Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
FHWA/TX-07/0-4703-4		

4. Title and Subtitle	5. Report Date
DEVELOPMENT OF TOOLS FOR EVALUATING THE SAFETY IMPLICATIONS OF HIGHWAY DESIGN DECISIONS	September 2006 Published: February 2007
	6. Performing Organization Code
7. Author(s) J. Bonneson, D. Lord, K. Zimmerman, K. Fitzpatrick, and M. Pratt	8. Performing Organization Report No. Report 0-4703-4
9. Performing Organization Name and Address Texas Transportation Institute The Texas A&M University System College Station, Texas 77843-3135	10. Work Unit No. (TRAIS) 11. Contract or Grant No. Project 0-4703
12. Sponsoring Agency Name and Address Texas Department of Transportation Research and Technology Implementation Office P.O. Box 5080 Austin, Texas 78763-5080	13. Type of Report and Period Covered Technical Report: September 2003-August 2006 14. Sponsoring Agency Code

15. Supplementary Notes

Project performed in cooperation with the Texas Department of Transportation and the Federal Highway Administration.

Project Title: Incorporating Safety Into the Highway Design Process

URL: http://tti.tamu.edu/documents/0-4703-4.pdf

16. Abstract

Highway safety is an ongoing concern to the Texas Department of Transportation (TxDOT). As part of its proactive commitment to improving highway safety, TxDOT is moving toward including quantitative safety analyses earlier in the project development process. The objectives of this research project are: (1) the development of safety design guidelines and evaluation tools to be used by TxDOT designers, and (2) the production of a plan for the incorporation of these guidelines and tools in the planning and design stages of the project development process.

This document summarizes the research conducted and the findings for the initial three years of the project. This research included a review of the TxDOT design and safety evaluation process, identification of the safety information sources and needs, identification of the data needed to use selected safety evaluation tools, assessment of the applicability of accident modification factors for design evaluation, and calibration of selected safety evaluation tools for Texas application.

17. Key Words	18. Distribution Statement
Highway Safety, Highway Design, Safety Management, Geometric Design	No restrictions. This document is available to the public through NTIS: National Technical Information Service Springfield, Virginia 22161 http://www.ntis.gov

19. Security Classif.(of this report)	20. Security Classif.(of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	144	

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

DEVELOPMENT OF TOOLS FOR EVALUATING THE SAFETY IMPLICATIONS OF HIGHWAY DESIGN DECISIONS

by

J. Bonneson, P.E.
Research Engineer
Texas Transportation Institute

D. Lord, P.E.
Assistant Professor
Texas A&M University

K. Zimmerman, P.E.
Assistant Research Engineer
Texas Transportation Institute

K. Fitzpatrick, P.E.
Research Engineer
Texas Transportation Institute

and

M. Pratt
Assistant Transportation Researcher
Texas Transportation Institute

Report 0-4703-4
Project 0-4703
Project Title: Incorporating Safety Into the Highway Design Process

Performed in cooperation with the
Texas Department of Transportation
and the
Federal Highway Administration

September 2006
Published: February 2007

TEXAS TRANSPORTATION INSTITUTE
The Texas A&M University System
College Station, Texas 77843-3135

DISCLAIMER

NOTICE

ACKNOWLEDGMENTS

This research project was sponsored by the Texas Department of Transportation and the Federal Highway Administration. The research was conducted by Dr. James Bonneson, Dr. Dominique Lord, Dr. Kay Fitzpatrick, Dr. Karl Zimmerman, and Mr. Michael Pratt. Dr. Lord is with Texas A&M University. The other researchers are with the Texas Transportation Institute.

The researchers would like to acknowledge the support and guidance provided by the Project Monitoring Committee:

- Ms. Aurora (Rory) Meza, Project Coordinator (TxDOT);
- Ms. Elizabeth Hilton, Project Director (TxDOT);
- Mr. David Bartz (FHWA);
- Mr. Mike Battles (TxDOT);
- Mr. Stan Hall (TxDOT);
- Mr. Richard Harper (TxDOT);
- Ms. Meg Moore (TxDOT); and
- Ms. Joanne Wright (TxDOT).

In addition, the researchers would like to acknowledge the valuable assistance provided by Dr. William Schneider IV during the development of several report chapters.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

Page

LIST OF FIGURES

LIST OF TABLES

LIST OF TABLES (Continued)

Table **Page**

CHAPTER 1. INTRODUCTION

OVERVIEW

There is a growing public demand for safer streets and highways. In response to this demand, state and national transportation agencies have developed safety programs that emphasize public education, accelerated highway renewal, community-sensitive street systems, and innovative technology to facilitate safe highway design.

Historically, information about the safety effect of a design component has been based on anecdotal evidence, laws of physics, before-after studies, or comparisons of site safety (i.e., sites with and without the design component). However, the accuracy of this information is suspect because of inherent random nature of crash data and the many factors (some of which pertain more to the driver and the vehicle than the roadway) that can lead to a crash at a specific location. As a result of this uncertainty, engineers have traditionally come to rely on design standards and policies to guide them in the design process, with the underlying premise that compliance with warrants and controls will yield a "safe" roadway.

In general, the safety experience with roadways built in compliance with warrants and controls has been good and the aforementioned premise largely validated. However, the weaknesses of this traditional design approach have become more apparent as traffic demands have increased over time, the performance of vehicles improved, and drivers became less patient. Points along the roadway having multiple, complex geometric components that tend to concentrate traffic and increase their interaction have started to show disproportionately high crash frequencies. Fortunately, in the past decade, new statistical analysis methods have been developed and the quality of crash data has improved. These advances have significantly increased both the accuracy and coverage of design-related safety information. Emerging technology is now making it possible to efficiently incorporate quantitative safety evaluations in the design process.

A significant amount of new safety information has been developed in recent years. The implementation of this information is now as pressing a problem as was the need for new research a decade ago. The forthcoming *Highway Safety Manual* (*HSM*) is expected to formalize the safety evaluation process; however, the *HSM* procedures will require local calibration to ensure that they accurately reflect the conditions and design practices in the jurisdiction for which they are used (*1*). It is likely that each agency will develop their own safety evaluation guidelines based on the *HSM* procedures, but it is also likely that they will tailor these guidelines to ensure their consistency with agency design policies.

OBJECTIVE AND SCOPE

Highway safety concerns are also evident in Texas. Crashes in Texas continue to increase and currently exceed 300,000 per year. Nearly 3800 motorists die annually on Texas highways. As part of its proactive commitment to improving highway safety, the Texas Department of Transportation is moving toward including quantitative safety analyses throughout the project development process. This research project has as its objectives: (1) the development of safety design guidelines and evaluation tools to be used by TxDOT designers, and (2) the production of a plan for the incorporation of these guidelines and tools in the planning and design stages of the project development process.

The products of this research project are intended to address: (1) new construction, major reconstruction, and 3R (resurfacing, restoration, rehabilitation) activities, (2) all highway facilities (freeway, arterial, collector), and (3) all highway travel modes.

RESEARCH APPROACH

A six-year program of research was developed to satisfy the stated objectives. The research approach consists of ten tasks that represent a logical sequence of needs assessment, research, evaluation, and workshop development. Six tasks have been undertaken in the first three years of the research; they include:

- Review the TxDOT design and safety evaluation processes,
- Identify safety information sources and needs,
- Determine the data needed for selected safety evaluation tools,
- Evaluate use of accident modification factors for design evaluation,
- Determine calibration factors for Texas application of safety prediction models, and
- Develop a *Roadway Safety Design Synthesis* and an *Interim Roadway Safety Design Workbook* (*2, 3*).

DEFINITIONS

This part of the chapter defines several terms related to highway safety. The definitions offered are consistent with their use in the safety-related literature; however, they may be refined for consistency with TxDOT design practice and the objectives of this research project.

Accident modification factor (AMF) is a constant or equation that represents the change in safety following a change in the character of a segment (or intersection). An AMF can be computed as the ratio $N_w/N_{w/o}$, where N_w represents the expected number of crashes experienced by a highway segment *with* one or more specified design components and $N_{w/o}$ represents the expected number of crashes experienced by the same segment *without* the specified components. AMFs are often used as multiplicative factors to adjust the estimate obtained from a safety prediction model to a value that reflects the safety of a specific segment.

AMFs typically range in value from 0.5 to 2.0, with a value of 1.0 representing no effect on safety. AMFs less than 1.0 indicate that the specified component is associated with fewer crashes.

To illustrate the concept of AMF, consider a road segment that has an expected crash frequency of 3.0 crashes/yr. A change is made to the road cross section and, after a period of time, a follow-up evaluation indicates that the change resulted in an expected crash frequency of 4.0 crashes/yr. The AMF for this change is 1.3 (= 4.0/3.0).

As a second illustration, consider that a safety prediction model is used to estimate the expected crash frequency of a typical two-lane highway with a specified average daily traffic volume (ADT) and length. The model was developed to reflect the following as "typical": 12-ft lanes, 6-ft shoulders, no grade, no horizontal curves, 10-ft horizontal clearance, 1V:4H side slope, and no vertical grades. This model estimates an expected crash frequency of 5.0 crashes/yr for the "typical" road segment. It is desired to estimate the crash frequency of a specific road segment for which all geometric elements are "typical" except that the clear zone is 20 ft wide. The AMF for horizontal clearance has a value of 0.93 when the clearance distance is 20 ft. Thus, the expected crash frequency for the specific road segment is estimated as 4.6 crashes/yr (= 5.0 × 0.93).

Crash reduction factor (CRF) is a constant that represents the proportion of crashes reduced as a result of a safety improvement at a specific location or along a specific road segment. CRFs typically range in value from 0.10 to 0.90. Larger CRFs in this range indicate a more significant reduction in crashes due to the improvement. To illustrate, consider a road segment that has a crash frequency of 3.0 crashes/yr. An improvement is made to the road's cross section and, after a period of time passes, a follow-up evaluation indicates that the change resulted in a crash frequency of 2.0 crashes/yr. The CRF for this improvement is 0.33 (= [3.0 − 2.0]/3.0) representing a 33 percent reduction in crashes.

Injury crash is a crash wherein one or more of the persons involved is injured. The injury severity is reported as "possible," "non-incapacitating," or "incapacitating."

Safety (or "substantive safety") is the expected crash frequency associated with a segment (or intersection) for a given set of design components, traffic control devices, and exposure conditions (e.g., traffic volume, segment length). Given that crashes are random events and that conditions can change over time, the safety of a specific segment is best conceptualized as the long-run average of the crash frequencies reported for a large group of segments with similar features and traffic conditions.

Safety evaluation tool is, at its simplest level, a set of equations that can be used to predict: (1) the safety of a given segment (or intersection), and (2) the safety effect associated with a change in its design features. At this "simple" level, a tool is equivalent to a model. However, complex tools can incorporate additional analysis techniques. For example, complex tools can include techniques for incorporating the reported crash history of a specific segment to improve the accuracy of the safety prediction. Complex tools can also include techniques for evaluating alternative designs using safety and other data (e.g., benefit-cost analysis). Tools are sometimes represented in software to facilitate their application.

Safety prediction model is an equation, or set of equations, that can be used to estimate the safety of a typical segment (or intersection). The model includes factors related to crash risk and exposure. A figure or table is sometimes used to portray the relationship (instead of an equation). A model can be derived to include one or more AMFs. Models intended for practical application have one or more empirically based factors that require calibration to local conditions to ensure accurate predictions.

Safety surrogate is any statistic that is directly related to crash frequency or severity (e.g., conflicts) and that quantifies the relative risk of collision or injury.

REFERENCES

1. Hughes, W., K. Eccles, D. Harwood, I. Potts, and E. Hauer. *Research Results Digest 286: Development of a Highway Safety Manual*. National Cooperative Highway Research Program, Transportation Research Board, Washington, D.C., March 2004.

2. Bonneson, J.A., K. Zimmerman, and K. Fitzpatrick. *Roadway Safety Design Synthesis.* FHWA/TX-05/0-4703-P1. Texas Department of Transportation, Austin, Texas, November 2005.

3. Bonneson, J.A., K. Zimmerman, and K. Fitzpatrick. *Interim Roadway Safety Design Workbook*. FHWA/TX-06/0-4703-P4, Texas Department of Transportation, Austin, Texas, April 2006.

CHAPTER 2. SAFETY CONSIDERATIONS IN ROADWAY DESIGN

OVERVIEW

Many analyses of crash causation indicate that "driver error" contributes to about 85 percent of all crashes. However, this finding should not be construed to mean there is little that the highway designer can do to improve the level of safety afforded motorists. In fact, a highway designed with explicit attention to safety can significantly reduce the frequency of crashes and their severity. Such a highway is called "forgiving" because its design mitigates the consequences of driver error.

The level of safety provided by a roadway is directly linked to the extent to which safety was explicitly considered throughout the planning, design, and construction stages. Although a single design exception for a specific highway element or the use of a minimum design value may result in an acceptably small reduction in safety, the net effect of several exceptions or the use of several minimum values can create an unsafe design condition. The failure to consciously and consistently consider safety during the design process can lead to safety problems that are not apparent until *after* the highway is open to traffic (*1*). The solution to these problems then comes under the purview of the hazard elimination (HES) program, and often at considerable additional cost to the highway agency.

Safety-conscious design is not limited to the design of new alignments or those highways undergoing major reconstruction. The report *Designing Safer Roads - Practices for Resurfacing, Restoration, and Rehabilitation* (*2*) promotes the use of a safety-conscious design process. Specifically, its authors state, "Significant improvements in safety are not automatic by-products of 3R projects; safety must be systematically engineered into each project. To do this, highway designers must deliberately seek safety opportunities specific to each project and apply sound safety and traffic engineering principles. Highway agencies must strengthen safety considerations at each major step in the design process, treating safety as an integral part of design and not as a secondary objective." (*2*, p. 5).

This chapter consists of four parts. The first part provides a brief overview of the types of design projects and the process by which they are developed from concept to construction. The second part describes recent efforts to incorporate a safety-conscious approach into the design process. Examples of safety-conscious design practices are summarized. The third part provides a review of safety considerations in the TxDOT design process. The last part summarizes the important findings from the previous three sections.

DESIGN PROJECT TYPE AND DEVELOPMENT PROCESS

Project Development Process

All proposed roadway projects involve a detailed, formalized procedure of planning, design, and construction that is intended to ensure the completed roadway represents a net benefit to the motoring public, is economical to maintain, and can operate in a safe and efficient manner. This procedure is called the "project development process." It consists of six stages: planning and programming; preliminary design; environmental; right-of-way and utilities; plans, specifications, and estimate (PS&E); and letting. The planning and programming, preliminary design, and PS&E development stages are the focus of this research. The sequence of these stages in the development process is shown in Figure 2-1.

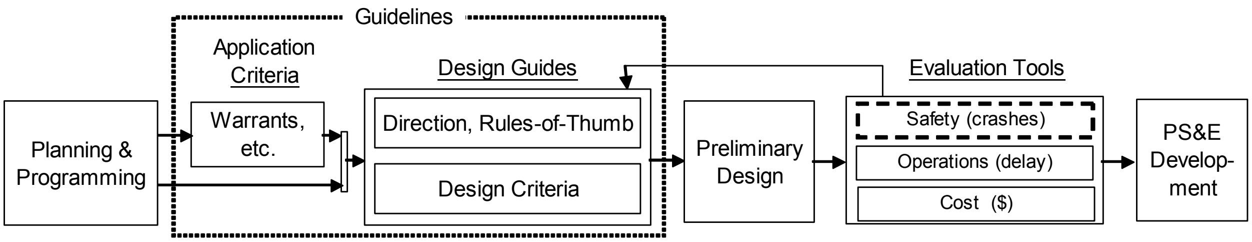


Figure 2-1. Components of the Project Development Process.

There are two stages of the project development process that embody the design process. These stages are: (1) Preliminary Design and (2) PS&E Development. During the preliminary design stage, the location of a facility (if it is new or being relocated) and its major design features are identified. Then, alternative locations and features are considered and the more promising ones are evaluated in greater detail.

The final design of the proposed facility is undertaken during the PS&E development stage. The environmental and right-of-way issues have generally been identified by the start of this stage and a "preferred" alignment has been identified. The product of this stage is a completed plan set with appropriate specifications for construction. Estimates of required quantities of materials needed for the construction bidding process are also provided.

The design process relies on the use of guidelines and evaluation tools to ensure the cost-effectiveness of the design feature (e.g., cross section), component (e.g., curve, tangent, etc.), and element (e.g., radius, lane width, etc.). Design guidelines typically consist of application criteria and design guides. Application criteria, when available, are consulted to determine when a configuration or element is viable. Application criteria represent a special type of guideline. They define conditions that justify, or warrant, the use of a design element and reflect a balance between operations, safety, comfort, convenience, and cost. The volume thresholds justifying left-turn lanes on two-lane highways in Table 3-11 of the *Roadway Design Manual* (*RDM*) (*3*) are an example of application criteria.

Design guides are used to design each alternative configuration to a level of detail sufficient to identify major cost items, environmental impacts, and right-of-way needs. These guides can include general design direction, rules-of-thumb, and criteria that limit design element sizes or their orientation. Through the adherence to these guides, the potential for the design to provide safe, efficient, comfortable, and convenient service in a cost-effective manner is greatly enhanced. Nevertheless, this performance cannot be confirmed unless a quantitative analysis is undertaken using various evaluation tools.

Evaluation tools are used by the designer to verify the performance potential of each configuration. The evaluation quantifies the design's performance in terms of the relevant measures of effectiveness (e.g., crash frequency, delay, cost, etc.). The objective of this evaluation is to ensure that the design offers a reasonable balance between cost and effectiveness. Tools are readily available to conduct level-of-service analyses and to estimate construction costs. Tools for quantifying an alternative's impact on safety are not as readily available. However, recent efforts by the Federal Highway Administration (FHWA) and the Transportation Research Board (TRB) have led to the development of new and better safety evaluation tools.

Design Project Types

Four project-type designations are used to describe the size and impact of a proposed roadway design project. These four project types are listed in Table 2-1.

Table 2-1. Design Project Types.

Project Type	Project Intent
New Location and Reconstruction	Enhanced mobility and safety that generally includes substantial changes in the geometry of the roadway.
Rehabilitation	Extend the service life and enhance the safety of a roadway.
Restoration or Resurfacing	Restore pavement structure, riding quality, or other necessary components, to their previous configuration.
Preventative Maintenance	Preserve, rather than improve, the structural integrity of the pavement, structure, or both.

Experience indicates that complexity and cost are highest for New Location and Reconstruction projects. They decrease significantly for Rehabilitation projects; decrease further for Restoration projects; and are the lowest for Preventative Maintenance projects. The complexity and cost of the first three project types listed are sufficient to justify consideration of these project types in the project development process. Preventative maintenance projects are not addressed in this process.

Facilities on new alignment are designed using criteria that are intended to provide operational efficiency, comfort, safety, and convenience for the motorist. Experience has shown that this practice typically yields roadways that provide an overall acceptable level of mobility and safety. These same criteria are also used when reconstructing roadway facilities and with similar results.

The reconstruction of existing facilities will not often be justified unless major realignment is needed. Resurfacing, restoration, and rehabilitation projects are often considered for these roadways because they enable highway agencies to improve the highway by selectively upgrading existing highway and roadside features without the cost of full reconstruction. One goal of these projects is to preserve the integrity of the pavement and bridge structures. Another goal is the identification and mitigation of potential safety and operational problems.

Two documents have been developed to encourage uniformity and consistency in highway design guidelines used by state departments of transportation (DOTs). One document, *A Policy on Geometric Design of Highways and Streets* (*Green Book*) (*4*), describes guidelines and criteria for New Location and Reconstruction projects. This document represents the basis for street and highway design used by most state DOTs.

A second document used by DOTs is *Special Report 214: Designing Safer Roads - Practices for Resurfacing, Restoration, and Rehabilitation* (*2*). This document describes recommended guidelines and criteria for Resurfacing, Restoration, and Rehabilitation projects. These guidelines are different from those described in the *Green Book* because of their greater focus on the cost-effective systemwide preservation of existing roadway infrastructure and the mobility they provide. A recent survey by Harwood et al. (*5*) indicates that almost all state DOTs have developed guidelines for these types of projects; however, few DOTs have directly adopted the criteria recommended in *Special Report 214.*

CONSCIOUS CONSIDERATION OF SAFETY

This part of the chapter summarizes the safety-conscious approach to highway planning and design. In the first section, the concepts of safety-conscious planning and design are described. In the second section, example guidelines that incorporate safety-conscious design are provided.

Overview of Safety-Conscious Planning and Design

In recent years, the safety provided by the Nation's highway system has come under increasing scrutiny. Federal legislation (i.e., TEA-21) required states and metropolitan planning organizations (MPOs) to emphasize projects that increase the safety of the transportation system for motorized and non-motorized users. To achieve further large reductions in crash frequency, research indicates that it will be necessary to continue to improve design policies and technologies that reduce the likelihood of a crash. The conscious consideration of safety in both the planning and design processes is one way to accomplish this goal. The relationship between safety-conscious planning and safety-conscious design is illustrated in Figure 2-2.

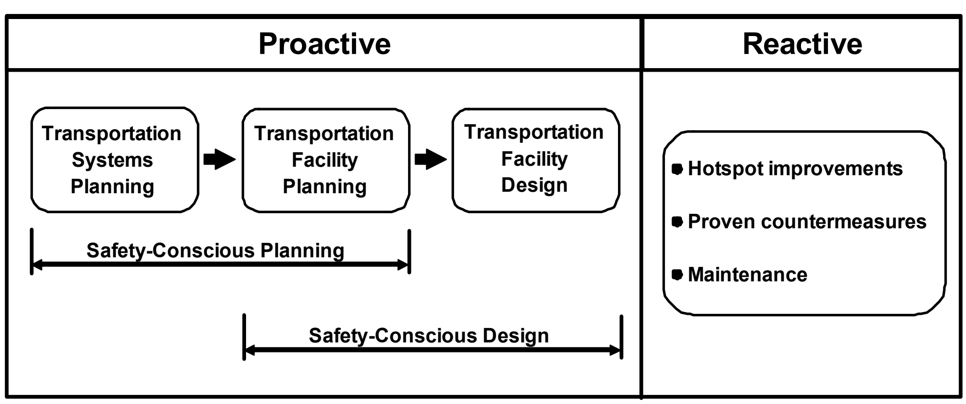


Figure 2-2. Safety-Conscious Planning and Design.

Safety-Conscious Planning

The FHWA Office of Planning, Environment, and Realty defines safety-conscious planning as "a proactive approach to the prevention of accidents and unsafe transportation conditions by establishing inherently safe transportation networks. Safety-conscious planning achieves road safety improvements through small quantum changes, targeted at the whole network" (*6*). According to Roberts (*7*), the goals of safety-conscious planning are to:

- Minimize exposure by providing a compact and efficient urban network;
- Minimize risk by minimizing intersection conflicts, separating travel modes, and using consistent (self-explaining) road forms; and
- Minimize the consequences of a crash by using road forms that reduce speeds at risky locations and providing efficient emergency response routes.

Safety-Conscious Design

Safety-conscious design represents the explicit evaluation of the safety consequences associated with design alternatives. It is incorporated at key points in the design process where changes necessary to accommodate safety considerations can be easily incorporated. The rationale for a safety-conscious design process is that the traditional approach of adherence to control values (i.e., minimum or maximum values) does not ensure a safe design, especially when consistently applied to each design element that comprises the highway segment or intersection. Historically, control values were developed to provide design guidance during a time of rapid growth in the highway system; a time when limited information was available about the safety consequences of design decisions. In these early days, limited data and engineering judgment were often subjectively combined to define critical control values that were rationalized to balance operational and safety benefits with the costs of construction and maintenance (*2*).

With the passage of time, new information about the safety effect of design elements has been acquired. Some design criteria have been updated in design policies and standards. However, there is increasing awareness that the use of design criteria alone is insufficient. The position taken by some transportation leaders is that the explicit consideration of design consistency, positive guidance, traffic flow efficiency, *and* safety is essential to ensuring an effective design (*1, 2, 8*). They also realized that explicitly balancing these considerations against the costs of constructing and maintaining the facility is necessary to obtain a cost-effective design.

The concept of "safety-conscious design" was first described in *Special Report 214* (*2*). More recently, the Transportation Association of Canada incorporated safety-conscious design in its design guide for new location and reconstruction projects (i.e., the *Geometric Design Guide for Canadian Roads*) (*8*). The justification offered for changing their design philosophy was the observation that: (1) the traditional approach to design has become less dependent on experience and judgment and more dependent on adherence to minimum criteria; (2) there is a belief among designers that safety is an automatic by-product of the design process; and (3) the difficulties associated with quantifying safety have relegated safety considerations to being only a secondary objective of the design process.

At its simplest level, the implementation of safety-conscious design involves the use of a "design domain" for key design elements. The design domain is effectively a lower and/or upper limit for a specific design element size. Domain limits are equivalent in concept to design criteria. However, domain limits are based on an evaluation of benefits and costs whereas design criteria tend to be based on experience, empirical evidence, and the physics of driver-vehicle-roadway interaction.

At a slightly higher level, safety-conscious design can be implemented by using safety evaluation tools to quantify the effect of alternative design choices on safety. At the highest level, safety-conscious design involves the use of evaluation tools and economic principles to evaluate the benefits and costs of design alternatives. In recognition of the time required to use the higher levels of safety-conscious design, the higher level evaluations tend to be reserved for more complex design conditions or those that involve high construction costs.

Hazard Elimination

Safety-conscious planning and design are complementary proactive processes intended to ensure that an acceptable level of safety is provided in a consistent and cost-effective manner. In contrast, efforts to improve the safety of collision-prone locations (i.e., "hot spots") constitute a reactive approach to achieving road safety. The FHWA's highway safety improvement program (HSIP) is an example of this approach. This program replaced FHWA's hazard elimination (HES) program as of October 2005. Each state is required to develop and implement a HSIP. They are also required to use benefit-cost analyses to evaluate and prioritize alternative safety improvement projects.

The HSIP is intended to reduce the number and severity of crashes at hazardous roadway locations. Typical projects include (*9*):

- intersection improvements (e.g., addition of channelization, traffic signal, etc.);
- pavement and shoulder widening;
- guardrail and barrier improvements;
- signing and pavement marking;
- breakaway utility poles and sign supports;
- pavement grooving and skid resistant overlays; and
- addition of shoulder rumble strips.

Summary

Collectively, the safety-conscious planning process, safety-conscious design process, and hazard elimination programs are intended to provide the mechanisms to ensure that a state's roadways are uniformly safe, efficient, and cost-effective. They reflect the thoughtful minimization of potential safety problems in future roadways, the development of designs that provide a uniform and affordable level of safety for new and existing facilities, and the immediate mitigation of safety problems at hazardous locations. The characteristics of these processes and programs are listed in Table 2-2.

Table 2-2. Characteristics of Safety-Conscious Planning and Design Processes.

Characteristic	Safety-Conscious Planning Process	Safety-Conscious Design Process	Hazard Elimination Program
Approach	Proactive	Proactive	Reactive[1]
Treatment Location	Network-wide	Facility	Spot
Localized Impact	Low	Moderate	High
Implementation	Long Term	Medium Term	Short Term

Note:
1 - Any project specifically implemented in response to crash experience or safety need is considered "reactive."

Examples of Safety-Conscious Design

This section provides a review of two authoritative reference documents that incorporate safety-conscious design principles. The first document reviewed is *Special Report 214: Designing Safer Roads - Practices for Resurfacing, Restoration, and Rehabilitation* (*2*). The second document reviewed is the *Geometric Design Guide for Canadian Roads* (*8*).

Special Report 214

In the mid-1980s, TRB commissioned a panel of experts to evaluate the 3R programs implemented by the state DOTs. One objective of this evaluation was to determine which minimum geometric criteria should be applied to 3R projects to both preserve the roadway and enhance safety. Among its many recommendations, the panel recommended that state DOTs develop a more safety-conscious design process for 3R projects. The steps they outlined for this process are listed in Table 2-3.

Two key tasks of safety-conscious design are to: (1) identify candidate geometric improvements to treat safety problems and (2) evaluate the cost, safety, and other impacts of each improvement. Together, these tasks represent a refinement to the traditional design process because the safety effect of alternative design elements and element sizes is explicitly quantified and compared to the associated construction cost. This activity is shown as the feedback loop between the application of Guidelines and Evaluation Tools in Figure 2-1. A similar type of loop is often used to evaluate the effect of alternative design elements on level of service and construction cost.

The types of improvements considered by state DOTs for 3R projects were identified by Harwood et al. (*5*) in a recent review of state 3R programs. The improvements that were most frequently identified are listed in Table 2-4. As indicated by the information in this table, improvements tend to involve shoulders, objects in the clear zone (i.e., horizontal clearance), and guardrail. The improvements identified with a plus (✚) symbol were identified by many states as not being considered unless there was clear evidence of a safety problem that could be mitigated by the improvement.

Table 2-3. Safety-Conscious Design Process for 3R Projects.

Step	Task
A. Assess Site Conditions Affecting Safety	1. Analyze accident and travel data. 2. Conduct a thorough site inspection. 3. Determine and verify existing geometry. 4. Determine prevailing speeds.
B. Determine Project Scope	1. Identify candidate geometric improvements to treat safety problems.
C. Document the Design Process	1. Identify applicable minimum design criteria.[1] 2. Describe design improvements considered. 3. Evaluate cost, safety, and other impacts of each improvement. 4. Identify proposed design exceptions. 5. Identify the recommended improvements in a report.
D. Review the Design by Traffic and Safety Engineers	1. Review the design report. 2. Review the design plans.

Note:
1 - Criteria likely to be considered would be those that are the subject of a design exception request, if exceeded.

Table 2-4. Typical Improvements Implemented in 3R Projects.

Category	Safety Improvement [1, 2]
Roadway Geometry	● Widen shoulders. ● Upgrade (buildup) shoulder structure. ● Add shoulders. ● Minor changes to horizontal and vertical alignment. ✚ ● Improve superelevation on horizontal curves.
Roadside Clear Zone (i.e., horizontal clearance)	● Increase clear zone. ● Remove objects from clear zone. ✚ ● Add guardrail and end treatments. ✚ ● Upgrade or adjust guardrail and end treatments. ✓ ● Flatten side slopes. ● Remove culvert headwalls. ✓ ● Install breakaway supports for roadside hardware.
Traffic Control Devices	● Install signs and markings. ✓
Intersection Geometry	● Add turn lanes at intersections.

Notes:
1 - ✚ improvements that many states do not implement unless there is a defined safety need.
2 - ✓ identified as "basic safety improvements" in the *Roadway Design Manual* (*3*).

Geometric Design Guide for Canadian Roads

The Transportation Association of Canada has recognized the need for a safety-conscious highway design process. This need was addressed in the production of the 1999 Edition of the *Geometric Design Guide for Canadian Roads* (*Guide*) (*8*). Historically, the guideline information in this document has paralleled that provided in the *Green Book*. However, the 1999 Canadian document was updated to include explicit quantitative information about the effect of select geometric elements on safety and to promote the need to consider the benefits and costs of design alternatives.

The newest edition of the *Guide* has been modified to support safety-conscious design by moving away from the concept of "strict adherence to criteria." The authors of this guide suggest that safety is a matter of degree. Any road with vehicular traffic will have crashes. The number of such crashes and their severity is influenced by the roadway's design. More generous designs will tend to have fewer crashes and more restricted designs will have more crashes. Hence, they contend that it is incorrect to assert that any road built in conformance with design criteria is "absolutely" safe (i.e., free of crashes).

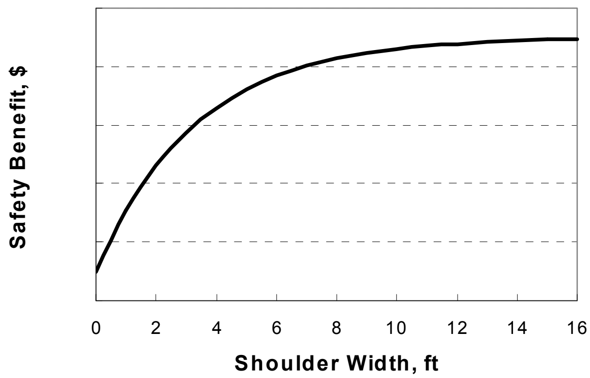
Roads built in conformance with design criteria may collectively provide an "acceptable" level of safety but the level of safety for any particular roadway is not truly known unless it is explicitly quantified. Without an examination of safety (including a forecast of crash frequency for all proposed designs), it cannot be known with reasonable certainty whether: (1) the road is over or under designed with respect to the desired acceptable level of safety, (2) it is uniformly safe, and (3) the design represents the best value for the invested funds.

Design Domain. The *Guide* has been developed using the "design domain" concept. This domain is represented as a range of controlling values that can be used to size a design element. The range is intended to limit the design element to values that have been found to be reasonably safe, efficient, and cost-effective for typical design applications. Within this range, the designer is encouraged to exercise judgment in selecting a reasonable design value.

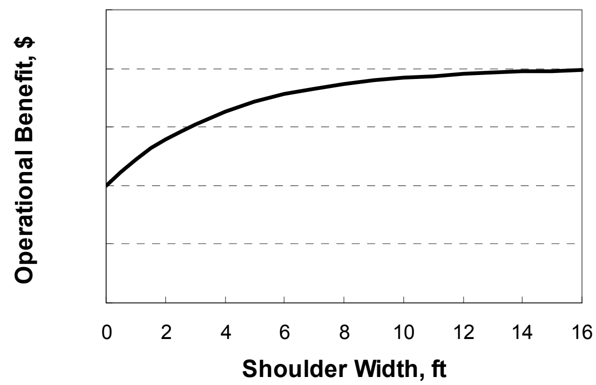
The approach described in the *Guide* is flexible in that designers can use their discretion to determine when a particular design choice requires an explicit evaluation of safety, operation, and/or cost. However, a more detailed evaluation is encouraged for controlling criteria in situations where atypical conditions exist, the design is complex, or construction costs are high. Through this process, the designer is provided with tangible information on which to make critical design decisions and should converge on a design that is uniformly safe and efficient as well as cost effective.

Illustrative Development of Design Domain for Typical Conditions. The use of benefit and cost information to quantitatively define the design domain for shoulder width is illustrated in Figure 2-3. In this figure, a quantitative relationship between crash frequency and shoulder width is used to develop a relationship between safety benefit (i.e., crash reduction) and width. A relationship between speed and width is similarly used to develop a relationship between operational benefit (i.e., speed increase) and width. Traditional cost estimation techniques are used to quantify maintenance and construction costs. Economic principles are then used to compute the present worth of the benefits relative to their construction cost.

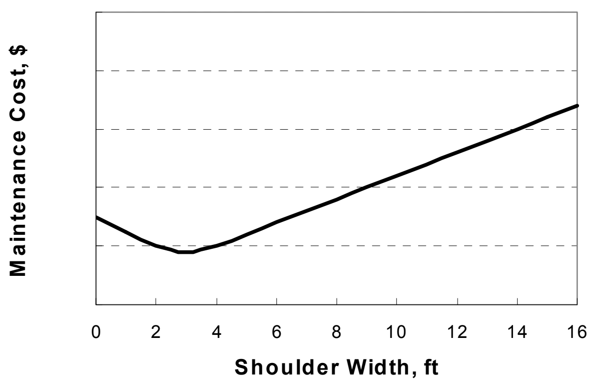
In this example, the design domain limits for shoulder width are identified as those widths associated with a benefit-cost ratio of 1.0. Shoulder widths between these limits are cost effective. Widths in the range of 2 to 4 ft are found to be the most cost effective. Because this is only an example application, other values may be found following a more realistic analysis.



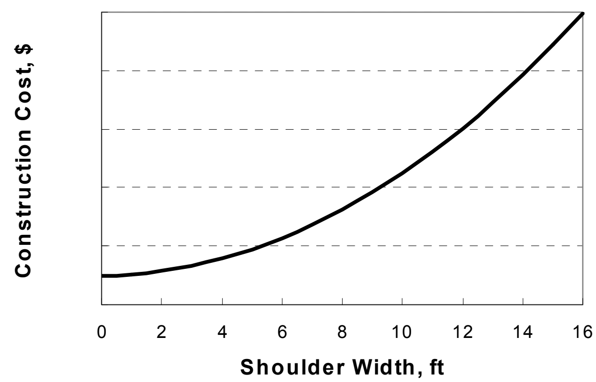
a. Safety Benefit.



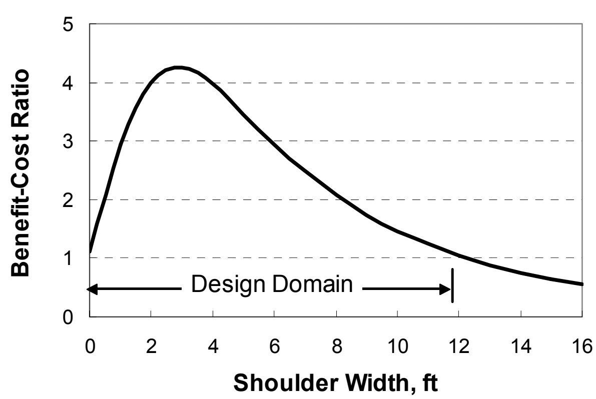
b. Operational Benefit.



c. Maintenance Cost.



d. Construction Cost.



e. Combined Benefits and Costs.

Figure 2-3. Design Domain Example.

This type of benefit-cost analysis can be used to develop domain limits for all controlling criteria. These limits would be defined using the characteristics of typical projects within the state. It is likely that they would be defined once by the transportation agency and, thereafter, used by engineers for design situations characterized as "typical" in that agency's jurisdiction.

Current Status of Design Domain Development. A review of the *Guide* indicates that the design domain limits identified for many of the design elements addressed are essentially equivalent to the criteria provided in the *Green Book*. The authors of the *Guide* acknowledge this tendency; however, they indicate that in future editions of the *Guide*, the domain limits will be revised in accordance with the aforementioned benefit-cost principles as new information and technology are made available.

The authors of the *Guide* recognize that their approach to safety-conscious design will likely increase design costs and the need for training on the use of evaluation tools. However, they rationalize that their approach will result in long-term savings to transportation agencies (by ensuring cost-effective designs) and society (by reducing crashes). Moreover, they offer that their new approach to design will: (1) provide a rational approach to design exception evaluation, (2) make value engineering a routine part of the design process, and (3) reduce the need for safety audits.

SAFETY CONSIDERATIONS IN THE TxDOT DESIGN PROCESS

This part of the chapter summarizes the safety considerations that have been incorporated in the TxDOT design process. Initially described are the findings from a review of various documents that guide roadway design in Texas. The objective of this review is to identify the guidelines provided in the documents that direct (or encourage) the engineer to identify and address safety issues during the various stages of the design process. Then, the manner in which these guidelines are used by TxDOT engineers is described. The information for this description was obtained through a series of interviews with engineers in various divisions and districts within TxDOT.

Review of Design Documents

Two TxDOT manuals are reviewed in this section; they are the *Project Development Process Manual* (*PDPM*) (*10*) and the *Roadway Design Manual* (*RDM*) (*3*).

Project Development Process Manual

As noted previously, the project development process consists of six stages; they are: planning and programming, preliminary design, environmental, right-of-way and utilities, PS&E development, and letting. The activities involved with each stage are described in the *PDPM*. This manual is comprehensive in that it addresses the activities associated with all project types, as listed in Table 2-1. Each stage is subdivided into a series of steps. The steps are further subdivided into tasks. All total, the *PDPM* describes the activities associated with nearly 200 tasks.

Guidance regarding safety assessment is provided in several locations in the *PDPM*. The guidance that is offered for use during the planning and programming, preliminary design, and PS&E development stages is the focus of this review. The tasks that specifically address safety in the *PDPM* are identified in Table 2-5. To be included in this table, the task had to describe an activity related to an investigation, analysis, or evaluation of safety associated with a proposed project. Indirect references to safety issues and general safety considerations are not included in Table 2-5.

Table 2-5. Location of Safety Guidance in the *Project Development Process Manual*.

Stage	Step	Safety-Related Task
Planning and Programming	1. Needs Identification	1000: Identify project need and scope
	2. Project Authorization	1210: Prepare programming assessment
	3. Compliance with Planning Requirements	--
	4. Study Requirements Determination	--
	5. Construction Funding Identification	--
Preliminary Design	1. Preliminary Design Conference	--
	2. Data Collection/Preliminary Design Prep.	2190: Obtain traffic accident data
	3. Public Meetings	--
	4. Preliminary Schematic	--
	5. Geometric Schematic	2550: Determine guide signing & operational controls
	6. Value Engineering	--
	7. Geometric Schematic Approval	--
PS&E Development	1. Design Conference	--
	2. Begin Detailed Design	--
	3. Final Alignments/Profiles	--
	4. Roadway Design	5240: Prepare cross sections & compute earthwork
	5. Operational Design	--
	6. Bridge Design	--
	7. Drainage Design	--
	8. Retaining/Noise Wall & Misc. Structures	--
	9. Traffic Control Plan	--
	10. PS&E Assembly/Design Review	--

Note:
"--" - no safety-related task identified in *Project Development Process Manual* (*10*).

As the information in column three of Table 2-5 indicates, safety-related guidance is included in five tasks of the project development process. A description of the guidance associated with each task is provided in Table 2-6. Tasks 1000 and 2190 focus on the evaluation of crash data to identify safety deficiencies associated with existing projects. Task 1210 is a multifaceted study that addresses issues related to the environment, safety, level of service, cost, and level of community support. Tasks 2550 and 5240 specifically focus on enhancements to guide sign and roadside design, respectively, to improve roadside safety.

Table 2-6. Safety Guidance Provided in the *Project Development Process Manual.*

Stage	Task	Applicable Project Types	Description of Safety Guidance
Planning and Programming	1000	All types	A review of traffic accident (or crash) information may alert the department to needed improvements.
	1210	Projects competing for statewide funds	The Transportation Planning and Programming (TPP) Division reviews a programming assessment study in determining approval of projects. One element addressed in this study is "safety issues." [A description of the safety issues to be addressed is not provided.]
Preliminary Design	2190	All except preventative maintenance	An accident/crash analysis is essential in the design process for a project involving an existing transportation facility.
	2550	Mainly urban freeways	Determine guide signs needed to increase roadway safety based on analysis of crash data.
PS&E Development	5240	All requiring earthwork	Evaluate and modify side slopes and ditch grades to provide for safety and economy of design.

Roadway Design Manual

Guidelines for the geometric design of streets and highways are provided in the *RDM* (*3*). This manual provides guidance and criteria to be used in the geometric design of roadway facilities. The content of the manual is divided into chapters that collectively address one of three project types; they include: new location and reconstruction (4R), 3R, and non-freeway restoration projects (2R). In general, a project's type dictates which chapter of the *RDM* is used to guide its design; however, some rehabilitation projects are required to use the guidelines developed for 4R projects. Table 2-7 illustrates the relationship between project type, facility type, and *RDM* chapter.

Table 2-7. *RDM* Chapter Associated with Specific Facility and Project Types.

Facility Type		Project Type [1]			
		New Location	Reconstruction	Rehabilitation	Restoration
National Highway System (NHS)	Freeway	Chapter 3 - 4R	Chapter 3 - 4R	Chapter 3 - 4R	n.a.
	Non-Freeway	Chapter 3 - 4R	Chapter 3 - 4R	Chapter 4 - 3R	n.a.
On-System Non-NHS	Freeway	Chapter 3 - 4R	Chapter 3 - 4R	Chapter 3 - 4R	n.a.
	Non-Freeway	Chapter 3 - 4R	Chapter 3 - 4R	Chapter 4 - 3R	Chapter 5 -2R[2]

Note:
1 - n.a.: not applicable.
2 - Restoration for this facility type is only allowed if current ADT is less than 3000 veh/d; otherwise use 3R criteria.

There are seven chapters in the *RDM*; they are listed in Table 2-8. The first chapter provides an overview of the manual and information on several administrative elements associated with the design process. The second chapter describes the basic design criteria that are applicable to all project types. Chapters 3, 4, and 5 describe design criteria specific to 4R, 3R, and 2R projects, respectively. Chapter 6 describes design criteria for all projects not addressed by Chapters 3, 4, and

5. Chapter 7 describes guidelines for miscellaneous design elements not addressed in any of the previous chapters.

Table 2-8. Contents of the *Roadway Design Manual*.

Chapter	Section
1. Design General	1. Overview 2. Design Exceptions, Design Waivers, and Design Variances 3. Schematic Layouts 4. Additional Access to the Interstate System 5. Preliminary Design Submissions 6. Maintenance Considerations in Design
2. Basic Design Criteria	1. Functional Classifications 2. Traffic Characteristics 3. Sight Distance 4. Horizontal Alignment 5. Vertical Alignment 6. Cross Sectional Elements 7. Drainage Facility Placement 8. Roadways Intersecting Dept. Projects
3. New Location and Reconstruction (4R) Design Criteria	1. Overview 2. Urban Streets 3. Suburban Roadways 4. Two-Lane Rural Highways 5. Multi-Lane Rural Highways 6. Freeways 7. Freeway Corridor Enhancements
4. Non-Freeway Rehabilitation (3R) Design Criteria	1. Purpose 2. Design Characteristics 3. Safety Enhancements 4. Frontage Roads 5. Bridges
5. Non-Freeway Resurfacing or Restoration Projects (2R)	1. Overview
6. Special Facilities	1. Off-System Bridge Replacement and Rehabilitation Projects 2. Historically Significant Bridge Projects 3. Texas Parks and Wildlife Department (Park Road) Projects 4. Bicycle Facilities
7. Miscellaneous Design Elements	1. Longitudinal Barriers 2. Fencing 3. Pedestrian Separations & Ramps 4. Parking 5. Shoulder Texturing 6. Emergency Median Openings 7. Minimum Designs for Truck and Bus Turns

Guidance regarding safety is provided in several locations in the *RDM*. The guidance that explicitly addresses safety is identified in Table 2-9. To be included in this table, guidance had to describe a technique for evaluating the safety of a design element, direction on the safety benefits of a specific design element, or instruct as to which design elements should be evaluated for their effect on safety. Passages that briefly mention general safety issues and considerations were not included in the table.

Chapter 4, Section 3 of the *RDM* explicitly deals with safety enhancements for 3R projects. The guidance in this section indicates that "Basic safety improvements will be required for all 3R projects" (*3*). These improvements include: upgrading guardrail, providing signing and markings, providing a skid resistant surface, and treating cross drainage pipe culverts. The discussion in this section also quotes from *Special Report 214: Designing Safer Roads - Practices for Resurfacing, Restoration, and Rehabilitation* (*2*). The quote reiterates the need for a safety-conscious design process.

Table 2-9. Safety Guidance Provided in the *Roadway Design Manual*.

Project Type	Intent of Project	Description of Safety Guidance
New Location and Reconstruction	Enhanced mobility, substantial changes in alignment, or improvements to pavement structure to provide long-term service.	Chapter 2, Section 6: Principles of safe roadside design outlined as follows: ● Provide horizontal clearance that is as wide as practical. ● Roadside obstacles should be eliminated, redesigned, relocated, made breakaway, protected, or delineated. ● Use higher than minimum design criteria when possible. ● Design should be consistent with driver expectancy. Chapter 3, Section 3: Research has shown that reducing the number of access points on suburban roadways will reduce the potential for crashes. In addition, accidents can be reduced with the use of turn bays or turn lanes. Chapter 3, Section 5: Safety benefits of specific facility types: ● If an undivided four-lane road is being considered, the impact of left-turn movements on safety should be examined. ● Grade separations or interchanges may be provided to increase safety at accident-prone crossings.
Rehabilitation	Extend the service life and enhance the safety of a roadway.	Chapter 4, Section 3: Guidance is provided on the conduct of an accident analysis, safety assessment, cost assessment of alternative design elements, and required inclusion of "basic safety improvements."[1]
Restoration	Restore pavement structure, riding quality, or other necessary components to their existing configuration.	Chapter 5, Section 1: An accident analysis should be conducted, high accident frequencies reviewed, and corrective measures taken where appropriate.

Note:
1 - Basic safety improvements include: upgrading guardrail, providing signing and markings, providing a skid resistant surface, and treating cross drainage pipe culverts.

With one exception, the basic safety improvements identified in the *RDM* are also used by other state DOTs in their 3R projects. The exception is "skid resistant surface." This option was not identified as a safety improvement used by DOTs in a recent survey by Harwood et al. (*5*).

The *RDM* does not describe a process for safety-conscious design nor does it provide tools to facilitate the explicit evaluation of safety. Guidelines in the chapters addressing 3R and 2R projects encourage the explicit examination of crash data. However, the need to evaluate crash history for reconstruction projects is not indicated. Similarly, the use of information describing the safety benefits associated with specific design choices for new location projects is not described.

State-of-the-Practice Survey

This section summarizes a review of the TxDOT design evaluation process, with an emphasis on the role of safety evaluation in this process. This review was undertaken through a series of meetings with design and operations engineers in two TxDOT divisions and in five TxDOT districts. The specific objectives of the meetings were to:

- identify methods (e.g., standards, guidelines, tools, procedures) by which safety evaluation is incorporated in the 3R and 4R design processes; and
- identify constraints and issues associated with making safety assessment a more explicit part of design.

The divisions and districts selected for inclusion in the meeting series are listed in Table 2-10. The districts were chosen such that they collectively offered a range of population density and geographic regions of the state. To some degree, this range is evidenced by the annual project and letting information provided in the table. Attendees at the district meetings included representatives from the project planning, design, and operations functions within the respective districts. A total of 24 engineers were interviewed.

Table 2-10. Division and District Personnel Interviewed.

Division or District	Annual Projects Let	Annual Letting, $ (millions)	Representation, persons		
			Planning & Programming	Design	Operations
Design	n.a.	n.a.	0	3	0
Traffic Operations	n.a.	n.a.	0	0	3
Tyler	30	90	1	2	1
Dallas	100	450	2	2	1
Waco	55	195	1	1	1
Houston	110	500	1	1	1
San Angelo	16	50	1	2	0

Note: n.a. - not applicable.

The remainder of this section describes safety issues and considerations during the early stages of the design process (i.e., the planning and programming stage and the preliminary design stage). The discussion focuses on the 3R and 4R projects. Some discussion of HES projects is incorporated where appropriate.

In general, the discussion regarding the manner in which safety information is incorporated into the 3R and 4R design processes yielded relatively consistent responses from the interviewees. The most significant opportunity for safety improvement was judged to be in the 3R process because of opportunities to integrate safety work into rehabilitation contracts. Projects completed in the 4R categories are considered to be acceptably safe (and, thereby, not generally needing additional safety evaluation) because they adhere to the design criteria in the *RDM* (*3*), the *Green Book* (*4*), or both.

Planning and Programming Stage

Safety considerations during the planning and programming stage of the project development process emphasize: (1) identification of projects based on safety concerns, and (2) the means by which these projects are prioritized. The insights provided by the interviewees in each of these areas are discussed in the following two subsections.

Needs Identification Step. The objective of this step is to identify and document the need for a project (*10*). This need can be identified in many ways, including suggestions from maintenance supervisors, area engineers, district staff, local elected officials, developers, and the local community. Many factors are considered in determining project need including crash frequency and severity, pavement condition, bridge condition, and conformance with current geometric standards. Once a project is identified, research is conducted to prioritize its need relative to other projects competing for limited funds.

Comments made by the interviewees indicate that considerations of safety at this step in the process are directed toward the benefits derived by changes to an *existing* facility. Safety considerations for a *new* facility tend to receive less deliberation because adherence to design criteria in the *RDM* or *Green Book* is generally believed to yield an acceptably safe design.

In general, the need for safety improvements to existing facilities is often identified using information obtained from one or more of the following sources (examples are identified by open circle bullet):

- Informal Interaction
  - Local law enforcement
  - Local city or county transportation agencies (with citizen input)
  - Maintenance supervisor
  - Area engineer
- Formal Interaction with Local Agencies through Periodic Meetings
  - Metropolitan planning organization
- Conduct System Evaluation Programs
  - Annual rehabilitation program developed by maintenance supervisor and area engineer
  - Pavement conditioning monitoring program
  - Wet weather crash monitoring program
  - District safety plan intended to identify safety projects and consistently apply solutions
  - Annual hazard elimination program (and associated crash summaries)
- Focused Crash Data Analysis
  - Master accident listing using mainframe computer
  - Hazard identification listing using PC-based software (i.e., Microsoft Access)

The combination of sources used by any one district is reported to vary as does the specific manner in which the information is used. The more rural districts tend to rely more on informal interaction and crash data analysis. In contrast, the more urban districts rely more on formal interaction with the MPO and the findings from various system evaluation programs.

With few exceptions, the information provided by the aforementioned sources is used to identify highway facilities that may derive a safety benefit from design improvements. The information provided can range from an awareness of recurring crashes at a location to a quantitative evaluation of crash patterns along a road system.

Engineers in some districts thoroughly review crash data during the planning and programming stage; however, most cite only a limited review. Engineers in these latter districts conduct a thorough review when they have prior knowledge of safety concerns. The Texas Department of Public Safety (DPS) crash database is not frequently used for the crash data analysis because it often lacks sufficient detail to identify crash location. This problem can be overcome by a detailed analysis of the officer crash reports. However, time is not generally available during the planning and programming stage for the acquisition and detailed analysis of the printed reports.

Once a project is identified, research is conducted to prioritize its need relative to others that are competing for limited funds. A site visit is often undertaken to properly assess project constraints and to assist in identifying needed safety, operation, and structural improvements.

Construction Funding Identification Step. The objectives of this step are to identify potential construction funding sources and to prioritize projects based on consideration of their cost and effectiveness (*10*).

Comments by the interviewees indicate that funding sources, time schedules, and workload often dictate which projects are completed and when. When these external influences are not present, the district directors review the scope of the various projects being considered and prioritize them based on a range of criteria, including pavement condition and safety concerns.

Project funding is a major determinant of construction priority. Project funding categories follow the 4R, 3R, 2R, PM, and HES designations that correspond to project type. Within each of these categories, different criteria are used to prioritize projects. The criteria offered by the interviewees are listed in Table 2-11. The criteria listed for each project vary in relation to its intent (or objective). The number of criteria listed reflects the extent of a project's impact and the benefits derived by the public.

The local MPOs also have some input to the project funding and prioritization process. Specifically, they initiate 4R projects for the Metropolitan Mobility/Rehabilitation program and the Congestion Mitigation and Air Quality Improvement program. In addition, the MPOs identify high crash locations and develop prioritized lists of recommended safety-related projects. These project recommendations are ultimately forwarded to TxDOT for consideration in the planning and programming stage.

Table 2-11. Design Project Types and Prioritization Criteria.

Project Type	Project Intent	Prioritization Criteria
New Location and Reconstruction (4R)	Enhanced mobility and safety that generally include substantial changes in the geometry of the roadway.	• Mobility benefit • Extent of environmental documentation required • Level of public interest • Right-of-way impact • Extent of utility adjustments • Number of environmental permits needed • Availability of funds (incl. local contribution) • Congestion and air quality
Rehabilitation (3R)	Extend the service life and enhance the safety of a roadway.	• Pavement condition • Presence of a safety concern • Time required to obtain right-of-way
Resurfacing or Restoration (2R)	Restore pavement structure, riding quality, or other necessary components, to their existing configuration.	• Pavement condition • Presence of a safety concern
Preventative Maintenance (PM)	Preserve, rather than improve, the structural integrity of the pavement, structure, or both.	• Pavement condition
Hazard Elimination (HES)	Reduce the number and severity of crashes at hazardous roadway locations.	• Presence of a safety concern • Safety improvement potential • Cost of improvement

Preliminary Design

As indicated in Table 2-5, the preliminary design stage consists of seven steps. Five of these steps include some consideration of safety issues, use of safety data, or assessment of safety impact. The insights provided by the interviewees, as related to these considerations, are discussed in the following subsections. The order of presentation follows the steps of the preliminary design stage.

Preliminary Design Conference Step. The objective of the preliminary design conference is to establish and agree on fundamental aspects, concepts, and preliminary design criteria associated with a 4R project (*10*). This conference can also be conducted for other project types, although it is not required. The agenda for this conference includes discussion of project background, scope, corridor issues, environmental issues, multimodal issues, alternatives, applicable design criteria, right-of-way requirements, and permits. A product of this conference is the Design Summary Report. This report documents the discussion and resolution of each agenda item.

Information provided by the interviewees indicates that the discussion of project background during the conference includes a review of the design elements of the existing roadway. A topic of discussion during this review is the findings from an analysis of crash history. The depth of this discussion is often dependent on the engineer's awareness of existing safety concerns.

The Design Summary Report was found to vary in content among districts because each district has tailored the report to suit their needs. The most current version of this report includes a check box to indicate whether a crash analysis has been performed. This version is being used by some districts. However, the reports provided by a few districts did not include this check box, which suggests they are using an older version of the document.

One interviewee offered that guidance was needed with regard to the investigations that should be conducted prior to the conference. Specifically, guidelines were believed needed describing the conduct of the "field review" (or site visit) that occurs during the needs identification step of the planning and programming stage. Guidelines are also believed needed on the process of obtaining and reviewing crash data (especially as it pertains to the recognition of crash trends).

Data Collection/Preliminary Design Preparation Step. One task in the data collection/preliminary design preparation step deals with the collection and analysis of crash data (*10*). This task is applicable to all project types except preventative maintenance. Data collected include crash data for the most recent three-year period and current traffic counts. The objective of the analysis is to look for similarities, patterns, or abrupt changes in crash trends over time. The findings from this analysis are then used to identify changes to design features that might reduce crash frequency or severity, or improve operations.

Discussions with the interviewees indicated that evaluations of overall and project-level crashes focus primarily on crash frequency. Crash rates are rarely used in the evaluation even though the effect of traffic volume on crash frequency is recognized by the interviewees.

Crash data are obtained in a variety of ways. For most projects, they are obtained from the database assembled by the DPS. This database is made available to TxDOT engineers through a mainframe computer. Software that provides access to the database is executed using remote terminals located throughout each district. The Traffic Engineering Section of the Traffic Operations Division staff maintain these software programs. The software is capable of producing reports that list crashes by location or by selected crash attributes (e.g., by vehicle movement, manner of collision, severity, contributing factor, etc.). Collision plot diagrams are also available.

Preliminary Schematic Step. Once the data collection step is substantially complete, the preliminary schematic step begins. Horizontal and vertical alignments are calculated for one or more feasible route alternative and are shown on preliminary schematics. The alignment is sufficiently well defined that its operational level of service, right-of-way impacts, and construction cost can be estimated (*10*).

The interviewees indicated that the level of service provided by a proposed alignment is frequently evaluated for 4R projects but very rarely for other project types. This analysis is perceived to be helpful when traffic delays are expected to be high.

Right-of-way impacts were a concern to most of the interviewees because they tend to have an adverse effect on project schedule and cost. These impacts are most prevalent in urban areas. Right-of-way constraints were cited as also having an adverse effect on safety because they often limited the type of design solution that could be applied. It was noted that the need to acquire right-of-way could slow the project development process considerably and occasionally delay application of some safety improvements. It was offered that a more quantitative understanding of the safety benefits of design alternatives could increase the likelihood of right-of-way being acquired (and thereby, a safety benefit realized) by city or county officials responsible for road-related right-of-way acquisition.

Value Engineering Step. The value engineering step is used to assess a project's overall cost-effectiveness or how well it meets the identified needs (*10*). Value engineering is designed to gather the expertise of many individuals to produce the most effective solution to the transportation need. The value engineering team is charged with recommending project alternatives that provide an economical, yet high quality product. This step is typically performed on high-cost and complex projects. In fact, projects in transportation corridors or federal-aid projects on the National Highway System with an estimated cost of $25 million or more require the conduct of a value engineering study. An evaluation of the safety benefits of alternative designs is not explicitly identified in the *PDPM* as a component of the value engineering step.

The objectives of the value engineering process include: (1) improve roadway efficiency, (2) honor commitments to the public, (3) improve the design, and (4) lower construction cost (but only if the cost savings do not compromise long-term benefits). The benefits of a change to the design are typically assessed using experience and judgment (as opposed to formulas and computations). A quantitative assessment of benefit is not routinely undertaken because alternative designs of entire alignments (or significant portions thereof) are rarely completed for the sole purpose of identifying the most cost-effective solution.

During the interviews, the value engineering step was investigated as a possible point in the process to formally evaluate the level of safety provided by a design. However, few value engineering studies are being conducted simply because the projects for which they are required are few in number.

Only HES projects routinely include an assessment of cost-effectiveness of safety improvement alternatives. In fact, the benefit-cost ratio that is computed for each HES project is used as a basis for project prioritization and selection.

Geometric Schematic Approval Step. One task of the geometric schematic approval step is to identify the need for a design exception or design waiver (*10*). Design exceptions or waivers are required any time design criteria do not meet the established minimums cited in the *RDM* or, when appropriate, the *Green Book*. When applying for a design exception, the nature of the design exception and an explanation of why it is needed is sent to the Design Division for review and approval.

The district interviewees reported that design exceptions were occasionally requested for 3R projects and were rarely requested for the design of new facilities. Typical design features or elements cited for recent design exception requests included:

- vertical clearance on structures,
- vertical alignment,
- lane width,

- shoulder width,
- design speed into stop conditions (e.g., button-hook designs where horizontal curvature is introduced to eliminate intersection skew), and
- horizontal alignment on off-system bridge projects.

The design exception process was recognized by the interviewees as an additional opportunity for safety evaluation. As told by one interviewee, the evaluation is currently based on an assessment of the crash history of the existing facility. If safety problems do not exist that are potentially related to the design feature for which an exception is requested, then it is inferred that the exception will not likely have an adverse impact on safety. For example, consider the reconstruction of an existing highway with 4-ft shoulders and for which at least 8-ft shoulders are required for compliance with the applicable design criteria. If the highway has negligible crashes related to narrow shoulders, then an exception requesting allowance for 6-ft shoulders may be granted because it is reasoned that widening to 8 ft at this location would not provide additional safety benefit.

Post Design Safety Evaluation

The interviewees were asked about follow-up evaluations of facility safety after construction was completed. The district interviewees reported that while a formal evaluation is not a requirement, they do informally monitor newly opened facilities. In this regard, a natural feedback loop exists between the districts and local law enforcement, local officials, and the motoring public. Rural districts also rely on feedback from the maintenance supervisor and area engineer about potential safety concerns. Urban districts also rely on local MPOs to track changes in the crash history of recently opened facilities. If a newly completed project exhibits safety problems, the engineers are confident they will learn of it in a timely manner.

The Traffic Operations Division conducts formal follow-up evaluations of HES projects. These evaluations are required by FHWA. However, the time lag in the DPS database makes the findings of such assessments very dated. Specifically, FHWA requires three years of "before" data and three years of "after" data for the evaluation. Given that it can sometimes take a year for crash data to be incorporated in the DPS database, the earliest that an "after" study could be completed is four years after the project is *reported* by the district as having been completed. It was noted that some districts are occasionally slow to report when their HES projects are completed, further extending the time by which a follow-up safety evaluation can be initiated.

SUMMARY OF FINDINGS

It is recognized that safety considerations underlie much of the content of all the documents reviewed in this chapter. It is also recognized that there is a general belief among designers that designs that adhere to the guidelines and policies described in these manuals will generally operate in an acceptably safe and efficient manner. However, the recurring need for retrofit and hazard elimination projects is a reminder that a design can occasionally have unique features that combine to produce an unacceptable level of safety. This realization highlights the need for a more proactive design process that explicitly considers safety, especially for atypical or complex design elements.

The traditional approach to design incorporates a nominal level of safety through adherence to minimum design criteria for key design elements. The level of safety provided is referred to as "nominal" because the correlation between actual crash frequency and these design elements is likely unknown to the designer (*11*). To achieve further improvement in highway safety, it will be necessary to focus on design policies and technologies that: (1) reduce the likelihood of a crash and (2) reduce the severity of the crashes that do occur. A safety-conscious design approach embodies these policies and technologies.

In some typical situations, use of warrants and design criteria may provide a safe and efficient design. However, in more complicated design situations, adherence to warrants and criteria may not be sufficient. In these instances, it may be more appropriate for the designer to explicitly evaluate the safety, operation, and/or cost associated with one or more design elements. As a minimum, evaluation tools would be used to quantify the safety impact associated with a design choice. If conditions are especially complex or construction costs are especially high, it may be helpful to use benefit-cost techniques to more fully evaluate alternative design choices.

Ideally, guidance and tools that relate a design choice to crash frequency (and severity) would be available to help guide the engineer during the design process such that:

- each design element (or design element size) selected provides an acceptable level of safety;
- the design features that comprise the design are consistent in the degree of safety provided;
- the combination of design elements for each design feature are cost-effective (i.e., neither over- nor under-designed); and
- the benefits derived from the resulting design can be shown to outweigh its costs and represent the best use of limited program funds.

REFERENCES

1. *Highway Safety Design and Operations Guide.* American Association of State Highway and Transportation Officials, Washington, D.C., 1997.

2. *Special Report 214: Designing Safer Roads - Practices for Resurfacing, Restoration, and Rehabilitation.* Transportation Research Board, Washington, D.C., 1987.

3. *Roadway Design Manual.* Texas Department of Transportation, Austin, Texas, February 2004.

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

5. Harwood, D.W., E.R. Kohlman Rabbani, K.R. Richard, H.W. McGee, and G.L. Gittings. *NCHRP Report 486: Systemwide Impact of Safety and Traffic Operations Design Decisions for 3R Projects.* National Cooperative Highway Research Program, Transportation Research Board, Washington, D.C., 2003.

6. Safety-Conscious Planning Website. FHWA Office of Planning, Environment, and Realty (HEP), Washington, D.C., 2003. http://www.fhwa.dot.gov/planning/scp/index.htm. Accessed April 2004.

7. Roberts, K. *Safety-Conscious Planning: The Development of the Safer Transportation Network Planning Process.* Materials provided on the Seminars website. Institute of Transportation Engineers, Washington, D.C., January 2001. http://www.ite.org/seminars/SCP.asp. Accessed April 2004.

8. *Geometric Design Guide for Canadian Roads.* Part 1 and Part 2. Transportation Association of Canada, Ottawa, Canada, 1999.

9. Epstein, K., G. Corino, and D. Neuman. "National Review of the Highway Safety Improvement Program." *Public Roads*. Vol. 65, No. 5. Federal Highway Administration, Washington, D.C., March 2002.

10. *Project Development Process Manual.* Texas Department of Transportation, Austin, Texas, August 2003.

11. *The Traffic Safety Toolbox: A Primer on Traffic Safety.* Institute of Transportation Engineers, Washington, D.C., 1999.

CHAPTER 3. SAFETY INFORMATION SOURCES AND NEEDS

OVERVIEW

This chapter summarizes a review of the various sources of safety-related data maintained by, or available to, TxDOT. It also describes the findings from an evaluation of the suitability of these data sources to the calibration and application of various safety evaluation tools. A focus of this review and evaluation relates to the specific types of data needed to implement two forthcoming safety evaluation tools: *SafetyAnalyst* and *Highway Safety Manual* (*HSM*).

This chapter consists of four main parts. Initially, several safety evaluation tools are reviewed and their data needs identified. Then, the databases maintained by TxDOT and the Texas Department of Public Safety (DPS) are reviewed and relevant attributes identified. Next, the data needed by the two evaluation tools are compared with that available in various DPS and TxDOT databases. Finally, the last part summarizes the key findings from the previous three parts.

SAFETY EVALUATION TOOLS

This part of the chapter provides a brief review of several safety evaluation tools. One goal of this review is to provide perspective for the subsequent review of TxDOT and DPS databases. However, prior to this review, several terms related to the topic of safety databases and evaluation tools are defined.

Terminology

This section defines several terms related to databases and their attributes. Some definitions related to roadway characteristics are also provided.

Database Terms

Relational Database. A database that is linkable to other databases using a common reference system. The reference system can be defined using temporal measures, spatial measures, or both. Spatial measures identify the entity by its physical location along the roadway. They can be specified in terms of a linear referencing system (e.g., control section and milepoint) or a geo-coordinate system (e.g., latitude and longitude).

Attribute. A fundamental data element in a database. An attribute can date, number, or describe the data. Examples of crash attributes include: crash number, date of crash, first harmful event, etc. Examples of roadway characteristic attributes include: location, surface type, functional class, average daily traffic volume (ADT), etc.

Reference System. An attribute (or set of attributes) that facilitates relating an observation in one database with data in another database (e.g., time/date, physical location, inventory number, license plate number, driver name, etc.).

Highway Safety Manual Terms

Homogeneous Segment. A highway, ramp, or frontage road segment where the following attributes are considered constant: ADT, lane width, shoulder width, shoulder type, driveway density, and roadside hazard rating. A new segment also starts at any of the following locations: intersection, beginning or end of a horizontal curve, point of vertical curvature, beginning of a passing lane, and beginning of a two-way left-turn lane (TWLTL). For brevity, a homogeneous segment is referred to herein as a "roadway segment," or just "segment."

Roadway Section. A continuous length of roadway that includes one or more road segments and may include one or more intersection legs. In the case of roadway construction or reconstruction, a roadway section is also referred to as a "project."

Interactive Highway Safety Design Model

The Interactive Highway Safety Design Model (IHSDM) is a road safety evaluation tool that is distributed as a software product (*1*). The Federal Highway Administration developed IHSDM with an initial focus on two-lane rural highways. FHWA research and development for a multilane rural highway application of IHSDM is now underway.

The 2003 release of IHSDM for two-lane rural highways is now available for testing and evaluation purposes. It consists of the following five modules:

- Crash Prediction Module - it estimates the number and severity of crashes on specified roadway segments.
- Design Consistency Module - it evaluates the operating speed consistency along a roadway.
- Intersection Review Module - it provides a structured process for evaluating the safety impact of intersection design alternatives using an expert system approach.
- Policy Review Module - it checks compliance with highway design policies.
- Traffic Analysis Module - it uses traffic simulation models to estimate the operational effects of road designs under current and projected traffic flows.

Resurfacing Safety Resource Allocation Program

The Resurfacing Safety Resource Allocation Program (RSRAP) was developed by Harwood et al. (*2*) to evaluate alternative improvements for 3R projects. The program was implemented as a software product. It can evaluate a specific set of improvement alternatives for each candidate roadway section including: do nothing, resurface only, and various types of safety improvements. Improvement types include the following:

- widen lanes,
- widen shoulders,
- pave shoulders,
- provide left-turn lanes at intersections,
- provide right-turn lanes at intersections,

- make minor changes in horizontal alignment to increase curve radii, and
- improve roadside conditions.

Several cost-estimation features are incorporated in the program including: estimating the construction cost of each improvement alternative, estimating the penalty for not resurfacing, and estimating the safety benefits for each improvement alternative. The potential safety benefits of an alternative are determined using the expected percentage reduction in crashes.

RSRAP was "developed to allow highway agencies to implement a process that selects a program of safety improvements, to be made in conjunction with resurfacing projects, that maximizes traffic safety and operational benefits resulting from the program (*2*)."

Roadside Safety Analysis Program

The *Roadside Safety Analysis Program* (RSAP) can be used to evaluate the cost-effectiveness of installing individual roadside safety treatments or appurtenances at specific roadway locations (*3, 4*). RSAP accepts, as input, information about road segment geometry and traffic characteristics. It also allows the analyst to describe the roadside cross section, location of fixed objects, and safety appurtenance design. These input variables are listed in Table 3-1. The output from RSAP includes an estimate of annual crash frequency as well as the road-user costs associated with these crashes.

Table 3-1. RSAP Input Data Requirements.

Design Category	Design Component	Design Element	
General	--	Area type (urban/rural) One-way/two-way Segment length	Functional class Speed limit
Traffic characteristics	--	Traffic volume (ADT) Traffic growth factor	Truck percentage
Geometric design	Horizontal alignment	Direction of curve	Radius
	Vertical alignment	Grade	
	Cross section	Divided/undivided Lane width Median type	Number of lanes Shoulder width Median width
Roadside design	Cross section	Foreslope Parallel ditches	Backslope Intersecting slopes
	Fixed object	Offset Type (wood pole, headwall, etc.) Width	Side of roadway Spacing
	Safety appurtenances	Offset Type (barrier, cushion, etc.) Length	Side of roadway Flare rate Width

Note:
"--" not applicable.

Safety Analyst

SafetyAnalyst is a suite of software-based safety evaluation tools that are intended to facilitate the safety management of a street or highway system (*5*). It includes six evaluation tools. Each tool corresponds to one step in the traditional highway safety improvement program; they are:

- network screening tool,
- diagnosis tool,
- countermeasure selection tool,
- economic appraisal tool,
- priority-ranking tool, and
- evaluation tool.

The purpose of the network screening tool is to use available highway safety data to review an entire road network and identify and prioritize those locations that have potential for safety improvement. These locations are then evaluated more formally using the remaining evaluation tools.

The purpose of the diagnosis tool is to guide the analyst in the evaluation of safety problems at a specific location. The countermeasure selection tool is intended to assist in the selection of viable countermeasures based on the problem diagnosis. Each of these countermeasures can be assessed using economic principles with the fourth tool. If several countermeasures are being considered at a location, the priority-ranking tool can be used to rank the various alternatives based on their benefits and costs. Lastly, the evaluation tool would be used after implementation to determine if a safety improvement has been realized and to quantify the magnitude of the crash reduction. This information can be used to refine the estimate of countermeasure effectiveness for subsequent applications of *SafetyAnalyst*.

If *SafetyAnalyst* development is on schedule, then beta testing of the software should begin during 2006, with a target release date for *SafetyAnalyst* set for sometime in 2008.

Highway Safety Manual

The development of the *HSM* is being overseen by a joint task force sponsored by the Transportation Research Board (*6*). The goals for the manual are "to provide the best factual information and tools, in a useful and widely accepted form, to facilitate roadway design and operational decisions based upon explicit consideration of their safety consequences." A draft chapter for two-lane rural highways is available and research is currently underway to develop methodologies for evaluating the safety of rural multilane highways, urban and suburban arterial highways, and the intersections associated with these highways. The first draft of the *HSM* will not include a methodology for evaluating limited access facilities, interchange ramps, or frontage roads.

SAFETY INFORMATION SOURCES

The evaluation tools described in a previous part of this chapter require two types of data for their successful application. These two types of data can be categorized as "roadway characteristics" and "crash history." Roadway characteristics data include the traffic control, traffic characteristics, and geometry for highways, intersections, interchange ramps, and frontage roads. These data are identified by their location and date of last change in condition or status. Crash history data include the location, date, and description of the crash as well as information about the vehicles involved and their occupants.

This section describes the safety-related databases that can be used to calibrate safety evaluation tools and their associated models. Initially, several of the more common safety applications are described. Then, the databases available to TxDOT are summarized.

Database Applications

Safety databases are used for a wide variety of safety management functions within TxDOT. Some of the more common applications are shown in Table 3-2. They include:

- crash attribute summary,
- project safety evaluation,
- hazardous location identification, and
- safety model calibration.

Each of these applications is briefly described in the following subsections in terms of their typical analysis objective and scope.

Crash Attribute Summary

As indicated in Table 3-2, a crash attribute summary describes crash events occurring in a specified region. The summary is specific to one or more crash attributes, but without association to a specific roadway section or highway. Examples of a crash attribute summary include: all alcohol-related crashes in Texas, or all injury crashes in Brazos County. A crash attribute summary is used for before-after studies of area-wide safety programs, evaluating overall highway system improvements, and assessing the effect of regional changes in economy, traffic enforcement, vehicle design, or driver education on safety. In these applications, only a crash history database is needed. The specific attributes of the associated road system are not necessary or considered.

Project Safety Evaluation

A project safety evaluation focuses on crashes that occur at a single, known section of roadway. The intent of this evaluation is to identify sections that may benefit from safety improvement, possibly as part of a reconstruction project. Because the location to be evaluated is known, roadway characteristic data for all components are readily available or can be easily obtained from agency files. Therefore, only a crash history database is needed for this application.

Table 3-2. Database Applications.

Application Type [1]	Typical Analysis Objective	Typical Analysis Scope			Data Available to TxDOT?
		Location Evaluated	Facility Component [2]	Databases Used	
Crash Attribute Summary	Summarize crashes of a common type, driver condition, weather, etc. in a specified region.	City, county, district, or state	Not applicable	Crash history	Yes
Project Safety Evaluation	Summarize crashes along a specific roadway section.	One road section	All components in section	Crash history	Yes
Hazardous Location Identification	Quantify crashes for specified components in a specified region and identify those that suggest the need for safety improvement.	City, county, district, or state	Specified subset of components	Roadway characteristics first, then Crash history	Yes, for highway segments
Safety Model Calibration	Quantify crashes for specified components in a specified region and use to calibrate model.	City, county, district, or state	Specified subset of components	Roadway characteristics first, then Crash history	Yes, for highway segments

Notes:
1 - All applications use data associated with a specified period of time.
2 - A facility component is a highway segment, intersection, ramp segment, ramp terminal, frontage road segment, or frontage road intersection.

Hazardous Location Identification

Hazardous location identification focuses on identifying facility components in a specified region that are experiencing an unusually large number of crashes. The analyst can evaluate all components or focus on a specified type of component (e.g., all highway segments). Roadway characteristics data are used to define a subset database that includes all components of interest. For example, an evaluation of hazardous signalized intersections in a specific district could start with the identification of all signalized intersections (and associated legs) in that district.

Once the components with the desired roadway characteristics have been identified, their crash history is extracted from the crash history database. This approach yields a subset database suitable for identifying atypical (or hazardous) locations because it includes all facility components and properly characterizes the full range of crash experience of these components. Continuing the example, the crash history database would be queried once for each signalized intersection in the district. For each query, all intersection-related crashes identified would be extracted and reviewed. Some intersections would have no crashes; some would have many. Statistics could be used to determine which intersections were atypical.

The aforementioned approach can be compared to one where the crash history database is queried first, before the roadway characteristics database. In this instance, all crashes with specified attributes would be identified first. Then, the facility components at which these crashes occurred would be identified. Statistics could be used to determine which components were atypical. There is an important limitation of this approach. Specifically, it does not identify the safest components (i.e., those with no crashes) because the crash database is used as the means of identifying the desired subset database. As a result, the statistics derived from this database to define "atypical" components are biased toward "problem" components and important information about the attributes of the safest facility components remain hidden in the road characteristics database.

As noted in the last column of Table 3-2, the safety databases available to TxDOT are restricted primarily to the analysis of highway segments. Road characteristics data for ramp segments, ramp terminals, frontage road segments, and frontage road intersections are not explicitly identified in these TxDOT safety databases. Data for intersection legs are only partially available. As a result, the identification of intersections, ramps, and frontage roads with potential for safety improvement will require the manual assembly of a roadway characteristics database.

Safety Model Calibration

The safety prediction models to be included in the forthcoming *SafetyAnalyst* and the *HSM* require calibration to "local" conditions to obtain accurate estimates of roadway safety. The calibration process requires the identification of crashes at specified facility components and in a specified region. For example, crashes at several four-leg, signalized intersections in East Texas would be needed to calibrate a signalized intersection crash prediction model for use in that region of Texas. Both a crash history and a road characteristic database are needed for this application.

For model calibration, as with hazardous location identification, roadway characteristic data are first used to define a subset database that includes a random sample of the facility components of interest. Then, the crash history is extracted from the crash history database for each of the identified components. This approach yields a safety database suitable for model calibration because it includes a representative sample of facility components and properly characterizes the full range of crash experience of these components.

As noted in the last column of Table 3-2, the safety databases available to TxDOT are restricted to the calibration of safety prediction models for highway segments. Key road characteristics data for intersections, ramp segments, ramp terminals, frontage road segments, and frontage road intersections are not explicitly identified in the TxDOT safety databases. As a result, the calibration of safety prediction models for these facility components will require the manual assembly of a roadway characteristics database.

Summary

The database suitable for use with safety evaluation tools should have three key features. First, it should have a temporal referencing system for linking the attributes of the crash history and roadway characteristic databases for a specific time period. Second, the roadway characteristics database should uniquely locate, and have information for, each of the following highway facility components:

- highway segments,
- intersections,
- ramp segments,

- ramp terminals,
- frontage road segments, and
- frontage road intersections.

Third, the database suitable for use with safety evaluation tools should have a spatial reference system. A database with these three features would allow for the examination of crash history associated with a specific facility component for a specific time period.

Texas Databases

This section describes three Texas databases. They include the DPS crash history database, the TxDOT Roadway Inventory File (RI-File), and the Texas Reference Marker System (TRM). The latter two databases contain roadway characteristics. Both TRM and RI-File are linked to the DPS database using a linear referencing system. Also noted in this section is the forthcoming Crash Record Information System (CRIS) being developed jointly by TxDOT and DPS. It is intended to replace the RI-File database and to update the list of attributes in the DPS crash history database.

Department of Public Safety Crash History Database

The DPS crash history database contains information about each reported crash in Texas. This information describes the vehicles, drivers, and vehicle occupants involved in the crash as well as the roadway characteristics at the site of the crash. The DPS database consists of three individual component databases; they are: accident database, driver/vehicle database, and casualty/occupant database.

The DPS database uses a unique crash report number assigned to each crash to link the three component databases. To link to roadway characteristics databases, the DPS database also includes a spatial reference system based on the control-section/milepoint system developed by TxDOT for the location of construction and maintenance activities. Every highway segment on the state highway system is identified using this system.

TxDOT Roadway Inventory File

The TxDOT Roadway Inventory File, or RI-File, is a roadway characteristics database developed by TxDOT for use with the DPS crash history database. The RI-File uses the same control-section/milepoint system as the DPS database, which allows crashes to be linked to roadway segments relatively easily. The RI-File includes information about the traffic characteristics and geometry for roadway segments for both state highways and county roads. It does not have sufficient information about intersections, interchange ramps, or frontage roads to fully support safety-related applications.

The RI-File contains relevant highway and county road information for the years 2000 and earlier. After 2000, information about state highway segments in the RI-File has been superseded by other databases, described in the following section. However, the RI-File continues to be the repository for county road information.

The RI-File can be described as a relational database. The reference systems used to link it to the crash history database are both spatial and temporal. However, the temporal referencing system for the RI-File is limited to calendar years. Through the use of both referencing systems, the crashes on a specific highway segment can be identified for a specified calendar year. A change in geometry during the calendar year is posted the following year; hence, there is some risk that the relationship between crash frequency and road characteristics will be misleading whenever changes to the roadway characteristics occur during the analysis period.

Texas Reference Marker System Database

TxDOT began development of the TRM database in 1988. It is based on roadside reference markers that are usually placed on route marker signs at approximately two-mile intervals. All database attributes are located relative to their distance from the nearest reference marker.

TRM consists of three subset databases: RHiNo, Geo-HiNI, and P-HiNI. RHiNo contains information about road segments. Geo-HiNI contains information about curves. P-HiNI contains information about point-specific features of the roadway (e.g., intersections). All three of these databases are linked to each other using the reference marker system.

TRM includes information about speed limit, traffic characteristics, and geometry for all state highway facilities and some limited information about intersections. However, data for some roadway characteristics, such as grade or vertical curvature, are not available. P-HiNI does not contain intersection traffic control type or information about the presence of turn bays.

TRM is a relational database. The reference system used is spatial and is based on reference markers. Spatial linkages between the three subset databases and the DPS crash database are based on the control-section/milepoint reference system, which is common to all databases.

Crash Record Information System

TxDOT and DPS have completed a joint project to develop a new crash records information system. CRIS is operational and is currently being populated with Year 2002 to Year 2006 backlog crash data.

When completed, the CRIS Data Warehouse will contain all crash attributes and the roadway characteristics in relation to both existing conditions and conditions at the time of the crash.

Currently the map layers that comprise the spatial component of CRIS are centerline only. Hence, the roadway characteristics are related to the main lanes and not ramps, frontage roads, etc. A future enhancement to the system will incorporate an intelligent roadbed map layer. This will allow a crash to be located on the actual part of the roadway that it occurred and consequently will retrieve the roadway characteristics for that part of roadway.

Other Available Data

Other sources of roadway characteristic data exist outside of the aforementioned TxDOT databases. These data include the Highway Pavement Management System (HPMS) and the various traffic count data assembled by the Transportation Planning and Programming Division (TPP). HPMS contains information at selected locations in Texas. For example, it includes information about the presence of turn lanes and traffic control type at intersections. TPP often collects detailed traffic counts for special projects around Texas. These data include: frontage road traffic counts, ramp traffic counts, traffic counts by movement at intersections, and transit information.

EVALUATION OF AVAILABLE DATA

This section provides a more detailed evaluation of the data required by *SafetyAnalyst* and *HSM* with a comparison to the data that are available to TxDOT. Following this evaluation, some challenges in using the available data for various safety-related applications are discussed.

Data Required by Safety Evaluation Tools

The data required by *SafetyAnalyst* and *HSM* are described in this section. These data were identified through a review of draft documentation for both tools. For *SafetyAnalyst*, the final version of the database design was obtained from FHWA. For the *HSM*, the prototype chapter for two-lane highways was obtained from the National Cooperative Highway Research Program.

The data required for the two evaluation tools were categorized by facility component and by attribute category (e.g., traffic characteristic, traffic control, geometry, etc.). The data needed for frontage roads were not specifically identified by any of the aforementioned documents so it was estimated by the authors. These data will likely be needed to evaluate frontage road safety for design and hazardous location identification applications.

The findings from the review of data needs are summarized in Tables 3-3 through 3-7. A solid check mark (✔) is used to denote data that are required by the corresponding tool. An open check mark (✅) is used to denote data that are accepted as inputs to the tool and can be used to improve the accuracy of the analysis. Hence, these data are considered "desirable" and default values are used in their absence.

As indicated by Tables 3-3 through 3-7, data needed for the *HSM* are focused on highway segments and intersections. *SafetyAnalyst* also addresses segments and intersections as well as interchange ramps. Neither of these tools is intended to address frontage roads. A count of the number of attributes identified for each facility component indicates that the most data are needed for highway and frontage road segments. Intersections also require a large amount of data and interchange ramps require the least data.

Table 3-3. Required Roadway Data for Highway Segments.

Attribute Category	Attribute	Safety Evaluation Tool [1]		Availability [2]		
		Highway Safety Manual	***Safety-Analyst***	**RI-File**	**TRM**	**Other Sources [3]**
Location	Route type (e.g., IH, US, SH, FM)		✔	**Yes**	**Yes**	
	Route name		✔	**Yes**	**Yes**	
	County number		✔	**Yes**	**Yes**	
	Route segment location (e.g., milepoint)	✔	✔	**Yes**	**Yes**	
General	Area type (e.g., urban, rural)	✔	✔	**Yes**	**Yes**	
	Roadway functional class	✔	✔	**Yes**	**Yes**	
Traffic control	Two-way vs. one-way operation	✔	✔	**Yes**	**Yes**	
Traffic characteristics	ADT (by specified year)	✔	✔	**Yes**	**Yes**	
Geometry	Number of through lanes (both directions)	✔	✔	**Yes**	**Yes**	
	Median type (e.g., raised, depressed)	✔	✔	No	**Yes**	
	Auxiliary lanes (e.g., TWLTL, HOV, bus)	✔	✅	No	Est.	Field
	Shoulder type (i.e., paved, turf, or curb)	✔	✔	**Yes**	**Yes**	
	Shoulder width	✔	✅	No	**Yes**	
	Lane width (surface width+paved median)	✔	✅	Yes	Est.	HPMS
	Length of curve (if segment is on curve)	✔		No	**Yes**	
	Radius of curve (if segment is on curve)	✔		No	**Yes**	
	Presence of spiral (if segment is on curve)	✔		No	**Yes**	
	Policy superelevation (if segment is on curve)	✔		No	No	Field
	Existing superelevation (if segment is on curve)	✔		No	No	Field
	Grade	✔		No	No	Field
	Driveway density	✔	✅	No	No	Field
	Presence of a passing lane	✔		No	Est.	Field
	Presence of a short 4-lane passing section	✔		No	Est.	Field
Roadside	Roadside hazard rating (1 best to 7 worst) [4]	✔		No	No	Field

Notes:

1 - ✅ - desirable data. ✔ - required data.

2 - Yes: Data are available; No: Data are not available; Est.: Data can be estimated but require field verification.

3 - Other sources: HPMS - Highway Pavement Management System sample sections; Field - a field study or other manual data collection method. Data from these sources may not be available for all locations.

4 - A rating system that characterizes the crash potential of the roadside adjacent to the highway. The system uses a seven-point scale with "1" used to denote wide horizontal clearance of 30 ft or more and side slopes flatter than 1:4. A "7" is used to denote a clearance of less than 5 ft and side slopes 1:2 or steeper.

Table 3-4. Required Roadway Data for Intersections.

Attribute Category	Attribute	Safety Evaluation Tool [1]		Availability [2]		
		Highway Safety Manual	*Safety-Analyst*	RI-File	TRM	Other Sources [3]
Location	Route type (e.g., IH, US, SH, FM)		✔	**Yes**	**Yes**	
	Route name		✔	**Yes**	**Yes**	
	County number		✔	**Yes**	**Yes**	
	Intersection location (e.g., milepoint)	✔	✔	No	**Yes**	
	Designated major-road direction (i.e., N-S, E-W)		✔	No	No[4]	HPMS
General	Area type (e.g., urban, rural)	✔	✔	**Yes**	**Yes**	
	Intersection type (i.e., four-leg, Tee, roundabout)	✔	✔	No	**Yes**	
Traffic control	Traffic control type (e.g., signal, sign)	✔	✔	No	No[4]	HPMS
Geometry	Skew angle	✔		No	**Yes**	
Leg location	Position code (e.g., 1 = major road, increasing milepoint) (specified by leg)	✔	✔	No	**Yes**	
Leg traffic characteristics	ADT (specified by leg and year)	✔	✔	**Yes[5]**	**Yes[5]**	HPMS
Leg geometry	Number of through lanes (specified by leg)	✔	✔	**Yes[5]**	**Yes[5]**	HPMS
	Number of left-turn lanes (specified by leg)	✔	✔	No	No	HPMS
	Number of right-turn lanes (specified by leg)	✔		No	No	HPMS
	Intersection sight distance adequacy (yes/no)	✔		No	No	Field
	Median type (e.g., raised, flush) (specified by leg)	✔	✔	No	**Yes[5]**	HMPS

Notes:
1 - ✔ - required data.
2 - Yes: Data are available; No: Data are not available.
3 - Other sources: HPMS - Highway Pavement Management System sample sections; Field - a field study or manual data collection.
4 - Data to be added to P-HiNI by TxDOT at a future time.
5 - Data available only for intersection legs on the state highway system. HPMS sample locations have information for all legs.

Table 3-5. Required Roadway Data for Interchange Ramps.

Attribute Category	Attribute	Safety Evaluation Tool [1]		Availability [2]		
		Highway Safety Manual	*Safety-Analyst*	RI-File	TRM	Other Sources [3]
Location	Route type (e.g., IH, US, SH, FM)	Model is not being developed at this time.	✔	**Yes**	**Yes**	
	Route name		✔	**Yes**	**Yes**	
	County number		✔	**Yes**	**Yes**	
	Ramp location (e.g., milepoint)		✔	No	**Yes**	
General	Area type (e.g., urban, rural)		✔	**Yes**	**Yes**	
	Ramp type (e.g., on ramp, off ramp, connector)		✔	No	**Yes**	
	Ramp configuration (e.g., parclo, free-flow loop)		✔	No	No	Field
Traffic characteristics	ADT (specified by year)		✔	No	No	TPP Counts

Notes:
1 - ✔ - required data.
2 - Yes: Data are available; No: Data are not available.
3 - Other sources: TPP Counts - traffic counts by TPP that are not in computerized databases; Field - a field study or manual data collection.

Table 3-6. Required Roadway Data for Frontage Roads.

Attribute Category	Attribute	Safety Evaluation Tool		Availability [1]		
		Highway Safety Manual	*Safety-Analyst*	RI-File	TRM	Other Sources [2]
Location	Route type of mainlanes (e.g., IH, US, SH, FM)	Model is not being developed at this time.	Model is not being developed at this time.	**Yes**	**Yes**	
	Route name of mainlanes			**Yes**	**Yes**	
	County number			**Yes**	**Yes**	
	Route segment location of mainlanes			**Yes**	**Yes**	
General	Area type (e.g., urban, rural)			**Yes**	**Yes**	
	Roadway functional class of mainlanes			**Yes**	**Yes**	
Traffic control	Two-way vs. one-way operation			No	**Yes**	
Traffic characteristics	ADT (by specified year)			No	**Yes**	
Geometry	Number of through lanes (both directions)			No	**Yes**	
	Auxiliary lanes (e.g., HOV, bus, bike)			No	Est.	Field
	Shoulder type (i.e., paved, turf, or curb)			No	**Yes**	
	Shoulder width			No	**Yes**	
	Lane width			No	Est.	Field
	Length of curve (if segment is on a curve)			No	No[3]	Field
	Radius of curve (if segment is on a curve)			No	No[3]	Field
	Presence of spiral (if segment is on a curve)			No	No[3]	Field
	Policy superelevation (if segment is on a curve)			No	No	Field
	Existing superelevation (if segment is on a curve)			No	No	Field
	Grade			No	No	Field
	Driveway density			No	No	Field
Roadside	Roadside hazard rating (1 best to 7 worst)			No	No	Field

Notes:
1 - Yes: Data are available; No: Data are not available; Est.: Data can be estimated but require field verification.
2 - Other sources: HPMS - Highway Pavement Management System sample sections; TPP Counts - traffic counts by TPP that are not in computerized databases; Field - a field study or other manual data collection method. Data from these sources may not be available for all locations.
3 - Curve data linked to roadway centerline only. No frontage road specific curve data available.

Also shown in the last few columns of each table is an indication as to whether a specific data attribute is available in the corresponding TxDOT database. The availability of an attribute is denoted by “Yes” in the appropriate table cell. The roadway characteristics databases considered include RI-File and TRM. CRIS is not listed separately because it is created from RHiNo, Geo-HiNI, and P-HiNI and contains all of the elements of those three databases.

A review of Tables 3-3 through 3-6 indicates that none of the available databases provides all of the roadway characteristics data needed by the evaluation tools. In contrast, all of the crash data needed are available in the DPS and CRIS databases. In general, the RI-File provides the fewest number of needed attributes while TRM provides many of the needed attributes. Some of the needed data may be available from other sources within TxDOT. Regardless, it appears that some roadway characteristic data may need to be collected manually.

Table 3-7. Required Data for Crash Database.

Attribute Category	Attribute	Safety Evaluation Tool		Availability
		Highway Safety Manual	***Safety-Analyst***	**DPS & CRIS**
Location	Accident case number		✔	**Yes**
	Route type (e.g., IH, US, SH, FM)	✔	✔	**Yes**
	Route name	✔	✔	**Yes**
	County number	✔	✔	**Yes**
	Accident location (e.g., milepoint)	✔	✔	**Yes**
	Relationship to junction (e.g., at intersection)	✔	✔	**Yes**
Description	Accident date	✔	✔	**Yes**
	Accident time		✔	**Yes**
	Accident type and manner of collision	Calib.[1]	✔	**Yes**
	Number of vehicles involved	Calib.[1]	✔	**Yes**
	Accident severity	Calib.[1]	✔	**Yes**
Vehicle description (repeat for each vehicle)	Initial direction of travel		✔	**Yes**
	Vehicle maneuver/action (e.g., straight, backing)		✔	**Yes**
	Vehicle configuration (e.g., pass. car, light truck)		✔	**Yes**
	First harmful event (e.g., overturn)		✔	**Yes**

Note:
1 - Calib.: data are needed only for model calibration.

About 80 percent of the needed attributes are available or can be estimated for highway segments. The percentage of available attributes for the other facility components ranges from about 60 to 75 percent; however, the attributes missing for intersections and ramps appear to be more critical in terms of their likely correlation to crash frequency.

As indicated in Table 3-4, data are only available for intersection legs on the state highway system. Thus, all of the needed data are available at intersections of two state highways. However, they are not available for the intersection legs that are on a city street or county road.

Additional safety-related documents that describe safety prediction models were reviewed to determine if additional roadway characteristics or crash history data would be useful to future safety evaluation tools. The results of this review indicated that several additional roadway characteristics would be desirable in an analysis of crash data. These data are listed in Table 3-8.

As indicated in Table 3-8, desirable roadway characteristics data were identified for highway segments, intersections, and interchange ramps. No additional data were identified for frontage roads. The last few columns in Table 3-8 indicate that a few of the data are available from the various Texas databases or from other sources.

Table 3-8. Desirable Data for Roadway Characteristics Database.

Facility Component	Attribute Category	Attribute	Availability [1]		
			RI-File	TRM	Other Sources [2]
Highway segment	General	Terrain type (e.g., level, rolling)	No	No	Field
	Traffic control	Posted speed limit	No	**Yes**	
		Presence of curb parking	No	**Yes**	
	Traffic characteristics	Percent heavy vehicles (specified by year)	**Yes**	**Yes**	
	Geometry	Median width	No	**Yes**	
		Bikeway (e.g., signed route) (specified by direction)	No	No	Field
		Curb type	No	**Yes**	
		Median opening density	No	**Yes**	
Intersection	Geometry	Crossroad leg offset distance	No	**Yes**	
	Leg traffic control	Left-turn phasing (if signalized)	No	No	HPMS
		Turn prohibitions	No	No	HPMS
		Presence of curb parking	No	**Yes**	
	Leg traffic characteristics	Turn movement ADT (specified by year)	No	No	TPP Counts
Interchange ramp	Geometry	Type of connection at freeway (e.g., acceleration lane)	No	No	Field
		Type of connection at crossroad (e.g., intersection)	No	Yes [3]	
		Number of lanes	No	No [4]	Field
		Length	No	No [4]	Field

Notes:

1 - Yes: Data are available; No: Data are not available.

2 - Other sources: HPMS - Highway Pavement Management System sample sections; TPP Counts - traffic counts by TPP that are not in computerized databases; Field - a field study or other manual data collection method. Data from other sources may not be available for all locations.

3 - Data available only if interchange ramp connects to the frontage road, or the crossroad is a state highway.

4 - Data to be added by TxDOT at a future time.

Challenges Using Available Data

As discussed in the previous section, some of the data necessary to use the *HSM* and *SafetyAnalyst* are not available in the Texas databases. Some of these data are more important than others and default values can be used to replace some attributes. The most important data (from the perspective of crash correlation) that are not available in an existing database are described in this section. Some options for acquiring the missing data and assembling a database suitable for safety applications are described in this section.

The Texas databases will need to be supplemented with several additional data attributes before the *HSM* and *SafetyAnalyst* can be used to evaluate design alternatives or identify hazardous locations. These attributes are identified in Table 3-9 by facility component and attribute category.

Table 3-9. Key Roadway Characteristics Data Needed for Safety Evaluation.

Facility Component	Attribute Category	Attribute
Highway segment	Geometry	Auxiliary lanes (e.g., TWLTL, HOV, bike)[1]
		Lane width[1]
		Driveway density
		Roadside hazard rating (1 best to 7 worst)
Intersection	Traffic control	Traffic control type (e.g., stop sign on major road)[2]
	Leg traffic characteristics	ADT for non-system legs
	Leg geometry	Number of through lanes on non-system legs
		Number of left-turn lanes all legs
Interchange ramp	Geometry	Ramp type (e.g., on ramp, off ramp)[1]
		Ramp configuration (e.g., parclo, free-flow loop)
	Traffic characteristics	ADT (specified by year)
Frontage road	Traffic control	Two-way vs. one-way operation
	Geometry	Lane width [1]
		Driveway density
		Roadside hazard rating (1 best to 7 worst)

Notes:
1 - Can be estimated from inventory data but requires field confirmation.
2 - To be added by TxDOT at a future time. Currently requires field data collection.

A variety of data sources can be used to obtain the needed supplemental data. One source for roadway and intersection data is HPMS, which has already been mentioned. The sources for traffic control and geometric data also include topographic maps, aerial photography, as-built plans, and a field survey. Traffic characteristics data can be obtained from one-day traffic counts, many of which may have been performed by TPP prior to construction projects. All of these sources reflect conditions recorded on a specified date and must be reconciled with the period of time represented by the crash data. The possibility of a change in traffic control or geometry over time will need to be evaluated and accommodated, if possible, such that the conditions at the time of the crash are known. Traffic counts will need to be adjusted backward to reflect the likely growth in traffic volume since the time period associated with the crash data.

The number of facility components (i.e., highway segments, intersections, ramp segments, ramp terminals, frontage road segments, and frontage road intersections) for which these supplemental data are collected is dependent on the application. If the evaluation tool is being used for project safety evaluation, then there are likely only a few components of interest. In this situation, the supplemental traffic control and geometric data would likely be collected during a field survey.

For safety model calibration, a randomly selected sample of facility components in a specified region would be identified, and depending on their number, the needed data would be obtained from a field survey of individual components or a review of maps, photographs, and plans. For hazardous location identification, the number of sites to be evaluated can be significant. In this instance the needed data would likely be obtained from a review of maps, photographs, and plans. In all cases, the effort to gather the supplemental data would be balanced by the cost of its collection and the desired accuracy of the safety evaluation.

SUMMARY OF FINDINGS

This part of the chapter summarizes a review of the various sources of safety-related data maintained by, or available to, TxDOT. It also summarizes the findings from an evaluation of the suitability of these data sources to the calibration and application of various safety evaluation tools.

The primary databases used for safety evaluation and management in Texas include the DPS crash history database (also called the crash report repository) and TRM. TRM is a roadway characteristics database where the location of various road features is identified using a reference marker system.

TRM has some limitations with regard to its use for safety applications. Specifically, some roadway characteristics related to intersections, ramps, and frontage roads are not included in these databases. Examples of this omitted data include: intersection traffic control type, ramp configuration, ramp traffic volume, and frontage road direction of travel (i.e., one-way vs. two-way).

Data needed for the *HSM* is focused on highway segments and intersections. *SafetyAnalyst* also addresses segments and intersections as well as interchange ramps. Neither of these tools is planning to address frontage roads. Regardless, the database suitable for use with these safety evaluation tools (and others) should uniquely locate, and have information for, each of the following highway facility components:

- highway segments,
- intersections,
- ramp segments,
- ramp terminals,
- frontage road segments, and
- frontage road intersections.

This "idealized" database can be contrasted with existing Texas databases which are focused almost exclusively in their coverage on highway segments (with some limited coverage of intersections). These databases are not directly applicable to the identification of ramp or frontage road components and their associated individual crash frequencies. Moreover, none of the available databases provides all of the roadway characteristics data needed by the evaluation tools.

REFERENCES

1. Krammes, R.A., and C. Hayden. "Making Two-Lane Roads Safer." *Public Roads*. Vol. 66, No. 4. Federal Highway Administration, U.S. Department of Transportation, Washington, D.C., January 2003.

2. Harwood, D.W., E.R. Kohlman Rabbani, K.R. Richard, H.W. McGee, and G.L. Gittings. *NCHRP Report 486: Systemwide Impact of Safety and Traffic Operations Design Decisions for 3R Projects.* National Cooperative Highway Research Program, Transportation Research Board, Washington, D.C., 2003.

3. Mak, K., and D.L. Sicking. *Roadside Safety Analysis Program (RSAP) - User's Manual*. NCHRP Project 22-9. National Cooperative Highway Research Program, Transportation Research Board, Washington, D.C., June 2002.

4. Mak, K., and D.L. Sicking. *NCHRP Report 492: Roadside Safety Analysis Program (RSAP) - Engineer's Manual.* Transportation Research Board, Washington, D.C., 2003.

5. *SafetyAnalyst*. Federal Highway Administration, U.S. Department of Transportation, Washington, D.C. http://www.safetyanalyst.org. Accessed April 2004.

6. Hughes, W., K. Eccles, D. Harwood, I. Potts, and E. Hauer. *Research Results Digest 286: Development of a Highway Safety Manual*. National Cooperative Highway Research Program, Transportation Research Board, Washington, D.C., March 2004.

CHAPTER 4. CALIBRATION OF AMFs FOR RURAL TWO-LANE HIGHWAY SEGMENTS

OVERVIEW

This chapter describes the calibration of selected accident modification factors (AMFs) using crash data for rural two-lane highways in Texas. The AMFs considered for this calibration were those described in the *Interim Roadway Safety Design Workbook* (*Workbook*) that are intended for rural highway safety evaluations (*1*). AMFs can be represented by a constant or as a function of other variables.

There were two objectives for this calibration task. One objective was to determine if AMFs developed with crash data from other regions of the U.S. were reasonably accurate when used to evaluate the safety of a Texas highway. Achievement of this objective would require a comparison of the original AMF to crash trends on Texas highways. Regression techniques were used for this comparison because they provide useful statistics describing the quality of fit and, if the fit were unacceptable, could also be used to "re-calibrate" the AMF and improve the fit. Regardless, if re-calibration is found to be unnecessary (in a statistical sense) then it could be claimed that the original AMF is suitable for use in Texas without adjustment. However, if re-calibration is needed, then it would have to remain in the AMF to allow for unbiased use in Texas.

The second objective of this task was to develop an appropriate method for calibrating (or re-calibrating) AMFs that are functions of other variables, such as lane width or curve radius. Currently, authoritative documents that describe AMFs, and their use for safety evaluation, do not describe guidelines for calibrating AMF functions.

A three-step procedure was developed to achieve the aforementioned objectives. Initially, the statistical fit of an existing AMF was assessed. If the fit was unsatisfactory, then the existing AMF was re-calibrated. If the fit of the re-calibrated AMF was unsatisfactory, then a new AMF function was derived and calibrated. Hereafter, this procedure is referred to as "AMF calibration."

This chapter is divided into four parts. The first part describes the rural highway segment AMFs listed in Chapter 3 of the *Workbook.* It also discusses the rationale used to select the specific AMFs for calibration. The second part describes the methodology that was developed for calibrating an AMF. The third part describes the findings from the calibration of the curve radius AMF. The fourth part describes the findings from the calibration of the lane width AMF and the shoulder width AMF.

RURAL HIGHWAY AMFS

This part of the chapter is divided into three sections. The first section describes the rural highway AMFs listed in the *Workbook*. The second section discusses the rationale used to select the specific *Workbook* AMFs for calibration. The last section identifies the process used to divide Texas into regions of similar climate and geography.

Candidate AMFs

The AMFs identified in Chapter 3 of the *Workbook* are listed in Table 4-1. All of these AMFs were developed for use in the evaluation of rural highway safety. Three of the 17 AMFs listed are constants and are identified by the shaded cells in the table. The remaining 14 AMFs are functions of one or more input variables. The table identifies each AMF by name and indicates the variables needed to estimate its value.

Table 4-1. AMFs for Rural Highway Segments.

Application	Accident Modification Factor	Input Variables	Effect on Safety
Geometric design	Horizontal curve radius	Radius Curve length	AMF increases with decreasing radius or decreasing length.
	Spiral transition curve	Radius Curve length	AMF increases with increasing radius or increasing curve length.
	Grade	Grade	AMF increases with increasing grade.
	Lane width	Lane width ADT	AMF increases with decreasing lane width. AMF increases with increasing ADT.
	Outside shoulder width	Shoulder width ADT	AMF increases with decreasing shoulder width. AMF increases with increasing ADT.
	Inside shoulder width	Shoulder width	AMF increases with decreasing shoulder width.
	Median width	Median width	AMF increases with decreasing median width.
	Shoulder rumble strips	AMF is a constant	Addition of rumble strips reduces crashes.
	Centerline rumble strips	AMF is a constant	Addition of rumble strips reduces crashes.
	Two-way left-turn lane median type	Driveway density	AMF decreases with increasing driveway density.
	Superelevation	Existing super. rate Design super. rate	AMF increases with increasing difference between design and existing rates.
	Passing lane	AMF is a constant	Addition of passing lane reduces crashes.
Roadside design	Horizontal clearance	Horizontal clearance	AMF increases with decreasing clearance.
	Side slope	Side slope	AMF increases with increasing slope.
	Utility pole offset	Pole density Pole offset	Pole density has negligible effect on AMF. AMF increases with decreasing offset.
	Bridge width	Bridge width Approach width	AMF increases with decreasing bridge width or increasing approach width.
Access control	Driveway density	Driveway density	AMF increases with increasing driveway density.

The AMFs provided in the *Workbook* were all derived from safety information presented in the literature. They are intended to be used to evaluate the effect of a geometric design element on the frequency of injury (plus fatal) crashes. Whenever possible, they were derived from injury (plus fatal) crash data or existing models that were calibrated using injury (plus fatal) crash data. However, in those instances where an AMF that explicitly addressed injury (plus fatal) crashes could not be derived from the literature, it was derived from *total* (i.e., property-damage-only, injury, and fatal) crash data or existing models that were calibrated using total crash data.

Selection of AMFs for Calibration

The AMFs identified in Table 4-1 were individually evaluated to determine which AMFs could be calibrated using data available in the Texas Reference Marker system (TRM) database. This database is maintained by TxDOT. It contains geometric and traffic attributes for the Texas state highway system. The database consists of several thousand highway segments, each of which is described in terms of its geometry, traffic, and location attributes. In fact, about 140 attributes are used to describe each highway segment in TRM. However, only about 20 attributes are used to describe segment geometry or traffic characteristics, and about 10 more attributes are used to describe road name and physical location. The remaining attributes describe administrative designations and road-management-related information that are not particularly helpful for safety analysis. The TRM database is described in more detail in Chapter 3.

A highway segment in TRM is defined as a length of road along which no one attribute changes. A change in any one attribute dictates the end of one segment and the start of a new segment. The average tangent segment length in TRM is about 0.4 mi and the average curve length in TRM is about 0.1 mi.

As described in Chapter 3, several geometric attributes in TRM are consistent with the inputs needed for several AMFs listed in Table 4-1. A review of the list of available attributes indicated that input variables for the following AMFs could be extracted from the TRM database:

- horizontal curve radius (i.e., curve length and radius);
- spiral transition curve (i.e., curve length, radius, and spiral presence);
- lane width (i.e., lane width is computed from the TRM variable for surface width);
- outside shoulder width;
- inside shoulder width; and
- median width.

Of the AMFs in the preceding list, the last two listed were not applicable to two-lane highway segments and were excluded from further consideration. Of those that remained, it was decided that the AMF for spiral presence would have a lower priority than the other three AMFs because few spiral transitions are being constructed in Texas. Thus, the following three AMFs were selected for calibration:

- horizontal curve radius (i.e., curve length and radius);
- lane width; and
- outside shoulder width.

Additional TRM attributes that would be used to identify rural two-lane highway segments included: functional system (i.e., all rural functional classes), number of lanes (i.e., two lanes), and median width (i.e., no median).

Texas Regions

One of the goals of the calibration task was to determine whether the AMFs developed from data sources external to Texas could provide accurate estimates when applied to Texas highways. The task was expanded to include possible AMF calibration on a regional level. To facilitate the need for possible regional calibration, the state of Texas was divided into six regions of similar geography and climate. The regional boundaries were defined to be consistent with the TxDOT district boundaries. These regions are shown in Figure 4-1.

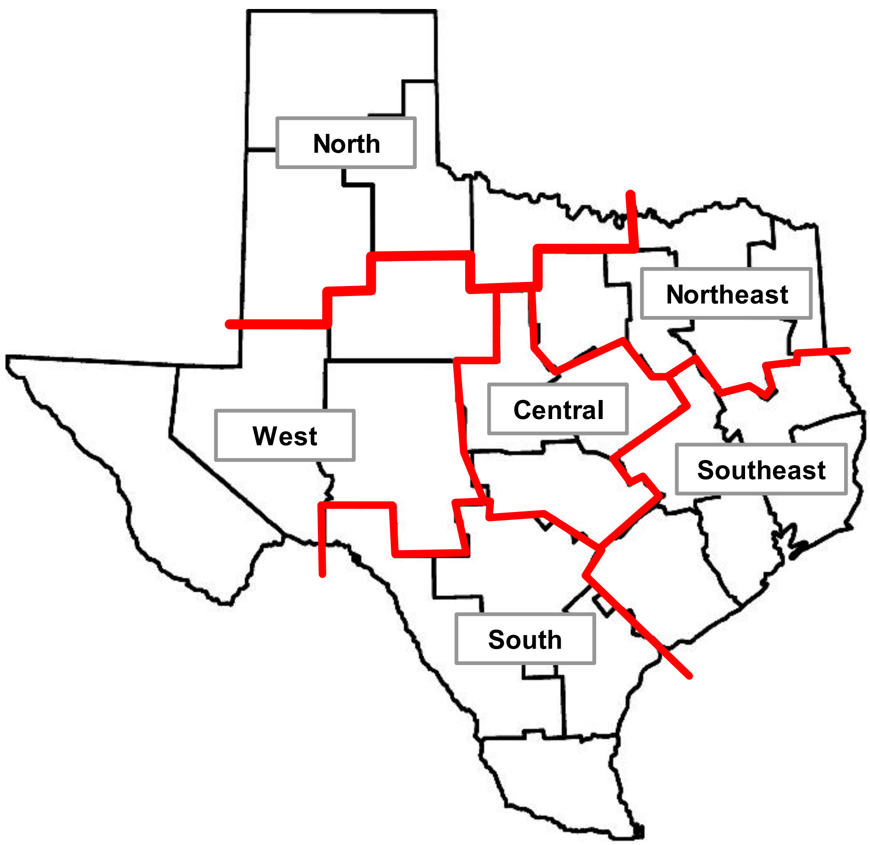


Figure 4-1. Regions Used for AMF Calibration.

As indicated in Figure 4-1, the state of Texas was divided into six regions: north, northeast, southeast, south, west, and central Texas. The TxDOT districts in each of these regions are listed in Table 4-2.

Table 4-2. Distribution of TxDOT Districts to Regions.

TxDOT District by Region [1]					
North	**Northeast**	**Southeast**	**South**	**West**	**Central**
Wichita Falls (3) Amarillo (4) Lubbock (5) Childress (25)	Paris (1) Fort Worth (2) Tyler (10) Dallas (18) Atlanta (19)	Lufkin (11) Houston (12) Yoakum (13) Bryan (17) Beaumont (20)	San Antonio (15) Corpus Christi (16) Pharr (21) Laredo (22)	Odessa (6) San Angelo (7) Abilene (8) El Paso (24)	Waco (9) Austin (14) Brownwood (23)

Note:
1 - Numbers in parentheses are the district number, as assigned by TxDOT.

CALIBRATION METHODOLOGY

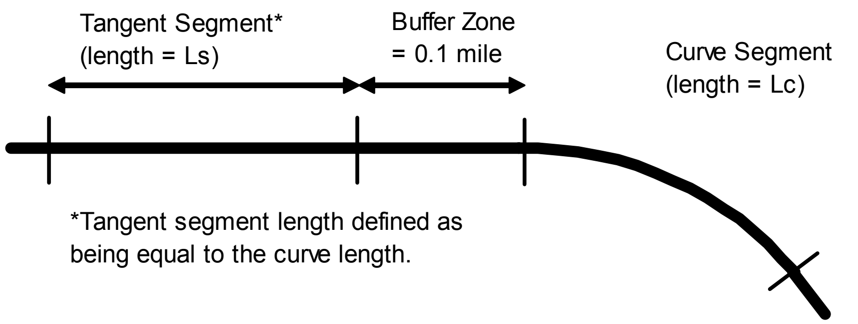
This part of the chapter describes the methodology developed for calibrating AMFs. The methodology consists of a segmentation procedure and a calibration procedure. The segmentation procedure describes the technique for defining the highway segment used for AMF calibration. The calibration procedure describes the steps involved in calculating the calibration coefficients and determining whether they are statistically significant. Each procedure is described in a separate section.

Segmentation Procedure

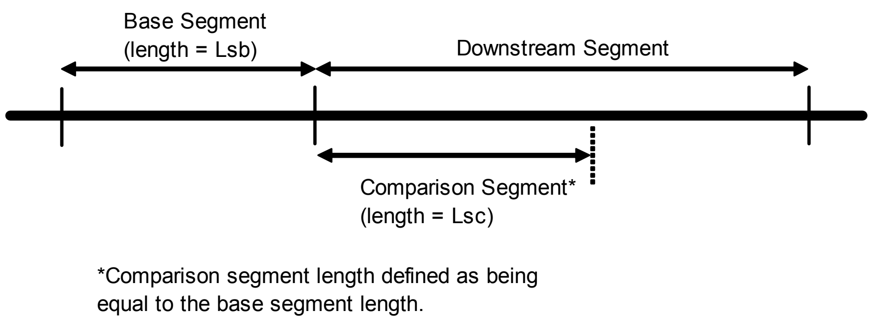
The calibration methodology is founded on the use of paired road segments as the basis of AMF calibration. The attributes of each segment pair are identical except for a difference in the attributes that are associated with the AMF input variables. By selecting pairs of matched segments, the effect of the selected attributes on safety is isolated and all other factors are controlled. To ensure that the pairs are as nearly matched as possible, the segments are selected to be adjacent to each other on the same roadway. Figure 4-2 illustrates how the matched pairs are physically related for the examination of curve radius, shoulder width, and lane width.

As shown in Figure 4-2, the segment pairs are located on adjacent segments of roadway to ensure similarity in environment as well as geometry and traffic stream. For each pair, the length of one segment has to be adjusted such that both segments have the same length. Thus, for the curve-tangent pairs, this requires using only a portion of the adjacent tangent segment. For the tangent-tangent pairs, the shorter segment defines the length for both segment pairs, in which case only a portion of the longer segment is used. Segment length is used primarily to define the boundaries within which segment crashes are identified.

For the curve-tangent pair shown in Figure 4-2a, there is a 0.1 mi buffer separating the tangent segment and the curve segment. This "buffer zone" is used to ensure that curve-related crashes are not inadvertently placed on the tangent segment and vice versa. The length of this buffer reflects a recognition of the precision of crash location in the Texas Department of Public Safety crash database (DPS). This database locates crashes to the nearest 0.1 mi.



a. Segment Pairs for Calibrating the Horizontal Curve Radius AMF.



b. Segment Pairs for Calibrating Shoulder Width and Lane Width AMFs.

Figure 4-2. Segment Pairing for AMF Calibration.

Calibration Procedure

The TRM database includes several thousand segments from which to select the segment pairs described in the previous section. However, there are relatively few pairs that have identical geometric and traffic attributes (with the exception that the one attribute of interest can change). Moreover, the length of these segments is relatively small such that most segments have no recorded crashes over a period of three years. As a result, the direct comparison of crash frequency on individual matched segment pairs tends to yield little useful information.

A procedure is described in this section that overcomes the aforementioned limitation of low crash count for matched segment pairs, yet still allows the calibration of an AMF. The procedure uses a multivariate model to estimate the expected crash frequency for one segment of each pair, as may be influenced by its geometric and traffic attributes. This estimate is then refined using the empirical Bayes (EB) technique developed by Hauer (*2*) to include information about the reported crash frequency for the segment. The segment for which the expected crash frequency is estimated is referred to as the "before" segment.

The second segment of each pair is considered to be the "after" segment. Its reported crash frequency is compared with the empirical Bayes estimate for the "before" segment during AMF calibration. To facilitate this calibration, the pairs are aggregated into groups where all members of a group have similar "before" values and similar "after" values for the geometric element of interest. For example, one group could consist of all segment pairs where the "before" lane width is 10 ft and the "after" lane width is 12 ft, a second group could consist of all pairs where the "before" lane width is 11 ft and the "after" lane width is 13 ft, etc. The number of groups for each possible lane width combination is factorial; however, many of these combinations are rarely found on actual highway segment pairs.

For each group, the reported crash frequency is summed for all "after" segments and the expected crash frequency is summed for all the "before" segments. Regression analysis is used to calibrate an AMF using the data for all groups. The computational steps in this procedure are described in the following paragraphs.

Step 1. Assemble Segment-Pair Database

The objective of this step is to identify segment pairs in the TRM database. The pairs should be defined using the methods described in the previous section and should be selected to facilitate the calibration of one specific AMF. The segment pairs should be identical, except that they should exhibit some variation in the input variables associated with the subject AMF. For example, if the AMF is for curve radius, then one segment should have a curve with a specified radius and the other segment should be on the adjacent tangent. If the AMF to be calibrated is for shoulder width, then the segment pairs should be identical except that they would not be required to have the same shoulder width.

The segment-pair (SP) database is used to form two additional databases. The first database is used for regression model calibration. It is referred to as the "SP regression" database. In this database, each individual segment represents one observation. Thus, if there are n segment pairs, there would be $2n$ observations in the SP regression database. The second database is used for AMF calibration. It is referred to as the "SP group" database. In this database, the number of observations (or groups) is equal to the number of unique combinations of the "before" and "after" values for the geometric element of interest.

If the SP regression database has too few observations for model calibration, then a non-paired (NP) regression database can be directly assembled for use in Step 2. The NP regression database does not require each of its segments to be part of a segment pair; however, they must satisfy all other site selection criteria (e.g., rural two-lane highway). This deviation increases the size of the regression database; however, it also weakens the similarity between the regression and SP group databases.

Step 2. Calibrate the Multivariate Model

The objective of this step is to calibrate a multivariate regression model using the regression database. The form of the model is:

$$E[N] = ADT^{b_1} \, L \quad e^{(b_0 + \sum b_i x_i)} \qquad \textbf{(1)}$$

where,

$E[N]$ = expected crash frequency, crashes/yr;
L = highway segment length, mi;
ADT = average daily traffic volume, veh/d;
b_i = calibration coefficients (i = 0, 1, 2, 3,...); and
x_i = variable describing various geometric or traffic-control-device attributes (i = 0, 1, 2, 3,...).

Two types of variables are included in the model: key variables and supplemental variables. The key variables include: traffic volume, segment length, and the input variables in the subject AMF function (e.g., curve radius, lane width, shoulder width, region, etc.). These variables are kept in the multivariate model regardless of whether they are found to be statistically significant.

All key variables that have continuous values, except volume and length, are converted to discrete values and included in the model as categorical variables. Discrete variables, such as spiral curve or turn bay presence, do not require this conversion and an indicator variable can be used (i.e., x_i = 1.0 if bay is present, 0.0 if it is not present). In contrast, variables that are continuous (e.g., radius, lane width, etc.) should be rounded to form scalar categories, with the value for each category representing a small range of values for the input variable. For example, to calibrate the lane width AMF, the lane width of each segment could be rounded to the nearest 1.0 ft and converted to scalar categories of 9, 10, 11, 12, and 13 ft. Each of these categories would have its own unique indicator variable and calibration coefficient b_i in the regression model. The number of categories used for lane width would depend primarily on the range of values associated with the variable and the number of observations in each category.

The advantage of using categorical variables in the regression model is that they do not impose any preconceived functional relationship on the estimates obtained from the model. If a functional relationship was used in the regression model (e.g., $b_i \times$ lane width), then this relationship could be reflected in the expected values used to calibrate the AMF. It follows that, if the functional relationship used is only a simple approximation of a more complicated relationship, then it could indirectly bias the AMF calibration. The use of categorical variables minimizes the potential for bias because it does not require the specification of a function for the subject AMF input variable. If the subject AMF has several continuous input variables, then each variable would be converted into discrete values and inserted in the multivariate model as a categorical variable.

The second type of variable to include in the multivariate model is the "supplemental" variable (e.g., speed limit). Ideally, the regression database would be subset such that all included segments have identical attributes (except for the key variables) and that these attributes would be considered "typical" for the type of highways being considered. This approach would eliminate the need for supplemental variables. However, it is not always possible to find a sufficient number of "identical" segments while maintaining the minimum sample size needed for statistical significance. In this case, the need for identical segments is relaxed and a categorical variable is included in the model to account for differences in this attribute among the sites. A supplemental variable is kept in the model if its category (i.e., the collective set of indicator variables) is statistically significant.

It is assumed that crash occurrence at a particular location is Poisson distributed and that the distribution of the mean crash frequency for a group of similar segments is gamma distributed. In this manner, the distribution of crashes for a group of similar locations can be described by the negative binomial distribution. For highway segments, the variance of this distribution is:

$$V[X] = y\,E[N] + \frac{(y\,E[N])^2}{k\,L} \qquad \textbf{(2)}$$

where,
$V[X]$ = crash frequency variance for a group of similar locations;
X = reported crash count for y years, crashes;
y = time interval during which X crashes were reported, yr; and
k = over-dispersion parameter, mi^{-1}.

Equation 2 includes a variable for the length of the segment. As demonstrated by Hauer (*3*), this variable should be added to ensure that the regression model coefficients are not biased by exceptionally short segments.

Step 3. Estimate the Segment Expected Crash Frequency

The objective of this step is to estimate the expected crash frequency for each "before" segment in the segment-pair database. For the analysis of curve radius, the "before" segments are those found on the tangents. The EB method developed by Hauer (*2*) is used to compute this estimate. The expected crash frequency *E[N]* from the multivariate regression model represents the average crash frequency for all segments that have the same volume, length, and value for any other variables in the regression model. The EB method combines the expected crash frequency *E[N]* with the observed count of crashes *X* for the subject segment to yield a best-estimate of the expected crash frequency for that specific segment. This estimate is obtained using the following equations:

$$E[N|X] = E[N] \times weight + \frac{X}{y} \times (1 - weight) \qquad \textbf{(3)}$$

with,

$$weight = \left(1 + \frac{E[N]\,y}{k\,L}\right)^{-1} \qquad \textbf{(4)}$$

where,
$E[N|X]$ = expected crash frequency given that X crashes were reported in y years, crashes/yr; and
weight = relative weight given to the prediction of expected crash frequency.

The variance of the expected crash frequency can be estimated using the following equation:

$$V[N|X] = (1 - weight)\,\frac{E[N|X]}{y} \qquad \textbf{(5)}$$

where,
$V[N|X]$ = variance of $E[N|X]$.

Step 4. Assemble Group Database

The objective of this step is to assemble the SP group database by aggregating the segment-pair data into groups. Each segment pair is assigned to the one group in which all members have a similar "before" and a similar "after" value for the geometric element of interest. The aggregation is based on the specified geometric elements (e.g., lane width, shoulder width, or curve radius) for which the AMF is being calibrated. For example, if five lane width categories are represented in the SP group database (i.e., 9, 10, 11, 12, and 13 ft), then there can be as many as 25 (= 5 × 5) unique combinations of lane width for the "before" and "after" conditions represented by each segment pair.

Two values are computed for each group. One value represents the sum of the expected crash frequency for each "before" segment $E[N|X]_{before}$ in the group. The second value represents the sum of the reported crash frequency for each "after" segment X_{after} in the group.

Step 5. Calibrate the AMF Model

The objective of this step is to calibrate the subject AMF model. This objective is achieved by using regression to relate the paired crash frequencies in the SP group database, where the regression model includes the subject AMF. The form of the regression model is:

$$X_{after} = c_0 \; y \; E[N|X]_{before} \; AMF_m \quad \textbf{(6)}$$

with

$$AMF_m = f(c_j, x_j) \quad \textbf{(7)}$$

where,

X_{after} = sum of reported crash frequency for "after" segments in a group, crashes;
$E[N|X]_{before}$ = sum of expected crash frequency for "before" segments in a group, crashes/yr;
c_0 = calibration coefficient; and
AMF_m = AMF for geometric element *m*, as a function of geometric variables x_j and calibration coefficients c_j.

The AMF model represented by Equation 7 can have the form of a function, such as the following AMF for shoulder width:

$$AMF_{SW} = e^{c_0 (SW_w - SW_{w/o})} \quad \textbf{(8)}$$

where,

AMF_{SW} = shoulder width accident modification factor;
c_0 = constant describing the relationship between a change in shoulder width and a change in crash frequency;
SW_w = shoulder width on segment where width is different from base segment, ft; and
$SW_{w/o}$ = shoulder width of base segment, ft.

The AMF could also be a constant. Regardless of whether it is a function or a constant, regression analysis is used to calibrate the AMF. The use of expected crash frequency $E[N|X]$ for the "before" segments in Equation 6 (instead of the reported crash frequency X for the before segments) is intended to minimize the variability associated with the independent variable.

If it is desired to test whether the calibration coefficient for shoulder width varies by region of the state, then the following variation of Equation 8 can be used:

$$AMF_{SW} = e^{(b_1 + b_2 I_R)\, c_o\, (SW_w - SW_{w/o})} \quad \textbf{(9)}$$

where,

I_R = indicator variable for region (= 1.0 if segment pairs are in a specified region of the state, 0.0 if they in a different region).

As with the multivariate model, a generalized linear model form is used for the regression analysis. However, a Poisson distribution is assumed for the dependent variable in Equation 6. The expected crash frequency $E[N|X]_{before}$ is used as an offset variable in this application. Model fit is based on maximum likelihood using the scaled deviance for the Poisson function. This function is:

$$SD = 2 \sum_i \left[y_i\, LN(y_i/u_i) - (y_i - u_i) \right] \quad \textbf{(10)}$$

where,

SD = sum of scaled deviance for all i observations;
y_i = i^{th} observation; and
u_i = model prediction for i^{th} estimate.

The goal of the AMF model calibration process is to estimate the calibration coefficients and to identify any additional variables (e.g., regional effect) that contribute to the explanatory power of the AMF model. Initially, the base model (i.e., $X_{after} = c_0 \times y \times E[N|X]_{before}$) is used to obtain a baseline estimate of the maximum scaled deviance. Then, the existing AMF model is added to the regression model (i.e., $X_{after} = c_0 \times y \times E[N|X]_{before} \times AMF_m$). If the scaled deviance for this model is lower than that for the base model and the regression coefficient c_0 is not significantly different from 1.0, then the existing model is deemed acceptable without re-calibration. If the scaled deviance exceeds that for the base model, then the AMF model form is suspect and alternative model forms are considered.

If an alternative AMF model form is considered, additional variables can be added to the model, one at a time, and evaluated. Only those variables that satisfy the following two conditions should be retained in the model: (1) the associated calibration coefficient is significant at a 95 percent confidence level, and (2) the model scaled deviance is reduced by at least 3.84 (= $\chi^2_{0.05,\,1}$). Of all candidate variables identified, the one that reduces the scaled deviance by the largest amount is added to the AMF model to yield an "enhanced" model. If the scaled deviance for the enhanced model is lower than that for the re-calibrated existing model, then the enhanced model is selected and recommended as a substitute for the re-calibrated existing model.

Step 6. Assess Goodness-of-Fit

Several statistics are available for assessing the fit of the models developed in Steps 2 and 5. One measure of model fit is the Pearson χ^2 statistic. This statistic is calculated as:

$$\chi^2 = \sum_{i=1}^{n} \frac{(X_i - y\,E[N]_i)^2}{Var[X]_i} \tag{11}$$

where,
n = number of observations.

McCullagh and Nelder (4) indicate that this statistic follows the χ^2 distribution with n-p degrees of freedom where n is the number of observations (i.e., segments) and p is the number of model variables. This statistic is asymptotic to the χ^2 distribution for larger sample sizes and exact for normally distributed error structures. As noted by McCullagh and Nelder, this statistic is not well defined in terms of minimum sample size when applied to non-normal distributions; therefore, it probably should not be used as an absolute measure of model significance. The Pearson χ^2 statistic is available from the non-linear regression procedure (NLIN) in the SAS statistical software as the weighted sum of squares for the residual.

The root mean square error s_e is a useful statistic for assessing the precision of the model estimate. It represents the standard deviation of the estimate when each independent variable is at its mean value. This statistic can be computed as:

$$s_e = \frac{1}{y} \sqrt{\frac{\sum_{i=1}^{n} (X_i - y\,E[N]_i)^2}{n - p}} \tag{12}$$

where,
s_e = root mean square error of the model estimate, crashes/yr.

Another, more subjective, measure of model fit can be obtained from a graphical plot of the Pearson residuals versus the expected value of the dependent variable (e.g., $E[N]$). This type of plot provides a graphical means of assessing the predictive capability of the model. A well-fitting model would have the residuals symmetrically centered around zero over the full range of the dependent variable, most clustered near zero, and with a spread ranging from about -3.0 to + 3.0. The Pearson residual PR_i for segment i can be computed as:

$$PR_i = (X_i - y\,E[N]_i) \sqrt{\frac{1}{Var[X]_i}} \tag{13}$$

Another measure of fit is the scale parameter φ. This parameter was noted by McCullagh and Nelder (4) to be a useful statistic for assessing the amount of variation in the observed data, relative to the specified distribution. This statistic can be calculated by dividing Equation 11 by the quantity n-p. It is also available from NLIN as the weighted mean square for the residual. A scale parameter near 1.0 indicates that the assumed distribution of the dependent variable is approximately equivalent to that found in the data (i.e., negative binomial or Poisson).

Another measure of model fit is the coefficient of determination R^2. This statistic is commonly used for normally distributed residuals. However, it has some useful interpretation when applied to data from other distributions when computed in the following manner (5):

$$R^2 = 1 - \frac{SSE}{SST} \tag{14}$$

with,

$$SSE = \sum_{i=1}^{n} (X_i - y\, E[N]_i)^2 \tag{15}$$

$$SST = \sum_{i=1}^{n} (X_i - X.)^2 \tag{16}$$

where,

$X.$ = average crash frequency for all n observations.

The R^2 statistic indicates the percentage of the variability in the crash data that is explained by the multivariate regression model. A value near 0.0 suggests a lack of correlation; a value of 1.0 suggests that the model estimates are in perfect agreement with the observed crash frequency.

The last measure of model fit is the dispersion-parameter-based coefficient of determination R_k^2. This statistic was developed by Miaou (*6*) for use with data that exhibit a negative binomial distribution. It is computed as:

$$R_k^2 = 1 - \frac{k_{null}}{k} \tag{17}$$

where,

k_{null} = over-dispersion parameter based on the variance in the observed crash frequency.

The two over-dispersion parameters in Equation 17 can be estimated using a variety of methods (*7, 8, 9*). The null over-dispersion parameter k_{null} represents the dispersion in the observed crash frequency, relative to the overall average crash frequency for all segments. This parameter can be obtained using a null model formulation (i.e., a model with no independent variables but with the same error distribution, link function, and offset as used with the full or base model).

HORIZONTAL CURVE RADIUS AMF CALIBRATION

This part of the chapter describes the findings from a calibration of the horizontal curve radius AMF described in the *Workbook.* The description is provided in five sections. The first section presents the horizontal curve radius AMF. The second section describes the criteria used to select the highway segment pairs to be analyzed and the method by which the crash data were acquired for each segment. The third section summarizes the crash data and the techniques used to develop the multivariate regression model. The fourth section describes the calibrated AMF and the last section examines the sensitivity of the calibrated AMF to curve radius.

Horizontal Curve Radius AMF

The horizontal curve radius AMF described in the *Workbook* is based on research conducted by Harwood et al. (*10*). The form of this AMF is:

$$AMF_{cr} = \frac{1.55\, L_c + \dfrac{80.2}{R_c}}{1.55\, L_c} \qquad \textbf{(18)}$$

where:

AMF_{cr} = horizontal curve radius accident modification factor;
L_c = length of horizontal curve (= $I_c\, R_c$ / 5280 / 57.3), mi;
I_c = curve deflection angle, degrees; and
R_c = curve radius, ft.

The formulation of Equation 18 indicates that the AMF increases with decreasing radius or decreasing curve length. Alternatively, it decreases as the radius increases, and converges to 1.0 as the radius increases to infinity (i.e., a tangent section).

Site Selection and Data Collection

The curve segments used to form the segment-pair database were required to satisfy the following criteria:

- Cross section: undivided, two through lanes, no median;
- Area type: rural;
- Minimum curve length: 0.1 mi;
- Intersection presence: no intersections;
- Shoulder type: shoulder present (no curbed cross sections);
- Shoulder width: 4 to 13 ft;
- Lane width: 11 to 13 ft; and
- Curve transition: tangent to circular curve (i.e., not spiral).

In addition to the above criteria, each segment pair was checked to ensure that neither an intersection nor another curve existed in the 0.1 mi buffer zone.

The curve segments were initially identified using the criteria above. Then, an adjacent tangent segment was identified that matched the characteristics of each curve segment, such that together they served as a matched pair. All tangent segments identified in this manner had to satisfy all but the last criterion in the preceding list. For similar reasons, all segments were required to have a minimum length of 0.1 mi.

Experience with the crash database indicated that many short segments tended to have no reported crashes during a period of several years. This trend was also exhibited by segments that had very low traffic volume. When one of these segments is associated with one or more crashes, it has a tendency to exhibit undue leverage on regression model coefficients and increases the Pearson χ^2 statistic in a disproportionate manner, relative to other segments. To avoid these issues, the segments were screened such that only those segments with a minimum level of exposure were included in the SP regression database .

To ensure that segments with a low level of exposure and one or more reported crashes did not exert an unreasonable leverage on the model coefficients, it was rationalized that the corresponding prediction ratio PR_i for a segment should not exceed a value of 3.0. In recognition of this desired limit, Equation 13 was used to derive the following equation for computing the minimum segment exposure:

$$E_{\min} = \frac{PR^2 + 2 - \sqrt{(PR^2 + 2)^2 - 4}}{2\ Base\ y} \tag{19}$$

where,

E_{min} = minimum segment exposure associated with prediction ratio PR = 3.0, million-vehicle-miles (mvm); and

$Base$ = injury (plus fatal) crash rate, crashes/mvm.

Equation 19 is based on the conservative assumption that $V[X]_i$ is equal to $E[N]$ and that X_i is equal to 1.0 crash. When the prediction ratio PR is set to 3.0, the estimated crash rate $Base$ is 0.23 cr/mvm, and the time interval for the crash data y is 3 years; the resulting minimum exposure is 0.13 mvm. Equation 20 was used to compute the exposure for all candidate segments. Only those segments exceeding 0.13 mvm were included in the database.

$$E = 0.000365\, ADT\ L \tag{20}$$

where,

E = segment exposure, mvm.

The "minimum exposure" criterion defined by Equation 19 was used only in the selection of sites for the SP regression database. This criterion was needed because the "segment" was the base observation for the multivariate model calibration in Step 2. Adherence to the criterion ensured that no one segment observation exerted undue leverage on the model. In contrast, the SP group database aggregated the segments into groups of similar segments and used the grouped data for AMF model calibration in Step 5. Observations in the SP group database represented the summation of many segments, such that no one segment could have undue leverage on the model, regardless of its exposure. As a result, the minimum exposure criterion was not used to assemble the SP group database.

Site Characteristics

As described in a previous section, two databases were assembled for the curve radius AMF calibration. The segment-pair database is summarized in the bottom half of Table 4-3. A total of 1757 curves (and associated tangent segments) were identified for the segment-pair database. The minimum exposure criterion was used with the segment-pair database to derive the SP regression database. This database is summarized in the top half of Table 4-3. A total of 691 curves (and associated tangent segments) were found to satisfy the minimum exposure criterion.

Table 4-3. Summary Characteristics for Tangent-Curve Segment Pairs.

Database	Region	Total Segment Pairs	Segment Length, mi	Volume Range, veh/d		Radius Range, ft	
				Minimum	Maximum	Minimum	Maximum
SP Regression	Central	92	18.4	1,300	7,770	1,146	34,400
	Northeast	157	31.2	1,060	14,700	1,146	34,400
	North	41	10.4	960	7,300	1,432	22,900
	Southeast	180	36.0	1,420	13,200	955	43,000
	South	152	33.8	980	17,600	955	17,200
	West	69	22.4	770	3,670	1,910	11,500
	Overall:	691	152.2	770	17,600	955	43,000
Segment Pair	Central	194	35.9	210	7,770	955	57,300
	Northeast	248	44.4	170	14,700	955	49,100
	North	242	44.4	50	7,300	955	22,900
	Southeast	273	49.7	110	13,200	955	43,000
	South	321	63.2	120	17,600	955	28,600
	West	479	97.8	70	3,670	955	11,500
	Overall:	1,757	335.4	50	17,600	955	57,300

Data Collection

Crash data were identified for each segment using the DPS database. Three years of crash data, corresponding to years 1999, 2000, and 2001 were identified for each segment. The ADT for each of these three years was obtained from the TRM database and averaged to obtain one ADT for each segment. Crash data prior to 1999 were not used due to the increase in speed limit that occurred on many Texas highways in 1997 and 1998. Crashes that were associated with intersections and driveways were identified as "non-curve-related" and excluded from the database.

Data Analysis

This section is divided into two subsections. The first subsection summarizes the crash data at the selected study sites. The second subsection describes the formulation of the multivariate and calibration models and summarizes the statistical analysis methods used to calibrate these models.

Database Summary

The crash data for each of the segment pairs are summarized in Table 4-4. The segments in the SP regression database were associated with 566 crashes, of which 257 occurred on the tangent segments and 309 occurred on the curved segments. The segments in the segment-pair database were associated with 822 crashes, of which 349 occurred on the tangent segments and 473 occurred on the curved segments. Segment crash rates are provided in the last two columns.

Table 4-4. Crash Data Summary for Tangent-Curve Segment Pairs.

Database	Segment Type	Region	Exposure,[1] mvm	Crashes / 3 years			Crash Rate, cr/mvm	
				PDO [2]	I+F [3]	Total	I+F [3]	Total
SP Regression	Tangent	Central	25.2	14	22	36	0.29	0.48
		Northeast	54.3	22	37	59	0.23	0.36
		North	9.9	4	4	8	0.13	0.27
		Southeast	60.9	33	55	88	0.30	0.48
		South	42.7	20	34	54	0.27	0.42
		West	16.2	7	5	12	0.10	0.25
		Overall:	209.2	100	157	257	0.25	0.41
	Curve	Central	25.2	12	23	35	0.30	0.46
		Northeast	54.3	45	36	81	0.22	0.50
		North	9.9	3	7	10	0.24	0.34
		Southeast	60.9	31	61	92	0.33	0.50
		South	42.7	21	37	58	0.29	0.45
		West	16.2	18	15	33	0.31	0.68
		Overall:	209.2	130	179	309	0.29	0.49
Segment Pair	Tangent	Central	32.9	21	34	55	0.34	0.56
		Northeast	61.3	26	45	71	0.24	0.39
		North	21.2	7	16	23	0.25	0.36
		Southeast	68.2	37	60	97	0.29	0.47
		South	53.7	28	41	69	0.25	0.43
		West	37.5	17	17	34	0.15	0.30
		Overall:	274.8	136	213	349	0.26	0.42
	Curve	Central	32.9	16	30	46	0.30	0.47
		Northeast	61.3	55	48	103	0.26	0.56
		North	21.2	17	21	38	0.33	0.60
		Southeast	68.2	38	77	115	0.38	0.56
		South	53.7	31	53	84	0.33	0.52
		West	37.5	36	51	87	0.45	0.77
		Overall:	274.8	193	280	473	0.34	0.57

Notes:
1 - mvm: million-vehicle-miles.
2 - PDO: property-damage-only crashes.
3 - I+F: injury plus fatal crashes.

Typical injury (plus fatal) crash rates for rural two-lane highways range from 0.2 to 0.3 crashes/mvm. In general, the injury (plus fatal) crash rates for the tangent segments listed in Table 4-4 are within this range. The tangent segments in the West region have a notably low crash rate. This trend could reflect the relatively flat terrain in the West region that allows highways to be constructed there with few vertical curves and little grade.

The typical range for total crash rate on tangents can vary widely, depending on the portion of PDO crashes that are reported in a region. It is estimated that PDO crashes in the DPS database represent about one-half of all the crashes that are reported for rural two-lane highway segments. If this estimate is correct, then the total crash rate on tangents should vary between 0.4 and 0.6 crashes/mvm. In fact, the total crash rates in Table 4-4 generally lie within this range. A notable exception is the West region. A possible reason for this trend was cited in the previous paragraph.

Model Development and Statistical Analysis Methods

This subsection describes the general form of the multivariate model and the AMF calibration model. A separate formulation was used for each model and was dictated by the variables designated as inputs to the subject AMF (i.e., the horizontal curve radius AMF). In addition, region was also included as a model variable in the two formulations to facilitate exploration of its influence on crash frequency.

The general form of the multivariate model is:

$$E[N] = ADT^{b_1} \; L \;\; e^{(b_0 + \sum_{2}^{6} b_i Deg_i + \sum_{7}^{11} b_i Region_i)} \qquad \textbf{(21)}$$

where,

$E[N]$ = expected crash frequency, crashes/yr;
L = highway segment length, mi;
ADT = average daily traffic volume, veh/d;
b_i = calibration coefficients (i = 1, 2, 3,...);
b_0 = calibration coefficient corresponding to West region and 6 degree curve;
$Region_i$ = categorical variable for Texas region (5 levels; see Table 4-2); and
Deg_i = categorical variable for degree of curve (= 5730/R_c) (5 levels: 0, 1, 2, 3, 4, 6 degrees).

As described in the section titled Calibration Procedure, the advantage of using categorical variables in the regression model is that they do not impose any preconceived notions about functional relationship on the estimates obtained from the model.

The Generalized Modeling procedure (GENMOD) in SAS was used to automate the regression analysis. This procedure estimates model coefficients using maximum-likelihood methods. GENMOD is particularly well suited to the analysis of models with additive terms that are either continuous or categorical. GENMOD automates the estimation of the over-dispersion parameter k when the variance function (i.e., Equation 2) does *not* include the segment length variable. However, given that the variance function does include segment length when applied to highway segments, the GENMOD procedure was modified using its Variance option such that Equation 2 was directly specified in the GENMOD code.

To estimate the over-dispersion parameter, two GENMOD procedures were applied in succession for each analysis iteration. In the first application, GENMOD used the calibration model and the specified variance function (i.e., Equation 2) with a specified value of k. In the second application, GENMOD used the natural log of the predicted crash frequency from the first application as an offset variable and the "internal" variance function (i.e., this function allows Equation 2 to be explicitly specified). The best-estimate of k from the second GENMOD application was divided by the average segment length and the quotient formed a new estimate of k for use in Equation 2. This new estimate was used to replace the value of k specified in the first application. This sequential use of two GENMOD procedures was repeated in an iterative manner until convergence was achieved between the k value used in the first GENMOD application and that obtained from the second GENMOD application. Convergence was typically achieved in two iterations.

The general form of the AMF calibration model is:

$$X_{curve} = c_0 \ y \ E[N|X]_{tangent} \ AMF_{cr} \tag{22}$$

with,

$$AMF_{cr} = 1 + \left(d_0 + \sum_{1}^{5} d_i \, Region_i \right) \frac{80.2}{1.55 \, L_c \, R_c} \tag{23}$$

where,

d_i = calibration coefficients (i = 1, 2, 3,...);
d_0 = calibration coefficient corresponding to Central region; and
$Region_i$ = categorical variable for Texas region (5 levels; see Table 4-2).

This equation was derived from Equation 18. Some of the original terms were simplified and several calibration coefficients were added.

The Nonlinear Regression procedure (NLIN) in the SAS software was used to estimate the calibration model coefficients. Like GENMOD, this procedure also estimates model coefficients using maximum-likelihood methods. The benefit of using this procedure is that it is not constrained to additive model terms. Rather, it can be used to evaluate complex AMF model forms such as Equation 23. The "loss" function associated with NLIN was specified to equal the scaled deviance for the Poisson distribution (i.e., Equation 10).

Model Calibration

Multivariate Model

The regression analysis of the multivariate model is presented in Table 4-5. Calibration of this model was based on injury (plus fatal) crash frequency. The SP regression database consisted of 1382 segments. The variables and coefficients listed in this table correspond to those identified in Equation 21. An over-dispersion parameter k of 11.3 mi^{-1} was found to yield a scale parameter φ of 1.00. The Pearson χ^2 statistic for the model is 1373, and the degrees of freedom are 1370 (= $n - p$ = 1382 − 12). As this statistic is less than $\chi^2_{0.05,\,1370}$ (= 1457), the hypothesis that the model fits the data cannot be rejected. The R^2 for the model is 0.05. An alternative measure of model fit that is better suited to the negative binomial distribution is R_k^2, as developed by Miaou (*6*). The R_k^2 for the calibrated model is 0.22. This statistic indicates that about 22 percent of the variability due to systematic sources is explained by the model.

Table 4-5. Multivariate Model Statistical Description–Curve Radius AMF.

Model Statistics		Value		
	R^2 (R_k^2):	0.05 (0.22)		
	Scale Parameter φ:	1.00		
	Pearson χ^2:	1373 ($\chi^2_{0.05,\ 1370}$ = 1457)		
	Over-Dispersion Parameter k:	11.3 mi^{-1}		
	Observations n_o:	1382 segments (336 injury + fatal crashes in 3 years)		
	Standard Deviation s_e:	±0.17 crashes/segment/yr		
Range of Model Variables				
Variable	**Variable Name**	**Units**	**Minimum**	**Maximum**
ADT	Segment ADT	veh/d	770	17,600
L	Segment length	miles	0.10	0.82
Deg	Degree of curvature (= 5730 / R_c)	degrees	0	6
Calibrated Coefficient Values				
Variable	**Definition**	**Value**	**Std. Dev.**	**t-statistic**
b_0	Intercept	-7.09	1.08	-6.5
b_1	Effect of segment ADT	0.847	0.119	7.1
b_2	Effect of 0 degree curvature (i.e., tangent)	-1.280	0.477	-2.7
b_3	Effect of 1 degree curvature	-1.355	0.487	-2.8
b_4	Effect of 2 degree curvature	-1.012	0.486	-2.1
b_5	Effect of 3 degree curvature	-0.803	0.519	-1.5
b_6	Effect of 4 degree curvature	-0.522	0.643	-0.8
in b_0	Effect of 6 degree curvature	0.000		
b_7	Effect of Central region	0.403	0.289	1.4
b_8	Effect of Northeast region	0.179	0.282	0.6
b_9	Effect of North region	-0.089	0.378	-0.2
b_{10}	Effect of Southeast region	0.524	0.269	1.9
b_{11}	Effect of South region	0.352	0.270	1.3
in b_0	Effect of West region	0.000		
Categorical Variable Statistics				
Category	**Definition**	**Deg. Freedom**	**Chi-Square**	**p-value**
Deg	Degree of curve	5	15.0	0.01
Region	Region	5	9.1	0.11

The statistics for each categorical variable, as a group, are listed in the last two rows of Table 4-5. The p-value of 0.11 suggests that there is not a significant difference between the regions. However, the Degree-of-Curve and Region variables are "key" categorical variables and were kept in the model, regardless of their significance. The influence of the Region variable was tested further during the analysis of the AMF model.

The coefficients in the middle section of Table 4-5 can be combined with Equation 21 to yield the calibrated crash estimation model. Because each categorical variable comprises several indicator variables, a unique model is derived for each combination of factors. For example, for tangent segments in the Central Texas region, the following model would be derived using the coefficients in the table:

$$E[N] = ADT^{0.847}\ L\ \ e^{(-7.09\ -\ 1.280\ +\ 0.403)}$$
$$= ADT^{0.847}\ L\ \ e^{(-7.97)} \tag{24}$$

The coefficients in Table 4-5 can also be used to evaluate the effect of any one variable in a given category, relative to any other variable in the same category. This relative effect represents the ratio of crash frequency on a segment with conditions consistent with one category (say, 6 degree curvature) to that on a segment with conditions consistent with a different category (say, a tangent segment), all other conditions being the same. For example, the "crash frequency ratio" for a segment with 6 degree curvature, relative to one with no curvature is computed using the following equation:

$$R_N = \frac{e^{0.000}}{e^{-1.280}}$$
$$= 3.60 \tag{25}$$

The value of 3.60 obtained for this example suggests that there are 3.60 crashes on a 6 degree curve for every 1.0 crash on a tangent, all other conditions being the same. Similar ratios can be computed for other curvature, speed, and region variable combinations. The crash frequency ratio for each curve indicator variable, relative to a tangent segment, is shown in Figure 4-3. Each curvature category has been converted to an equivalent radius for presentation purposes. The trend in the data points in the figure suggests that crash frequency on curves is higher than on tangents, especially for sharper curves.

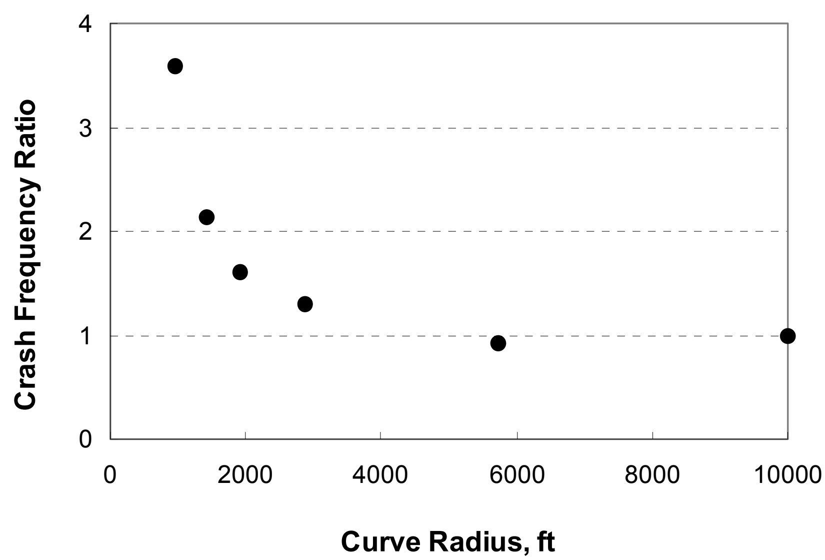


Figure 4-3. Crash Frequency Ratio for Curve Radius.

Calibration Model

The regression analysis of the AMF calibration model revealed that the original AMF model (i.e., Equation 18) underestimated the effect of curvature. In fact, the AMF regression model (i.e., Equation 23) was found to have formulation problems that resulted in a poor fit to the data, relative to the base model (i.e., $X_{curve} = c_0 \times y \times E[N|X]_{tangent}$). Also, the correlation between curve length and the AMF value was found to be relatively weak. A similar lack of correlation was found for the various Texas regions. The revised AMF model form that reflects these findings and overcomes the observed problems is (c_0 = 1.0):

$$AMF_{cr} = 1 + d_0 \left(\frac{5730}{R_c} \right)^2 \tag{26}$$

The statistics related to the calibrated model are shown in Table 4-6. The calibration is based on injury (plus fatal) crash frequency. The SP group database consisted of 56 unique combinations of curve radius and segment length (as obtained from the segment-pair database). The variables and coefficients listed in this table correspond to those identified in Equation 26. The Pearson χ^2 statistic for the model is 52.1, and the degrees of freedom are 55 (= n − p = 56 − 1). As this statistic is less than $\chi^2_{0.05,\,55}$ (= 73.3), the hypothesis that the model fits the data cannot be rejected. The R^2 for the model is 0.91.

Table 4-6. Calibration Model Statistical Description–Curve Radius AMF.

Model Statistics		Value		
R^2:		0.91		
Scale Parameter φ:		0.95		
Pearson χ^2:		52.1 ($\chi^2_{0.05,\,55}$ = 73.3)		
Observations n_o:		56 curve radius and length combinations (493 injury + fatal crashes in 3 years)		
Standard Deviation s_e:		±0.62 crashes/segment/yr		
Range of Model Variables				
Variable	**Variable Name**	**Units**	**Minimum**	**Maximum**
R_c	Curve radius	ft	955	57,300
Calibrated Coefficient Values				
Variable	**Definition**	**Value**	**Std. Dev.**	**t-statistic**
d_0	Effect of radius	0.133	0.020	6.7

The fit of the model to the data is shown in Figure 4-4. This figure relates the Pearson residuals to the model-estimated AMF value. The spread in the data is approximately centered on the horizontal line at y = 0.0 for the range of estimated AMFs and suggests that there is no bias in the model estimates.

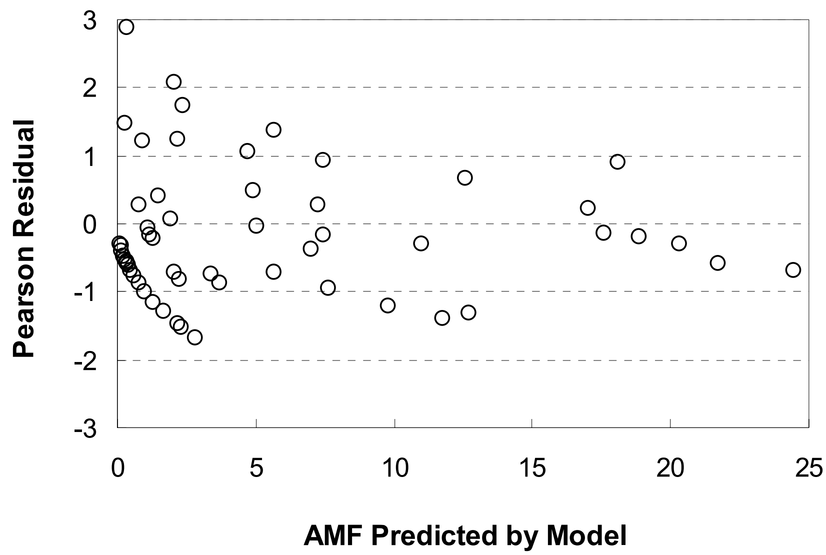


Figure 4-4. Fit of Revised Curve AMF Model.

The crash data assembled for the curve radius AMF analysis excluded driveway-related crashes. An examination of the database with, and without, these crashes indicated that they constitute about 20 percent of all segment crashes, for segments located in the vicinity of a curve. Thus, Equation 26 was adjusted to reflect its focus on non-driveway-related crashes. The adjustment was based on a weighed average technique, where the AMF_{cr} from Equation 26 (for non-driveway crashes) was weighted by 0.80 and an AMF_{cr} = 1.0 (for driveway crashes) was weighted by 0.20. The adjusted AMF is:

$$AMF_{cr} = 1 + 0.106 \left(\frac{5730}{R_c} \right)^2 \qquad \textbf{(27)}$$

This equation is recommended for estimating the effect of radius on a segment's crash frequency.

Sensitivity Analysis

The revised AMF model is shown in Figure 4-5 for a range of radii. The values obtained from the revised model are shown with a solid trend line. The values obtained from Equation 18 are shown using two dashed lines. This equation is sensitive to curve length; however, it was converted to include a sensitivity to curve deflection angle I_c instead by using the relationship between curve length and deflection angle provided in the variable definitions associated with Equation 18. This conversion was performed to facilitate a more equitable presentation of Equation 18 for the range of radii shown in Figure 4-5.

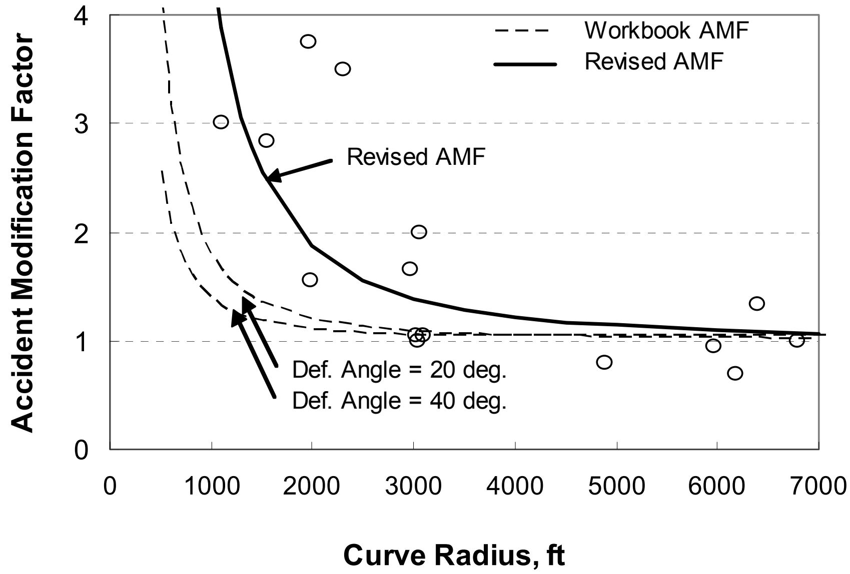


Figure 4-5. Relationship between Radius and AMF Value.

The revised AMF values are larger than those obtained from Equation 18, although the difference diminishes with increasing radius. The reason for this difference may be that the data used to develop Equation 18 were obtained from Washington state. The terrain and climate of this state are quite different from that found in Texas.

The data points shown in Figure 4-5 represent the actual reported injury (plus fatal) crash frequencies for the individual curve and tangent segments in the database. To obtain these data points, the sites were sorted by radius and grouped such that each group included between 10 and 15 crashes for the tangent segments. The AMF value was estimated for each group as the ratio of the reported crash frequency for the curve and tangent segments. The average radius was also computed for each group. The trend in the data confirms that the revised AMF provides a good fit to the data, and that Equation 18 underestimates the AMF value for Texas highways.

LANE AND SHOULDER WIDTH AMF CALIBRATION

This part of the chapter describes the findings from a calibration of the lane width and the shoulder width AMFs described in the *Workbook*. The description is provided in five sections. The first section presents the two AMF models. The second section describes the criteria used to select highway segment pairs to be analyzed and the method by which the crash data were acquired for each segment. The third section summarizes the crash data and the techniques used to develop the multivariate regression model. The fourth section describes the calibrated AMF and the last section examines the sensitivity of the calibrated AMF to lane width and shoulder width.

Lane and Shoulder Width AMF

The lane width AMF described in the *Workbook* is based on research conducted by Harwood et al. (*10*). Guidance in the *Workbook* indicates that the following form of this AMF is applicable to rural two-lane highways:

$$AMF_{lw} = (e^{-0.047\,(W_l - 12)} - 1.0)\,\frac{0.42}{0.36} + 1.0 \qquad \textbf{(28)}$$

where:
AMF_{lw} = lane width accident modification factor; and
W_l = lane width, ft.

The shoulder width AMF described in the *Workbook* is also based on research conducted by Harwood et al. (*10*). Guidance in the *Workbook* indicates that the following form of this AMF is applicable to rural two-lane highways:

$$AMF_{osw} = (e^{-0.021\,(W_s - 8)} - 1.0)\,\frac{0.34}{0.16} + 1.0 \qquad \textbf{(29)}$$

where:
AMF_{osw} = outside shoulder width accident modification factor; and
W_s = outside shoulder width, ft.

The formulation of both equations indicates that the AMF increases with decreasing lane or shoulder width. The lane width AMF value is equal to 1.0 when the lane width equals 12 ft. Similarly, the shoulder width AMF value is equal to 1.0 when the shoulder width equals 8 ft. Both of these widths represent the "base condition" for the AMF. A base condition is determined to be a condition considered typical of most rural two-lane highways.

Site Selection and Data Collection

The highway segments used to form the segment-pair database were required to satisfy the following criteria:

- Cross section: undivided, two through lanes, no median;
- Area type: rural;
- Minimum segment length: 0.1 mi;
- Intersection presence: no intersections;
- Curve presence: no curves; and
- Shoulder type: shoulder present (no curbed cross sections).

In addition, only those segments with an exposure exceeding 0.13 mvm were included in the SP regression database. The rationale for this criterion was described in the previous part addressing the horizontal curve radius AMF. Adherence to the criterion ensured that no one segment exerted undue leverage on the model.

One segment of each pair was initially identified using the criteria above. Then, the second segment was identified using the characteristics of the first segment, such that together they served as a matched pair of segments from the same roadway. To maximize the potential for finding matched pairs, the second segment was located as near as possible to the first segment. As noted in the section titled Segmentation Procedure, segment length was initially specified to equal the length defined in the TRM database. However, if the two segments did not have the same length, then the length of the longer segment was set equal to that of the shorter segment. All segments were required to have a minimum length of 0.1 mi. The value of 0.1 mi reflects a recognition of the precision of crash location in the DPS crash database.

The segment pairs were subsequently screened to ensure that they were matched as closely as could be determined using the attributes in the TRM database. The specific attributes considered, and the criteria used, include:

- Shoulder composition: paved, same on both segments; and
- ADT: no more than 10 percent change in ADT between segments pairs.

The resulting segment-pair database developed using the aforementioned criteria yielded data for 658 segment pairs. However, an initial attempt to calibrate the multivariate model indicated that the effect of lane width and shoulder width was so subtle that additional segments would be needed to accurately quantify their correlation with crash frequency. Therefore, a non-paired (NP) regression database was assembled that included individual segments, without the requirement for finding a matched segment. This database includes data for 2864 segments. It was used to calibrate the multivariate model.

A segment-pair database was assembled to develop the SP group database (as needed to calibrate the AMF model). However, as explained in the previous part, the minimum exposure criterion is not required when developing a SP group database. The segment-pair database assembled for this purpose includes data for 1116 segment pairs.

Site Characteristics

As described in a previous section, two databases were assembled for the curve radius AMF calibration. The NP regression database is summarized in the top half of Table 4-7. The segment-pair database is summarized in the bottom half of Table 4-7. All total, 2864 viable segments were identified for the NP regression database. In contrast, 1116 segment pairs were identified for the segment-pair database. Segments with 8 ft lanes were very few in number and, for the subsequent analyses, were combined with those segments with 9 ft lanes.

Data Collection

Crash data were identified for each segment using the DPS database. Three years of crash data, corresponding to years 1999, 2000, and 2001 were identified for each segment. The ADT for each of these three years was obtained from the TRM database and averaged to obtain one ADT for each segment. Crash data prior to 1999 were not used due to the increase in speed limit that occurred on many Texas highways in 1997 and 1998. Crashes that were associated with intersections and driveways were identified as "non-segment-related" and excluded from the database.

Table 4-7. Summary Characteristics for Lane-Shoulder Segment Pairs.

Database	Region	Total Segment Pairs	Segment Length, mi	Volume Range, veh/d		Shoulder Width, ft		Lane Width, ft	
				Min.	Max.	Min.	Max.	Min.	Max.
NP Regression	Central	341	137	500	12,400	1	12	9	13
	Northeast	789	267	560	19,200	1	16	9	13
	North	356	169	460	9,400	1	10	9	13
	Southeast	745	286	400	23,700	1	14	9	13
	South	382	149	390	18,400	1	16	9	13
	West	251	126	490	20,800	1	11	9	13
	Overall:	2,864	1,134	390	23,700	1	16	9	13
Segment Pair	Central	135	39	90	9,500	1	11	9	13
	Northeast	234	58	220	9,200	1	14	9	13
	North	177	53	50	4,400	1	10	9	13
	Southeast	271	85	200	10,200	1	14	9	13
	South	149	52	70	18,400	1	13	9	13
	West	150	58	140	4,600	1	16	9	13
	Overall:	1,116	345	50	18,400	1	16	9	13

Data Analysis

This section is divided into two subsections. The first subsection summarizes the crash data at the selected study sites. The second subsection describes the formulation of the multivariate and calibration models and summarizes the statistical analysis methods used to calibrate these models.

Database Summary

The crash data for each of the databases are summarized in Table 4-8. The segments in the NP regression database were associated with 1964 crashes, of which 1125 were injury or fatal crashes. The segments in the segment-pair database were associated with 998 crashes, of which 561 were injury or fatal crashes. The trends in crash rate are provided in the last two columns.

Typical injury (plus fatal) crash rates for rural two-lane highways range from 0.2 to 0.3 crashes/mvm. Comparison of this range with the injury (plus fatal) crash rates for tangents in Table 4-8 suggests that the segments in the West region have a low crash rate. This trend was noted previously with respect to Table 4-4 and could reflect the relatively flat terrain in the West region that allows highways to be constructed with few vertical curves and little grade.

Table 4-8. Crash Data Summary for Lane-Shoulder Segment Pairs.

Database	Region	Exposure,[1] mvm	Crashes / 3 years			Crash Rate, cr/mvm	
			PDO [2]	I+F [3]	Total	I+F [3]	Total
NP Regression	Central	148	88	104	192	0.23	0.43
	Northeast	480	319	449	768	0.31	0.53
	North	116	58	70	128	0.20	0.37
	Southeast	377	260	327	587	0.29	0.52
	South	194	70	139	209	0.24	0.36
	West	80	44	36	80	0.15	0.33
	Overall:	1,395	839	1,125	1,964	0.27	0.47
Segment Pair	Central	62.9	38	40	78	0.21	0.41
	Northeast	200.6	160	190	350	0.32	0.58
	North	51.6	20	50	70	0.32	0.45
	Southeast	184.9	140	176	316	0.32	0.57
	South	85.8	36	64	100	0.25	0.39
	West	57.0	43	41	84	0.24	0.49
	Overall:	642.7	437	561	998	0.29	0.52

Notes:
1 - mvm: million-vehicle-miles.
2 - PDO: property-damage-only crashes.
3 - I+F: injury plus fatal crashes.

The typical range for total crash rate can vary more widely, depending on the portion of PDO crashes that are reported in a region. It is estimated that PDO crashes in the DPS database represent about one-half of all the crashes that are reported for rural two-lane highway segments. If this estimate is correct, then the total crash rate should vary between 0.4 and 0.6 crashes/mvm. In fact, the total crash rates in Table 4-8 generally lie within this range.

Model Development and Statistical Analysis Methods

This subsection describes the general form of the multivariate model and the AMF calibration model. A separate formulation was used for each model and was dictated by the variables designated as inputs to the subject AMFs (i.e., lane width AMF and shoulder width AMF). In addition, Region was also included as a model variable in the two formulations to facilitate exploration of its influence on crash frequency. The two AMFs were combined in a single regression analysis to explore possible interaction between lane width and shoulder width on crash frequency.

The general form of the multivariate model is:

$$E[N] = ADT^{b_1} \, L \, e^{\left(b_0 + \sum_{2}^{5} b_i LW_i + \sum_{6}^{10} b_i SW_i + \sum_{11}^{15} b_i Region_i + b_{16} W_l W_s\right)} \qquad \textbf{(30)}$$

where,

$E[N]$ = expected crash frequency, crashes/yr;
L = highway segment length, mi;

ADT = average daily traffic volume, veh/d;
b_i = calibration coefficients (i = 1, 2, 3,...);
b_0 = calibration coefficient corresponding to West region, 11.5 ft paved shoulder width, 13 ft lane width;
$Region_i$ = categorical variable for Texas region (5 levels; see Table 4-2);
LW_i = categorical variable for lane width (4 levels: 9, 10, 11, 12, 13 ft);
SW_i = categorical variable for paved shoulder width (5 levels: 1.5, 3.5, 5.5, 7.5, 9.5, 11.5 ft);
W_l = lane width, ft; and
W_s = paved shoulder width, ft.

As described in the section titled Calibration Procedure, the advantage of using categorical variables in the regression model is that they do not impose any preconceived notions about functional relationship on the estimates obtained from the model. A categorical lane-and-shoulder width interaction term was also included in the model, but its coefficients could not be derived due to sparse data. As an alternative model formulation to explore this important interaction, a multiplicative term was added to the model. The use of a continuous term in the model violated the desire for only categorical variables, but it was the only option available for exploring lane width and shoulder width interaction.

The Generalized Modeling procedure (GENMOD) in SAS was used to automate the regression analysis. This procedure estimates model coefficients using maximum-likelihood methods. GENMOD is particularly well suited to the analysis of models with additive terms that are either continuous or categorical. GENMOD automates the estimation of the over-dispersion parameter k when the variance function (i.e., Equation 2) does *not* include the segment length variable. However, given that the variance function does include segment length when applied to highway segments, the GENMOD procedure was modified using its Variance option such that Equation 2 was directly specified in the GENMOD code.

To estimate the over-dispersion parameter, two GENMOD procedures were applied in succession for each analysis iteration. In the first application, GENMOD used the calibration model and the specified variance function (i.e., Equation 2) with a specified value of k. In the second application, GENMOD used the natural log of the predicted crash frequency from the first application as an offset variable and the "internal" variance function (i.e., this function allows Equation 2 to be explicitly specified). The best-estimate of k from the second GENMOD application was divided by the average segment length and the quotient formed a new estimate of k for use in Equation 2. This new estimate was used to replace the value of k specified in the first application. This sequential use of two GENMOD procedures was repeated in an iterative manner until convergence was achieved between the k value used in the first GENMOD application and that obtained from the second GENMOD application. Convergence was typically achieved in two iterations.

The general form of the combined lane and shoulder width AMF calibration model is:

$$X_a = y\ E[N|X]_b\ AMF_{lw}\ AMF_{osw} \qquad \textbf{(31)}$$

with,

$$AMF_{lw} = \left(e^{-0.047\,(d_0 + \sum_1^5 d_i\,Region_i)\,(W_{l,a} - W_{l,b})} - 1.0\right)\frac{0.42}{0.36} + 1.0 \qquad \textbf{(32)}$$

$$AMF_{osw} = \left(e^{-0.021\,(c_0 + \sum_1^5 c_i\,Region_i)\,(W_{s,a} - W_{s,b})} - 1.0\right)\frac{0.34}{0.16} + 1.0 \qquad \textbf{(33)}$$

where:

$W_{l,a}$ = lane width for segment *a*, ft;
$W_{l,b}$ = lane width for segment *b*, ft;
$W_{s,a}$ = shoulder width for segment *a*, ft;
$W_{s,b}$ = shoulder width for segment *b*, ft;
d_i = calibration coefficients for lane width (i = 1, 2, 3,...);
d_0 = calibration coefficient for lane width corresponding to West region;
c_i = calibration coefficients for shoulder width (i = 1, 2, 3,...);
c_0 = calibration coefficient for shoulder width corresponding to West region; and
$Region_i$ = categorical variable for Texas region (5 levels; see Table 4-2).

The Nonlinear Regression procedure (NLIN) in the SAS software was used to estimate the calibration model coefficients. Like GENMOD, this procedure also estimates model coefficients using maximum-likelihood methods. The benefit of using this procedure is that it is not constrained to additive model terms. Rather, it can be used to evaluate complex AMF model forms such as Equations 32 and 33. The "loss" function associated with NLIN was specified to equal the scaled deviance for the Poisson distribution (i.e., Equation 10).

Model Calibration

Multivariate Model

The regression analysis of the multivariate model is presented in Table 4-9. The calibration is based on injury (plus fatal) crash frequency. The variables and coefficients listed in this table correspond to those identified in Equation 30. An over-dispersion parameter k of 9.05 mi^{-1} was found to yield a scale parameter φ of 0.97. The Pearson χ^2 statistic for the model is 2758, and the degrees of freedom are 2847 (= $n - p$ = 2864 − 17). As this statistic is less than $\chi^2_{0.05,\,2847}$ (= 2972), the hypothesis that the model fits the data cannot be rejected. The R^2 for the model is 0.24. An alternative measure of model fit that is better suited to the negative binomial distribution is R_k^2, as developed by Miaou *(6)*. The R_k^2 for the calibrated model is 0.82. This statistic indicates that about 82 percent of the variability due to systematic sources is explained by the model.

The statistics for each categorical variable, as a group, are listed in the last three rows of Table 4-9. These variables are "key" categorical variables and are kept in the model, regardless of whether they are significant. Their influence was tested further during the analysis of the AMF model.

Table 4-9. Multivariate Model Statistical Description–Lane and Shoulder Width AMFs.

Model Statistics		Value		
R^2 (R_k^2):		0.24 (0.82)		
Scale Parameter φ:		0.97		
Pearson χ^2:		2758 ($\chi^2_{0.05,\ 2847}$ = 2972)		
Over-Dispersion Parameter k:		9.05 mi^{-1}		
Observations n_o:		2864 segments (1125 injury + fatal crashes in 3 years)		
Standard Deviation s_e:		±0.23 crashes/segment/yr		
Range of Model Variables				
Variable	**Variable Name**	**Units**	**Minimum**	**Maximum**
ADT	Segment ADT	veh/d	390	23,700
L	Segment length	miles	0.10	1.00
LW	Lane width	ft	9	13
SW	Paved shoulder width	ft	1.0	16.0
Calibrated Coefficient Values				
Variable	**Definition**	**Value**	**Std. Dev.**	**t-statistic**
b_0	Intercept	-11.63	0.92	-12.7
b_1	Effect of segment ADT	0.999	0.058	17.4
b_2	Effect of 9 ft lane width	0.725	0.442	1.6
b_3	Effect of 10 ft lane width	0.171	0.197	0.9
b_4	Effect of 11 ft lane width	0.181	0.163	1.1
b_5	Effect of 12 ft lane width	0.008	0.079	0.1
in b_0	Effect of 13 ft lane width	0.000		
b_6	Effect of 1.5 ft paved shoulder	1.763	0.708	2.5
b_7	Effect of 3.5 ft paved shoulder	1.304	0.617	2.1
b_8	Effect of 5.5 ft paved shoulder	0.969	0.506	1.9
b_9	Effect of 7.5 ft paved shoulder	0.724	0.401	1.8
b_{10}	Effect of 9.5 ft paved shoulder	0.515	0.340	1.5
in b_0	Effect of 11.5 ft paved shoulder	0.000		
b_{11}	Effect of Central region	0.463	0.204	2.3
b_{12}	Effect of Northeast region	0.720	0.188	3.8
b_{13}	Effect of North region	0.292	0.211	1.4
b_{14}	Effect of Southeast region	0.655	0.187	3.5
b_{15}	Effect of South region	0.487	0.198	2.5
in b_0	Effect of West region	0.000		
b_{16}	Lane and shoulder width interaction	0.011	0.0047	2.3
Categorical Variable Statistics				
Category	**Definition**	**Deg. Freedom**	**Chi-Square**	**p-value**
LW	Lane width	4	3.7	0.44
SW	Paved shoulder width	5	7.7	0.17
Region	Region	5	24.0	0.00

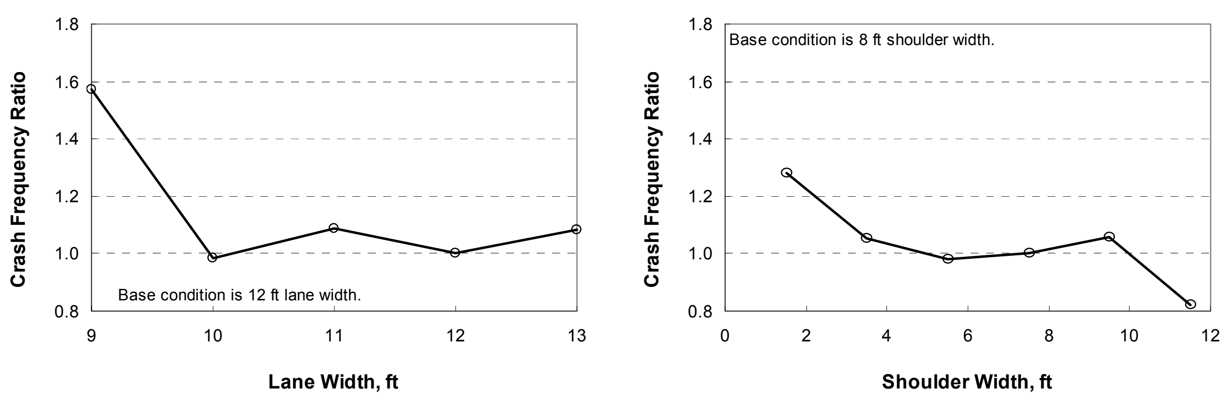
The coefficients in the middle section of Table 4-9 can be combined with Equation 30 to yield the calibrated crash estimation model. Because each categorical variable comprises several indicator variables, a unique model is derived for each combination of factors. For example, for segments in the Central Texas region with 12 ft lanes and 5.5 ft paved shoulders, the following model would be derived using the coefficients in the table:

$$\begin{aligned} E[N] &= ADT^{0.999}\ L \quad e^{(-11.63\ +\ 0.463\ +\ 0.008\ +\ 0.969\ +\ 0.011\,[12\times 5.5])} \\ &= ADT^{0.999}\ L \quad e^{(-9.46)} \end{aligned} \qquad \textbf{(34)}$$

The coefficients in Table 4-9, combined with the lane-and-shoulder width interaction term, can also be used to evaluate the effect of any one variable in a given category, relative to any other variable in the same category. This relative effect represents the ratio of crash frequency on a segment with conditions consistent with one category (say, 9 ft lane width) to that on a segment with conditions consistent with a different category (say, 12 ft lane width), all other conditions being the same. For example, the "crash frequency ratio" for a segment with 9 ft lanes and an 8 ft shoulder, relative to one with 12 ft lanes and an 8 ft shoulder is computed using the following equation:

$$\begin{aligned} R_N &= \frac{e^{0.725\ +\ 0.011\,[9\times 8]}}{e^{0.008\ +\ 0.011\,[12\times 8]}} \\ &= 1.57 \end{aligned} \qquad \textbf{(35)}$$

The value of 1.57 obtained for this example suggests that there are 1.57 crashes on a segment with 9 ft lanes for every one crash on a segment with 12 ft lanes, all other conditions being the same. Similar ratios can be computed for other lane width, shoulder width, and region variable combinations. In fact, the crash frequency ratio for all of the lane width and shoulder width indicator variables, relative to base conditions of a 12 ft lane width and 8 ft shoulder width, are illustrated in Figure 4-6.



a. Lane Width. **b. Shoulder Width.**

Figure 4-6. Crash Frequency Ratio for Lane Width and Shoulder Width.

The trends in Figure 4-6 were derived using the lane and shoulder width categorical variable coefficient values in Table 4-9 along with the lane-and-shoulder width interaction term. The trend in the data points in Figure 4-6a suggests that crash frequency is higher on segments with lanes that are 9 ft in width, but is relatively insensitive to lane widths in the range of 10 to 13 ft. The trend in Figure 4-6b suggests that crash frequency is similarly influenced by shoulder width. That is, the sensitivity to shoulder width is largest for the smallest and largest widths.

A more detailed analysis of the categorical variable coefficient values in Table 4-9 was undertaken to determine their correlation with lane or shoulder width. These values are shown in Figure 4-7. The data points shown as open circles correspond to the coefficients associated with the lane width category. The data points shown as solid diamonds correspond to the coefficients associated with the shoulder width category. The best-fit trend lines are also shown in the figure.

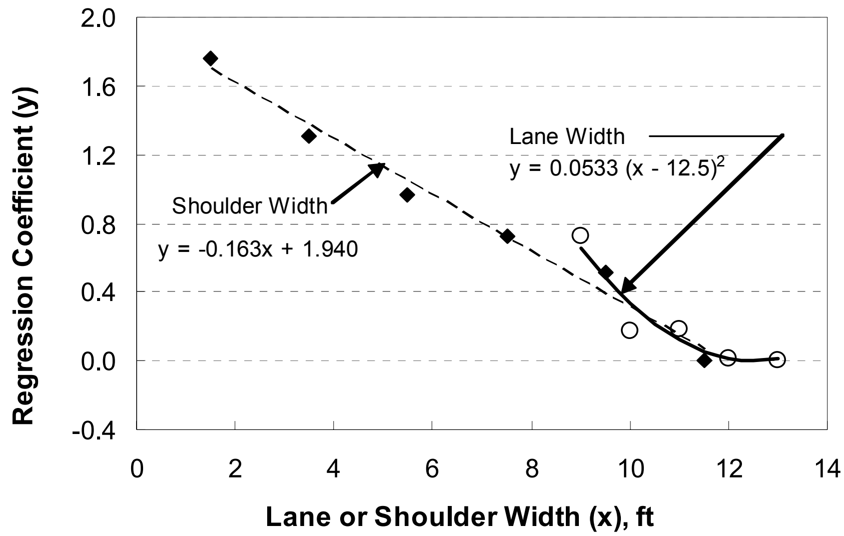


Figure 4-7. Correlation between Coefficients and Lane or Shoulder Width.

Calibration Model

The SP group database was assembled for all possible combinations of lane and shoulder width category. A total of 90 unique groups were found in the segment-pair database. However, closer examination of group sample sizes indicated that only 10 percent of the segment pairs reflected a change in lane width, shoulder width, or both. Hence, many of the 90 combinations were represented by only one or two segment pairs, which tended to make AMF model calibration more difficult and limited the examination of regional influence.

The regression analysis revealed that the original AMF model for lane width (i.e., Equation 32) underestimated the effect of lane width. The calibration coefficient for the shoulder width model indicated an illogical trend of increasing crash risk with increasing shoulder width; however, the coefficient value was not statistically significant. In fact, the AMF model was found to have formulation problems that resulted in a poor fit to the data, relative to the base model (i.e., $X_a = c_0 \times y \times E[N|X]_b$). Most notably, it did not account for the lane-and-shoulder interaction suggested in the multivariate model. Based on these findings, the following revised AMF model was formulated:

$$AMF_{lw,osw} = e^{d_0([W_{l,a} - 12.5]^2 - [W_{l,b} - 12.5]^2) + d_1(W_{s,a} - W_{s,b}) + d_2(W_{s,a}W_{l,a} - W_{s,b}W_{l,b})} \quad \textbf{(36)}$$

The terms in Equation 36 are based on the trends found in Figure 4-7 and the interaction term found in the multivariate model. The values for the regression coefficients d_0 and d_1 suggested by the equations in Figure 4-7 are 0.0533 and -0.163, respectively. The value for the d_2 coefficient is suggested by coefficient b_{16} in Table 4-9 to be 0.011.

The regression analysis of Equation 36 revealed an improved fit to the data, relative to the original AMF. However, the small group sample sizes noted previously frustrated the effort to quantify the three calibration coefficients. The values obtained for the three coefficients were consistent with the values suggested in the previous paragraph but not statistically significant. As a result, a model using the "suggested" coefficients was evaluated. The form of this model is:

$$AMF_{lw,osw} = e^{c_0 + 0.0533([W_{l,a} - 12.5]^2 - [W_{l,b} - 12.5]^2) - 0.163(W_{s,a} - W_{s,b}) + 0.011(W_{s,a}W_{l,a} - W_{s,b}W_{l,b})} \quad \textbf{(37)}$$

The statistics related to the calibrated version of Equation 37 are shown in Table 4-10. The calibration is based on injury (plus fatal) crash frequency. The variables and coefficients listed in this table correspond to those identified in Equation 37. The Pearson χ^2 statistic for the model is 91.8, and the degrees of freedom are 89 (= n − p = 90 − 1). As this statistic is less than $\chi^2_{0.05,\,89}$ (= 112), the hypothesis that the model fits the data cannot be rejected. The R^2 for the model is 0.98. The calibration coefficient c_0 is not significantly different from 0.0, and was subsequently eliminated from the model.

Table 4-10. Calibration Model Statistical Description–Lane and Shoulder Width AMFs.

Model Statistics		Value		
	R^2:	0.98		
	Scale Parameter φ:	1.03		
	Pearson χ^2:	91.8 ($\chi^2_{0.05,\,89}$ = 112)		
	Observations n_o:	90 lane width and shoulder width combinations (561 injury + fatal crashes in 3 years)		
	Standard Deviation s_e:	±0.46 crashes/segment/yr		
Range of Model Variables				
Variable	**Variable Name**	**Units**	**Minimum**	**Maximum**
$W_{l,a}$ - $W_{l,b}$	Lane width change	ft	-4	3
$W_{s,a}$ - $W_{s,b}$	Shoulder width change	ft	-14	7
Calibrated Coefficient Values				
Variable	**Definition**	**Value**	**Std. Dev.**	**t-statistic**
c_0	Calibration coefficient	0.031	0.063	0.5

The fit of the model to the data is shown in Figure 4-8. This figure relates the Pearson residuals to the model-estimated AMF value. The spread in the data is approximately centered on the horizontal line corresponding to a residual value of 0.0 for the range of estimated AMFs.

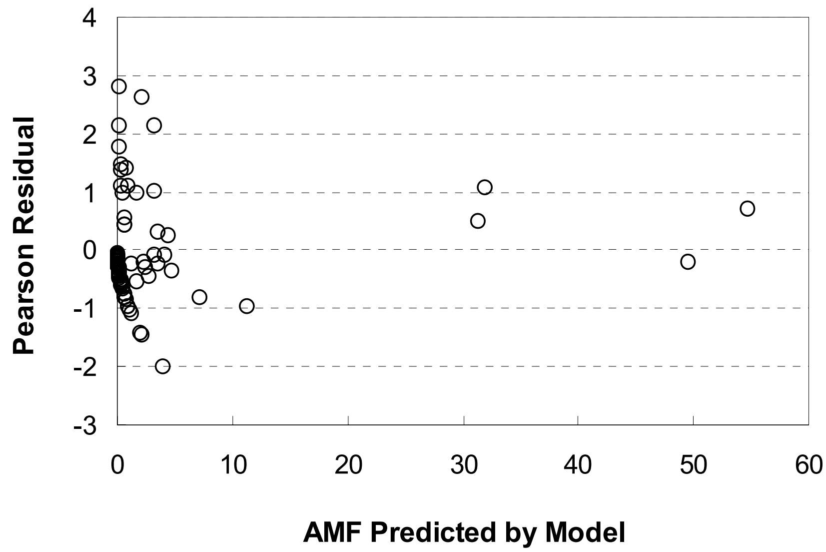


Figure 4-8. Fit of Combined Lane Width and Shoulder Width AMF Models.

The crash data assembled for this analysis excluded driveway-related crashes. An examination of the database with, and without, these crashes indicated that they constitute about 46 percent of all segment crashes, for segments not located in the vicinity of a curve. Thus, Equation 37 was adjusted in the following manner to reflect its focus on non-driveway-related crashes.

$$AMF_{lw,osw} = \left(e^{0.0533\,([W_{l,a} - 12.5]^2 - [W_{l,b} - 12.5]^2)\; -\; 0.163\,(W_{s,a} - W_{s,b})\; +\; 0.011\,(W_{s,a}W_{l,a} - W_{s,b}W_{l,b})} - 1\right)0.54 + 1.0 \quad \textbf{(38)}$$

Equation 38 is recommended for estimating the effect of lane and shoulder width on a segment's crash frequency. The interaction term requires the use of a combined AMF, as opposed to the use of separate lane width and shoulder width AMFs.

Sensitivity Analysis

The revised AMF model (i.e., Equation 38) is shown in Figure 4-9 for a range of lane widths. The values obtained from Equation 38 are shown with a thick trend line. The values obtained from the original AMF model (i.e., Equation 28) are shown with a thin line. The AMF values from the revised model with an 8 ft shoulder width are comparable to those from the original AMF model. The two thick trend lines indicate that the relationship between AMF value and lane width varies, depending on the shoulder width. Lane width is indicated to have a larger effect on safety when the shoulder is narrow, which is logical.

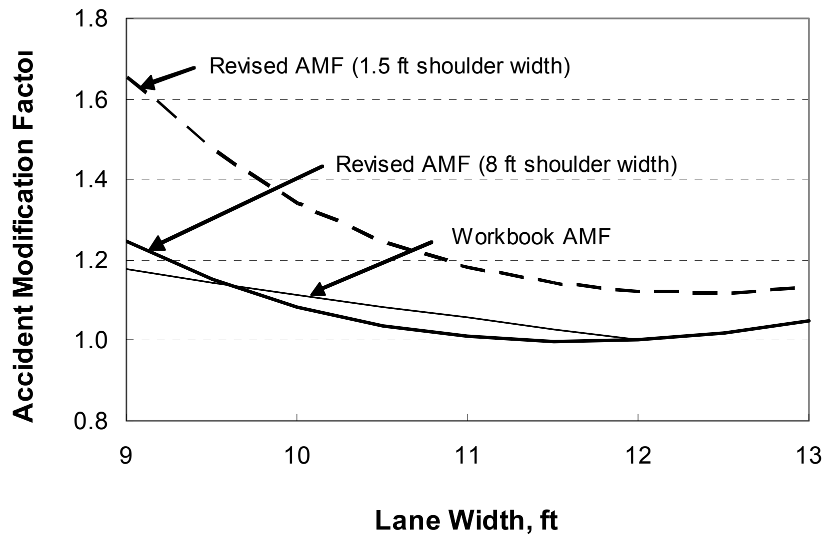


Figure 4-9. Relationship between Lane Width and AMF Value.

The trends in Figure 4-9 were compared with AMF models derived from three other sources (*10, 11, 12*). These AMF models are illustrated in Figure 4-10. They are also based on linear model forms, in contrast to the revised model which has a curved shape. Comparison of the trend lines in Figure 4-10 suggests that the revised AMF is consistent with the AMF developed by Harwood et al. (*10*) when the shoulder width is about 8 ft. However, the AMF models attributed to the other two researchers indicate that lane width has a more significant effect on crash frequency. These models yield AMF estimates that are more consistent with those obtained from the revised model when the shoulder width is less than 8 ft.

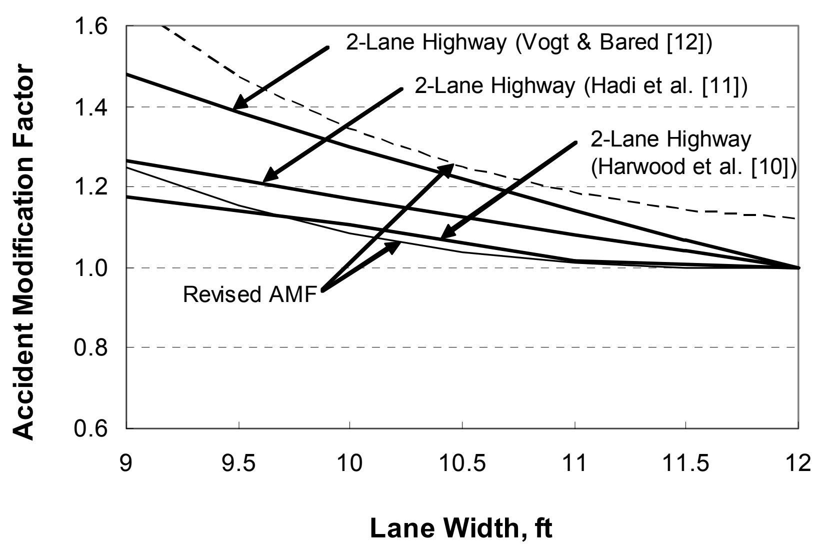


Figure 4-10. Lane Width AMFs from Other Sources.

It should be noted that the AMFs shown in Figure 4-10 that are attributed to Hadi et al. (*11*) and Vogt and Bared (*12*) were both derived using injury (plus fatal) crash data, as was Equation 38. In contrast, the AMF developed by Harwood et al. (*10*) is based on "total" crashes (i.e., property-damage-only, injury, and fatal crashes). The trends in Figure 4-10 may also be suggesting that injury (plus fatal) crash frequency is more sensitive to changes in lane width than is total crash frequency.

The revised AMF model is shown in Figure 4-11 for a range of shoulder widths. The values obtained from Equation 38 are shown with the two thick trend lines. The thick dashed trend line corresponds to shoulders adjacent to 10 ft lanes. The thick solid trend line corresponds to shoulders adjacent to 12 ft lanes. The values obtained from the original AMF model (i.e., Equation 29) are shown with a thin solid line.

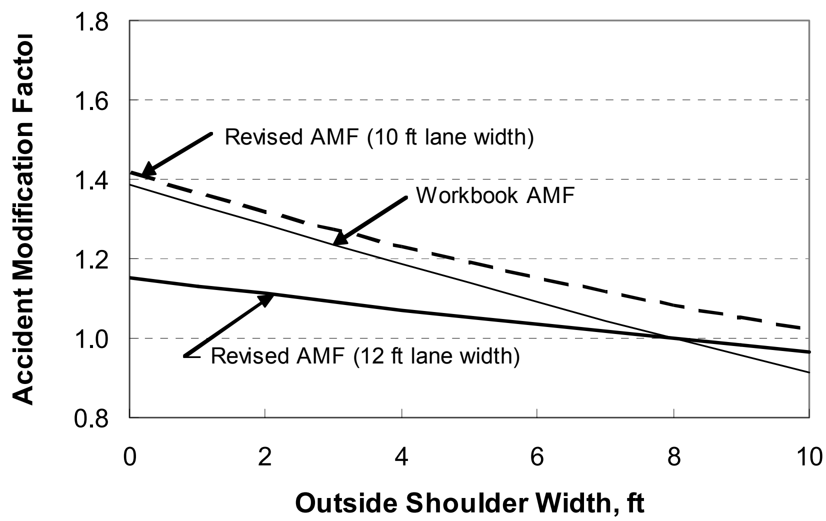


Figure 4-11. Relationship between Shoulder Width and AMF Value.

As shown in Figure 4-11, the revised AMF yields values similar to those obtained from Equation 29 for roads with narrow shoulders and a lane width of about 10 ft. They are also similar for roads with wide shoulders and a lane width of about 12 ft. These two combinations are likely to be found on the rural two-lane highways of most states and reflect a correlation between lane and shoulder width that is more accurately modeled by the revised AMF. In general, shoulder width is indicated to have a larger effect on safety when the lane width is narrow, which is logical.

REFERENCES

1. Bonneson, J.A., K. Zimmerman, and K. Fitzpatrick. *Interim Roadway Safety Design Workbook.* FHWA/TX-06/0-4703-P4. Texas Department of Transportation, Austin, Texas, April 2006.

2. Hauer, E. *Observational Before-After Studies in Road Safety.* Pergamon Press, Elsevier Ltd., Oxford, Great Britain, 1997.

3. Hauer, E. "Over-dispersion in Modeling Accidents on Road Sections and in Empirical Bayes Estimation." *Accident Analysis and Prevention.* Vol. 33. Pergamon Press, 2001, pp. 799-808.

4. McCullagh, P., and J.A. Nelder. *Generalized Linear Models.* Chapman and Hall, New York, N.Y., 1983.

5. Kvalseth, T.O. "Cautionary Note About R^2." *The American Statistician.* Vol. 39, No. 4, (Part 1), American Statistical Association, November 1985, pp. 279-285.

6. Miaou, S.P. *Measuring the Goodness-of-Fit of Accident Prediction Models.* FHWA-RD-96-040. Federal Highway Administration, Washington, D.C., 1996.

7. Hauer, E., and B.N. Persaud. "How to Estimate the Safety of Rail-Highway Grade Crossings and the Safety Effects of Warning Devices." *Transportation Research Record 1114*. Transportation Research Board, National Research Council, Washington, D.C., 1987, pp. 131-140.

8. Kulmala, R. *Safety at Rural Three- and Four-Arm Junctions: Development and Application of Accident Prediction Models*. VTT Publications, Espoo, Finland, 1995.

9. Cameron, A.C., and P.K. Trivedi. *Regression Analysis of Count Data*. Econometric Society Monograph No. 30, Cambridge University Press, Cambridge, U.K., 1998.

10. 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.* FHWA-RD-99-207. Federal Highway Administration, Washington, D.C., 2000.

11. Hadi, M.A., J. Aruldhas, L-F. Chow, and J.A. Wattleworth. "Estimating Safety Effects of Cross-Section Design for Various Highway Types Using Negative Binomial Regression." *Transportation Research Record 1500.* Transportation Research Board, National Research Council, Washington, D.C., 1995, pp. 169-177.

12. Vogt, A., and J. Bared. *Accident Models for Two-Lane Rural Roads: Segments and Intersections.* FHWA-RD-98-133. Federal Highway Administration, Washington, D.C., October 1998.

CHAPTER 5. CALIBRATION OF AMFs FOR RURAL FRONTAGE ROAD SEGMENTS

OVERVIEW

This chapter describes the development of quantitative tools for evaluating the safety of rural frontage-road segments in Texas. These tools include a "base" model for estimating the crash frequency of typical frontage-road segments and several accident modification factors (AMFs). The focus of this development is the basic geometric design factors that influence the safety of frontage-road segments in rural areas. The tools described herein are intended to help the designer evaluate alternative geometric design element dimensions, such as a lane width of 11 ft versus 12 ft, in terms of their impact on safety. These tools do not address the safety of ramp/frontage-road terminals or the safety of frontage-road/crossroad intersections. Moreover, they do not directly address the safety of one-way frontage-road operation versus two-way frontage-road operation.

No frontage-road-based AMFs were specifically identified in a review of the literature. AMFs that have been developed for rural two-lane highways could arguably be used for rural frontage roads given their similar area-type, high-speed character, and two-lane cross section. AMFs for rural two-lane highways have been documented in several reports (*1, 2*). However, the frontage road is different from the two-lane highway because it has restricted access along at least one side of the road, a higher percentage of turning traffic, and periodic ramp-frontage-road terminals. As a result of these differences, a given design element is likely to have a different effect on frontage-road safety than on two-lane highway safety. For this reason, there is a need to develop AMFs for rural frontage roads.

The remainder of this chapter is divided into four main parts. The first part describes the frontage-road segment selection and data collection activities. The second part explains the statistical analysis procedure used to develop the safety evaluation tools. The third part describes the frontage-road AMFs derived from the data. The last part describes a model for estimating the crash frequency of typical frontage-road segments.

DATA COLLECTION ACTIVITIES

This part of the chapter describes the data collection activities undertaken to assemble a database suitable for developing frontage-road safety evaluation tools. The first section outlines the criteria used in the segment selection process. The second section describes the characteristics of the crash data. The last section describes the process used to collect traffic flow and geometry data.

Selection Process

Figure 5-1 shows the frontage-road segment that formed the basis for the analysis. All frontage-road segments considered for inclusion in the database were located between successive interchanges. Also, each segment selected was required to have at least one ramp terminal along its length (however, the ramp terminal area was not considered to be part of the segment).

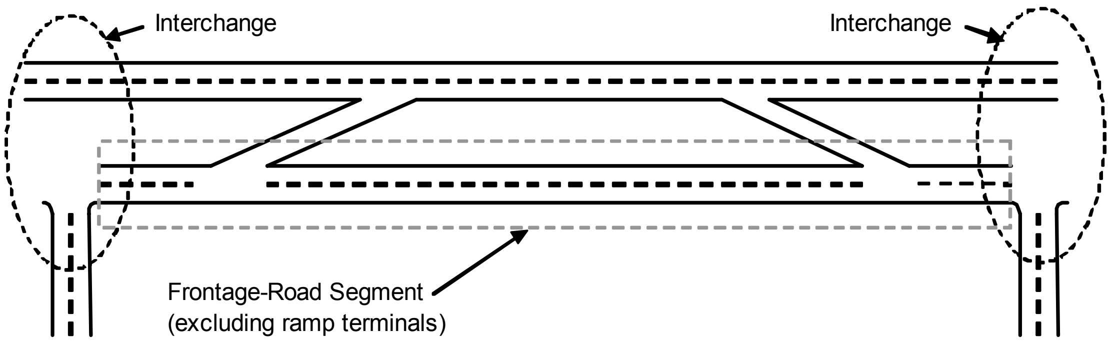


Figure 5-1. Frontage-Road Analysis Segment.

Four Texas highway corridors were considered for segment selection. In all cases, only those frontage-road segments in rural areas were considered for inclusion in the database. The first corridor was located along I-35 between the city of Georgetown and the location where I-35 splits between I-35E and I-35W north of Waco (segments in the vicinity of the cities of Temple and Waco were excluded). The second corridor was located along SH 6/US 190 near the city of Bryan. The third study corridor was located on I-10 between the cities of Glidden and Brookshire. The last study corridor was located on I-45 between the cities of Willis and Madisonville. The study corridors collectively contained a mix of one-way and two-way frontage roads.

All total, 141 segments were ultimately identified from a review of various maps and aerial photos. Each segment was subsequently visited to collect additional data not available from other sources. A distance measuring instrument was used to measure segment length and the location of intersection roads and ramp terminals. After screening the initial 141 segments for data availability and construction activity, the sample size was reduced to 123 segments. The characteristics of these segments are summarized in Table 5-1.

Crash Data

Crash data for each frontage-road segment were extracted from the Department of Public Safety (DPS) electronic database. Five years of crash data (1997 to 2001) were used. Only crashes that were "segment-related" were included in the database assembled for this research. Crashes that were related to the ramp/frontage-road terminals, and those that were related to the frontage-road/crossroad intersections, are referred to as "non-segment-related crashes" and were excluded. These crashes were rationalized to be strongly influenced by the design of the terminal, or intersection, and would not be helpful in identifying correlation between segment design and segment crash frequency. Non-segment-related crashes were defined to be those crashes identified as "at-intersection" or "intersection-related" in the DPS database.

Table 5-1. Frontage-Road Segment Physical Characteristics.

Highway	Operation	Number of Segments	Average Daily Traffic Volume, veh/d			Percent Segments with Edge Delineation		Segment Length, mi		
			Ave.	Min.	Max.	Left	Right	Ave.	Min.	Max.
SH 6/ US 190	Two-way	11	2360	110	6168	73	73	2.00	1.06	2.66
	One-way	20	2550	140	5270	100	10	1.12	0.78	1.89
I-10	Two-way	16	675	168	1585	25	25	2.74	1.36	5.34
I-35	Two-way	57	575	125	2199	33	33	1.97	0.69	3.76
	One-way	6	790	361	1046	100	100	1.93	1.13	2.64
I-45	Two-way	10	1990	218	1988	50	40	2.40	0.79	4.21
	One-way	3	4470	3093	5766	100	100	1.26	1.00	1.77
Summary	Two-way	94	2385	110	6168	38	37	2.15	0.69	5.34
	One-way	29	940	140	5766	100	38	1.30	0.78	2.64
	Overall	123	1230	110	6168	53	37	1.92	0.69	4.21

Crashes for all types of severity were extracted from the DPS database. A summary of the crash data characteristics is presented in Table 5-2. The data in this table correspond to crashes that were reported to have occurred on the frontage-road segment and were determined to be segment-related crashes. The crash frequencies listed correspond to a five-year period.

Table 5-2. Frontage-Road Segment Crash Characteristics.

Highway	Operation	Crash Frequency [1, 2]		Total Crashes Per Segment [1, 2]			Total Crash Rate, [3] cr/mvm
		I+F [4]	Total	Ave.	Min.	Max.	
SH 6/ US 190	Two-way	11	19	1.73	0	3	0.20
	One-way	23	33	1.65	0	7	0.31
I-10	Two-way	8	15	0.94	0	3	0.34
I-35	Two-way	44	72	1.26	0	5	0.60
	One-way	8	9	1.50	0	3	0.54
I-45	Two-way	17	21	2.10	0	6	0.59
	One-way	13	17	5.67	4	8	0.59
Summary	Two-way	80	127	1.35	0	6	0.43
	One-way	44	59	2.03	0	8	0.39
	Overall	124	186	1.51	0	8	0.42

Notes:
1 - Crash data apply to frontage-road segments and do not include crashes that may have occurred at ramp/frontage-road terminals or at frontage-road/crossroad intersections.
2 - Crash frequencies listed in columns three through seven are based on a five-year period.
3 - Crash rate has units of total crashes per million-vehicle-miles (cr/mvm).
4 - I+F: injury plus fatal crashes.

As indicated in Table 5-2, there were 186 total crashes that occurred on the 123 frontage-road segments during the five years for which crash data were available. Many of the segments experienced no crashes during this time period. One segment experienced eight crashes during this period. Injury (plus fatal) crashes accounted for 124 of the 186 crashes, or about 67 percent of all crashes. Although not shown in the table, the injury (plus fatal) crash rate for frontage-road segments is 0.28 cr/mvm. This rate is slightly higher than the injury (plus fatal) crash rate of 0.20 cr/mvm found for typical rural two-lane highways (*2*).

Supplemental Data Collection

Supplemental data were collected to facilitate the statistical examination of factors that may influence segment crashes. The data collected include: traffic counts, lane width, paved shoulder width, presence of pavement edge line markings, presence of curb, and the number of private and commercial driveways. The characteristics of these data are summarized in Table 5-3.

Table 5-3. Frontage-Road Segment Supplementary Data.

Statistics	Private Driveways, dr/mi		Commercial Driveways, dr/mi		Lane Width, ft		Paved Right-Shoulder Width, ft		Paved Left-Shoulder Width, ft	
	Two-Way	One-Way	Two-Way	One-Way	Two-Way	One-Way	Two-Way	One-Way	Two-Way	One-Way
Average	2.10	1.44	2.55	5.50	10.5	11.7	1.3	1.3	1.1	2.4
Minimum	0.00	0.00	0.00	0.53	9	10	0	0	0	0
Maximum	9.34	12.33	10.63	21.04	13	13	9	8	9	7
Sum	421	54	466	207						

Estimates of segment ADT were obtained from a variety of sources. Many of the estimates were extracted from the Texas Reference Marker (TRM) system database for the years 1997 to 2001. For those segments for which ADTs were not available from TRM, automatic traffic recorders (ATR) were used to obtain counts. All ATR counts were collected for a 24-hour time period. They were subsequently adjusted to yield an estimate of the ADT for 1999, which corresponds to the midpoint year for the period corresponding to 1997 to 2001.

STATISTICAL ANALYSIS

The statistical analysis of the database consisted of developing a safety prediction model relating the reported crash frequency to the measured site characteristics. Several alternative model forms were tested. However, only the model that provided the best combination of good fit to the data and a logical relationship between the independent and dependent variables is described in this section.

The number of crashes at segment i, N_i, when conditional on its mean μ_i, is assumed to be Poisson distributed and independent over all segments as:

$$N_i \mid \mu_i \cong Poisson(\mu_i) \qquad i = 1,2,...,I \tag{1}$$

The mean of the Poisson distribution is structured as:

$$m_i = f(X_i;\boldsymbol{b})\, e^{error_i} \tag{2}$$

where,

$f(.)$ = predictive model represented as a function of variables X_i;
$\boldsymbol{b}$ = a vector of unknown coefficients; and
$error_i$ = random error for segment i.

It is usually assumed that e^{error} is independent and gamma distributed with a mean equal to 1.0 and a variance equal to $1/k$ for all i (with $k > 0.0$). With this characteristic, it can be shown that N_i, conditional on $f(.)$ and k, is distributed as a negative binomial random variable with a mean equal to $f(.)$ and a variance equal to $f(.)\,[1 + f(.)\,/k]$. The variable k is usually defined as the over-dispersion parameter for the negative binomial distribution. If k approaches infinitely large values, then the distribution converges to a Poisson distribution.

An important element of the development of a multivariate regression model is its structural form and the relationship of model variables. At the start of the modeling effort, several regression model forms were attempted using one-way and two-way frontage roads together in one model, and then in separate models. Due to the low sample mean values and small sample size, some models did not provide reasonable results. As explained by Lord *(3)*, databases with these characteristics may not exhibit the large variability associated with the negative binomial distribution. Rather, the regression model may "over-explain" some of the random variability in a small database, or the low sample mean may introduce an instability in the model coefficients *(3)*. Preliminary analyses revealed that the model was over-explaining some of the random variability in the data and, as a result, it was necessary to specify a Poisson distribution in the regression analysis.

During model development, each variable was added to the model one at the time. Initially, models were developed separately for the one-way and two-way frontage road types. However, it was found that the regression coefficients for common variables in these two models were not significantly different. For this reason, the data were combined and one model was developed. Indicator variables were used in this model whenever the effect of a specific variable was found to be correlated with frontage road type. The form ultimately selected is:

$$E[N] = ADT^{\,b_1}\, L\, e^{\,b_0 + b_2 W_l + b_3 (W_{s,r} + W_{s,l}) + b_4 EM\, I_2} \tag{3}$$

where,

$E[N]$ = expected crash frequency, crashes/yr;
ADT = average daily traffic volume, veh/d;
L = segment length, mi;
W_l = average lane width, ft;
$W_{s,r}$ = paved shoulder width on the right side of the frontage road, ft;
$W_{s,l}$ = paved shoulder width on the left side of the frontage road, ft;
EM = proportion of the segment with pavement edge markings (both directions);

I_2 = indicator variable (= 1.0 for two-way operation, 0.0 for one-way operation); and
b_i = calibration coefficients (i = 0, 1, 2, 3,...).

It was decided to use segment length as an offset variable in Equation 3, as opposed to having it adjusted by an exponential regression coefficient (e.g., as is the case for the ADT variable). This approach was taken because segment length is considered to be directly related to segment crash frequency (i.e., the number of crashes on a segment increases in direct proportion to the increase in its length). Hence, empirical adjustment is not believed to be needed for the length variable. The coefficients in the model were estimated using the Genstat statistical analysis software (*4*).

Table 5-4 summarizes the coefficient values associated with the calibrated model. These coefficients correspond to a model that predicts total crashes on rural frontage-road segments. An additional model was fit to the injury (plus fatal) crash data; however, the elimination of property-damage-only crashes from the database contributed further to the low sample mean and small sample size problems noted previously. Further efforts to calibrate a model using only injury and fatal crash data were abandoned. However, as described in a subsequent section, the model listed in Table 5-4 was used to derive a model for estimating injury (plus fatal) crash frequency.

Table 5-4. Multivariate Model Statistical Description–Frontage-Road Segments.

Model Statistics		**Value**		
Range of Model Variables				
Variable	**Variable Name**	**Units**	**Minimum**	**Maximum**
ADT	Segment ADT	veh/d	110	6,168
L	Segment Length	miles	0.69	4.21
W_l	Lane width	ft	9	13
$W_{s,l}+W_{s,r}$	Paved shoulder width (left + right shoulder)	ft	0	17
EM	Proportion with pavement edge markings	--	0	1
I_2	Two-way operation indicator variable	--	0	1
Calibrated Coefficient Values				
Variable	**Definition**	**Value**	**Std. Dev.**	**t-statistic**
b_0	Intercept	-3.85	1.11	-3.5
b_1	Effect of segment ADT	0.641	0.086	7.5
b_2	Effect of lane width	-0.188	0.104	-1.8
b_3	Effect of paved shoulder width	-0.035	0.028	-1.3
b_4	Effect of pavement edge marking presence	-0.518	0.203	-2.6

The coefficient values listed in Table 5-4 indicate the nature of the correlation between the corresponding variable and crash frequency. Specifically, positive coefficient values indicate that an increase in the variable value correlates with an increase in crash frequency (and negative values correlate with a decrease in crashes). For example, the coefficient of -0.188 associated with the lane width variable indicates that an increase in lane width is associated with a decrease in the number of crashes. The interaction variable (Edge Marking Presence × Two-Way Operation) shows that the presence of a pavement edge line is negatively associated with the number of crashes for two-way frontage roads. It suggests that the addition of edge lines to a two-way frontage road reduces crash frequency. A similar effect was not found for one-way frontage-road segments in the assembled database; however, intuition would suggest it is likely to exist.

The coefficient values listed in Table 5-4 can be substituted into Equation 3 to yield the following calibrated model:

$$E[N] = 0.021 \ ADT^{0.641} \ L \ e^{-0.188\, W_l \, - \, 0.035\,(W_{s,r} \, + \, W_{s,l}) \, - \, 0.518\, EM\, I_2} \tag{4}$$

The variable *EM* used in Equation 4 is a proportion that varies from 0.0 to 1.0. It represents the length of edge line on the right side of the roadway plus the length of edge line on the left side of the roadway, all of which is divided by twice the segment length. Thus, a 1.0 mi frontage road with edge lines only on the right side would have $EM = 0.5$ (= [1.0 + 0.0]/[2 ×1.0]).

ACCIDENT MODIFICATION FACTORS

Three AMFs were derived from the frontage-road segment database. Alternative approaches for developing AMFs from cross section data, such as that proposed by Washington et al. (*5*), were examined and ruled out due to the small number of crashes in the database. Consequently, the AMFs were estimated directly from the coefficients of the model, as listed in Table 5-4. This approach for AMF development assumes that each model variable is independent and, thus, not influenced by the value of any other variable. It also assumes that the relationship between the change in the variable value and the change in crash frequency is exponential (as suggested by Equation 3). A more rigorous study design and a larger database (i.e., one with more segments) would be needed to test the validity of these assumptions. However, experience in deriving AMFs in this manner indicates that the assumptions are reasonable and, with thoughtful model development, the resulting AMFs can yield useful information about the first-order effect of a given variable on safety.

AMF - Lane Width

The recommended AMF for frontage-road lane width is:

$$AMF_{LW} = e^{-0.188\,(W_l \, - \, 12.0)} \tag{5}$$

where,
AMF_{LW} = lane width accident modification factor.

The average lane width used in Equation 5 represents the total width of all through traffic lanes on the frontage road divided by the number of through lanes. The value of 12.0 in Equation 5 reflects the base, or typical, lane width condition. By definition, it is associated with an AMF value of 1.0.

The graphical representation of the lane width AMF is shown in Figure 5-2. The relationship between lane width and AMF value shown in this figure suggests that crash frequency is reduced

17 percent for a 1 ft increase in lane width. Based on the range of lane widths in the database, the lane width AMF is applicable to lane widths ranging from 9 to 12 ft.

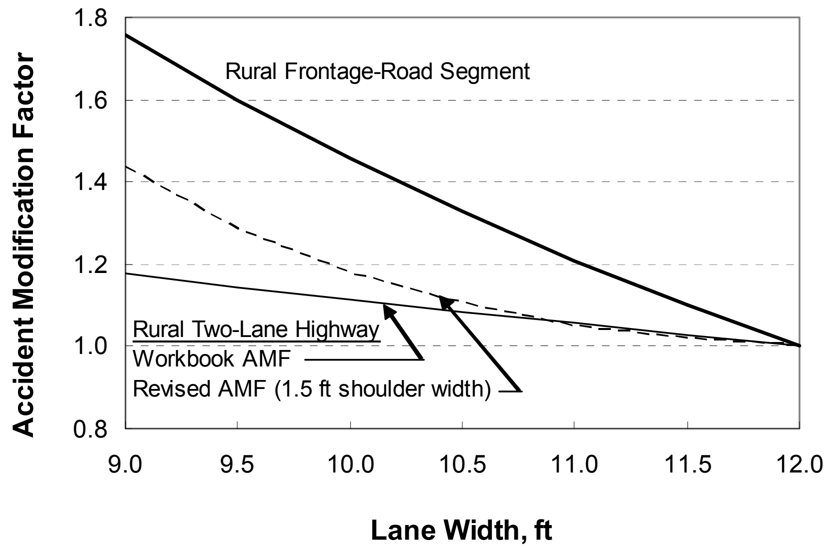


Figure 5-2. AMF for Lane Width.

Also shown in Figure 5-2 is the lane width AMF for rural two-lane highways from the *Interim Roadway Safety Design Workbook* (*Workbook*) (*2*) and from Chapter 4. A comparison of these AMFs with the lane width AMF for frontage roads suggests that lane width on a frontage road has a greater impact on crash frequency than it does on a two-lane highway. It is possible that this trend stems from the relatively high percentage of turning traffic and the considerable weaving activity that occurs on frontage roads (between the ramp terminals and the crossroad intersection), relative to a two-lane highway. Wider lanes on frontage-road segments may provide some additional room for recovery when these turning and weaving-related conflicts occur.

AMF - Shoulder Width

The recommended AMF for frontage-road shoulder width is:

$$AMF_{SW} = e^{-0.070\,(W_s - 1.5)} \qquad \textbf{(6)}$$

where,

AMF_{SW} = shoulder width accident modification factor and

W_s = average paved shoulder width (= $[W_{s,r} + W_{s,l}]/2$), ft.

This AMF is derived from Equation 4; however, the associated regression coefficient (i.e., -0.035) has been doubled such that the resulting AMF is based on the average paved shoulder width. This average is computed as the sum of the left and right shoulder widths divided by 2.0. The value of 1.5 in Equation 6 reflects the base, or typical, average shoulder width condition for frontage roads.

The graphical representation of the shoulder width AMF is shown in Figure 5-3. The relationship between frontage-road shoulder width and the AMF value suggests that crash frequency is reduced 7 percent for a 1 ft increase in average shoulder width. Based on the range of shoulder widths in the database, the AMF is applicable to shoulder widths ranging from 0 to 5 ft.

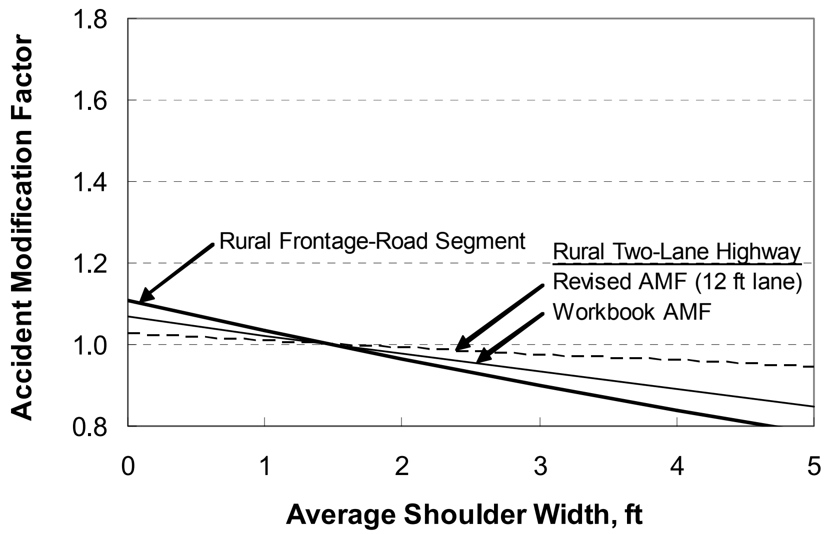


Figure 5-3. AMF for Shoulder Width.

Also shown in Figure 5-3 is the shoulder width AMF for rural two-lane highways from the *Workbook* (*2*) and from Chapter 4. The base condition for the *Workbook* AMF has been changed to a shoulder width of 1.5 ft to facilitate its comparison with Equation 6. As suggested by the trend lines in this figure, shoulder width has a slightly larger impact on frontage-road safety than on rural two-lane highways. This trend is consistent with that found for the lane width AMF.

AMF - Edge Marking Presence on Two-Way Frontage Roads

The AMF that was derived from variables associated with the presence of edge line delineation is:

$$AMF_{EM} = e^{-0.518\,EM} \tag{7}$$

where,

AMF_{EM} = edge marking presence accident modification factor.

This AMF was derived using data for two-way frontage roads. An equivalent AMF for one-way frontage roads could not be derived. The AMF likely explains the effect of edge line delineation *and* other traffic control devices that are used to highlight the two-way operation (relative to the more common one-way operation) and to ensure correct driving behavior.

Equation 7 reflects a base frontage road condition where there are no pavement edge lines. As shown in Table 5-1, this base condition reflects the lack of marking presence found at slightly more than one-half of the frontage-road segments in the database. A graphical representation of the edge marking AMF is shown in Figure 5-4.

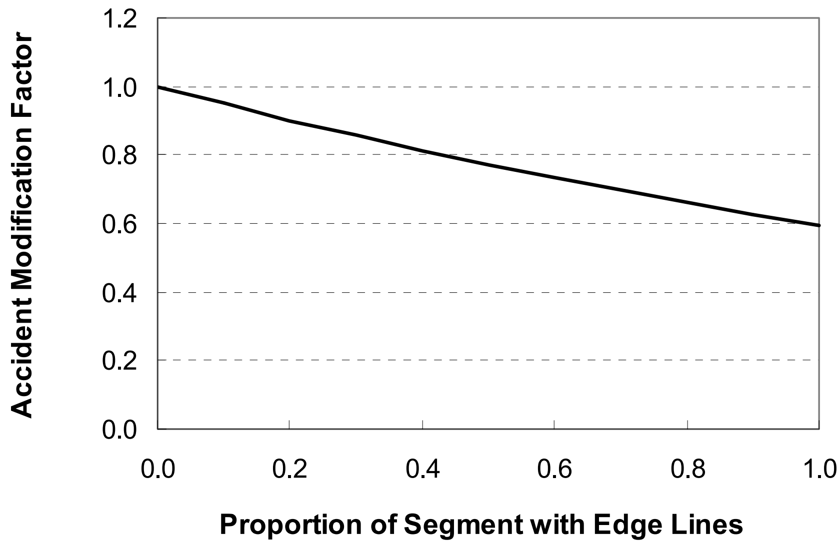


Figure 5-4. AMF for Presence of Edge Line Delineation on Two-Way Rural Frontage Roads.

The trend shown in Figure 5-4 suggests that edge markings can reduce crashes on two-way frontage roads by 40 percent if placed fully along both sides of the frontage road. This reduction is relatively large and suggests that this AMF is explaining more than just the effect of pavement edge marking presence on crash frequency. The presence of pavement edge markings is also likely to be accompanied by additional warning signs and centerline markings that denote two-way operation. Hence, the 40 percent reduction noted previously is likely a reflection of the effectiveness of the full complement of traffic control devices often deployed on two-way frontage roads to mitigate the increased potential for wrong-way driving. Given this speculation and the lack of corroborating evidence from other research projects, this AMF cannot be recommended for evaluating edge marking presence. However, it does provide some validation to the belief that a full complement of traffic control devices on two-way frontage roads will reduce crashes, although the amount of reduction is uncertain at this time.

BASE MODEL

This part of the chapter describes the methods used to develop a base model for frontage-road segments. A base model is used to estimate the expected crash frequency for a typical frontage-road segment. It would be used with the AMFs described in the previous part of the chapter to estimate the crash frequency for a specific frontage-road segment.

The base model was developed from the regression model shown as Equation 4. Two adjustments were made to this equation to obtain the base model. First, the variables in the exponential term were removed because they had been incorporated into the aforementioned AMFs and the coefficient of 0.021 was modified to reflect the base conditions of 12 ft lane width, an average paved shoulder width of 1.5 ft, and no pavement markings. The resulting coefficient was computed to be 0.0020. Second, this coefficient was adjusted such that the base model could be used to estimate *injury plus fatal* (as opposed to total) crash frequency. This latter adjustment entailed multiplying the coefficient 0.0020 by 0.67 (= 0.00134), where the multiplier 0.67 reflects the fact that 67 percent of the crashes in the database correspond to injury or fatal crashes (as noted previously in the discussion of Table 5-2). The resulting base model for frontage roads is:

$$E[N]_b = 0.00134 \ ADT^{0.641} \ L \tag{8}$$

where,
$E[N]_b$ = expected base crash frequency, crashes/yr.

Equation 8 predicts the injury (plus fatal) crash frequency that would be estimated for a two-way or one-way frontage-road segment with 12 ft lanes and an average paved shoulder width of 1.5 ft. In application, the crash frequency predicted by Equation 8 would be multiplied by the AMFs for lane width and shoulder width to estimate the injury (plus fatal) crash frequency for a given segment with a specified lane and shoulder width.

The estimate obtained from Equation 8 does not include the crashes that would be attributed to the ramp/frontage-road terminal or the frontage-road/crossroad intersection. It also does not include any crashes that may occur on the main lanes that may indirectly be attributed to the frontage-road operation or its ramp design.

Equation 8 is compared in Figure 5-5 with the rural two-lane highway base model included in the *Workbook* (*2*). This model is based on an injury (plus fatal) crash rate of 0.20 cr/mvm. The trend lines in the figure indicate that a frontage road experiences slightly more injury (or fatal) crashes than a rural two-lane highway for ADTs less than 3500 veh/d. The reverse trend applies for ADTs greater than 3500 veh/d. It is possible that the increased turning and weaving activity associated with the frontage road (relative to the two-lane highway) may explain the slightly higher crash frequency on frontage roads for ADTs less than 3500 veh/d. As ADT exceeds 3500 veh/d, there may be less opportunity for turning (i.e., fewer gaps) and the weaving activity may be more constrained (i.e., lower speed) on the frontage road, such that frontage-road crash frequency is lower than that found on two-lane highways.

The predicted values from the two models shown in Figure 5-5 are not statistically different from each other when a 95th percentile confidence interval is used. Hence, there is a small chance that the trend shown in Figure 5-5 is a result only of random variation in the data and that the two facility types actually have a similar crash frequency for a given ADT and segment length.

During the regression analysis, the small sample size required specification of a Poisson distribution for the dependent variable. The need for this adjustment was based partly on an examination of the Pearson χ^2 statistic and the corresponding scale parameter φ, as described in

Chapter 4. It is recognized that crash data from a collection of road segments have, in reality, a negative binomial distribution. However, the small sample size allowed the regression model to over-explain the data and reduce the random variability in the data such that it was a better fit to the Poisson distribution.

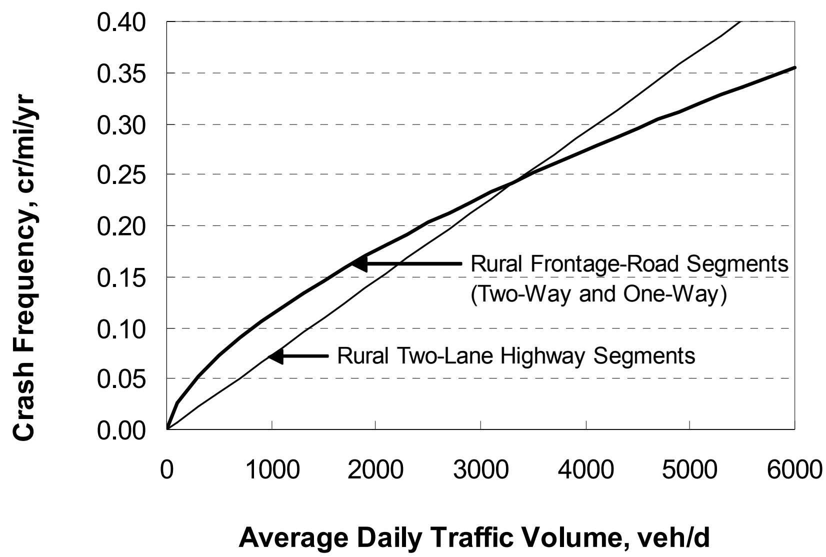


Figure 5-5. Comparison between Frontage-Road Segment Model and Rural Two-Lane Highway Segment Model.

Equation 9 below describes the variance of the crash data when it has a negative binomial distribution. This distribution has a larger variance than the Poisson distribution, which has its variance equal to the mean (i.e., $V[X] = y\,E[N]$). In fact, the variability computed using Equation 9 converges to that of the Poisson distribution when the over-dispersion parameter k has an infinite value such that the last term in the equation equals 0.0. This latter term is important because it is used to estimate the variance of the base model estimate (i.e., $V(E[N]) = \{y\,E[N]\}^2 / \{k\,L\}$).

$$V[X] = y\,E[N] + \frac{(y\,E[N])^2}{k\,L} \tag{9}$$

where,

X = reported crashes for a segment with an expected annual crash frequency of $E[N]$;
y = time interval during which X crashes were reported, yr; and
k = over-dispersion parameter, mi^{-1}.

Given that the frontage-road crash database should theoretically have a negative binomial distribution, it is desirable to have an estimate of k for the data to permit computation of the variance of the base model estimates. The regression analysis suggests that k is infinite, but this is a gross overestimate of k and is due to the small sample size. A lower bound on k was obtained by conducting a regression analysis using the null model (i.e., $E[N] = b_0$) and a negative binomial distribution. The results from this analysis indicated that a k of 1.37 mi^{-1} yielded the best fit to the data. The true value of k for this database is likely to be somewhat larger than this value (but not equal to infinity). For this reason, the value 1.37 mi^{-1} should yield a conservatively large estimate of the variance of the expected base crash frequency.

REFERENCES

1. 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.* FHWA-RD-99-207. Federal Highway Administration, Washington, D.C., 2000.

2. Bonneson, J.A., K. Zimmerman, and K. Fitzpatrick. *Interim Roadway Safety Design Workbook.* FHWA/TX-06/0-4703-P4. Texas Department of Transportation, Austin, Texas, April 2006.

3. 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, Elsevier Ltd., Oxford, Great Britain, July 2006, pp. 751-766.

4. Payne, R.W. (ed.) *The Guide to Genstat.* Lawes Agricultural Trust, Rothamsted Experimental Station, Oxford, Great Britain, 2000.

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

CHAPTER 6. CALIBRATION OF SAFETY PREDICTION MODELS FOR RURAL TWO-LANE HIGHWAYS

OVERVIEW

This chapter describes the activities undertaken to re-calibrate existing safety prediction models for rural two-lane highway segments and intersections in Texas. The existing models considered for re-calibration are described in the *Interim Roadway Safety Design Workbook* (*Workbook*) (*1*). The *Workbook* describes separate models for the following components of rural two-lane highways:

- highway segments (excluding intersections),
- three-leg intersections with two-way stop control,
- three-leg intersections with signal control,
- four-leg intersections with two-way stop control, and
- four-leg intersections with signal control.

Each model provides an estimate of the expected crash frequency for the associated component, given specified traffic volume and geometric design conditions. However, each model was initially calibrated using data from locations outside of Texas. To ensure that the estimate is not biased for Texas applications, the model was re-calibrated using data specific to Texas.

The objective of this research was to re-calibrate models for each of the aforementioned components using crash data for Texas highways. However, if the effort to re-calibrate indicated that an alternative model structure was more appropriate, then the option to develop (and calibrate) a new model was available.

This chapter is divided into four parts. The first part provides a review of the literature on the topic of safety prediction model re-calibration. The next part summarizes the method used to re-calibrate the existing models. The third part describes the analysis and findings from the re-calibration of the highway segment model. The last part describes the analysis and findings from the re-calibration of the intersection models.

LITERATURE REVIEW

This part of the chapter describes a framework for safety prediction model re-calibration, as described in the literature. It consists of seven sections. The first section provides a brief overview of safety prediction and the role of "base" prediction models. The second section describes two alternative techniques used to assemble a calibration database. The third section compares the goals of initial calibration with those of re-calibration. The fourth section illustrates how the calibration methods and databases are combined to yield four different types of calibrated models. The fifth section summarizes the base models described in the *Workbook*. The sixth section reviews the data available in Texas' road inventory database. The last section identifies the process used to divide

Texas into regions of similar climate and geography for the purpose of exploring regional influences on the calibration coefficients.

Safety Prediction Models

The expected crash frequency for a highway segment (or intersection) with specified attributes is computed using a safety prediction model. This model represents the combination of a "base" model and one or more accident modification factors (AMFs). The base model is used to estimate the expected crash frequency for a typical segment. The AMFs are used to adjust the base estimate when the attributes of the specific segment are not considered typical. For highway segments, the safety prediction model is shown below as Equation 1 and its base model component is shown as Equation 2.

$$E[N] = E[N]_b \times AMF_{lw} \times AMF_{dd} \cdots \qquad (1)$$

with,

$$E[N]_b = a\ ADT^{\,b}\ L \qquad (2)$$

where,
$E[N]$ = expected crash frequency, crashes/yr;
$E[N]_b$ = expected base crash frequency, crashes/yr;
AMF_{lw} = lane width accident modification factor;
AMF_{dd} = driveway density accident modification factor;
a, b = calibration coefficients;
ADT = average daily traffic volume, veh/d; and
L = highway segment length, mi.

When the coefficient b for the ADT variable is equal to 1.0, then the coefficient a in Equation 2 is effectively equal to the crash rate, with units of "crashes per million-vehicle-miles."

Two variations of Equation 1 are used to estimate the expected crash frequency for intersections, where the AMFs included in it are applicable to intersection design elements. A common form for the intersection base model is:

$$E[N]_b = a\ ADT_{major}^{b_1}\ ADT_{minor}^{b_2} \qquad (3)$$

where,
a, b_1, b_2 = calibration coefficients;
ADT_{major} = average daily traffic volume on the major road, veh/d; and
ADT_{minor} = average daily traffic volume on the minor road, veh/d.

This model form predicts 0.0 crashes when either ADT variable is equal to 0.0. This boundary condition is illogical because some types of crashes (e.g., rear-end) are still likely to occur when one of the ADT variables is 0.0 and the other is nonzero. An alternative form for the intersection base model that does not share this limitation is:

$$E[N]_b = a\ (ADT_{major} + ADT_{minor})^b \qquad (4)$$

When the coefficient b in this equation is equal to 1.0, then the coefficient a is effectively equal to the crash rate, with units of "crashes per million-entering-vehicles."

The selected intersection base model form is typically used for all intersection configurations (i.e., three-leg, four-leg, etc.) and control types (i.e., signalized, two-way stop control, etc.). However, the model is separately calibrated for each configuration and control type combination. This approach yields unique calibration coefficients for each combination.

Model Development Databases: Full versus Base

This section describes two techniques used to develop a base model and the type of database needed for each technique. One technique is based on the derivation of a base model from a "full" model previously calibrated using regression analysis. For highway segments, the full model form is shown in Equation 5.

$$E[N] = a\ ADT^{b}\ L\ e^{c_1 x_1 + c_2 x_2 + c_3 x_3 \dots} \qquad \textbf{(5)}$$

where,

x_i = variable describing geometric design element i (e.g., lane width); and
c_i = calibration coefficient for design element i.

As suggested by Equation 5, a full model includes all of the variables that explain the systematic variation in the crash data. The base model is then derived from the full model by: (1) setting each full model variable (except the *ADT* and *L* variables) at a value that is typical of segments in the region represented in the database and, (2) algebraically reducing the full model to a form consistent with Equation 2, 3, or 4. This technique is described by Harwood et al. (*2*).

With this technique, the database required to develop the base model is the same as that needed to develop a full model. For this reason, the database is referred to as a "full" database because it contains a full range of data for each segment (or intersection) it includes. Thus, a full database for rural two-lane highway segments could include data for ADT, length, lane width, shoulder width, shoulder type, driveway density, horizontal curvature, vertical curvature, horizontal clearance, side slope, and grade. This list is illustrative; additional variables may be needed if they are believed to have some effect on crash frequency.

The second technique for database development is based on the desire to directly calibrate a base model. The database assembled for this purpose is referred to as a "base" database. This database includes only those segments for which their design attributes are considered to be typical. This database is likely to be a subset of a full database. In this instance, the full database would be screened to include only those segments whose attributes were considered to be typical. This technique is described by Washington et al. (*3*).

In practice, strict compliance with the need for every attribute to be equal to the designated typical value can result in there being a relatively small number of segments represented in the base database. This limitation is overcome by including additional segments in the database that have attributes that are *nearly* equal to the designated typical value. This refinement is most appropriate for continuous variables (e.g., lane width) as opposed to discrete variables (e.g., with/without left-turn bay). For example, the typical shoulder width may be designated as 8 ft, yet segments with shoulder widths in the range of 7 to 9 ft may also be included in the base database. The premise of this deviation is that the effect of the attribute, when within the allowable range, is either negligible or offsetting on an aggregate basis and that the increase in sample size will offset any added variability that is introduced.

Calibration Methods: Initial Calibration versus Re-Calibration

This section discusses the differences between initial model calibration and model re-calibration. All newly developed models of the type used to estimate crash frequency are initially calibrated using data for a specific location (or region). Initial calibration is usually based on a database that includes a large number of observations to ensure accurate calibration coefficients.

Re-calibration of an existing model is required when it is desired to use the model to estimate the safety of segments located in a region that is different from that represented in the existing model's calibration coefficients. The number of segment observations in the database used for re-calibration can vary, depending on the number of coefficients being re-calibrated. If only one coefficient is being re-calibrated, then the number of observations in the database is typically smaller than that needed for initial calibration. When all of the model coefficients are being re-calibrated, the number of observations needed for re-calibration is similar to that needed for initial calibration.

There is some disagreement in the literature regarding the number of model coefficients that should be adjusted during re-calibration. The re-calibration coefficients recommended by four different groups of researchers are listed in Table 6-1. The table cells indicate the recommendation made by each group of researchers.

Table 6-1. Candidate Re-Calibration Coefficients.

Researchers	Coefficients Recommended for Adjustment during Re-Calibration [1, 2]		
	a **(constant)**	***b*** **(ADT exponent)**	***k*** **(over dispersion)**
Harwood et al. (*2*)	Yes	no	no
Lyon et al. (*4*)	Yes	Yes	no
Sawalha & Sayed (*5*)	Optional	no	Yes
Washington et al. (*3*)	Yes	Yes	Yes

Notes:
1 - Coefficients *a* and *b* are defined in Equations 2, 3, and 4.
2 - Over-dispersion parameter *k* is used to describe the variance *V* of the prediction estimate $E[N]$ (i.e., $V[N] = E[N]^2/k$).

The re-calibration procedure developed by Harwood et al. (*2*) effectively calls for the adjustment of the constant *a*. It does not allow for adjustment to the ADT exponent *b* or the over-dispersion parameter *k*. The procedure is used with a full database. In application, data for a sample of segments representative of the new region are assembled. The initially calibrated Equation 1 is used to estimate the expected crash frequency for each segment *i*. A calibration factor *f* is then computed as the sum of the reported crashes for the sample divided by the sum of the expected crashes. The following equation illustrates this computation.

$$f = \frac{\sum Reported\ crashes_i}{\sum Expected\ crashes_i} \tag{6}$$

where,

f = calibration constant.

A new constant a is then computed as:

$$a_r = f\, a_i \tag{7}$$

where,

a_r = calibration constant for the re-calibrated model and
a_i = calibration constant in the initially calibrated model.

The advantage of this procedure is that it can be applied without the use of statistical analysis software.

There is some question about the accuracy of this procedure (*4, 5*). Lyon et al. (*4*) demonstrated that re-calibration based only on adjustment of the constant may not accurately reflect differences in design practice between regions because such differences may be correlated with traffic volume. Sawalha and Sayed (*5*) found that this procedure can yield a re-calibrated model whose fit to the data is not as good as that of the initial model.

The procedure described by Lyon et al. (*4*) and, subsequently by Washington et al. (*3*), is based on the re-calibration of both the constant and the ADT exponent. The coefficient values are quantified using regression based on maximum likelihood methods and a specified Poisson, or negative binomial, distribution for the dependent variable. The regression analysis is implemented in statistical analysis software.

An advantage of the procedure described by Lyon et al. (*4*) is that it provides a more accurate fit to the data for the full range of ADTs. A disadvantage of this procedure is that its implementation requires analysis software that is not readily available to most practitioners. Although not discussed by Lyon et al. (*4*), Washington et al. (*3*) also recommend use of the over-dispersion parameter obtained from the re-calibration analysis (in replacement of the over-dispersion parameter from the initial model).

A third procedure is described by Sawalha and Sayed (*5*). Like Harwood et al. (*2*), they recommend re-calibrating the constant, although it is not a requirement. And, like Washington et al. (*3*), they recommend the use of statistical analysis software to estimate best-fit values for this constant, as well as the over-dispersion parameter. They do not recognize a need to re-calibrate the ADT exponent.

It should be noted that this section discusses model calibration, as opposed to model validation. Model validation is a process of comparing a model calibrated with data from a specific region to additional data collected from different segments (or a different period of time) in that same region. None of the model coefficients are changed in this process; however, the quality of fit to the validation data is an indication of the accuracy and precision of the calibrated model.

Model Development Matrix

The concepts described in the previous two sections are combined in this section to illustrate how the calibration method relates to the calibration database. This relationship is shown in Table 6-2. It identifies the variables for which values are typically estimated during calibration, as related to the calibration method and the type of database assembled. For example, the initial calibration of a full model (i.e., Equation 5) requires the use of a full database to establish the value for each coefficient a, b, c_i, k. In contrast, the re-calibration of a model using a full database generally requires the c_i coefficients associated with the design variables x_i to remain fixed at the values established during initial calibration and only the values for a, b, or k are adjusted. Similarly, re-calibration of a safety prediction model (e.g., Equation 1) using a full database requires the AMFs associated with the design elements in the database to be used to compute appropriate AMF values for each observation and only the values for a, b, or k are adjusted.

Table 6-2. Association between Calibration Methods and Databases.

Calibration Method	Coefficients for Which Values are Estimated [1, 2]	
	Full Database	Base Database
Initial Calibration	a, b, c_i, k	a, b, k
Re-Calibration [3]	a, b, k (all c_i are held fixed)	a, b, k

Notes:
1 - Coefficients a and b are used in Equations 2, 3, and 4.
2 - Over-dispersion parameter k is used to describe the variance V of the prediction estimate $E[N]$ (i.e., $V[N] = E[N]^2/k$).
3 - As noted in Table 6-1, there is no consensus on whether b, k, or both should be re-calibrated.

When a base database is assembled, there is little difference between initial calibration and re-calibration as related to the treatment of the c_i coefficients. The initial calibration of a base model (i.e., Equation 2, 3, or 4) requires the use of a base database to establish the values for a, b, and k. The re-calibration of a base model also requires a base database; however, the analyst can choose to adjust any combination of the variables a, b, or k, depending on which of the procedures described in the previous section they find most plausible.

Base Models

This section describes base models derived by Bonneson et al. (*6*) for rural two-lane highway segments and intersections. These models are based on a review and synthesis of various models described in the research literature. The recommended models are summarized in the *Workbook* (*1*).

Base Model for Highway Segments

The base model in the *Workbook* for rural two-lane highway segments is:

$$E[N]_b = 0.000365\, Base\ ADT\ L \qquad \textbf{(8)}$$

where,

Base = injury (plus fatal) crash rate (= 0.20), crashes/million-vehicle-miles.

Comparison of Equation 8 with Equation 2 indicates that the constant *a* is equal to the product of the base crash rate *Base* and "0.000356." The latter constant represents a conversion factor to convert the ADT into "millions of vehicles per year." Also, the ADT exponent *b* is equal to 1.0 in Equation 8.

The base model in Equation 8 is applicable to two-lane highways that have typical geometric design conditions. These conditions are identified in the *Workbook* and are restated in column two of Table 6-3. With the exception of grade, the base conditions listed are most representative of roads classified as rural principal arterial highways. The terrain in Texas is expected to yield an average grade of 2 percent. The base model is combined with the AMFs described in the *Workbook* to yield the safety prediction model for rural two-lane highway segments.

Table 6-3. Base Conditions for Rural Two-Lane Highway Segments.

Characteristic	Base Condition	Sensitivity of Continuous Variables
Horizontal curve radius	tangent (no curve)	--
Spiral transition curve	not present	--
Grade	flat (0% grade)	-3 to +3%
Lane width	12 ft	11 to 12 ft
Paved Shoulder width	8 ft	7 to 9 ft
Shoulder rumble strips	not present	--
Centerline rumble strips	not present	--
Superelevation	not deficient	0 to 1% deficiency
Passing lane	not present	--
Horizontal clearance	30 ft	20 ft or more
Side slope	1V:4H	1V:2.5H or flatter
Utility pole density and offset	25 poles/mi, 30 ft average offset	any density, 20 ft offset or more
Relative bridge width	12 ft	2 ft or more
Driveway density	5 driveways/mi	0 to 12 driveways/mi

Several of the AMFs listed in Table 6-3 correspond to a characteristic that is represented as a continuous variable. These AMFs were evaluated for a range of values to examine the sensitivity of the AMF to a change in the continuous variable. The results of this sensitivity analysis are listed in column three of Table 6-3. The ranges listed in this column equate to variable values that yield an AMF between 0.95 and 1.05. For example, lane widths between 11 and 12 ft yield AMF values of 1.05 and 1.0, respectively. Lane widths less than 11 ft yield an AMF value in excess of 1.05. The ranges listed in column three provide some indication of the effect of the characteristic on crash risk. The wide range associated with some variables (e.g., side slope) suggests that the corresponding design characteristic has a modest effect on safety for most design situations.

Base Models for Intersections

The base model in the *Workbook* for rural intersections is:

$$E[N]_b = 0.000365\, Base\ (ADT_{major} + ADT_{minor})^b \tag{9}$$

where,

Base = injury (plus fatal) crash rate (see Tables 6-4 and 6-5), crashes/million-entering-vehicles.

Table 6-4. Crash Rates for Three-Leg Rural Intersections.

Control Mode	Major-Road Volume, veh/d	Crash Rate, injury + fatal crashes per million-entering-vehicles				
		Ratio of Minor-Road to Major-Road Volume				
		0.05	**0.10**	**0.15**	**0.20**	**0.25**
Unsignalized [1]	5000	0.10	0.14	0.16	0.18	0.19
	10,000	0.13	0.18	0.21	0.23	0.25
	15,000	0.15	0.20	0.24	0.26	0.28
	20,000	0.17	0.23	Intersection very likely to meet signal warrants		
	25,000	0.18				
Signalized	5000	0.08	0.11	0.14	0.16	0.17
	10,000	0.10	0.15	0.18	0.20	0.22
	15,000	0.12	0.17	0.21	0.23	0.25
	20,000	0.13	0.19	0.23	0.26	0.28
	25,000	0.14	0.20	0.25	0.28	0.30
	30,000	0.15	0.22	0.26	0.30	0.33
	40,000	0.17	0.24	0.29	0.33	0.36
	≥50,000	0.18	0.26	0.32	0.36	0.39

Note:
1 - Unsignalized intersections have an uncontrolled major road and a stop-controlled minor road.

Equation 9 was developed from the following, generalized equation for estimating intersection crash frequency:

$$E[N]_b = a \left(\frac{ADT_{major}}{1000}\right)^{b_1} \left(\frac{ADT_{minor}}{1000}\right)^{b_2} \tag{10}$$

The constant and coefficients listed in Table 6-6, if used in Equation 10, yield the same expected crash frequency as obtained from Equation 9 (with Table 6-4 or 6-5).

Table 6-5. Crash Rates for Four-Leg Rural Intersections.

Control Mode	Major-Road Volume, veh/d	Crash Rate, injury + fatal crashes per million-entering-vehicles				
		Ratio of Minor-Road to Major-Road Volume				
		0.10	**0.30**	**0.50**	**0.70**	**0.90**
Unsignalized [1]	5000	0.18	0.26	0.30	0.31	0.32
	10,000	0.20	0.30	0.34	0.36	0.36
	15,000	0.22	0.33	0.37	0.39	0.40
	20,000	0.23	Intersection very likely to meet signal warrants			
	25,000	0.25				
Signalized	5000	0.15	0.24	0.28	0.30	0.31
	10,000	0.17	0.28	0.32	0.35	0.36
	15,000	0.18	0.30	0.35	0.38	0.39
	20,000	0.20	0.32	0.37	0.40	0.42
	25,000	0.20	0.33	0.39	0.42	0.44
	30,000	0.21	0.35	0.41	0.44	0.45
	40,000	0.23	0.37	0.43	0.46	0.48
	≥50,000	0.24	0.38	0.45	0.49	0.50

Note:
1 - Unsignalized intersections have an uncontrolled major road and a stop-controlled minor road.

Table 6-6. Intersection Base Model Coefficients.

Control Mode	Three Intersection Legs			Four Intersection Legs		
	a	b_1	b_2	a	b_1	b_2
Unsignalized [1]	0.0973	0.863	0.497	0.166	0.692	0.514
Signalized	0.0973	0.782	0.577	0.166	0.611	0.595

Note:
1 - Unsignalized intersections have an uncontrolled major road and a stop-controlled minor road.

Either of the two aforementioned base models is applicable to intersections that have typical geometric design conditions. These conditions are identified in the *Workbook* and are restated in columns two and four of Table 6-7. The base conditions listed are representative of intersections on roads classified as rural principal arterial highways.

AMFs are available in the *Workbook* to assess each of the design characteristics listed in Table 6-7. For routine applications, the base model is combined with the AMFs in the safety prediction model (i.e., Equation 1) to estimate the crash frequency for a particular intersection that has design elements not in agreement with base conditions.

Table 6-7. Base Conditions for Rural Intersections.

Characteristic	Two-Way Stop Control		Signal Control	
	Base Condition	Sensitivity of Continuous Variables	Base Condition	Sensitivity of Continuous Variables
Left-turn lanes on major road	not present	--	present	--
Right-turn lanes on major road	not present	--	not present	--
Number of lanes on major road	2	--	2	--
Number of lanes on minor road	2	--	2	--
Shoulder width on major road	8 ft	6 to 10 ft	not applicable	--
Median presence on major road	not present	--	not applicable	--
Alignment skew angle	no skew	-2 to +2 degrees	no skew	any angle
Sight distance restrictions	none	up to 1 quadrant	not applicable	--
Driveways within 250 ft of int.	0 driveways	0 to 1 driveways	3 driveways	2 to 4 driveways
Truck presence	9 % trucks	7 to 11%	9 % trucks	7 to 11%

Several of the AMFs in Table 6-7 correspond to a characteristic that is represented as a continuous variable. These AMFs were evaluated for a range of values to examine the sensitivity of the AMF to a change in the continuous variable. The results of this sensitivity analysis are listed in columns three and five of Table 6-7. The ranges listed in these columns equate to variable values that yield an AMF between 0.95 and 1.05. For example, shoulder widths between 6 and 10 ft yield AMF values of about 1.05 and 0.95, respectively. Shoulder widths less than 6 ft yield an AMF value in excess of 1.05. The ranges listed provide some indication of the effect of the characteristic on crash risk.

Texas Highway Data

The geometric and traffic attributes for the Texas state highway system were obtained from the Texas Reference Marker (TRM) system highway database. This database is maintained by TxDOT and contains data for several thousand highway segments, each of which is described in terms of its geometry, traffic, and location attributes. In fact, about 140 attributes are used to describe each highway segment in TRM. However, only about 20 attributes are used to describe segment geometry or traffic characteristics, and about 10 more attributes are used to describe road name and physical location. The remaining attributes describe administrative designations and road-management-related information that is not particularly helpful for safety analysis.

A highway segment in TRM is defined as a length of road along which no one attribute changes. A change in any one attribute dictates the end of one segment and the start of a new segment. The average tangent segment length in TRM is about 0.4 mi.

As noted in the discussion associated with Table 6-3, several segment characteristics are needed in the calibration database. Those characteristics that are available in the TRM database are identified in the last column of Table 6-8.

Table 6-8. TRM Data for Rural Two-Lane Highway Segments and Intersections.

Facility Component	Characteristic	Availability in TRM Database
Highway segment	Horizontal curve radius	Yes
	Spiral transition curve	Yes
	Grade	No
	Lane width	Extracted from other data.
	Shoulder width	Yes
	Shoulder rumble strips	No
	Centerline rumble strips	No
	Superelevation	No
	Passing lane	Extracted from other data.
	Horizontal clearance	No
	Side slope	No
	Utility pole density & offset	No
	Relative bridge width	No
	Driveway density	No
Intersection	Left-turn lanes on major road	No
	Right-turn lanes on major road	No
	Number of lanes on major road	Yes
	Number of lanes on minor road	Yes [1]
	Shoulder width on major road	Yes
	Median presence on major road	Yes
	Alignment skew angle	Yes
	Sight distance restrictions	No
	Driveways within 250 ft of intersection	No
	Truck presence	Yes [1]

Note:
1 - Data are available for the minor road only if it is a state-maintained highway.

The information in Table 6-8 indicates that data for several segment characteristics are not available in the TRM database. The consequences of not having these data can vary, depending on whether a model is undergoing initial calibration or re-calibration. For an initial calibration activity, the "missing" characteristic data would need to be collected by site survey to ensure that their influence on safety was accurately quantified. For a re-calibration activity, it may be possible to assume that the base condition values are representative of the collective set of segments. This assumption results in the corresponding AMF value equaling 1.0, which effectively cancels out the effect of the characteristic in the calibration process.

Intersections are defined in TRM as point entities, having a unique reference number for each intersecting roadway. The combination of data for both intersecting roadways in the vicinity of their point of intersection yields the desired characteristics for the intersection. Data in TRM only apply to roadways on the state highway system. Hence, TRM only has information about both intersecting roadways when these two roadways are part of the state highway system.

As noted in the discussion associated with Table 6-7, several intersection characteristics are needed in the calibration database. Those characteristics that are available in the TRM database are identified in the last column of Table 6-8.

Texas Regions

One of the goals of the calibration activity was to determine whether the calibration coefficients varied within Texas on a regional level. To facilitate the need for possible regional calibration, the state was divided into six regions of similar geography and climate. The regional boundaries were defined to be consistent with the TxDOT district boundaries. These regions are shown in Figure 6-1.

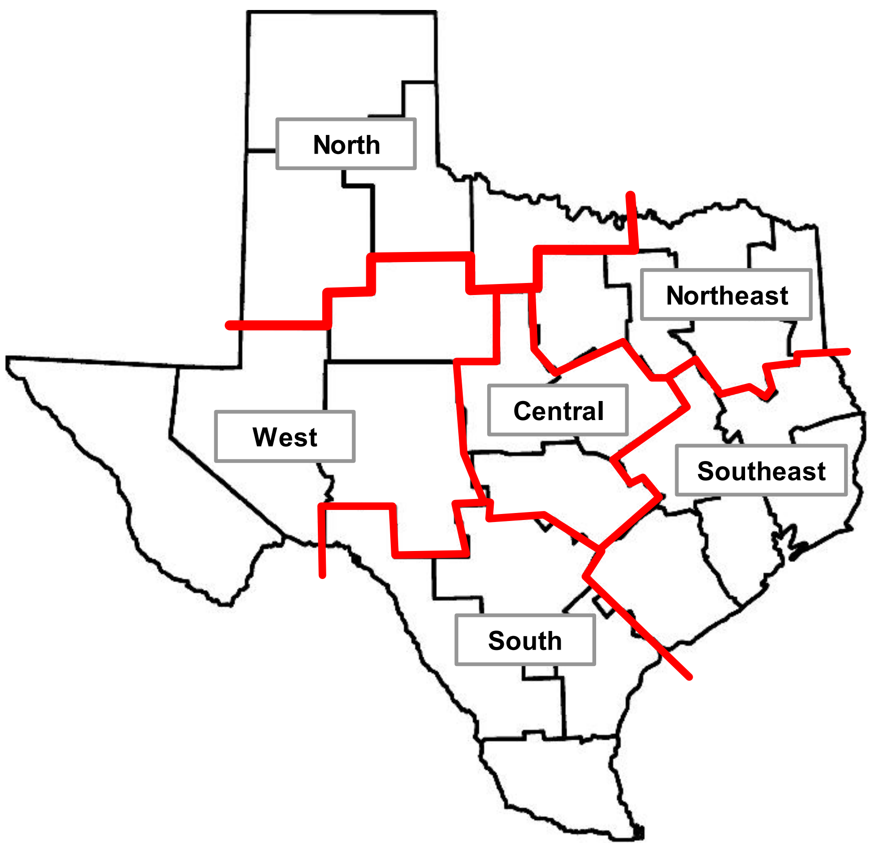


Figure 6-1. Regions Used for Base Model Calibration.

As indicated in Figure 6-1, the state of Texas was divided into six regions: north, northeast, southeast, south, west, and central Texas. The TxDOT districts in each of these regions are listed in Table 6-9.

Table 6-9. Distribution of TxDOT Districts to Regions.

TxDOT District by Region [1]					
North	**Northeast**	**Southeast**	**South**	**West**	**Central**
Wichita Falls (3) Amarillo (4) Lubbock (5) Childress (25)	Paris (1) Fort Worth (2) Tyler (10) Dallas (18) Atlanta (19)	Lufkin (11) Houston (12) Yoakum (13) Bryan (17) Beaumont (20)	San Antonio (15) Corpus Christi (16) Pharr (21) Laredo (22)	Odessa (6) San Angelo (7) Abilene (8) El Paso (24)	Waco (9) Austin (14) Brownwood (23)

Note:
1 - Numbers in parentheses are the district number, as assigned by TxDOT.

CALIBRATION METHODOLOGY

This part of the chapter describes the methodology used to re-calibrate the rural two-lane highway safety prediction models described in the *Workbook* (*1*). The objective of this activity was to adjust one or more model coefficients such that the resulting model provides an accurate estimate of the frequency of crashes on highway segments or intersections in Texas. The sensitivity of the model coefficients to various regions in Texas was also examined to determine if unique calibration coefficients were needed.

This part is divided into five sections. The first section describes the method used to re-calibrate the highway segment model. The second section describes the method used to re-calibrate the intersection model. The third section describes the technique used to quantify the calibration coefficients. The fourth section describes the technique used to adust the over-dispersion parameter for bias due to a small sample size. The last section summarizes the goodness-of-fit statistics used to assess the accuracy of the re-calibrated model.

Methodology for Highway Segments

A base database was assembled from the TRM database for segment re-calibration. In this type of database, only those segments that have characteristics equal to the base conditions were included in the database. However, as previously noted in the discussion associated with Table 6-8, attributes associated with some of the relevant characteristics are not included in the TRM database and cannot be used to form a subset of the database. Measurement of these characteristics in the field was not a feasible option. Fortunately, the sensitivity analysis described in Table 6-3 suggested that most of the missing characteristics had a modest effect on safety. Based on this observation, it was determined that the base condition values listed in Table 6-3 for the missing characteristics could be assumed to be representative of the collective set of segments in the database such that the re-calibrated base model would reflect typical conditions with reasonable accuracy, provided that the database included a large number of segments.

There were three refinements to the approach described in the preceding paragraph. First, to ensure that the segments extracted from the TRM database had superelevation, horizontal clearance, side slope, pole offset, and bridge width that were consistent with the base conditions, only those segments designated as principal arterial highways were included in the database.

Second, the base condition for grade is 0.0 percent; however, it is rationalized that most rural two-lane highways in Texas are not flat and that a nominal grade of 2 percent (sign does not matter for the grade AMF) is likely reflected in the segments extracted from TRM. Thus, the grade AMF from the *Workbook* (*1*) was used to estimate the AMF value corresponding to a grade of 2 percent and this value was used for all segments in the analysis.

Third, given the highly variable nature of driveway density among segments, it was decided that the best approach to account for this effect was to exclude all driveway-related crashes from the database. Then, the driveway density AMF was used to estimate the AMF value corresponding to 0.0 driveways per mile and this value was used for all segments in the analysis.

Methodology for Intersections

In contrast to the methodology used for segments, a full database was assembled from the TRM database for intersection model re-calibration. In this type of database, all of the relevant intersection characteristics are represented as attributes in the database. A base database could not be formed from the full database, as was done for highway segments, because several of the desired intersection characteristics are not included in the TRM database.

The full database was assembled using data from the TRM database combined with data obtained from aerial photography. These photos were acquired for the time period coincident with the crash data (i.e., circa 2000) and used to determine turn bay presence and intersection traffic control. They were also used to quantify the number of driveways near the intersection and to make some judgment about sight distance adequacy. Given the time-consuming nature of data collection from aerial photos, it was not feasible to selectively acquire only data for "typical" intersections. Thus, a base database could not be formed using manual methods because of time limitations.

As noted in the discussion associated with Table 6-2, re-calibration of the base model component of a safety prediction model using a full database requires the use of AMFs. Specifically, the AMFs associated with the design elements in the database are used to compute appropriate AMF values for each intersection. These values are then used in the full model during re-calibration such that only the values for *a*, *b*, and *k* are adjusted.

Re-Calibration Technique

The re-calibration activity consisted of a two-step process using statistical analysis software that employs maximum likelihood methods based on a negative binomial distribution of crash frequency. In the initial analysis, the constant *a*, the ADT exponent *b*, and the over-dispersion parameter *k* were quantified. If *b* was found to be significantly different from its counterpart in the existing model, then the new value was used in replacement of the existing coefficient. If *b* was not significantly different, then the coefficient in the existing model was retained. After a decision was made about coefficient *b*, a similar process was followed to define coefficient *a*. If the new value for coefficient *b* was used, then the new value for *a* was also used. If the existing value for *b* was retained, then the decision regarding the new *a* coefficient was based on whether it was significantly different from the existing coefficient.

During the second step, differences among Texas regions were explored. A unique constant a_r was inserted in the model for each region and used to compare each region to the combined set of remaining regions a_o. In this manner, if a_r was found to be significantly different from a_o then a regional influence was assumed to exist and the subject region variable was retained in the model.

Over-Dispersion Parameter Adjustment

It was assumed that crash occurrence on a segment (or at an intersection) is Poisson distributed and that the distribution of the mean crash frequency for a group of similar segments is gamma distributed. In this manner, the distribution of crashes for a group of similar segments can be described by the negative binomial distribution. The variance of this distribution is:

$$V[X] = y\,E[N] + \frac{(y\,E[N])^2}{k\,L} \qquad \textbf{(11)}$$

where,

X = reported crash count for y years, crashes;
y = time interval during which X crashes were reported, yr; and
k = over-dispersion parameter, mi^{-1}.

Equation 11 includes a variable for the length of the segment. As demonstrated by Hauer (*7*), this variable should be added to ensure that the regression model coefficients are not biased by exceptionally short segments. When Equation 11 is applied to intersections, the variable for segment length L is set to 1.0.

Research by Lord (*8*) has indicated that databases with low sample mean values and small sample size may not exhibit the variability associated with crash data from multiple, similar highway segments, as suggested by Equation 11. Rather, the regression model may "over-explain" some of the random variability in a small database, or the low sample mean may introduce an instability in the model coefficients (*8*). These issues are especially present in model re-calibration because of the inherent use of small databases and the desire to re-calibrate several model coefficients. To minimize the adverse impact of these issues, the number of coefficients that are re-calibrated is kept to a reasonable minimum. Also, the over-dispersion parameter k obtained from the regression analysis is adjusted downward, such that it yields a more reliable estimate of the variance of the crash distribution. The adjustment technique used is described in the remainder of this section.

Lord (*8*) explored the effect of sample size on the variability of crash data through the use of simulation. Specifically, he simulated the crash frequency for three database sizes (i.e., 50, 100, and 1000 sites), each with three different average crash frequencies (i.e., 0.5, 1.0, and 10 crashes per site), and three over-dispersion parameters (i.e., 0.5, 1, and 2). One complete set of simulations consisted of the nine combinations of average crash frequency and over-dispersion parameter, each simulated once for each site yielding a total of 10,350 (= 3×3×[50 + 100 + 1000]) site crash frequency estimates. A total of 30 replications was conducted yielding 310,500 site estimates. A maximum likelihood technique was used to estimate the over-dispersion parameter for each of the 27 combinations. This process was repeated 30 times to yield 30 estimates of k for each of the 27 combinations. The average value of k for each combination is shown in Figure 6-2.

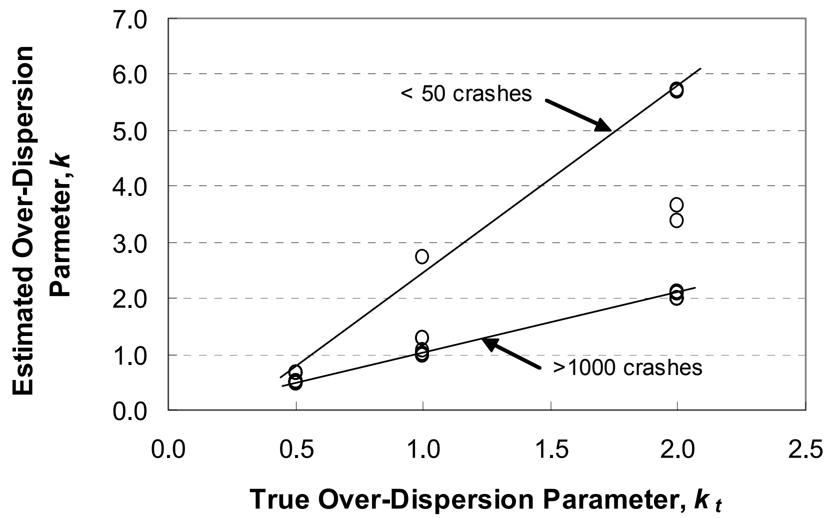


Figure 6-2. Simulation Results for the Over-Dispersion Parameter.

As shown in Figure 6-2, the estimated over-dispersion parameter for a given database was about equal to the specified (i.e., true) over-dispersion parameter, *provided* that there were 1000 or more crashes in the database. However, the estimated over-dispersion parameter was larger than the true value when the number of crashes was less than 1000. The following relationship was derived to relate the estimated and true over-dispersion parameters based on observed trends in the data:

$$k_r = k_t + \frac{17.2 \; k_t^2}{(n - p)\, m} \qquad \textbf{(12)}$$

where,

k_r = estimated over-dispersion parameter obtained from database analysis;
k_t = true over-dispersion parameter;
n = number of observations (i.e., segments or intersections in database);
p = number of model variables; and
m = average number of crashes per observation (= total crashes in database / n).

The constant "17.2" in Equation 12 represents an empirical adjustment derived through weighted regression analysis ($s_k = 2.45$, $p = 0.0001$). Equation 12 was algebraically manipulated to yield the following relationship for estimating the true over-dispersion parameter, given the other variables as input values.

$$k_t = \frac{\sqrt{(n - p)^2\, m^2 + 69\, k_r\, L.\, (n - p)\, m} - (n - p)\, m}{34.5 \; L.} \qquad \textbf{(13)}$$

where,

$L.$ = average segment length for all n observations (= 1.0 for intersection applications).

Equation 13 was used to estimate the true over-dispersion parameter for each of the models described in this chapter. All subsequent references to the over-dispersion parameter k in this chapter denote the estimated true parameter obtained from Equation 13 (i.e., hereafter, $k = k_t$) .

Goodness-of-Fit Statistics

Several statistics are available for assessing model fit and the significance of model coefficients. One measure of model fit is the Pearson χ^2 statistic. This statistic is calculated as:

$$\chi^2 = \sum_{i=1}^{n} \frac{(X_i - y\, E[N]_i)^2}{Var[X]_i} \tag{14}$$

where,
n = number of observations.

McCullagh and Nelder (*9*) indicate that this statistic follows the χ^2 distribution with n-p degrees of freedom. This statistic is asymptotic to the χ^2 distribution for larger sample sizes and exact for normally distributed error structures. As noted by McCullagh and Nelder, this statistic is not well defined in terms of minimum sample size when applied to non-normal distributions; therefore, it probably should not be used as an absolute measure of model significance. The Pearson χ^2 statistic is available from the non-linear regression procedure (NLIN) in the SAS statistical software as the weighted sum of squares for the residual.

The root mean square error s_e is a useful statistic for assessing the precision of the model estimate. It represents the standard deviation of the estimate when each independent variable is at its mean value. This statistic can be computed as:

$$s_e = \frac{1}{y} \sqrt{\frac{\sum_{i=1}^{n} (X_i - y\, E[N]_i)^2}{n - p}} \tag{15}$$

where,
s_e = root mean square error of the model estimate, crashes/yr.

Another, more subjective, measure of model fit can be obtained from a graphical plot of the Pearson residuals versus the expected value of the dependent variable (e.g., *E[N]*). This type of plot can be used to graphically assess the predictive capability of the model. A well-fitting model would have the residuals symmetrically centered around zero over the full range of the dependent variable, most clustered near zero, and with a spread ranging from about -3.0 to + 3.0. The Pearson residual PR_i for segment i can be computed as:

$$PR_i = (X_i - y\, E[N]_i) \sqrt{\frac{1}{Var[X]_i}} \tag{16}$$

Another measure of fit is the scale parameter φ. This parameter was noted by McCullagh and Nelder (*9*) to be a useful statistic for assessing the amount of variation in the observed data, relative to the specified distribution. This statistic can be calculated by dividing Equation 14 by the quantity *n-p*. It is also available from NLIN as the weighted mean square for the residual. A scale parameter near 1.0 indicates that the assumed distribution of the dependent variable is approximately equivalent to that found in the data (i.e., negative binomial).

Another measure of model fit is the coefficient of determination R^2. This statistic is commonly used for normally distributed residuals. However, it has some useful interpretation when applied to data from other distributions (*10*). This statistic can be computed as:

$$R^2 = 1 - \frac{SSE}{SST} \tag{17}$$

with,

$$SSE = \sum_{i=1}^{n} w_i \, (X_i - y\, E[N]_i)^2 \tag{18}$$

$$SST = \sum_{i=1}^{n} w_i \, (X_i - X.)^2 \tag{19}$$

where,

$X.$ = average crash frequency for all n observations (= $\Sigma\, [X_i] / n$).

The R^2 statistic indicates the percentage of the variability in the crash data that is explained by the regression model. A value near 0.0 suggests a lack of correlation; a value of 1.0 suggests that the model estimates are in perfect agreement with the observed crash frequency.

The last measure of model fit is the dispersion-parameter-based coefficient of determination R_k^2. This statistic was developed by Miaou (*11*) for use with data that exhibit a negative binomial distribution. It is computed as:

$$R_k^2 = 1 - \frac{k_{null}}{k} \tag{20}$$

where,

k_{null} = over-dispersion parameter based on the variance in the observed crash frequency.

The over-dispersion parameters in Equation 20 can be estimated using a variety of methods (*5, 8*). The null over-dispersion parameter k_{null} represents the dispersion in the observed crash frequency, relative to the overall average crash frequency for all segments. This parameter was obtained using a null model formulation (i.e., a model with no independent variables but with the same error distribution, link function, and offset as used with the full or base model) in a subsequent regression analysis.

The R_k^2 statistic indicates the proportion of the variability due to systematic sources that is explained by the model. It does not include the random variability component and, thereby, is not as limited as R^2. The R_k^2 statistic has a value of 0.0 when no independent variables are included in the model and a value of 1.0 when these variables explain all of the systematic variability in the data.

HIGHWAY SEGMENT MODEL

This part of the chapter describes the calibration of the highway segment model. Initially, the process used to select the highway segment sites is identified. Then, the characteristics of the sites are summarized. Next, the techniques used to assemble the crash data for each site are described. Finally, the findings from the model calibration are discussed.

Site Selection and Data Collection

The highway segments used to calibrate the safety prediction model were required to satisfy the following criteria:

- Cross section: undivided, two through lanes, no median;
- Area type: rural;
- Functional class: principal arterial;
- Segment length: 0.1 to 1.0 mi;
- Intersection presence: no intersections;
- Curve presence: no curves;
- Lane width: 11 to 12 ft;
- Shoulder width: 7 to 9 ft; and
- Shoulder type: paved.

All segments were also required to have a minimum length of 0.1 mi. The value of 0.1 mi reflects a recognition of the precision of crash location in the Texas Department of Public Safety (DPS) crash database. A maximum length of 1.0 mi was imposed as a practical upper limit for segment length; however, this limit had negligible effect on sample size because only a few segments were found to have a length of 1.0 mi or more.

Experience with the crash database indicated that many short segments tended to have no reported crashes during a period of several years. This trend was also exhibited by segments that had very low traffic volume. When one of these segments is associated with one or more crashes, it has a tendency to exhibit undue leverage on model coefficients and increases the Pearson χ^2 statistic in a disproportionate manner, relative to other segments. To avoid these issues, the segments were screened to include only those with a minimum level of exposure.

To ensure that segments with a low level of exposure and with one or more reported crashes did not exert an unreasonable leverage on the model coefficients, it was rationalized that the corresponding prediction ratio PR_i for a segment should not exceed a value of 3.0. In recognition of this desired limit, Equation 16 was used to derive the following equation for computing the minimum segment exposure:

$$E_{\min} = \frac{PR^2 + 2 - \sqrt{(PR^2 + 2)^2 - 4}}{2\ Base\ y} \qquad \textbf{(21)}$$

where,

E_{min} = minimum segment exposure associated with prediction ratio; PR = 3.0, mvm; and

$Base$ = injury (plus fatal) crash rate, crashes/mvm (million-vehicle-miles).

Equation 21 is based on the conservative assumption that $V[X]_i$ is equal to $E[N]$ and that X_i is equal to 1.0 crash. When the prediction ratio PR is set to 3.0, the estimated injury (plus fatal) crash rate $Base$ is 0.23 cr/mvm, and the time period for the crash data y is 3 years, the resulting minimum exposure is 0.13 mvm. Equation 22 was used to compute the exposure for all candidate segments. Only those exceeding 0.13 mvm were included in the database.

$$E = 0.000365\, ADT\ L \qquad \textbf{(22)}$$

where,

E = segment exposure, mvm.

Site Characteristics

Traffic and geometry data were identified for several thousand highway segments using the TRM database for year 2003, the first year for which a complete database was available. All total, 567 segments satisfied the site selection criteria. This number constitutes less than 3 percent of the total number of rural two-lane highway curves in the state. Selected characteristics are provided for these segments in Table 6-10.

Table 6-10. Summary Characteristics for Highway Segments.

Region	Total Segment Pairs	Segment Length, mi	Segment Length Range, mi		Volume Range, veh/d	
			Minimum	Maximum	Minimum	Maximum
Central	58	22.2	0.13	0.9	1600	6700
Northeast	109	38.0	0.10	0.9	1800	19,200
North	76	39.8	0.15	1.0	650	4500
Southeast	158	63.1	0.10	1.0	930	23,700
South	91	35.2	0.11	0.9	970	7500
West	75	35.4	0.15	1.0	730	3600
Overall:	567	233.7	0.10	1.0	650	23,700

As indicated in Table 6-10, the segments total 233.7 mi of rural highway. Segment length ranged from 0.1 mi to 1.0 mi, with an average length of 0.41 mi. The ADT ranges from 650 to 23,700 veh/d for the collective set of segments.

Data Collection

Crash data were identified for each segment using the DPS database. Three years of crash data, corresponding to years 1999, 2000, and 2001 were identified for each segment. The ADT for each of these three years was obtained from the TRM database and averaged to obtain one ADT for each segment. Crash data prior to 1999 were not used due to the increase in speed limit that occurred on many Texas highways in 1997 and 1998. For reasons cited in a previous section discussing the calibration methodology, crashes that were associated with intersections and driveways were excluded from the database.

Data Analysis

This section is divided into two subsections. The first subsection summarizes the crash data at the selected study sites. The second subsection describes the formulation of the calibration model and summarizes the statistical analysis methods used to calibrate it.

Database Summary

The crash data for each of the 567 segments are summarized in Table 6-11. These segments were associated with 405 crashes, of which 214 were injury or fatal crashes. The trends in crash rate are provided in the last two columns of the table.

Table 6-11. Crash Data Summary for Highway Segments.

Region	Exposure,[1] mvm	Crashes / 3 years			Crash Rate, cr/mvm	
		PDO [2]	F+I [3]	Total	F+I [3]	Total
Central	23.9	21	16	37	0.22	0.52
Northeast	94.1	59	80	139	0.28	0.49
North	23.5	13	9	22	0.13	0.31
Southeast	96.1	66	78	144	0.27	0.50
South	45.1	19	22	41	0.16	0.30
West	19.9	13	9	22	0.15	0.37
Overall:	302.6	191	214	405	0.24	0.45

Notes:
1 - mvm: million-vehicle-miles per year.
2 - PDO: property-damage-only crashes.
3 - F+I: injury plus fatal crashes.

The *Workbook* (*1*) indicates that the injury (plus fatal) crash rate for rural two-lane highways is 0.20 crashes/mvm. Comparison of this value with the injury (plus fatal) crash rates in Table 6-11 suggests that the various regions in Texas are consistent with this range. The variation in crash rates may be due to regional influences; however, it may also be due to random variation in crash data. The significance of the differences in crash rate in Table 6-11 are explored further in the next section.

Model Development and Statistical Analysis Methods

This section describes the segment calibration model and the methods used to re-calibrate it. The form of this model is:

$$E[N] = a \left(\frac{ADT}{1000}\right)^{b} L \times AMF_{g} \times AMF_{dd} \qquad \textbf{(23)}$$

where,

a, b = calibration coefficients;
AMF_g = grade accident modification factor for 2 percent grade ($= e^{0.016\,g} = 1.033$); and
AMF_{dd} = driveway density accident modification factor for 0 driveways/mi ($= e^{0.007\,(Dd-5)} = 0.966$).

Equation 23 is effectively equal to the base model, but it has two AMFs to account for variables that are not included in the TRM database. As noted in the discussion associated with Equation 8, the existing base model calibration coefficients in the *Workbook* are equivalent to a = 0.073 (= 0.365 ×0.20) and b = 1.0. As discussed in the section titled Calibration Methodology, the two AMFs were added to the model to eliminate bias in the calibration coefficients.

The Generalized Modeling procedure (GENMOD) in SAS was used to automate the regression analysis. This procedure estimates model coefficients using maximum-likelihood methods. GENMOD is particularly well suited to the analysis of models with additive terms that are either continuous or categorical. GENMOD automates the estimation of the over-dispersion parameter k when the variance function (i.e., Equation 11) does *not* include the segment length variable. However, given that the variance function does include segment length when applied to highway segments, the GENMOD procedure was modified using its Variance option such that Equation 11 was directly specified in the GENMOD code.

To estimate the over-dispersion parameter, two GENMOD procedures were applied in succession for each analysis iteration. In the first application, GENMOD used the calibration model and the specified variance function (i.e., Equation 11) with a specified value of k. In the second application, GENMOD used the natural log of the predicted crash frequency from the first application as an offset variable and the "internal" variance function (i.e., this function allows Equation 11 to be explicitly specified). The best-estimate of k from the second GENMOD application was divided by the average segment length and the quotient formed a new estimate of k for use in Equation 11. This new estimate was used to replace the value of k specified in the first application. This sequential use of two GENMOD procedures was repeated in an iterative manner until convergence was achieved between the k value used in the first GENMOD application and that obtained from the second GENMOD application. Convergence was typically achieved in two iterations. The value of k obtained from this iterative method was used in Equation 13 to estimate the true over-dispersion parameter.

Model Calibration

The results of the model calibration are presented in Table 6-12. Calibration of this model focused on injury (plus fatal) crash frequency. The Pearson χ^2 statistic for the model is 497, and the degrees of freedom are 565 (= $n - p$ = 567 − 2). As this statistic is less than $\chi^2_{0.05,\,565}$ (= 621), the hypothesis that the model fits the data cannot be rejected; however, it is noted that there is slightly less variability in the data than suggested by Equation 11. The R^2 for the model is 0.26. An alternative measure of model fit that is better suited to the negative binomial distribution is $R_k^{\,2}$, as developed by Miaou (*11*). The R_k^2 for the calibrated model is 0.90. This statistic indicates that about 90 percent of the variability due to systematic sources is explained by the model.

Table 6-12. Multivariate Model Statistical Description–Highway Segments.

Model Statistics		Value		
	R^2 (R_k^2):	0.26 (0.90)		
	Scale Parameter φ:	0.88		
	Pearson χ^2:	497 ($\chi^2_{0.05,\,565}$ = 621)		
	Over-Dispersion Parameter k:	15.3 mi^{-1}		
	Observations n_o:	567 segments (214 injury + fatal crashes in 3 years)		
	Standard Deviation s_e:	±0.22 crashes/yr		
Range of Model Variables				
Variable	**Variable Name**	**Units**	**Minimum**	**Maximum**
ADT	Segment ADT	veh/d	650	23,700
L	Segment length	miles	0.10	1.0
Calibrated Coefficient Values				
Variable	**Definition**	**Value**	**Std. Dev.**	**t-statistic**
a	Constant	0.0537	0.0097	5.5
b	Effect of segment ADT	1.30	0.10	12.6

The ADT coefficient *b* has a value of 1.30 which is much larger than 1.0, as assumed in the existing base model. In fact, the 95th percentile confidence interval for *b* is 1.10 to 1.50 which excludes 1.0. Thus, the coefficient values for both *a* and *b* were retained in the base model.

As a second step in the process, the base model was used to determine if there are differences in crash trend among the various Texas regions. Indicator variables for *a* were used to facilitate this analysis. The findings indicated that there are no significant differences among regions and that the coefficients in Table 6-12 can be used for all regions.

The coefficients in Table 6-12 can be combined with Equation 23 to obtain the calibrated crash estimation base model. The form of this model is:

$$E[N]_b = 0.0537 \left(\frac{ADT}{1000}\right)^{1.30} L \tag{24}$$

Sensitivity Analysis

The relationship between segment injury (plus fatal) crash frequency and traffic volume is shown in Figure 6-3. The thick trend line corresponds to the re-calibrated model (i.e., Equation 24). The thin dashed trend line corresponds to the existing base model in the *Workbook*. The data points represent the observed crash frequency for the 567 highway segments used to re-calibrate the base model. Each data point shown represents the average ADT and average observed crash frequency for a group of ten segments. The data were sorted by ADT to form groups of segments with similar ADT. The purpose of this grouping was to reduce the number of data points shown and, thereby, to facilitate an examination of trend in the data. This averaging technique was used to improve the graphic portrayal of trend for the sensitivity analysis; the individual segment observations were used for model calibration.

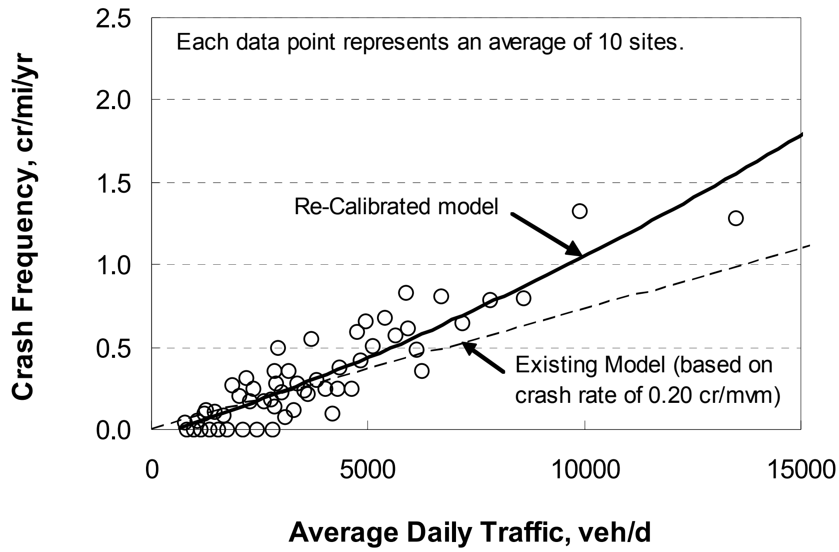


Figure 6-3. Relationship between Traffic Volume and Segment Crash Frequency.

The two trend lines in Figure 6-3 indicate that the re-calibrated model and the existing model yield similar estimates of crash frequency for ADTs of 5000 veh/d or less. However, at ADTs above 5000 veh/d, the estimated crash frequency for the re-calibrated model tends to exceed that obtained from the existing model. This trend suggests that the crash rate for higher ADT highways exceeds 0.20 cr/mvm.

INTERSECTION MODELS

This part of the chapter describes the re-calibration of three intersection models. Initially, the process used to select the intersection sites is identified. Then, the characteristics of the sites are summarized. Next, the techniques used to assemble the crash data for each site are described. Finally, the findings from the model calibration are discussed.

Site Selection and Data Collection

The intersections used to calibrate the safety prediction model were required to satisfy the following criteria:

- Cross section: undivided, two through lanes (one lane on each approach), no median;
- Area type: rural;

- Curve presence: radius of 5730 or larger on the major road at the intersection; and
- Intersection spacing: 650 ft or more between adjacent intersections.

All intersections were required to have both of the intersecting highways on the state highway system. This requirement ensured that data for both highways could be found in the TRM database. Also, the intersections selected were required to be within the regions of Texas for which high-resolution aerial photography was available on the Internet.

Guidance provided by Harwood et al. (*2*) was used to establish the desired sample size. Specifically, they indicate that a desirable minimum number of 25 four-leg signalized intersections are needed for model re-calibration. They also suggest that the desirable minimum number of unsignalized three-leg intersections is 100 and the desirable minimum number of four-leg unsignalized intersections is 100. Based on typical crash frequencies for each of these intersection types, the minimum number of sites equates to a minimum of 150 reported crashes for each intersection type. Harwood et al. (*2*) did not define a desired minimum number of three-leg signalized intersections.

Site Characteristics

Traffic and geometry data were identified for several thousand highway intersections using the TRM database for year 2003, the first year for which a complete database was available. A total of 1393 intersections were identified that satisfied the selection criteria (except the need for high-resolution aerial photography). A review of several sources of aerial photography on the Internet revealed that only about 20 percent of the identified intersections were in areas that the photography was of sufficient resolution that lane markings could be identified. Of those intersections that were in areas of high-resolution photography, about 50 percent were found to be undesirable for a variety of reasons (e.g., lane markings suggest a type of control that is contradictory to that found in the crash data, intersection was in a small town, intersection had unusual geometry, etc.). As a result of the screening process, a total of 208 intersections were identified that satisfied all of the site selection criteria. Selected characteristics are provided for these intersections in Table 6-13.

Table 6-13. Summary Characteristics for Intersections.

Legs	Control Type	Total Intersections	Major-Road Volume, veh/d		Minor-Road Volume, veh/d	
			Minimum	Maximum	Minimum	Maximum
Three	Two-way stop	123	80	19,200	30	6200
Four	Signal	20	4300	19,200	1800	11,100
	Two-way stop	65	120	12,800	30	4000

The number of three-leg unsignalized intersections listed in Table 6-13 exceeds the desired minimum number, as discussed in the previous section. However, the number of four-leg signalized and unsignalized intersections is less than the desired minimum number. Nevertheless, as the discussion in a subsequent section will show, this number was believed sufficient to re-calibrate the corresponding base model.

The small sample size imposed two limitations. First, there were too few intersections to allow for an investigation of regional trends in the crash data. Second, only seven three-leg signalized intersections were found to satisfy the site selection criteria. This number was determined to be too small to accurately re-calibrate the three-leg signalized intersection model.

Data Collection

Crash data were identified for each intersection using the DPS database. Three years of crash data, corresponding to years 1999, 2000, and 2001 were identified for each intersection. The ADT for each of these three years was obtained from the TRM database and averaged to obtain one ADT for each intersecting roadway. Only those crashes that occurred within 325 ft of the intersection, and that were identified as "intersection-related," were included in the database. A distance of 325 ft was used in recognition of the 0.1 mi crash location precision used by the DPS. Use of a shorter distance significantly increased the probability that some intersection-related crashes would not be identified for any one intersection.

Data Analysis

This section is divided into two subsections. The first subsection summarizes the crash data at the selected study sites. The second subsection describes the formulation of the calibration model and summarizes the statistical analysis methods used to calibrate it.

Database Summary

The crash data for each of the 208 intersections are summarized in Table 6-14. These intersections were associated with 430 crashes, of which 282 were injury or fatal crashes. The trends in crash rate are provided in the last two columns of the table.

Table 6-14. Crash Data Summary for Intersections.

Legs	Control Type	Exposure,[1] mev	Crashes / 3 years			Crash Rate, cr/mev	
			PDO [2]	I+F [3]	Total	I+F [3]	Total
Three	Two-way stop	151.2	51	100	151	0.22	0.33
Four	Signal	105.4	67	132	199	0.42	0.63
	Two-way stop	60.7	30	50	80	0.27	0.44
		Total:	148	282	430		

Notes:
1 - mev: million-entering-vehicles per year.
2 - PDO: property-damage-only crashes.
3 - I+F: injury plus fatal crashes.

Model Development and Statistical Analysis Methods

This section describes the intersection calibration model and the methods used to re-calibrate it. The general form for this model is shown below:

$$E[N] = a\left(\frac{ADT_{major}}{1000}\right)^{b_1}\left(\frac{ADT_{minor}}{1000}\right)^{b_2} \times AMF_{LT} \times AMF_{RT} \times AMF_{SW} \times AMF_{skew} \times AMF_{nd} \times AMF_{tk} \qquad \textbf{(25)}$$

where,

a, b_1, b_2 = calibration coefficients;
AMF_{LT} = major-road left-turn lane accident modification factor;
AMF_{RT} = major-road right-turn lane accident modification factor;
AMF_{SW} = major-road shoulder width accident modification factor;
AMF_{skew} = alignment skew angle accident modification factor;
AMF_{nd} = major-road driveway frequency accident modification factor; and
AMF_{tk} = truck presence accident modification factor.

Equation 25 is effectively a safety prediction model for rural intersections and requires a full database for re-calibration. As discussed in the section titled Calibration Methodology, AMFs must be included in the model during the re-calibration process. In this manner, only the coefficients associated with the base model component of Equation 25 (i.e., a, b_1, b_2) are re-calibrated. The AMFs listed in Equation 25 are described in the *Workbook* (*1*).

The Generalized Modeling procedure (GENMOD) in SAS was used to automate the regression analysis. This procedure estimates model coefficients using maximum-likelihood methods. GENMOD is particularly well suited to the analysis of models with additive terms that are either continuous or categorical. GENMOD automates the estimation of the over-dispersion parameter k when the variance function (i.e., Equation 11) does not include the segment length variable, which is the case when it is used to calibrate intersection models. The value of k obtained from this iterative method was used in Equation 13 to estimate the true over-dispersion parameter.

Model Calibration

This section describes the results of the model re-calibration analysis. For each of the three facility components considered, the ADT coefficients for the re-calibrated model (i.e., b_1, b_2) were not significantly different from those in the existing model (as listed previously in Table 6-6). For this reason, the ADT coefficients in the existing model were retained and only the constant a was re-calibrated for each model.

The results of the re-calibration analysis are presented in three subsections. The first subsection describes the results for three-leg unsignalized intersections. The second subsection describes the results for four-leg signalized intersections. The last subsection describes the results for four-leg unsignalized intersections.

Three-Leg Unsignalized Intersection

The results of the three-leg unsignalized intersection model calibration are presented in Table 6-15. Calibration of this model focused on injury (plus fatal) crash frequency. The Pearson χ^2 statistic for the model is 93.5, and the degrees of freedom are 122 (= $n - p$ = 123 − 1). As this statistic is less than $\chi^2_{0.05,\,122}$ (= 149), the hypothesis that the model fits the data cannot be rejected. The R^2 for the model is 0.29. However, an alternative measure of model fit that is better suited to the negative binomial distribution is R_k^2, as developed by Miaou (*11*). The R_k^2 for the calibrated model is 0.80. This statistic indicates that about 80 percent of the variability due to systematic sources is explained by the model.

Table 6-15. Multivariate Model Statistical Description–Three-Leg Unsignalized Intersection.

Model Statistics		Value		
R^2 (R_k^2):		0.29 (0.80)		
Scale Parameter φ:		0.77		
Pearson χ^2:		93.5 ($\chi^2_{0.05,\,122}$ = 149)		
Over-Dispersion Parameter k:		2.59		
Observations n_o:		123 intersections (100 injury + fatal crashes in 3 years)		
Standard Deviation s_e:		±0.39 crashes/yr		
Range of Model Variables				
Variable	**Variable Name**	**Units**	**Minimum**	**Maximum**
ADT_{major}	Major-Road ADT	veh/d	80	19,200
ADT_{minor}	Minor-Road ADT	veh/d	30	6200
Calibrated Coefficient Values				
Variable	**Definition**	**Value**	**Std. Dev.**	**t-statistic**
a	Constant	0.109	0.014	7.8

The a coefficient in Table 6-15 is 0.109. The corresponding 95th percentile confidence interval for the a coefficient is 0.08 to 0.14. This range does not exclude the corresponding a coefficient in Table 6-6 (i.e., 0.0973) and thus, it is recommended that the a coefficient in Table 6-6 be retained in the model. The form of this model is:

$$E[N]_{3LST} = 0.0973 \left(\frac{ADT_{major}}{1000} \right)^{0.863} \left(\frac{ADT_{minor}}{1000} \right)^{0.497} \qquad \textbf{(26)}$$

where,

$E[N]_{3LST}$ = expected crash frequency for a three-leg unsignalized intersection, crashes/yr.

The ADT coefficients in Equation 26 were obtained from Table 6-6.

Four-Leg Signalized Intersection

The results of the four-leg signalized intersection model calibration are presented in Table 6-16. Calibration of this model focused on injury (plus fatal) crash frequency. The Pearson χ^2 statistic for the model is 15.2, and the degrees of freedom are 19 (= $n - p$ = 20 − 1). As this statistic is less than $\chi^2_{0.05,\,19}$ (= 30.1), the hypothesis that the model fits the data cannot be rejected. The R^2 for the model is 0.50. However, an alternative measure of model fit that is better suited to the negative binomial distribution is R_k^2, as developed by Miaou (*11*). The R_k^2 for the calibrated model is 0.36. This statistic indicates that about 36 percent of the variability due to systematic sources is explained by the model.

Table 6-16. Multivariate Model Statistical Description–Four-Leg Signalized Intersection.

Model Statistics		Value		
	R^2 (R_k^2):	0.50 (0.36)		
	Scale Parameter φ:	0.80		
	Pearson χ^2:	15.2 ($\chi^2_{0.05,\,19}$ = 30.1)		
	Over-Dispersion Parameter k:	3.15		
	Observations n_o:	20 intersections (132 injury + fatal crashes in 3 years)		
	Standard Deviation s_e:	±1.2 crashes/yr		
Range of Model Variables				
Variable	**Variable Name**	**Units**	**Minimum**	**Maximum**
ADT_{major}	Major-Road ADT	veh/d	4300	19,200
ADT_{minor}	Minor-Road ADT	veh/d	1800	11,100
Calibrated Coefficient Values				
Variable	**Definition**	**Value**	**Std. Dev.**	**t-statistic**
a	Constant	0.221	0.031	7.1

The a coefficient in Table 6-16 is 0.221. The corresponding 95th percentile confidence interval for this coefficient is 0.16 to 0.28. This range includes the corresponding a coefficient in Table 6-6 (i.e., 0.166); however, it is just at the limit of this range and would be excluded if the confidence interval is relaxed to 92 percent. This relaxed percentage is relatively large and is a good indication that the actual coefficient is most likely larger than 0.166. Thus, the value of 0.221 is recommended for inclusion in the re-calibrated model. The form of this model is:

$$E[N]_{4LSG} = 0.221 \left(\frac{ADT_{major}}{1000}\right)^{0.611} \left(\frac{ADT_{minor}}{1000}\right)^{0.595} \qquad \textbf{(27)}$$

where,

$E[N]_{4LSG}$ = expected crash frequency for a four-leg signalized intersection, crashes/yr.

The ADT coefficients in Equation 27 were obtained from Table 6-6.

Four-Leg Unsignalized Intersection

The results of the four-leg unsignalized intersection model calibration are presented in Table 6-17. Calibration of this model focused on injury (plus fatal) crash frequency. The Pearson χ^2 statistic for the model is 52.5, and the degrees of freedom are 64 (= $n - p$ = 65 − 1). As this statistic is less than $\chi^2_{0.05,\,64}$ (= 83.7), the hypothesis that the model fits the data cannot be rejected. The R^2 for the model is 0.25. However, an alternative measure of model fit that is better suited to the negative binomial distribution is R_k^2, as developed by Miaou (*11*). The R_k^2 for the calibrated model is 0.72. This statistic indicates that about 72 percent of the variability due to systematic sources is explained by the model.

Table 6-17. Multivariate Model Statistical Description–Four-Leg Unsignalized Intersection.

Model Statistics		Value		
	R^2 (R_k^2):	0.25 (0.72)		
	Scale Parameter φ:	0.82		
	Pearson χ^2:	52.5 ($\chi^2_{0.05,\,64}$ = 83.7)		
	Over-Dispersion Parameter k:	1.61		
	Observations n_o:	65 intersections (50 injury + fatal crashes in 3 years)		
	Standard Deviation s_e:	±0.38 crashes/yr		
Range of Model Variables				
Variable	**Variable Name**	**Units**	**Minimum**	**Maximum**
ADT_{major}	Major-Road ADT	veh/d	120	12,800
ADT_{minor}	Minor-Road ADT	veh/d	30	4000
Calibrated Coefficient Values				
Variable	**Definition**	**Value**	**Std. Dev.**	**t-statistic**
a	Constant	0.235	0.044	5.3

The a coefficient in Table 6-17 is 0.263. The corresponding 95th percentile confidence interval for this coefficient is 0.15 to 0.32. This range includes the corresponding a coefficient in Table 6-6 (i.e., 0.166); however, it is just at the limit of this range and would be excluded if the confidence interval is relaxed to 88 percent. This relaxed percentage is relatively large and is a good indication that the actual coefficient is most likely larger than 0.166. Thus, the value of 0.235 is recommended for inclusion in the re-calibrated model. The form of this model is:

$$E[N]_{4LST} = 0.235 \left(\frac{ADT_{major}}{1000}\right)^{0.692} \left(\frac{ADT_{minor}}{1000}\right)^{0.514} \qquad \textbf{(28)}$$

where,

$E[N]_{4LST}$ = expected crash frequency for a four-leg unsignalized intersection, crashes/yr.

The ADT coefficients in Equation 28 were obtained from Table 6-6.

Sensitivity Analysis

This section examines the injury (plus fatal) crash rates that are estimated from the re-calibrated models with those obtained from the existing models. The crash rates were computed by dividing the expected crash frequency from the base model by the sum of the major-road and minor-road ADTs (with conversion to million-entering-vehicles). The rates computed from the existing models were previously shown in Tables 6-4 and 6-5.

The injury (plus fatal) crash rates computed with the re-calibrated models are listed in Tables 6-18 and 6-19 for three-leg and four-leg intersections, respectively. The rates for three-leg intersections in Table 6-18 are the same as in Table 6-4. The rates for three-leg signalized intersections were not modified because insufficient data were available for calibrating this intersection configuration. The rates for three-leg unsignalized intersections were not modified because the re-calibration analysis indicated that the resulting calibration coefficients were not significantly different from those in the existing model.

Comparison of the rates in Tables 6-5 and 6-19 indicate that the re-calibrated crash rates for four-leg signalized intersections are 33 percent larger than those from the existing model. By comparison, the crash rates for four-leg unsignalized intersections are 42 percent larger than those from the existing model. This trend, where the existing models underestimate injury (plus fatal) crash frequency in Texas, was also noted in the re-calibration of the segment model. On average, Texas highway segments were found to have an injury (plus fatal) crash rate that is 20 percent larger than the highways, and states, represented in the rate cited in the *Workbook* (*1*).

A comparison of the four-leg signalized intersection rates with those for four-leg unsignalized intersections suggests that signalization reduces crashes from 7 to 22 percent. A similar level of crash reduction is suggested for three-leg intersections. The amount of reduction varies with the ratio of minor-road to major-road volume, with reductions of about 22 percent when the ratio of minor- to major-road volume is 0.10. This range of crash reduction (i.e., 7 to 22 percent) is consistent with the range found by McGee, et al. (*12*) for the conversion of four-leg intersections from unsignalized to signalized control. Specifically, they found an 8 percent reduction for intersections with an ADT of 20,000 veh/d or more, and a 37 percent reduction for intersections with an ADT less than 20,000 veh/d.

Table 6-18. Re-Calibrated Crash Rates for Three-Leg Rural Intersections.

Control Mode	Major-Road Volume, veh/d	Crash Rate, injury + fatal crashes per million-entering-vehicles				
		Ratio of Minor-Road to Major-Road Volume				
		0.05	**0.10**	**0.15**	**0.20**	**0.25**
Unsignalized [1]	5000	0.10	0.14	0.16	0.18	0.19
	10,000	0.13	0.18	0.21	0.23	0.25
	15,000	0.15	0.20	0.24	0.26	0.28
	20,000	0.17	0.23	Intersection very likely to meet signal warrants		
	25,000	0.18				
Signalized	5000	0.08	0.11	0.14	0.16	0.17
	10,000	0.10	0.15	0.18	0.20	0.22
	15,000	0.12	0.17	0.21	0.23	0.25
	20,000	0.13	0.19	0.23	0.26	0.28
	25,000	0.14	0.20	0.25	0.28	0.30
	30,000	0.15	0.22	0.26	0.30	0.33
	40,000	0.17	0.24	0.29	0.33	0.36
	≥50,000	0.18	0.26	0.32	0.36	0.39

Note:
1 - Unsignalized intersections have an uncontrolled major road and a stop-controlled minor road.

Table 6-19. Re-Calibrated Crash Rates for Four-Leg Rural Intersections.

Control Mode	Major-Road Volume, veh/d	Crash Rate, injury + fatal crashes per million-entering-vehicles				
		Ratio of Minor-Road to Major-Road Volume				
		0.10	**0.30**	**0.50**	**0.70**	**0.90**
Unsignalized [1]	5000	0.25	0.37	0.42	0.44	0.45
	10,000	0.29	0.43	0.48	0.51	0.52
	15,000	0.31	0.47	0.52	0.55	0.56
	20,000	0.33	Intersection very likely to meet signal warrants			
	25,000	0.35				
Signalized	5000	0.20	0.32	0.37	0.40	0.42
	10,000	0.23	0.37	0.43	0.46	0.48
	15,000	0.24	0.40	0.47	0.50	0.52
	20,000	0.26	0.42	0.50	0.53	0.56
	25,000	0.27	0.44	0.52	0.56	0.58
	30,000	0.28	0.46	0.54	0.58	0.60
	40,000	0.30	0.49	0.57	0.62	0.64
	≥50,000	0.31	0.51	0.60	0.65	0.67

Note:
1 - Unsignalized intersections have an uncontrolled major road and a stop-controlled minor road.

REFERENCES

1. Bonneson, J.A., K. Zimmerman, and K. Fitzpatrick. *Interim Roadway Safety Design Workbook*. FHWA/TX-06/0-4703-P4, Texas Department of Transportation, Austin, Texas, April 2006.

2. 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.* FHWA-RD-99-207. Federal Highway Administration, Washington, D.C., 2000.

3. 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., July 2005.

4. Lyon, C., J. Oh, B. Persaud, S. Washington, and J. Bared. "Empirical Investigation of Interactive Highway Safety Design Model Accident Prediction Algorithm." *Transportation Research Record 1840.* Transportation Research Board, National Research Council, Washington, D.C., 2003, pp. 78-86.

5. Sawalha, Z., and T. Sayed. "Transferability of Accident Prediction Models." Vol. 44. *Safety Science.* Elsevier Ltd., Oxford, Great Britain, 2006, pp. 209-219.

6. Bonneson, J.A., K. Zimmerman, and K. Fitzpatrick. *Roadway Safety Design Synthesis.* FHWA/TX-05/0-4703-P1. Texas Department of Transportation, Austin, Texas, November 2005.

7. Hauer, E. "Over-dispersion in Modeling Accidents on Road Sections and in Empirical Bayes Estimation." *Accident Analysis and Prevention.* Vol. 33. Pergamon Press, 2001, pp. 799-808.

8. 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, Elsevier Ltd., Oxford, Great Britain, July 2006, pp. 751-766.

9. McCullagh, P., and J.A. Nelder. *Generalized Linear Models.* Chapman and Hall, New York, N.Y., 1983.

10. Kvalseth, T.O. "Cautionary Note About R^2." *The American Statistician.* Vol. 39, No. 4, (Part 1), American Statistical Association, November 1985, pp. 279-285.

11. Miaou, S.P. *Measuring the Goodness-of-Fit of Accident Prediction Models.* FHWA-RD-96-040. Federal Highway Administration, Washington, D.C., 1996.

12. McGee, H., S. Taori, and B. Persaud. *NCHRP Report 491: Crash Experience Warrant for Traffic Signals*. Transportation Research Board, National Research Council, Washington, D.C., 2003.