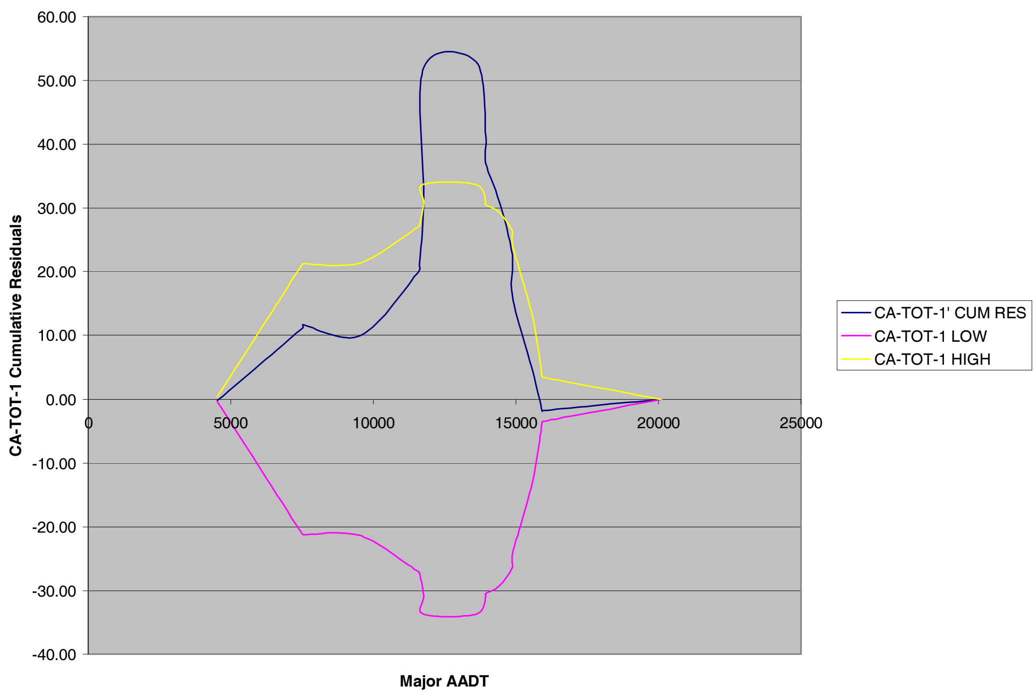


Figure D17 CURE Plot for the California Model and Total Crashes (Major AADT)

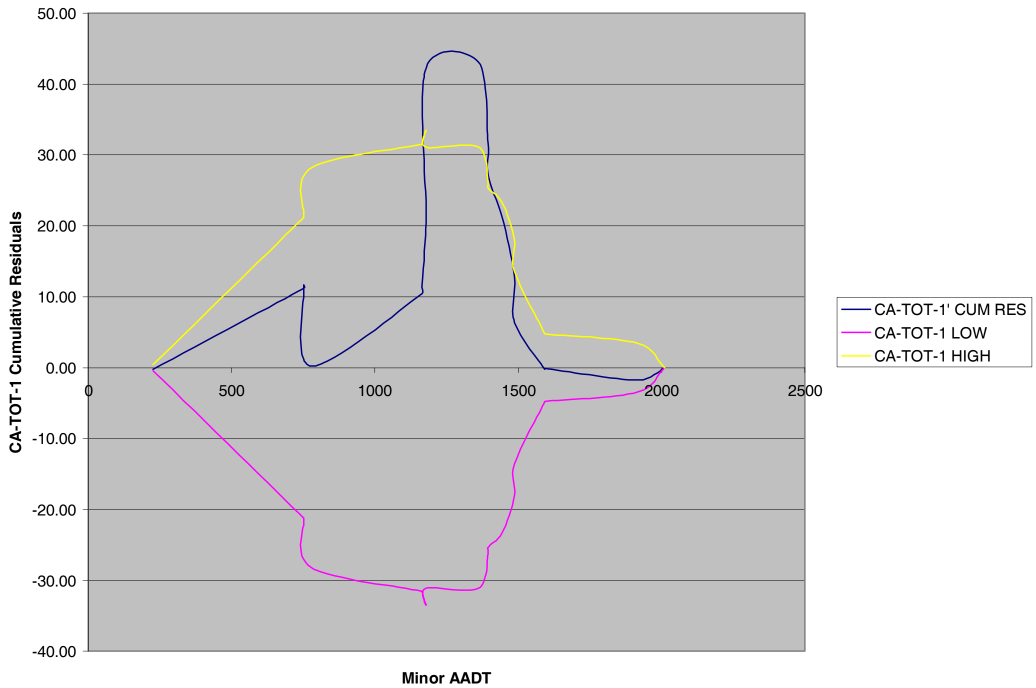


Figure D18 CURE Plot for the California Model and Total Crashes (Minor AADT)

4-LEGGED STOP-CONTROLLED INTERSECTIONS

Table D4 Summary Results for 4-Legged Stop-Controlled Intersections

New York Data 71 Sites 472 Total Crashes 299 Injury Crashes					
SPF	**Pearson's Correlation Coefficient**	**Mean Absolute Deviation**	**Mean Squared Prediction Error**	**Maximum Deviation From Cure Plot – Major AADT**	**Maximum Deviation From Cure Plot – Minor AADT**
MN-TOT-1	0.47	4.47	45.19	72.26	78.04
CA-TOT-1	0.49	4.41	41.21	46.59	53.31
MN-INJ-1	0.01	1.82	6.46	23.76	23.27
CA-INJ-1	0.00	1.73	5.75	15.47	14.94

Figure D19 CURE Plot for the Minnesota Model and Total Crashes (Major AADT)

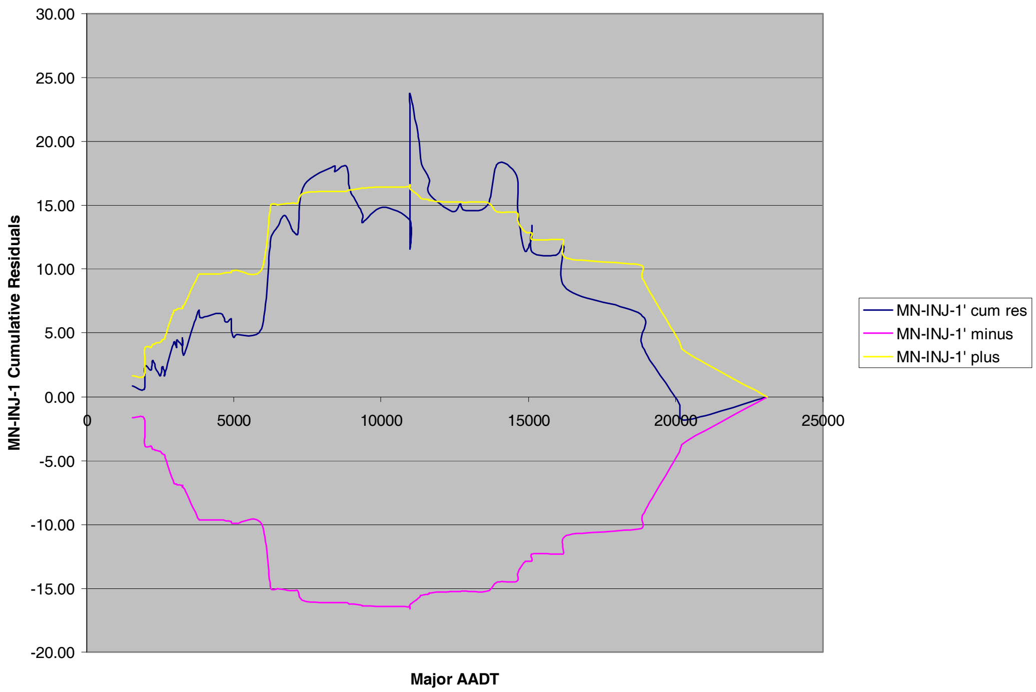


Figure D20 CURE Plot for the Minnesota Model and Injury Crashes (Major AADT)

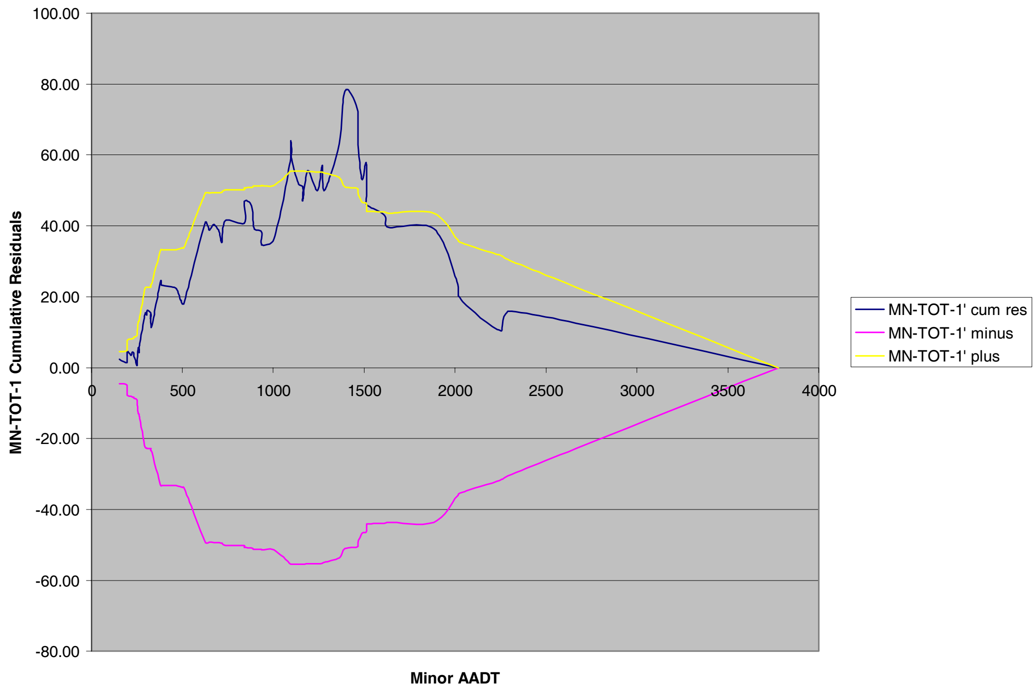


Figure D21 CURE Plot for the Minnesota Model and Total Crashes (Minor AADT)

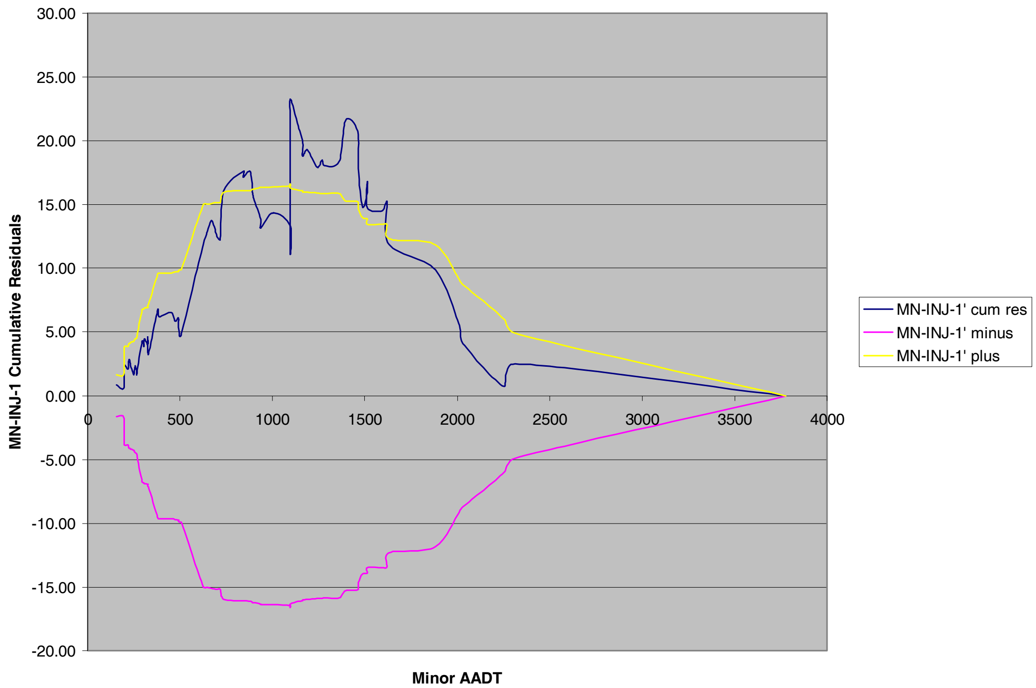


Figure D22 CURE Plot for the Minnesota Model and Injury Crashes (Minor AADT)

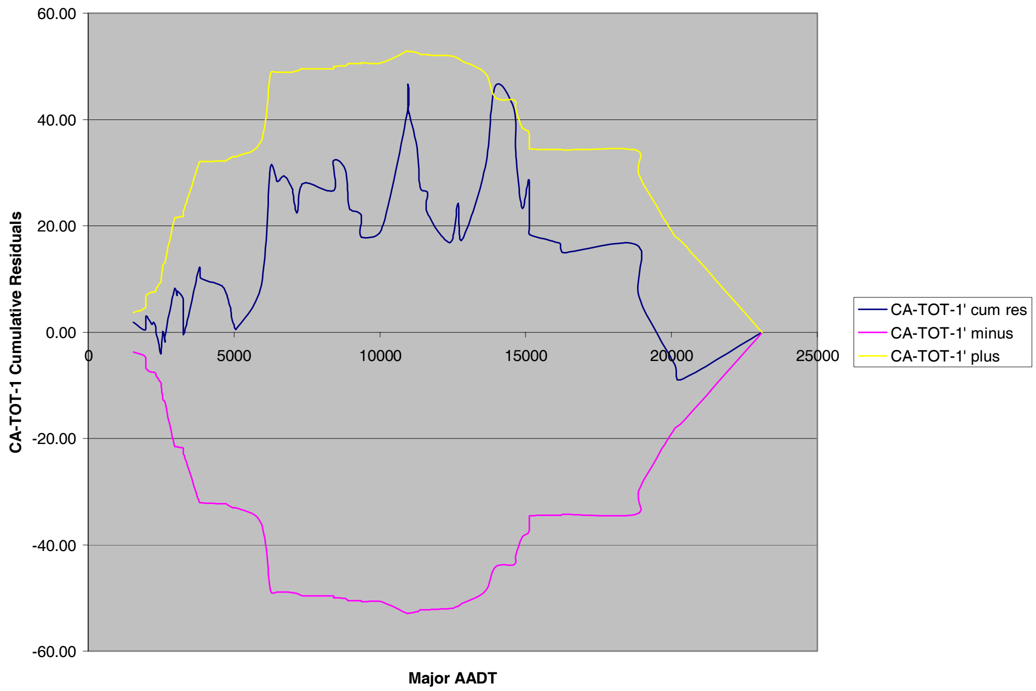


Figure D23 CURE Plot for the California Model and Total Crashes (Major AADT)

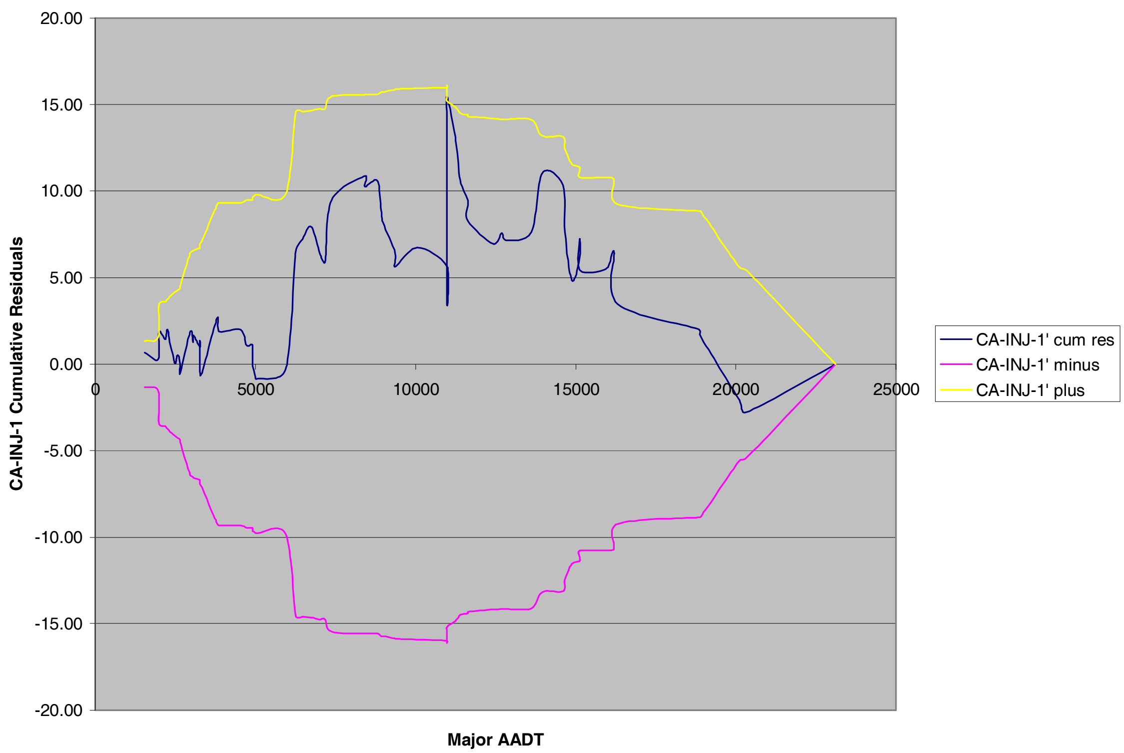


Figure D24 CURE Plot for the California Model and Injury Crashes (Major AADT)

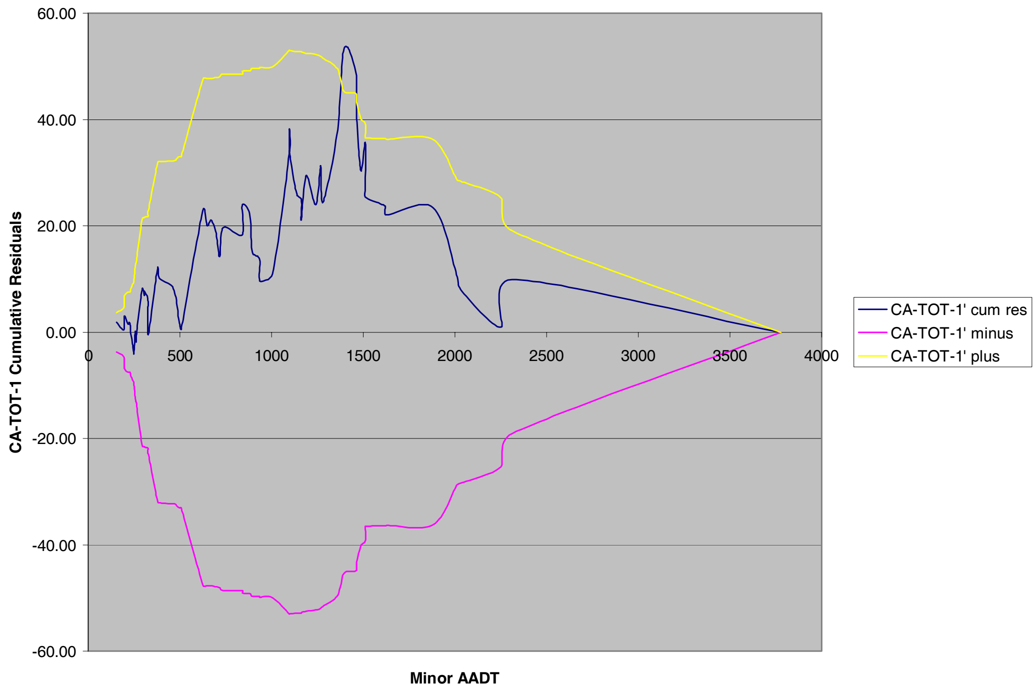


Figure D25 CURE Plot for the California Model and Total Crashes (Minor AADT)

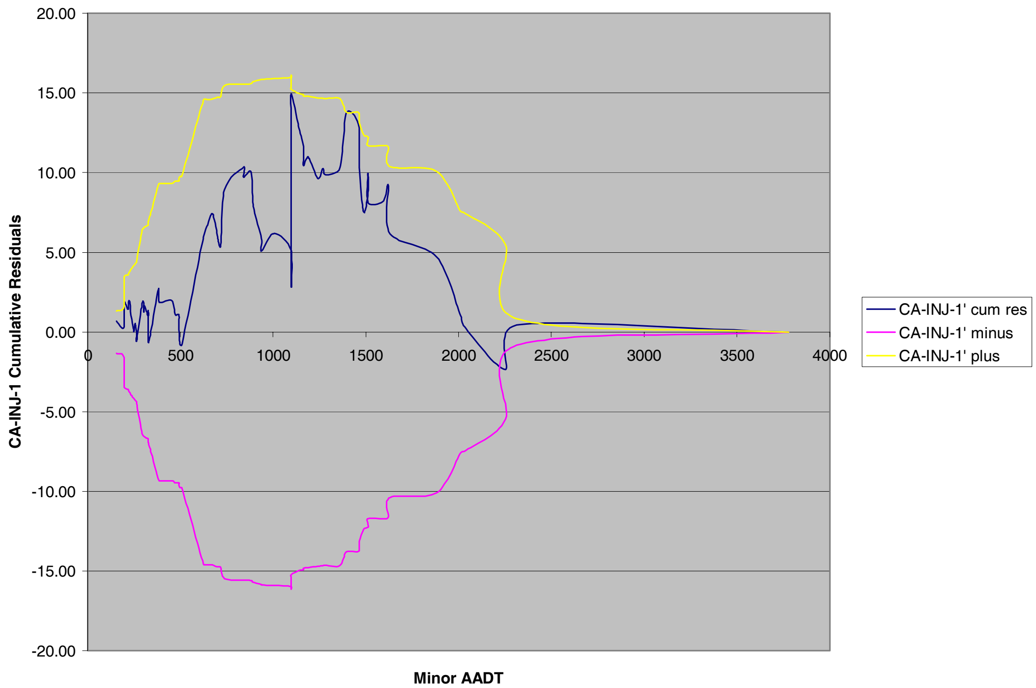


Figure D26 CURE Plot for the California Model and Injury Crashes (Minor AADT)

SEGMENT MODELS

Table D5 Summary Results for 4-Lane Undivided Segments

New York Data 199 Sites 2048 Total Crashes (7 years)				
Statistical Model	**Pearson's Correlation Coefficient**	**Mean Absolute Deviation**	**Mean Squared Prediction Error**	**Maximum Deviation From Cure Plot**
General ADT – Total (all states)	0.874	0.711	1.564	18.529
California –Total	0.757	0.936	12.308	49.310
Minnesota – Total	0.865	0.687	1.326	11.947

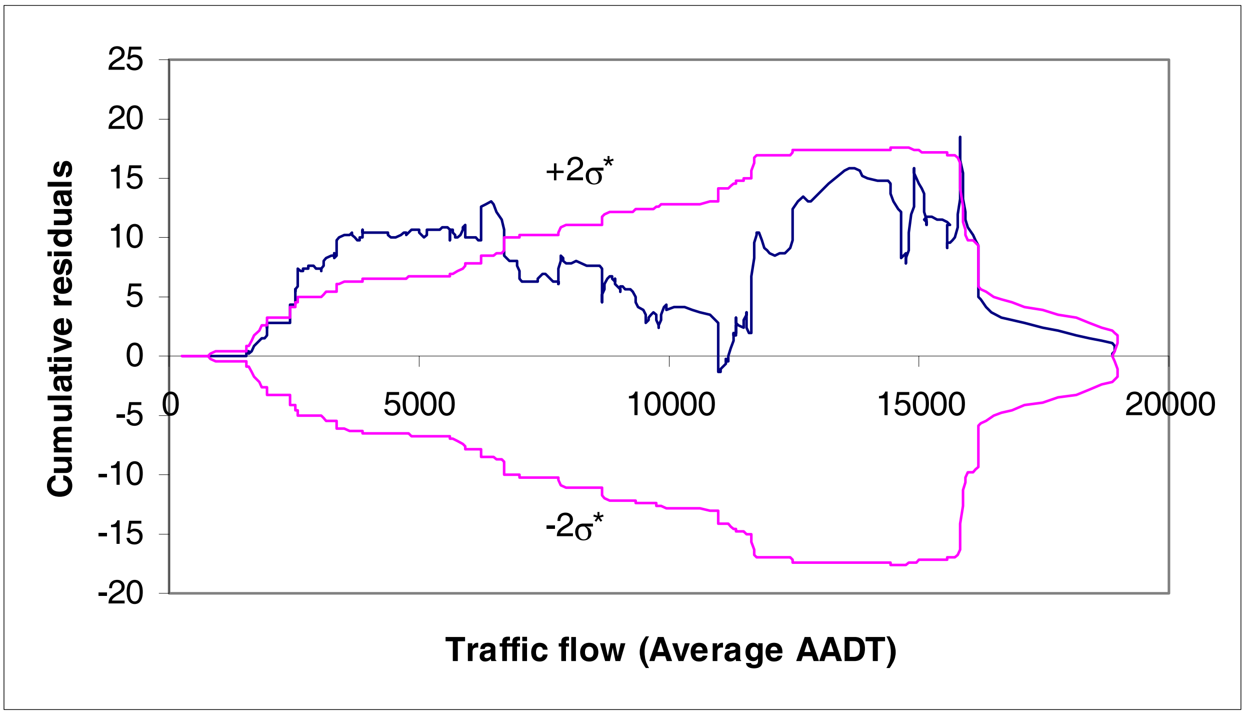


Figure D27 CURE Plot for the General ADT Model and Total Crashes

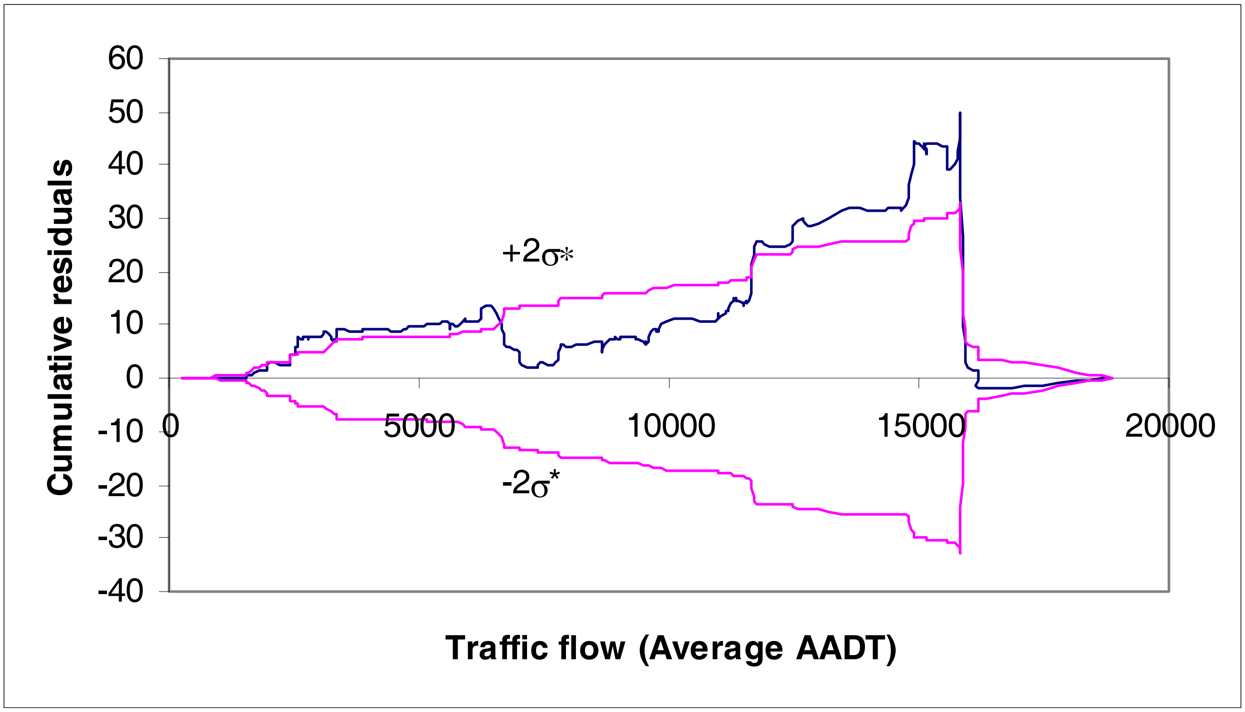


Figure D28 CURE Plot for the California Model and Total Crashes

Minnesota Model

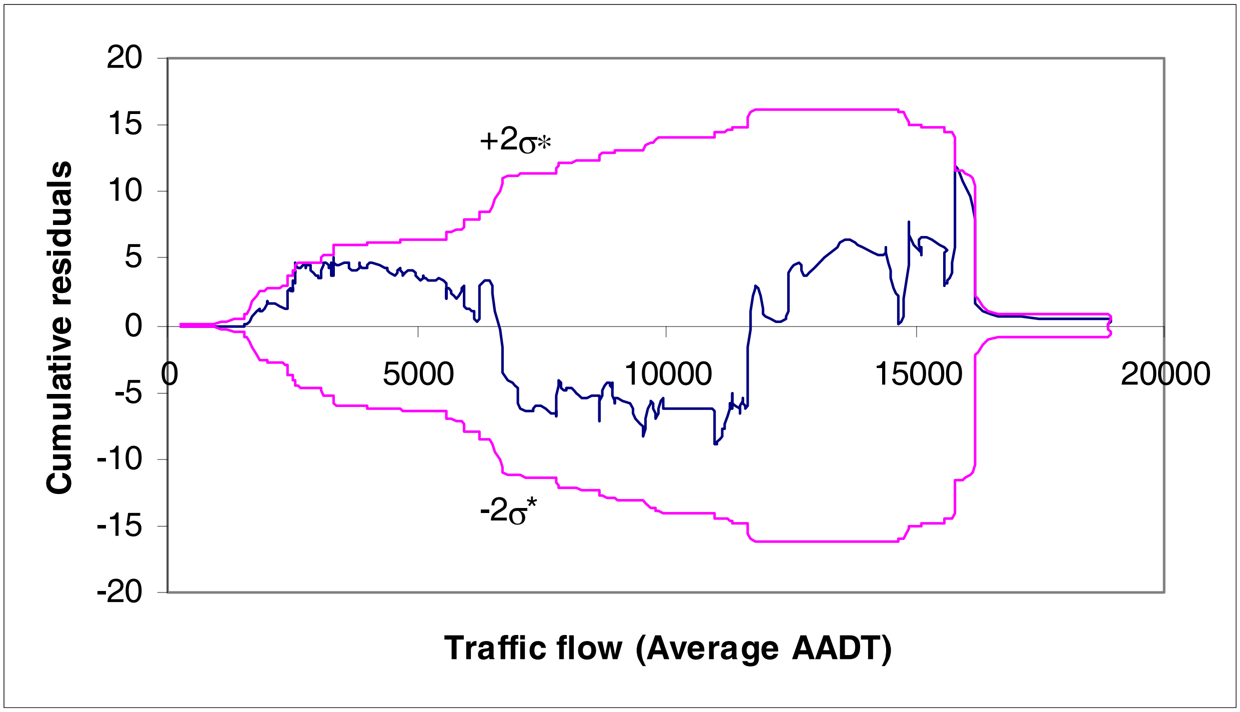


Figure D29 CURE Plot for the Minnesota Model and Total Crashes

APPENDIX E

Crash Severity Models

This appendix was originally prepared by Drs. Eun-Sug Park and Dominique Lord. A full paper has been published as a peer-reviewed publication in a Transportation Research Record Series: Park, E.S., and D. Lord (2007) Multivariate Poisson-Lognormal Models for Jointly Modeling Crash Frequency by Severity. *Transportation Research Record 2019*, pp. 1-6.

This appendix describes a new multivariate approach for jointly modeling crash counts by severity data based on Multivariate Poisson-Lognormal (MVPLN) models. This approach was evaluated as part of this research project, but will not be included in HSM Chapter 9. Although it offers potential for modeling motor vehicle crashes, further work is needed on this topic.

Even though the crash frequency by severity data are multivariate in nature, they have often been analyzed by modeling each severity level separately without taking into account correlations that exist among different severity levels. Traditionally, count data models have always been estimated for different severity levels (KAB, KABCO, etc.) separately, as done in this work. In other words, a different model is estimated for each severity level. In recent years, mixed logit (Milton et al., 2007) and ordered logit or probit models (Khattak, 2001; Kockelman and Kweon, 2002; Kweon and Kockelman, 2003) have been proposed for estimating risk of sustaining injuries by severity level, given that a crash occurred. The output of these models can be used for estimating crash count by severity level when it is multiplied by the output of crash count models for all crash severities (Miaou et al., 2005). However, even by combining the output of these two models, the correlation between the severity levels is not taken into account in the modeling process.

The new MVLPN approach builds upon the Multivariate Poisson regression (MVP) model (Tsionas, 2001; Tunaru, 2002; Bijleveld, 2005; Miaou and Song, 2005; Karlis and Meligkotsidou, 2005; Song et al., 2006, Ma and Kockelman, 2006). MVLPN can cope with either over-dispersion or a fully general correlation structure in the data as opposed to the recently suggested Multivariate Poisson regression approach that allows for neither over-dispersion nor a general correlation structure in the data. For additional details about the theoretical framework of these models, the reader is referred to Chib and Winkelmann (2001) and Park and Lord (2007).

Implementation of MVPLN models is not straightforward. It needs to be noted that no commercial statistical software packages have the ability to estimate these models as built-in functions. As mentioned in Chib and Winkelmann (2001), it is necessary to adapt simulation-based methods such as a MCMC simulation method (see, e.g., Tierney et al., 1994; Gilks et al., 1996; Lui, 2001) to cope with the multiple integral in the likelihood function. To estimate MVPLN models, the MATLAB (The MathWorks, 2006) codes tailored to multivariate crash data modeling have been developed according to the MCMC algorithm of Chib and Winkelmann (2001).

The MVPLN models were applied to the crash count data of five different severity levels described above: fatal (K), incapacitating-injury (A), non-incapacitating injury (B), minor injury (C), property damage only (PDO or O) collected from the 451 three-legged unsignalized intersections in California described above. Although the original data contained the crash counts from 537 intersections, only the intersections having 10 years of crash data history were retained (resulting in 451 intersections).

Table E1 contains the summary statistics of the variables of interest. In the table, the unit of crash frequency is the number of crashes per intersection for 10 years. The major and minor roads of the intersection are defined as a function of the entering traffic flow. The legs with the highest entering flows are defined as major AADT.

Table E1 Summary Statistics of the Variables for California 3-Legged Intersection Data

Variable Name	Mean	Std Dev	Min.	Max.
Dependent Variables				
Fatal	0.1707	0.5204	0	5
Injury A	0.4479	0.9609	0	6
Injury B	1.6364	2.5159	0	20
Injury C	1.9180	3.5571	0	28
PDO	6.3348	9.9493	0	88
Independent Variables				
Lighting (1= yes)	0.3525	0.4783	0	1
Painted Left Turn (1= yes)	0.3925	0.4888	0	1
Curb Med Left Turn (1=yes)	0.1330	0.3340	0	1
Right Turn Channel (1=yes)	0.1397	0.3470	0	1
ML Lanes (Nb of Main Lanes)	3.6851	0.7292	2	4
Mountain Terrain (1=yes)	0.1397	0.3470	0	1
Rolling Terrain (1=yes)	0.3570	0.4796	0	1
Logmaj (Logarithm of major AADT)	9.4195	0.7514	7.7956	11.2683
Logmin (Logarithm of minor AADT)	4.9193	1.5148	2.3026	10.0481

Table E2 gives the estimates (posterior means and standard deviations) of the regression coefficients $\boldsymbol{\beta}$ based on a MVPLN model implemented by MCMC using the MATLAB (The MathWorks, 2006) codes specifically developed for this research. Recall that the functional form used for the models was described in Equations 5.1 to 5.7. The dependent variable is defined as the number of crashes per 10 years. To ensure that the chain has converged to the posterior distribution by the end of the burn-in period, trace plots and the autocorrelation function plots of posterior sample values were inspected although those plots are not presented in the paper due to space limitations.

For comparison purposes, Table E2 reports the estimates obtained by applying the univariate Poisson regression model and the univariate NB regression model implemented in SAS (SAS, 2002) as well.

Table E2 Modeling Output for Models by Severity

Severity	Variable	Multivariate Poisson-Lognormal Model	Univariate Poisson Regression	Univariate Negative Binomial Regression
Fatal	Constant	**-13.0261** (1.6854)	**-15.2279** (2.0375)	**-14.9638** (2.2200)
	Lighting	-0.5544 (0.3229)	-0.5955 (0.3204)	-0.5704 (0.3470)
	Painted Left Turn	0.5349 (0.2859)	0.5032 (0.2886)	0.5158 (0.3138)
	Curb Med Left Turn	0.4994 (0.3534)	0.6221 (0.3446)	0.6228 (0.3884)
	Rhgt Trn Channel	0.2777 (0.3156)	0.3752 (0.2870)	0.2991 (0.3356)
	ML Lanes	0.2934 (0.2764)	0.2815 (0.3045)	0.2714 (0.3152)
	Mountain	-0.1367 (0.3720)	-0.3431 (0.3764)	-0.1864 (0.4232)
	Rolling	-0.3916 (0.2733)	**-0.5641** (0.2689)	-0.5400 (0.3005)
	Logmaj ADT	**0.8818** (0.1698)	**1.1537** (0.1894)	**1.1188** (0.2088)
	Logmin ADT	**0.2069** (0.0873)	**0.2052** (0.0810)	**0.2223** (0.0921)
				Dispersion: 0.7059
	Pearson Chi-Square/DF		1.2232	1.0667
Inj A	Constant	**-12.5689** (1.2596)	**-13.2302** (1.1873)	**-13.4023** (1.4116)
	Lighting	0.2345 (0.1993)	0.2997 (0.1733)	0.2844 (0.2072)
	Painted Left Turn	**0.5569** (0.2031)	**0.4796** (0.1706)	**0.5572** (0.2023)
	Curb Med Left Turn	0.1780 (0.2856)	0.2229 (0.2431)	0.2290 (0.2882)
	Rhgt Trn Channel	0.2285 (0.2379)	0.3425 (0.1821)	0.2686 (0.2408)
	ML Lanes	0.1625 (0.1737)	0.1571 (0.1563)	0.1490 (0.1703)
	Mountain Terrain	0.3866 (0.2667)	0.3106 (0.2294)	0.4187 (0.2762)
	Rolling Terrain	**0.4564** (0.1918)	**0.4112** (0.1611)	**0.4710** (0.1910)
	Logmaj	**0.9097** (0.1336)	**1.0435** (0.1186)	**1.0548** (0.1422)
	Logmin	**0.2331** (0.0612)	**0.1899** (0.0492)	**0.1952** (0.0619)
				Dispersion: 0.6070
	Pearson Chi-Square/DF		1.2699	1.0042
Inj B	Constant	**-9.8505** (0.8479)	**-9.9059** (0.5815)	**-10.1854** (0.8482)
	Lighting	0.2081 (0.1360)	**0.2315** (0.0907)	0.2025 (0.1321)
	Painted Left Turn	0.1088 (0.1388)	0.0648 (0.0844)	0.1206 (0.1271)
	Curb Med Left Turn	0.0560 (0.1875)	0.0780 (0.1188)	0.0896 (0.1811)
	Rhgt Trn Channel	0.0793 (0.1619)	**0.2511** (0.1002)	0.0499 (0.1655)
	ML Lanes	0.0417 (0.0995)	0.0404 (0.0692)	0.0491 (0.0911)
	Mountain	**0.4458** (0.1650)	**0.3636** (0.1074)	**0.5708** (0.1691)
	Rolling	0.0734 (0.1329)	0.0447 (0.0846)	0.0885 (0.1257)
	Logmaj	**0.8936** (0.0907)	**0.9463** (0.0604)	**0.9645** (0.0881)
	Logmin	**0.1789** (0.0419)	**0.1608** (0.0262)	**0.1670** (0.0393)
				Dispersion: 0.6048
	Pearson Chi-Square/DF		2.0799	1.0555
Inj C	Constant	**-11.9536** (0.8721)	**-12.4660** (0.5726)	**-11.4316** (0.8863)
	Lighting	**0.5212** (0.1409)	**0.5264** (0.0845)	**0.5422** (0.1394)
	Painted Left Turn	0.0119 (0.1485)	-0.0357 (0.0774)	0.0169 (0.1354)
	Curb Med Left Turn	-0.1958 (0.1990)	-0.1487 (0.1172)	-0.1396 (0.1984)
	Rhgt Trn Channel	0.2490 (0.1789)	**0.2908** (0.0917)	0.3392 (0.1743)
	ML Lanes	0.0134 (0.1007)	0.0140 (0.0649)	0.0093 (0.0966)
	Mountain	**0.4015** (0.1790)	**0.3253** (0.1007)	**0.4683** (0.1837)
	Rolling	0.0518 (0.1451)	0.0569 (0.0787)	0.0536 (0.1353)
	Logmaj	**1.0857** (0.0926)	**1.2034** (0.0593)	**1.0921** (0.0938)
	Logmin	**0.2317** (0.0442)	**0.1982** (0.0240)	**0.1997** (0.0417)
				Dispersion: 0.8015
	Pearson Chi-Square/DF		2.8881	1.2074
	Constant	**-9.9596** (0.6670)	**-10.1806** (0.3065)	**-9.6546** (0.6358)

PDO	Lighting	**0.4203** (0.1051)	**0.3544** (0.0465)	**0.4881** (0.1049)
	Painted Left Turn	**-0.2159** (0.1127)	**-0.2326** (0.0420)	**-0.2327** (0.1027)
	Curb Med Left Turn	-0.1494 (0.1482)	**-0.1836** (0.0611)	-0.2024 (0.1471)
	Rhgt Trn Channel	0.0715 (0.1263)	**0.1864** (0.0525)	0.1016 (0.1311)
	ML Lanes	0.1257 (0.0723)	**0.1041** (0.0373)	**0.1423** (0.0692)
	Mountain	**0.5337** (0.1347)	**0.5352** (0.0533)	**0.5966** (0.1376)
	Rolling	0.1260 (0.1046)	**0.1403** (0.0437)	0.0699 (0.1004)
	Logmaj	**0.9777** (0.0717)	**1.0593** (0.0315)	**0.9829** (0.0676)
	Logmin	**0.2493** (0.0333)	**0.2193** (0.0132)	**0.2291** (0.0321)
				Dispersion: 0.6225
	Pearson Chi-Square/DF		5.4932	1.1823

1. Multivariate Poisson-Lognormal model was implemented by MCMC coded in MATLAB (MathWorks, 2006).
2. Univariate Poisson regression and Univariate Negative Binomial regression were implemented in SAS (27).
3. Numbers in parentheses represent uncertainty estimates; posterior standard deviations under Multivariate Poisson lognormal model and standard errors under Univariate Poisson regression model and Univariate Negative binomial regression model, respectively.
4. Significant (at α=0.05) effects are shown in bold.

It needs to be noted that for an objective comparison the prior distributions of the parameters and the starting values in MCMC implementation have been obtained independently of the SAS results. Here, the research team used vaguer priors not requiring much prior knowledge on the parameters to illustrate that the suggested MVPLN models can be applied even without precise prior knowledge. When there exists good prior information on the parameters, however, it can be incorporated by the use of more informative (precise) prior distribution, and it may further improve the precision of the MVPLN models. A more comprehensive discussion on elicitation of priors in crash data analysis can be found in literature such as Schluter et al. (1997). Finally, all the variables described in Table E1 were included in the models to facilitate the comparison between the multivariate and univariate models.

It can be observed from Table E2 that for fatal and injury A, all three models give similar results in terms of point estimates and their uncertainty estimates except for *Rolling Terrain* for fatal crashes (which was significant only under the univariate Poisson regression model). For injuries A and B, and PDO crashes, however, univariate Poisson regression models give significantly different results (in terms of both point estimates and uncertainty estimates) from those of MVPLN models or univariate Negative Binomial regression models. For injuries A and B, and PDO crashes, it appears that under the univariate Poisson regression model, the standard errors are seriously underestimated and, as a result, many of the covariates are incorrectly declared to be significant. Note that the values of Pearson's Chi-Square divided by degrees of freedom for univariate Poisson regression models are considerably greater than 1 for injuries A and B, and PDO crashes, which indicates an apparent over-dispersion problem. It is well-known that the larger the over-dispersion, the more severe the underestimation of the standard errors, in which case those standard errors are not correct estimates of true uncertainties and the corresponding interval estimates will not be able to capture the true parameter values. This problem cannot be overcome with MVP regression models either because over-dispersion is not accounted for by those models. It needs to be emphasized that for the unbiased estimates, the small standard errors (more precise estimates), only when they are not underestimated, lead to more accurate parameter estimates.

MVPLN models and univariate NB models give, in general, consistent results in terms of significance of model coefficients. Notice, however, that for fatal crashes the uncertainty estimates from the MVPLN model are noticeably smaller than those from the univariate NB model. Unlike univariate Poisson regression models, both MVPLN models and univariate NB regression models are able to account for over-dispersion, and their standard errors can serve as good estimates of true uncertainties. This supports that by accounting for correlation in multivariate crash frequency by severity a MVPLN model leads to more precise parameter estimates than a univariate Negative Binomial model does.

Table E3 and Table E4 contain the MCMC estimates of the covariance matrix and correlation matrix of the latent effects (generating the correlation structure in the multivariate crash counts) of the MVPLN model, respectively.

Table E3 Posterior Means of the Covariance Matrix (Σ) of the Latent Effects

	Fatal	**Injury A**	**Injury B**	**Injury C**	**PDO**
Fatal	0.6592	0.4884	0.4743	0.5487	0.4638
Injury A	0.4884	0.7408	0.5251	0.6054	0.4998
Injury B	0.4743	0.5251	0.7224	0.6595	0.5424
Injury C	0.5487	0.6054	0.6595	0.9357	0.6760
PDO	0.4638	0.4998	0.5424	0.6760	0.6651

Table E4 Posterior Means of the Correlation Matrix of the Latent Effects

	Fatal	**Injury A**	**Injury B**	**Injury C**	**PDO**
Fatal	1.0000				
Injury A	0.7035	1.0000			
Injury B	0.6904	0.7203	1.0000		
Injury C	0.7030	0.7297	0.8035	1.0000	
PDO	0.7043	0.7152	0.7834	0.8575	1.0000

Recall that MVP regression models suggested by other researchers (e.g., Ma and Kockelman, 2006) are very restrictive in the sense that they assume the covariances for different severity levels are all identical and non-negative as well as no over-dispersion is found in the data. On the other hand, the new MVPLN models that can be implemented by MCMC allow for a fully general correlation structure as well as allow for over-dispersion in the crash data. From Tables 6.51 and 6.52, it can be observed that there is a positive correlation between each of the latent effects in the crash counts of five severity types but the correlations for different severity levels are not identical. Thus, as with any statistical models, this correlation needs to be incorporated in the estimation of the model.

References

Bijleveld, F. D. The Covariance Between the Number of Accidents and the Number of Victims in Multivariate Analysis of Accident Related Outcomes. *Accident Analysis and Prevention*, Vol. 37, No. 4, 2005, pp. 591-600.

Chib, S., and R. Winkelmann. Markov Chain Monte Carlo Analysis of Correlated Count Data. *Journal of Business & Economic Statistics*, Vol. 19, 2001, pp. 428-435.

Gilks, W. R., S. Richardson, D. J. Spiegelhalter. *Markov chain Monte Carlo in practice*. Chapman & Hall, London, 1996.

Karlis, D., and L. Meligkotsidou. Multivariate Poisson Regression with Covariance Structure, *Statistics and Computing*, Vol. 15, 2005, pp. 255-265.

Khattak, A.J., Injury Severity in Multivehicle Rear-End Crashes, *Transportation Research Record 1746*, 2001, pp. 59-68.

Kockelman, K.M. and Y. Kweon, Driver Injury Severity: An Application of Ordered Probit Models, *Accident Analysis & Prevention*, Vol. 34, No. 2, 2002, pp. 313-321.

Kweon Y., and K.M. Kockelman, Overall Injury Risk to Different Drivers: Combining Exposure, Frequency, and Service Models, *Accident Analysis & Prevention*, Vol. 35, No. 4, 2003, pp. 441-450.

Ma, J., and K. M. Kockelman. Bayesian Multivariate Poisson Regression for Models of Injury Count, by Severity. *Transportation Research Record 1950*, 2006, pp. 24-34.

Miaou, S.-P., and J. J. Song. Bayesian ranking of sites for engineering safety improvements: Decision parameter, treatability concept, statistical criterion and spatial dependence. *Accident Analysis and Prevention*, Vol. 37, No. 4, 2005, pp. 699-720.

Milton, J.C., V.N. Shankar, and F.L. Mannering. Highway Accident Severities and the Mixed Logit Model: An Exploratory Empirical Analysis, *Accident Analysis & Prevention*, 2007, in press, doi:10.1016/j.aap.2007.06.006

Park, E.S., and D. Lord. Multivariate Poisson-Lognormal Models for Jointly Modeling Crash Frequency by Severity. *Transportation Research Record 2019*, 2007, pp. 1-6.

Schluter, P. J., J. J. Deely, and A. J. Nicholson. Ranking and Selecting Motor Vehicle Accident Sites by Using a Hierarchical Bayesian Model, *The Statistician*, Vol. 46, No. 3, 1997, pp. 293-316.

SAS. Version 9 of the SAS System for Windows. SAS Institute Inc., Cary, NC, 2002.

Song, J. J., M. Ghosh, S. Miaou, and B. Mallick. Bayesian Multivariate Spatial Models for Roadway Traffic Crash Mapping. *Journal of Multivariate Analysis*, Vol. 97, 2006, pp. 246-273.

The MathWorks, Inc. MATLAB Neural Network Toolbox 5. Natick, MA, 2006.

Tierney, L. Markov Chains for Exploring Posterior Distributions. *Annals of Statistics*, Vol. 22, No. 4, 1994, pp. 1701-1762.

Tsionas, E. G. Bayesian Multivariate Poisson Regression. *Communications in Statistics—Theory and Methods*, Vol. 30, No. 2, 2001, pp. 243-255.

Tunaru, R. Hierarchical Bayesian Models for Multiple Count Data, *Austrian Journal of Statistics*, Vol. 31, No. 2&3, 2002, pp.221-229.

APPENDIX F

Methodology for Estimating the Variance and Confidence Intervals for the Estimate of the Product of Baseline Models and AMFs

This document was originally prepared by Dr. Dominique Lord. It has been published as a peer-reviewed publication in Accident Analysis & Prevention (in press, as of February, 2008):

INTRODUCTION

This appendix describes a methodology for estimating the variance and 95% confidence intervals (CI) for the estimate of the product between baseline models and Accident Modification Factors (AMFs). This methodology is provided for the upcoming Highway Safety Manual (HSM) (see Hughes et al., 2005, for additional information) currently in development in the United States (U.S.). The HSM, which is near completion, is a document that will serve as a tool to help practitioners make planning, design, and operations decisions based on safety. The document will serve the same role for safety analysis that the Highway Capacity Manual (HCM) (TRB, 2000) serves for traffic-operations analyses. The purpose of the HSM is to provide the best factual information and tools in a useful and widely accepted form to facilitate explicit consideration of safety in the decision making process.

The technique described in the HSM consists of first developing baseline models using data that meet specific nominal conditions, such as 12-ft lane and 8-ft shoulder widths for segments or no turning lanes for intersections. These conditions usually reflect design or operational variables most commonly used by state transportation agencies (defined as state DOTs). Consequently, baseline models typically only include traffic flow as covariates (e.g., $\mu = \beta_0 F_1^{\beta_1} F_2^{\beta_2}$). The second component of the technique consists of multiplying the output of such models by AMFs to capture changes in geometric design and operational characteristics (Hughes et al., 2005). An important assumption about using this technique is that the AMFs are considered independent, which may not always be true in practice. The formulation of the technique is given by the following:

$$\mu_{final} = \mu_{baseline} \times AMF_1 \times \ldots \times AMF_n \tag{1}$$

Where,

μ_{final} = final predicted number of crashes per unit of time;

$\mu_{baseline}$ = baseline predicted number of crashes per unit of time (via a regression model); and

$AMF_1 \times \ldots \times AMF_n$ = accident modification factors assumed to be independent.

Recent discussions at various meetings related to the production of the Manual have shown that estimating the uncertainty associated with baseline models, AMFs, and the estimate of the product between the two have become very important in the eyes of the Task Force members, the committee responsible for the implementation of the Manual, as well as for potential HSM users. So far, the work in this area has only focused on estimating the uncertainty associated with AMFs (Bahar et al., 2007) and, to a lesser degree, with regression models (Wood, 2005; Agrawal and Lord, 2006) (the latter not in the context of the HSM however). There is therefore a need to fill this gap by providing a methodology for estimating the variance of the estimate of the product of baseline models and AMFs.

This appendix is divided into three sections. The second section describes the first part of the methodology, which consists of estimating the variance of the product of baseline models and AMFs. This section covers the background material on the product of random materials and presents two examples describing different combinations of baseline models, AMFs, and their associated uncertainties. The third section describes the second part of the methodology and explains how to compute the variance of baseline models. An example illustrates the computation of the variance of baseline models as well as the 95% CI for the final product.

ESTIMATING THE VARIANCE OF BASELINE MODELS AND AMFs

This section describes the first part of the methodology and is divided into two sub-sections. The first sub-section provides details about the theory behind the multiplication of independent random variables. The second sub-section presents the application of the proposed method for estimating the variance of the estimate of the product of baseline models and AMFs. Two examples are provided.

Computing the Product of Random Variables

The estimation of the variance can be accomplished using the theory behind the multiplication of independent random variables (Ang and Tang, 1975; Browne, 2000). This theory states that the equations presented below will be exactly independent of the type of distribution to which each random variable belongs. For the purpose of this description, we will define z as the product of independent random variables:

$$z = x_1 x_2 x_3 \cdots \tag{2}$$

where,

z = the product of independent random variables; and
$x's$ = random variable taken from any kind of distribution.

It should be pointed out that the mean and variance estimates are defined as $E[x] = \lambda$ and $E\left[(x-\lambda)^2\right] = \nu$ (second central moment), respectively.

Mean of a Product

The mean of the product is the direct application of Equation (2):

$$\begin{aligned} z &= x_1 x_2 x_3 \cdots \\ E[z] &= E[x_1]E[x_2]E[x_3]\cdots \\ \lambda_z &= \lambda_{x1}\lambda_{x2}\lambda_{x3}\cdots \end{aligned} \tag{3}$$

The mean or average of the product is simply the product of the mean value of the random variables.

Variance of a Product

The variance of a product is obtained by taking the expectation square of z:

$$\begin{aligned} z^2 &= x_1^2 x_2^2 x_3^2 \cdots \\ E\left[z^2\right] &= E\left[x_1^2\right]E\left[x_2^2\right]E\left[x_3^2\right]\cdots \\ \left(\lambda_z^2+\nu_z\right) &= \left(\lambda_{x1}^2+\nu_{x1}\right)\left(\lambda_{x2}^2+\nu_{x2}\right)\left(\lambda_{x3}^2+\nu_{x3}\right)\cdots \end{aligned} \tag{4}$$

Note: $E\left[x^n\right]=E\left[\left(\left(x-\lambda\right)+\lambda\right)^n\right]$;

for $n=2$, $E\left[x^2\right]=E\left[\left(\left(x-\lambda\right)+\lambda\right)^2\right]=E\left[\left(x-\lambda\right)^2\right]+E\left[2\lambda\left(x-\lambda\right)\right]+E\left[\lambda^2\right]=\nu+\lambda^2$.

The variance ν_z is computed by first calculating the product on the right hand side and then subtracting the square of the mean λ_z^2 computed in Equation (4):

$$\nu_z=\left(\lambda_{x1}^2+\nu_{x1}\right)\left(\lambda_{x2}^2+\nu_{x2}\right)\left(\lambda_{x3}^2+\nu_{x3}\right)\cdots-\lambda_z^2 \tag{5}$$

Note that if all $\nu_x's$ equal zero, the variance ν_z will also equal zero:

$$\begin{aligned} \nu_z &= \left(\lambda_{x1}^2\right)\left(\lambda_{x2}^2\right)\left(\lambda_{x3}^2\right)\cdots-\lambda_z^2 \\ \nu_z &= \lambda_z^2-\lambda_z^2=0 \end{aligned} \tag{6}$$

Application of the Theory

This section describes the application of the theory behind the multiplication of independent random variables. Two examples describing different values of predicted values, AMFs, and associated uncertainties are presented. The uncertainty associated with the baseline models can be estimated using the method described in the next section. For estimating the uncertainty related to AMFs, the reader is referred to the work of Bahar et al. (2007), which will be incorporated into the forthcoming HSM.

Example 1 – One AMF

This example shows the application of a single AMF. Let x_1 represent the predicted value of a baseline model and x_2 an AMF:

$x_1 = 5.0$ crashes/year (standard deviation or SD = 2.0 crashes/year)
$x_2 = 0.80$ (SD = 0.10)

The mean is given by:

$$\lambda_z = 5.0 \times 0.80 = 4.0$$

The variance is given by:

$$v_z = \left(5.0^2 + 4.0\right)\left(0.80^2 + 0.01\right) - 4.0^2$$
$$v_z = (29)(0.65) - 16.0 = 18.85 - 16.0 = 2.85$$

The final value is estimated to be:

4.0 crashes/year (SD = 1.69 crashes/year)

Example 2 – Two AMFs

In this example, two AMFs are used. Let x_1 represent the predicted value of a baseline model and x_2 and x_3 independent AMFs:

$x_1 = 5.0$ crashes/year (stand. dev. = 2.0 crashes/year)
$x_2 = 0.95$ (SD = 0.10)
$x_3 = 0.90$ (SD = 0.20)

The mean is given by:

$$\lambda_z = 5.0 \times 0.90 \times 0.95 = 4.275$$

The variance is given by:

$$v_z = \left(5.0^2 + 4.0\right)\left(0.95^2 + 0.01\right)\left(0.90^2 + 0.04\right) - 4.275^2$$
$$v_z = (29)(0.912)(0.85) - 18.276 = 22.481 - 18.276 = 4.205$$

The final value is estimated to be:

4.275 crashes/year (SD = 2.05 crashes/year)

It should be pointed out that adding (independent) AMFs in Equation (1) increases the uncertainty associated with the final estimate, especially if the uncertainty for each AMF is large. This will have an influence for the comparison of different highway design alternatives based on safety, as discussed below. The next section explains how to estimate the variance associated with baseline models.

ESTIMATING THE VARIANCE OF BASELINE MODELS

This section briefly describes the second part of the methodology, which consists of estimating the variance of predictive baseline models. It should be pointed out that the methodology is not limited to baseline models, since it can also be applied to any type of regression models (e.g., general ADT models, models with covariates, etc.). The first sub-section describes the basic method for estimating the variance. The second sub-section presents an application of the method and provides details about estimating 95% CI after the product of baseline models and AMFs.

Method for Computing the Variance

There are difference methods for estimating the variance (and confidence intervals) of predicted values generated from generalized linear models (Cameron and Trivedi, 1998; Myers et al., 2002). The most recent and relevant method has been proposed by Wood (2005), who specifically developed a procedure for computing the variance and CI for crash prediction models. He worked out the procedure for both Poisson and Poisson-gamma models. Table F1 shows the variance of the expressions for estimates of the Poisson mean μ, gamma mean m, and predicted response y for Poisson and Poisson-gamma models. This table shows that confidence intervals used to estimate the uncertainty of the gamma mean and the predicted response for the Poisson-gamma model both incorporate the inverse dispersion parameter, ϕ. It should be pointed out that in most cases the predicted response will be of interest, since the model will be applied to an observation (or site) that was not used for developing the statistical models.

Table F1 Variance Estimation for Poisson and Poisson-gamma Models (Wood, 2005)

Parameter	Variance		
		Poisson	Poisson-gamma
μ	ν_{μ}	$\hat{\mu}^2 Var(\hat{\eta})$	$\hat{\mu}^2 Var(\hat{\eta})$
m	ν_m	----	$\left\{\hat{\mu}^2 \operatorname{var}(\hat{\eta}) + \frac{\hat{\mu}^2 \operatorname{var}(\hat{\eta}) + \hat{\mu}^2}{\phi}\right\}$
y	ν_y	$\left\{\hat{\mu}^2 Var\left(\hat{\eta}\right) + \hat{\mu}\right\}$	$\left\{\hat{\mu}^2 Var(\hat{\eta}) + \frac{\hat{\mu}^2 Var(\hat{\eta}) + \hat{\mu}^2}{\phi} + \hat{\mu}\right\}$

Note:

$Var(\hat{\eta}) = \mathbf{x}_0 (\mathbf{XWX}')^{-1} \mathbf{x}_0'$

$\hat{\mu}$ = mean estimate of the Poisson or Poisson-gamma model (see Equation 1 above)

ϕ = inverse dispersion parameter of the Poisson-gamma model

ν_{μ} = variance of the Poisson mean

ν_m = variance of the gamma mean

ν_y = variance of the predicted response

Application of the Method

For this application, data collected at 800 4-legged signalized intersections in Toronto were used (a subset of the original data found in Lord, 2000). The data included intersection and intersection-related crashes as well as entering flows at major and minor approaches for the year 1995. These data have been used extensively by previous researchers and have been found to be of relatively good quality (Lord, 2000; Miaou and Lord, 2003; Miranda-Moreno and Fu, 2007). Table F2 describes the summary statistics of the data.

Table F2 Summary Statistics

Variable	Minimum	Maximum	Average (Std Dev)
F_{Major} (AADT)	5,296	72,310	27,471 (14,299)
F_{Minor} (AADT)	52	42,644	10,768 (1,779)
Crashes per year	0	44	9.49 (7.82)

In this application, the model developed is classified as a general ADT model. These kinds of model are flow-only models estimated using average conditions reflected in the database. General ADT models have the same functional form as baseline models and are given by the following:

$$\mu = \beta_0 F_{Major}^{\beta_1} F_{Minor}^{\beta_2} \tag{7}$$

Where,

μ = the mean number of crashes per year;
F_{Major} = entering vehicle per day (AADT) for the major approaches;
F_{Minor} = entering vehicle per day (AADT) for the minor approaches; and
$\beta_0, \beta_1, \beta_2$ = regression coefficients.

As reported by Miaou and Lord (2003), the functional form described above is not the most adequate for describing the relationship between crashes and exposure since the form does not appropriately fit the data near the boundary conditions (i.e., when $F \rightarrow 0$ and for $F_{\max}$). Nonetheless, it is still relevant for this application, as it is considered an established functional form in the highway safety literature. In addition, the most adequate functional form proposed by Miaou and Lord (2003), a model with two distinct mean functions, cannot be estimated via a generalized linear modeling (GLM) framework, as it was done herein.

Tables F3 and F4 summarize the modeling output and variance-covariance matrix, $(\mathbf{XWX'})^{-1}$, respectively. The data were fitted using a Poisson-gamma regression model with a fixed inverse dispersion parameter (e.g., $Var(Y) = \mu + \mu^2/\phi$) [see Miaou and Lord (2003), and Mitra and Washington (2007) about this assumption]. The coefficients of the model were estimated using Genstat (Payne, 2000).

Table F3 Model Output

Variable	Coefficients (std. err.)
Intercept ($\ln \beta_0$)	-11.120 (0.204)
F_{Major} (β_1)	0.558 (0.020)
F_{Minor} (β_2)	0.650 (0.009)
Inverse dispersion parameter (ϕ)	6.46 (0.22)
Goodness-of-fit Statistics	Deviance: 5602 (F=3,807) Deviance Ratio: 1.076

Table F4 Variance-Covariance Matrix

Variable	Intercept	F_{Major}	F_{Minor}
Intercept	0.041513	-0.003735	-0.000377
F_{Major}	-0.003735	0.0000404	-0.000043
F_{Minor}	-0.000377	-0.000043	0.000089

In this application, we can assume that the signalized intersection investigated, which has not been used for developing the model, has the following characteristics: F_{Major} = 35,500 veh/day ($\ln 35{,}500 = 10.48$) and F_{Minor} = 5,000 veh/day ($\ln 5{,}000 = 8.52$). The mean number of crashes is estimated to be 1.30 crashes per year ($\mu = 0.00001481 \times 35{,}500^{0.558} \times 5{,}000^{0.650}$).

The variance $Var(\hat{\eta})$ can be computed the following way:

$$Var(\hat{\eta}) = \mathbf{x}_0 (\mathbf{XWX'})^{-1} \mathbf{x}_0'$$
$$Var(\hat{\eta}) = \begin{pmatrix} 1 & 10.48 & 8.52 \end{pmatrix} \begin{pmatrix} 0.041513 & -0.003735 & -0.0000377 \\ -0.003735 & 0.000404 & -0.000043 \\ -0.0000377 & -0.000043 & 0.000089 \end{pmatrix} \begin{pmatrix} 1 \\ 10.48 \\ 8.52 \end{pmatrix} \quad (8)$$
$$Var(\hat{\eta}) = 0.000166$$

Using the value of $Var(\hat{\eta})$ computed above, the variance for the Poisson mean μ, gamma mean m, and predicted response y can be calculated using the equations in Table F1 (using the column for Poisson-gamma models). The results are shown in Table F5.

Table F5 Estimated Variance for the Poisson Mean, gamma Mean, and Predicted Response

Parameter	Variance	
μ	ν_μ	0.000281
m	ν_m	0.2622
y	ν_y	1.5639
Note: $\hat{\mu} = 1.30$ $Var(\hat{\eta}) = 0.000166$		

Now, let us assume that two AMFs (i.e., hypothetical values) can be used to estimate changes associated with the introduction of left-turn and right-turn lanes, both located on the major approaches (both sides), respectively: AMF_1 = 0.90 (SD = 0.05) and AMF_2 = 0.95 (SD = 0.10). Also assume that the model represents a baseline model, where the sites do not have right- and left-turning lanes. Using Equations (4) and (5), one can, for instance, estimate the predicted value and associated uncertainty the following way [note: we changed the notation from μ_{final}, as shown in Equation (1), to λ_z for this example]:

$$\lambda_z = \mu \times \lambda_{AMF_1} \times \lambda_{AMF_2}$$
$$\lambda_z = 1.30 \times 0.90 \times 0.95 = 1.1115 \quad (9)$$

$$\nu_{zy} = \left(\mu^2 + \nu_y\right)\left(\lambda^2_{AMF_1} + \nu_{AMF_1}\right)\left(\lambda^2_{AMF_2} + \nu_{AMF_2}\right) - \lambda^2_z$$
$$\nu_{zy} = \left(1.30^2 + 1.564\right)\left(0.90^2 + 0.0025\right)\left(0.95^2 + 0.01\right) - 1.1115^2 \qquad (10)$$
$$\nu_{zy} = (3.2540)(0.8125)(0.9125) - 1.2354 = 1.1771$$

[Note: ν_{zy} denotes the variance was estimated using the variance associated with the predicted response of the baseline model (ν_y) and the variance of AMFs.]

Therefore, the estimated value after the AMFs are multiplied with the baseline model output is 1.112 crashes/year for the signalized intersection under investigation. The standard deviation of the estimate of the product is equal to 1.084.

If the 95% CI is the value of central interest, it can be computed using the equations listed in Table F6 (see Wood, 2005, for additional information). Confidence intervals can be used for screening highway design alternatives based on safety (e.g., Kononov and Allery, 2004) or identifying hazardous sites (Miranda-Moreno et al., 2005). For the value above, the 95% (predicted) confidence interval is located between 0 and 5 $\left(\left[0, \left\lfloor 1.112 + \sqrt{19 \times 1.177} = 5.841 \right\rfloor\right]\right)$. In this example, the confidence interval boundaries appear to be very wide given the estimated value, but the range is what one would expect for the predicted response (see Wood, 2005; Agrawal and Lord, 2006; and Geedipally and Lord, 2008, for other examples). As noted in Wood (2005), tighter CI for y could be computed using the equations presented in Appendix A of his paper when $\lambda_z \leq 1$.

Table F6 95% Confidence Interval for the Estimate of the Product of Baseline Model and AMFs (adapted from Wood, 2005)

Parameter	Intervals
Poisson Model	
μ	$\left[\dfrac{\lambda_z}{e^{1.96\sqrt{v_{z\mu}}}}, \lambda_z e^{1.96\sqrt{v_{z\mu}}}\right]$
y	$\left[0, \left\lfloor \lambda_z + \sqrt{19 v_{zy}} \right\rfloor\right]$
Poisson-gamma Model	
μ	$\left[\dfrac{\lambda_z}{e^{1.96\sqrt{v_{z\mu}}}}, \lambda_z e^{1.96\sqrt{v_{z\mu}}}\right]$
m	$\left[\max\left\{0, \lambda_z - 1.96\sqrt{v_{zm}}\right\},\ \lambda_z + 1.96\sqrt{v_{zm}}\right]$
y	$\left[0, \left\lfloor \lambda_z + \sqrt{19 v_{zy}} \right\rfloor\right]$

Note:

- λ_z = baseline model output × AMF_1 × …. × AMF_n or $\lambda_z = \mu_{baseline} \times \lambda_{AMF_1} \times \ldots \times \lambda_{AMF_n}$
- $v_{z\mu}$ = variance estimated from the variance of the Poisson mean (v_μ) (1st row in Table 1) and the variance of AMFs.
- v_{zm} = variance estimated from the variance of the gamma mean (v_m) (2nd row in Table 1) and the variance of AMFs.
- v_{zy} = variance estimated from the variance of the predicted response (v_y) (3rd row in Table 1) and the variance of AMFs; this is dependent upon whether a Poisson or a Poisson-gamma model was used for estimating the baseline model.
- $\lfloor x \rfloor$ denotes the largest integer less than or equal to x.
- For computing any other % CI for y, $\sqrt{19 v_{zy}}$ can be substituted with $\sqrt{\dfrac{1.0-\alpha}{\alpha} v_{zy}}$, where α = 1 – percentile in %; For instance, for estimating the upper boundary 90% CI, the upper value becomes $\sqrt{\dfrac{1.0-0.1}{0.1} v_{zy}} = \sqrt{9 v_{zy}}$.

SUMMARY

This manuscript has presented a methodology for computing the variance and 95% CI for the estimate of the product between baseline models and AMFs when the uncertainty is known both for models and the modification factors. The first section explained the technique proposed in the HSM to estimate the safety performance of segments or intersections using baseline models and AMFs. The second section described the first part of the methodology. This section covered the background material on the product of independent random variables. The third section presented the second part of the methodology, which focused on estimating the variance of baseline models. This section also explained how to compute the 95% CI when the product involved the estimate of the Poisson mean, gamma mean, and predicted response.

ACKNOWLEDGMENTS

The author would like to thank Drs. Simon Washington and Graham Wood for providing very useful comments for an earlier draft of this document. The paper also benefited from the input of two anonymous reviewers. This work was performed as part of the research project NCHRP 17-29 funded by the National Academy of Science.

REFERENCES

Agrawal, R., and D. Lord (2006) Effects of Sample Size on the Goodness-of-fit Statistic and Confidence Intervals of Crash Prediction Models Subjected to Low Sample Mean Values. *Transportation Research Record 1950*, pp. 35-43.

Ang, A. H.-S., and W.H. Tang (1975) Probability Concepts in Engineering Planning and Design: Basic Principles (Vol. 1), John Wiley & Sons, New York, N.Y.

Bahar, G., M. Parkhill, E. Hauer, F.M. Council, B. Persaud, C. Zegger, R. Elvik, A. Smiley, and B. Scotty (2007) Prepare Parts I and II of the Highway Safety Manual. Final Report for NCHRP Project 17-27, iTRANS Consulting, Richmond Hill, Ont. (note: future Parts I and II of the HSM)

Browne, J.M. (2000) Probabilistic Design. Class Notes, School of Engineering and Science Swinburne University of Technology, Hawthorn, Australia. http://www.ses.swin.edu.au/homes/browne/probabilisticdesign/coursenotes/notespdf/4ProbabilisticDesign.pdf

Cameron, A.C., and P.K. Trivedi (1998) Regression Analysis of Count Data. Cambridge University Press, Cambridge, U.K.

Geedipally, S.R., and D. Lord (2008) Effects of the Varying Dispersion Parameter of Poisson-gamma Models on the Estimation of Confidence Intervals of Crash Prediction Models. Paper presented at the 87th Annual Meeting of the Transportation Research Board, Washington, D.C.

Hughes, W., K. Eccles, D. Harwood, I. Potts, and E. Hauer (2005) Development of a Highway Safety Manual. Appendix C: Highway Safety Manual Prototype Chapter: Two-Lane Highways. NCHRP Web Document 62 (Project 17-18(4)). Washington, D.C. (http://onlinepubs.trb.org/onlinepubs/nchrp/nchrp_w62.pdf accessed October 2007)

Kononov, J., and B.K. Allery (2004) Explicit Consideration of Safety in Transportation Planning and Project Scoping. *Transportation Research Record 1897*, pp. 116-125.

Lord, D. (2000) The Prediction of Accidents on Digital Networks: Characteristics and Issues Related to the Application of Accident Prediction Models. Ph.D. Dissertation. Department of Civil Engineering, University of Toronto, Toronto, Ontario.

Miaou, S.-P., and D. Lord (2003) Modeling Traffic-Flow Relationships at Signalized Intersections: Dispersion Parameter, Functional Form and Bayes vs Empirical Bayes. *Transportation Research Record 1840*, pp. 31-40.

Miranda-Moreno, L.F., and L. Fu (2007) Traffic Safety Study: Empirical Bayes or Full Bayes? Paper 07-1680. Presented at the 84th Annual Meeting of the Transportation Research Board, Washington, D.C.

Miranda-Moreno, L.F., L. Fu, F.F. Saccomanno, and A. Labbe (2005) Alternative risk models for ranking locations for safety improvement. *Transportation Research Record 1908*, pp. 1-8.

Mitra, S., and S. Washington (2007) On the nature of over-dispersion in motor vehicle crash prediction models. *Accident Analysis & Prevention*, Vol. 39, No. 3, pp. 459-568.

Myers, R.H., D.C. Montgomery, and G.G. Vining (2002) Generalized Linear Models: with Applications in Engineering and the Sciences. Wiley Publishing Co., New York, N.Y.

Payne, R.W. (ed.), 2000. The Guide to Genstat. Lawes Agricultural Trust, Rothamsted Experimental Station, Oxford, U.K.

TRB (2000) Highway Capacity Manual 2000, Transportation Research Board, Washington, D.C.

Wood, G.R. (2005) Confidence and prediction intervals for generalized linear accident models. *Accident Analysis & Prevention*, Vol. 37, No. 2, pp. 267-273.