University of Alberta

The Relationship Between Accident Rates and Cross-Sectional Geometric Design Features of Rural Primary Highways in Alberta

by

Raymond Yuan Dai ©

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Master of Science

in

Transportation Engineering

Department of Civil Engineering

Edmonton, Alberta

Fall 1996

ISBN 0-612-18248-7

University of Alberta

Library Release Form

Name of Author: Raymond Yuan Dai

Title of Thesis: The Relationship Between Accident Rates and Cross-Sectional Geometric Design Features of Rural Primary Highways in Alberta

Degree: Master of Science

Year this Degree Granted: 1996

Raymond Yuan Dai
1107 - 111 A Street
Edmonton, Alberta
T6J 6R9

August 27, 1996
Date Submitted

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled ***The Relationship Between Accident Rates and Cross-Sectional Geometric Design Features of Rural Primary Highways in Alberta*** submitted by ***Raymond Yuan Dai*** in partial fulfillment of the requirements for the degree of Master of Science in Transportation Engineering.

Dr. S. Teply

Dr. L. R. Rilett

Prof. A. E. Peterson

Dr. M. G. Faulkner

Abstract

This thesis quantifies the relationship between accident rate and cross-sectional highway geometric factors in Alberta. A key aspect of highway rehabilitation projects is the cost-effective application of project dollars towards a quantifiable safety benefit. Measuring the effects of incremental cross-sectional geometric improvements on accident rate is crucial in gauging the magnitude of this safety benefit. Accidents rates are used as a measure of highway safety and essentially quantify the risk component of safety.

This thesis provides the framework, methodology, and tools to model the relationship between accident rates and cross-sectional highway geometric design elements in Alberta. Two models developed from accident and geometric data in Alberta are presented: one for a two-lane, rural, primary highway model, and one for a multi-lane, rural, primary highway model. These accident rate models are compared with the existing accident rate model to identify a correction coefficient. It is now possible to compare historical and current accident rates by applying this coefficient to the existing model.

Acknowledgments

I wish to show my appreciation by acknowledging the individuals and organizations who have help me to complete this thesis.

- Foremost, I wish to acknowledge the assistance, guidance, and insight that my advisor and supervisor, Dr. Laurence Rilett has provided me in the past two years. His contributions to the development and refinement of this thesis were invaluable.
- I further wish to acknowledge Dr. Stan Teply and Dr. William Sproule for showing me the opportunities and possibilities in Transportation Engineering.
- I wish to thank and acknowledge the generous funding and technical support that Allan Kwan and Alberta Transportation and Utilities have provided me for this thesis. I wish to specially thank Bill Kenny of Alberta Transportation and Utilities for his assistance and insights on 3R/4R concepts and geometric design.
- A special thanks to Li-ping Fu, Darryl Mullen, Chris Jordan, Christine Benedek and the rest of the ITE Student Chapter for all the times when we weren't working on our theses.
- For Courtney Claire Kennedy, whose unswerving faith and support constantly guided me in my studies and my writing.
- And finally, to my parents, who helped me dream of this goal and made all of this possible.

Table of Contents

Page

Page

List of Tables

Page

Chapter 7

List of Figures

Page

Chapter 7

Chapter 1

Introduction

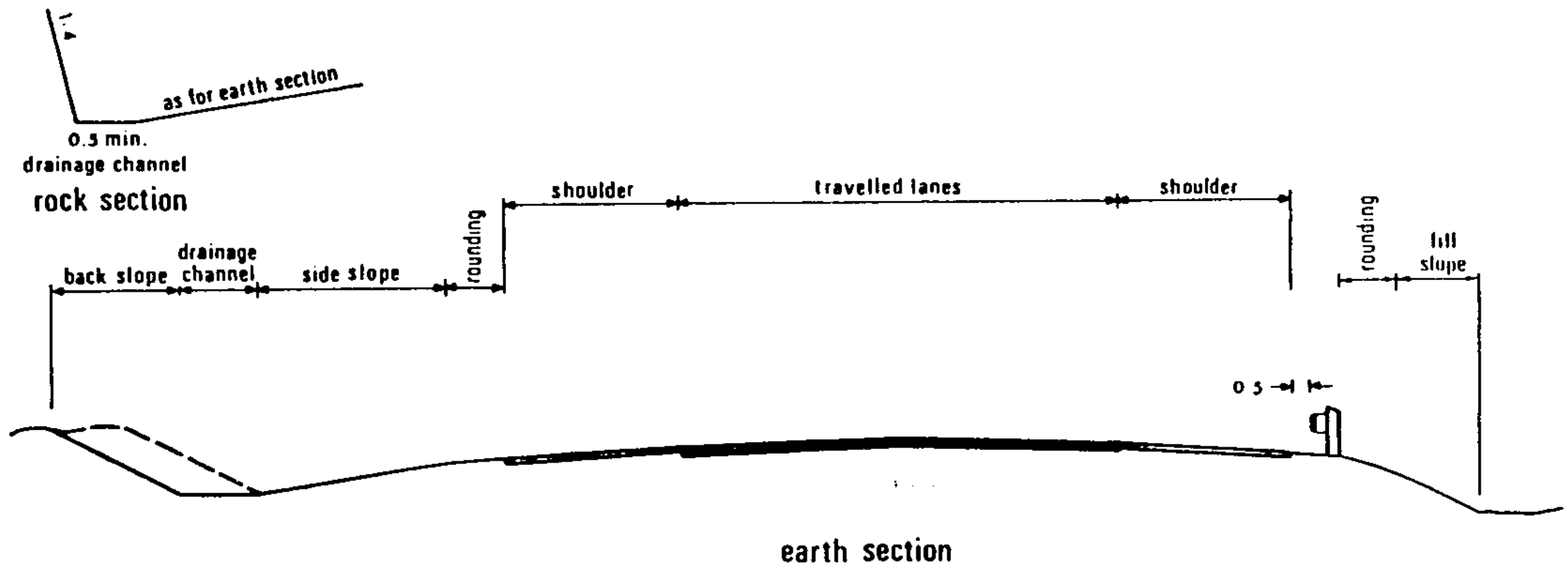
1.0 OVERVIEW

The objective of this research project was to quantify the effects of cross-sectional highway geometric design features on accident rates in Alberta. The role of cross-sectional geometric design features such as lane and shoulder width are seen to influence both the frequency and the magnitude of accident events occurring on highways in Alberta. However, highway accidents do not occur strictly as a result of cross-sectional geometric factors. Accidents will naturally and randomly occur irregardless of the cross-sectional geometric condition of the highway, due in part to a multitude of other geometric and non-geometric factors that may potentially contribute to a highway accident. These factors may include driver behavioral factors, environmental factors, ambient light factors, other geometric factors such as roadway curvature, stopping sight distances, and pavement surface conditions, as well as the cross-sectional factors examined in this research project. In essence, accidents are an undesired result of the interaction between the road, vehicle, environment, and the driver. Because any or all of these four factors may play a role in any given accident, it is easy to understand the complex nature of highway accidents and why it is difficult to isolate any one of the contributing factors or the effects of cross-sectional geometrics, in particular.

1.1 SCOPE OF THE ANALYSIS

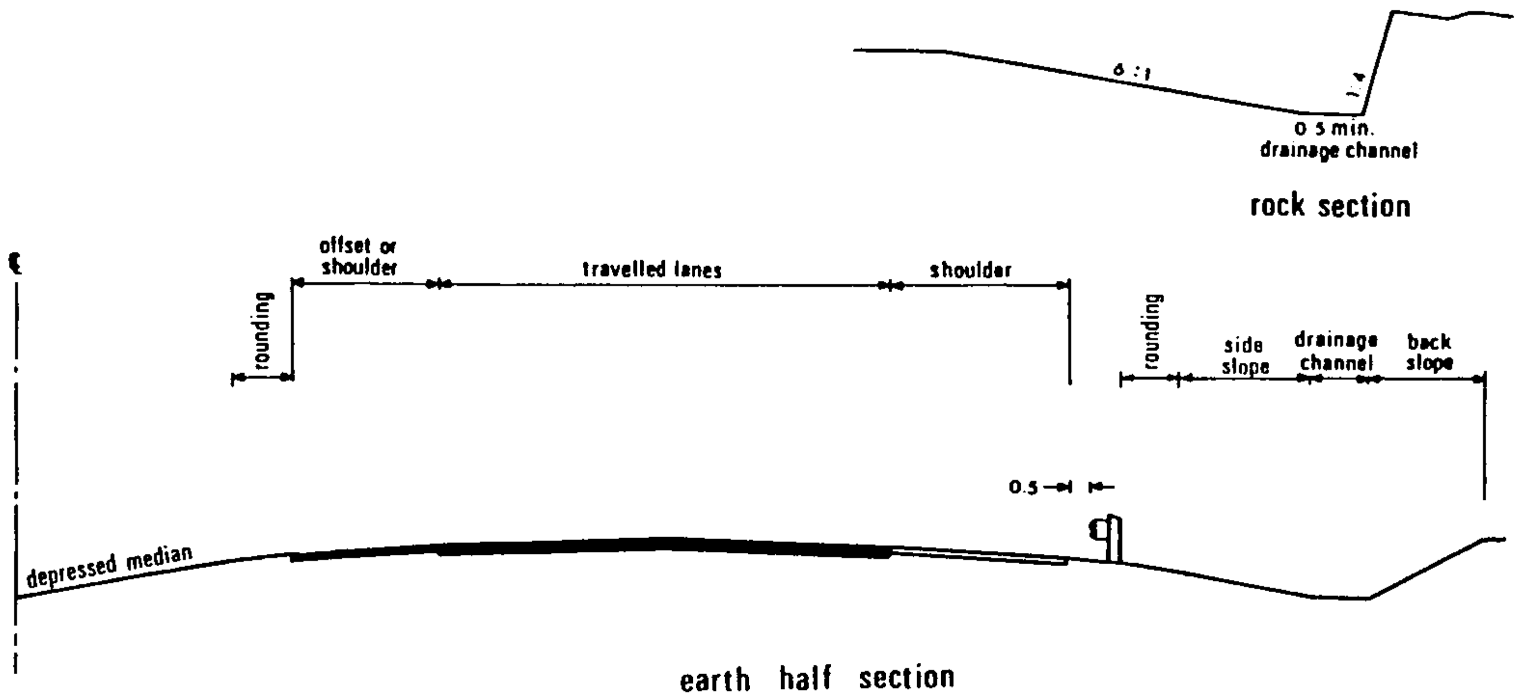
The scope of this thesis concentrated on two distinct elements of highway geometric design and highway accidents. This limitation was imposed on the objectives in order to concisely examine a specific area of the entire field of highway safety engineering. The two key elements of this scope are:

1. RURAL, TWO-LANE AND MULTI-LANE PRIMARY HIGHWAYS- The analysis was limited to these types of roadways due to the focus of prior research. Because highways in Alberta predominantly consist of rural, primary sections, a comparison between Alberta's roadway network and other prior research justifies this limitation of scope. An additional factor in this scope is that the analysis concentrated primarily on two-lane cross-sections, mainly due to the predominance of prior research into this area.
2. CROSS-SECTIONAL ELEMENTS- These elements were the only geometric elements examined in this thesis for two reasons: (1) the sizable number of cross-sectional elements that may potentially contribute to an accident was found to be adequate for model performance, and (2) analysis of horizontal and vertical alignment elements, cross-sectional elements and their effects on accident rates are best done separately, and not combinatively due to limitations in Alberta data structure and geometric models. Specifically, the cross-section geometric elements examined are: (1) lane width, (2) shoulder width, (3) lane and shoulder pavement



class RAU	D. H. V.	earth cut slopes back	earth cut slopes side	rn'd	shoulder	travelled lanes	shoulder	rn'd	fill slopes with guide rail earth	fill slopes with guide rail rock	fill slopes no guide rail
120&130		2:1	6:1	1.0	3.0	each lane 3.7	3.0	1.0	2:1	1.5:1	6:1
110	>450	2:1	6:1	1.0	3.0	each lane 3.7	3.0	1.0	2:1	1.5:1	6:1
110	<450	2:1	6:1	1.0	2.5	each lane 3.7	2.5	1.0	2:1	1.5:1	6:1
100				1.0				1.0			
90	>450	2:1	6:1	0.5	3.0	each lane 3.7	3.0	0.5	2:1	1.5:1	6:1
80				0.5				0.5			
100											
90	<450	2:1	4:1	0.5	2.5	each lane 3.5	2.5	1.0	2:1	1.5:1	6:1
80											

Figure 1-1 Typical Rural Undivided Cross-Section
(Source: RTAC, 1986)



class RAD	rn'd	offset or shoulder		travelled lanes	shoulder	rn'd	cut side slope	drainage channel	cut back slope	fill slope		
		*	+							with guide rail		no guide rail
										earth	rock	
130 120 110 100	1.0	1.5	2.5	each lane 3.7	3.0	1.0	6:1	varies	2:1	2:1	1.5:1	6:1
90 80	0.5	1.5	2.5	each lane 3.5	3.0	0.5	6:1	varies	2:1	2:1	1.5:1	6:1

* 2 lanes
\+ 3 or more lanes

Figure 1-2 Typical Rural Divided Cross-Section (Source: RTAC, 1986)

type, (4) sideslope ratio (mean and extreme values), (5) ditch width (mean and extreme values), (6) back slope ratio (mean and extreme values), and (7) a roadside hazard rating (a subjective rating based on sideslope ratio, ditch width, and possibly backslope ratio). Sample cross-sections of two-lane and four-lane rural primary highways in Alberta are given in Figure 1-1 and Figure 1-2.

While this research does not address the multitude of other factors influencing highway accidents, this does not in any way discount their effects on highway safety. What the scope of this research project serves to accomplish (that is, limiting the research to the quantification of the effects of cross-sectional design features on accident rates) is to firm up relationships between design features and highway safety, that in the past, relied on engineering judgment, and identify relationships that may not be applicable for use in Alberta.

1.2 RESEARCH IMPETUS

The impetus behind this work is twofold. First, clearly defined relationships between cross-sectional geometric design factors and accident rates are of key importance to the benefit-cost assessment of highway improvements. Simply put, these accident-geometric relationships can be used to accurately gauge the resulting societal gains from safety by improving the highway, as well as the potential costs (in terms of lives lost, for example) by not improving cross-sectional geometrics.

The second driving force behind this work revolved around the lack of quantifiable relationships for specific use in 3R/4R highway infrastructure improvements projects Alberta. By definition, a 3R (Restoration, Rehabilitation, or Repaving) project involves the work required to return an existing section of highway to an adequate structural and safety condition or sufficiently sound to support additional surfacing (***TRB, 1987***). Generally, intensive geometric improvements are not undertaken within a 3R project's scope. Corrections involving regrading or curve realignments are generally undertaken in 4R (Restoration, Rehabilitation, Repaving, or Reconstruction) projects only. However, relatively minor corrections to geometrics such as pavement widening, or correction of roadside sideslopes are readily undertaken in 3R projects and may enhance highway safety through a reduction in the probability of accident occurrence. These 3R project enhancements are designed to not only extend the service life of a highway, but to improve overall safety concurrently.

As result, 3R/4R guidelines have been developed by federal, state, and provincial agencies that addressed these issues. Crucial to the development of 3R/4R guidelines is the identification of the relationships between accident rates and highway cross-section geometric design. Currently, the Alberta Transportation and Utilities (AT&U) uses a cross-sectional geometric-accident rate relationship developed in the United States by the Transportation Research Board (TRB). However, this relationship or model, was developed using U.S. highway data, and has not been calibrated for use within Alberta. In essence, the design of highways in Alberta may be based on a relationship that does not reflect the actual relationship seen in Alberta between accident rates and cross-section. The impact of this is that highways in Alberta may be potentially over or under designed, depending on the predicted accident rate difference between the Alberta and the TRB model relationships. The need to quantify and to establish this relationship in Alberta by the AT&U was the second motivating factor in this research project.

1.3 RESEARCH OBJECTIVES

In broad terms, the objectives of the research project encompassed by this thesis is to develop a cross-section geometric accident rate model that can be calibrated and validated for use within Alberta. This model will eventually form the basis of estimating accident rates for roadway sections in Alberta and be incorporated into a benefit-cost analysis methodology of determining roadway sections requiring 3R/4R work. The research consisted of five distinct objectives:

1. Identifying and collating prior engineering and scientific research in determining the relationship between accident rates and geometric design elements. Concurrent to this objective, present methods of incorporating safety into geometric design and 3R/4R design work within transportation agencies were also evaluated to identify possible methodologies applicable for Alberta conditions.
2. Determining whether theoretical or empirical assumptions used in prior research, models, and methodologies are valid with respect to Alberta accident and geometric elements.

3. highway accident databases with their primary highway cross-section geometric databases.
4. Identification of the relationship between accident rates and cross-sectional geometric elements in Alberta, and the development of a resulting Alberta accident rate model.
5. Calibration and validation of the Alberta accident rate model through a comparison with both observational accident and geometric data, and applicable other models.

1.4 KEY RESEARCH FINDINGS

The key finding developed from this research project centres on the applicability of the TRB accident rate model to predict the probability of accident occurrence in Alberta. Based on the findings of Chapter 6 in this thesis, the magnitude of the TRB model appears to differ significantly from observed Alberta accident conditions. However, the findings also indicate that the log-linear form of the accident rate model appears to be applicable for use in Alberta, despite the difference in magnitude.

A secondary result of this research project identified the lack of accident and geometric data for use in developing a multi-lane accident rate model for use in Alberta. This result was not entirely unexpected, as the uniformity of cross-sectional geometrics on Alberta's primary multi-lane highway network was pre-identified as a potentially confounding factor to the analysis.

1.5 ORGANIZATION OF THIS THESIS

Chapter 2 of this thesis provides background information into prior research attempts at quantifying various aspects of the relationship between highway accident rates and cross-sectional geometric design. This chapters includes a detailed examination of the relationship between highway safety (in terms of accident frequency and rate), and cross-sectional elements of highway geometric design. This chapter also helps in identifying additional avenues of research into geometric factors to further expand the knowledge base of highway safety and geometric design in Alberta.

To augment the literature review conducted in Chapter 2, Chapter 4 consists of a survey of North American state and provincial transportation agencies conducted to provide details of current standards and methodologies in highway safety engineering. Questions polled from the survey recipients concentrated on identifying current techniques in using accident statistics to identify roadway sections requiring 3R/4R work, and ongoing research into improving this identification process.

Chapter 4 introduces the concept of using generalized linear regression and regression modeling techniques in developing an accident rate model for Alberta. Furthermore, this chapter examines in detail, the underlying assumptions concerning the validity of these types of models, as well as the situations permitting the valid application of them.

The methodology to develop a model representing the relationship between highway safety (through accident rate) and cross-sectional geometric factors is presented in Chapter 5. A preliminary analysis of both Alberta accident and geometric data is performed in this chapter, as well as an outline of the technique to be used in developing the Alberta accident rate model.

Chapter 6 concentrates on the development and validation of an accident rate model that relates accident rate with various cross-sectional geometric factors. This model specifically targets two-lane, rural primary highway sections and their corresponding accident rates only. Issues encountered concerning the availability and suitability of various data used in the analysis is also examined in this chapter. Finally, conclusions drawn from the analysis and comparison of the Alberta accident rate model and other models currently used are presented.

Chapter 7 briefly examines the potential effects of cross-sectional factors of multi-lane (four or more lanes), rural, primary highways in Alberta on accident rates. This chapter utilizes the same methodology as used in Chapter 6 to arrive at a similar accident rate model, with few modifications.

Conclusions reached from the analysis performed in Chapter 6 and Chapter 7 form the basis of Chapter 8. This chapter also presents possible alternatives in approaching the accident-geometrics relationship in terms of model development and validity. A summary of model limitations and issues surrounding input data integrity is also discussed in detail in this chapter. Finally, recommendations developed from the conclusions and identification of problems in the analysis are addressed.

Chapter 2

Relationship Between Highway Safety And Geometric Design: A Literature Search

2.0 INTRODUCTION

In order to identify the relationship between highway safety and geometric design, the type, magnitude and frequency of safety gains obtained from geometric improvements need to be identified and quantified. These safety gains can arise from such improvements as lane and shoulder widening, straightening sharp horizontal curves, reducing vertical curvature, identifying and removing roadside hazards, and improving sideslopes. Because project costs greatly affect the scope of rural roadway infrastructure improvement projects, the balance between the dollar cost of improvement and incremental safety gains resulting from the improvement is frequently the key issue in 3R projects. Addressing this issue requires that quantitative knowledge be identified and developed into the relationship between safety and the various geometric design factors in Alberta

The random and often confounding nature of accidents both in magnitude and frequency can greatly complicate the quantifying of the relationship between safety and geometrics. Nonetheless, models representing the relationship between highway safety and key geometric factors must be developed as they are crucial in understanding the safety-geometric relationship and assist in answering how and why accidents occur.

These relationships can then be used in benefit-cost comparisons to determine the cost effectiveness of an incremental safety improvement within Alberta. In turn, benefit-cost analysis (among other methods of gauging effectiveness) allows an engineer to make the correct design decision.

Because the majority of the North American roadway network is comprised of 2-lane roadways, research into these safety-geometric relationships pertain mainly to 2-lane, rural highways. In the United States, 2-lane, rural highways comprise approximately 75% of all U.S. federal highway mileage, 25% of all vehicle miles traveled and 35% of all highway fatalities in the United States (***TRB, 1987***). In comparison, 2-lane, rural, primary highways are the predominant highway type in Alberta, and comprise nearly 85% of the total primary highway network length. Nearly two-thirds of all vehicle accidents on the network occur on 2-lane, rural, primary highways.

AASHTO, (the American Association of State Highway Transportation Officials), which historically has assumed responsibility in the United States for setting out guidelines on safe geometric design, uses a committee of experienced design engineers to quantify the link between safety and geometrics. While noteworthy because the committee relies heavily on professional judgment in determining standards, the process has traditionally lacked rigorous or methodical statistical backing of their views concerning safety. This point is extremely important because unlike new roadway infrastructure projects, the costs involved in making an incremental geometric improvement are often large compared to other project costs. If safety is considered from the start of a new roadway infrastructure project, the cost of those safety improvements is only incidental compared to the entire construction cost (***TRB, 1987***). In essence, a more quantifiable method of determining these relationships is both necessary and economically feasible to incremental roadway improvement projects in Alberta.

2.0.1 RISK AND EXPOSURE CONCEPT OF HIGHWAY SAFETY

Highway safety is influenced by two independent variables: (1) risk and (2) exposure. The following equation represents the relationship between safety and these two variables (***Chan, 1994***):

$$S = R \times E \qquad \textbf{(EQN 2.1)}$$

where,
S = The measure of non-safety over a period of time
R = Measure of risk presented to drivers during the same time period
E = Measure of exposure (the number of vehicles on the highway section in the time period)

Accident rates are frequently used as a quantifiable and surrogate measure of highway safety (***TRB, 1987***). As a result, accident rates depend on the interaction of both risk and exposure. The risk component of accident rates are quantified in terms of the various geometric and non-geometric factors having a causal relationship with highway accidents. The exposure component of accident rates are quantified in terms of the traffic volume of a given highway section. The product of these two factors yields the probability of accident occurrence.

2.1 FACTORS IN HIGHWAY SAFETY

However, geometrics are clearly not the only factor affecting highway safety. Road environment conditions (weather, pavement condition, lighting and traffic volumes for example), driver behavioral, physiological and psychological characteristics (driver temperament, alcohol, age, and driver visual acuity for example), and vehicle characteristics (size, weight, braking capability and tire condition for example) also may play a major role in the cause of highway accidents. Because of the multitude and the often confounding and interrelated effects of these factors in highway accidents, it is difficult to isolate and quantify the effects of geometric factors on accidents. This interaction between road, car, environment and driver can obscure the ability to gauge the effectiveness of a single safety improvement through mathematical models, especially in light of the fact that much of this non-geometric data cannot be collected.

Nonetheless, the goal of this thesis is not necessarily the identification of causal relationships between accidents and highway factors, but rather the identification of correlative relationships between accident rates and cross-sectional geometric factors. By examining correlation and not causality, the onus to include these other non-geometric factors in the relationship analysis is no longer present. It becomes safe to assume that these non-geometric factors will play a role in highway accidents, but that their effects lie outside of the scope of this thesis.

2.2 EFFECT OF GEOMETRIC FEATURES ON SAFETY

Highway geometric features can affect highway safety in several ways. For example, lane width, sight distances, alignments and pavement surface characteristics can influence the driver's vehicle control and hazard identification. Through elements such as intersection design, access control point design and passing lanes, geometrics can influence the number and types of potential vehicle conflicts. The outcome of run-off-the-road and out-of-control vehicle accidents can be greatly affected by roadside physical characteristics such as pavement drops, shoulder widths, guardrails and sideslopes. As well, the behaviour (particularly alertness) of a driver can be influenced through geometric related characteristics such as speed limits. In effect, driver behaviour is directly affected, in some part, by all roadway characteristics.

While the presence of the other confounding influences on highway safety may make geometric improvements appear to be less significant than other factors, it remains that improvements in road geometrics can prevent accidents from either occurring or reducing the severity of accidents once they do occur. While little can be done from the standpoint of a highway design engineer in altering the behaviour characteristics of drivers, or affecting the environmental factors associated with safety, highway geometrics is one area that engineering can play a direct role in providing drivers with a safer operating environment.

2.3 QUANTIFYING SAFETY-GEOMETRIC RELATIONSHIPS

One of the problems associated with quantifying relationships between highway geometric factors and accident rates is a traditional lack of background knowledge on the characteristics of various factors involved in an accident. This lack of knowledge and a lack of quantitative equations are due, in part, to the following issues (***TRB, 1987***):

2.3.1 PROBLEMS WITH THE NATURE OF ACCIDENTS

The random and infrequent nature of accidents traditionally require that consistent statistical data be collected over relatively long periods of time to build confidence in the data. In many cases, the time frame for the amount of data required can be longer than what is available. This issue is particularly relevant in regression modeling as the inadequate data can result in poorly defined relationships.

The increasing use of Bayesian statistics in highway accident analysis enables past accident rates at specific locations to be integrated into a regional historic background accident rate such that random elements present in all roadway accidents can be mathematically defined and isolated. New elements are incorporated into the analysis as data becomes available to update and refine the estimation of predicted accident rates. These predicted accident rates are generated by combining the base background rate with a site and geometric specific accident rate to yield a total accident rate for a given roadway section with specific geometrics.

A key advantage of the Bayesian approach to modeling accident rates lies in identifying hazardous sites by comparing the probability of that site's total accident rate exceeding a pre-determined background level (***Higle and Witkowski, 1988***). Should a roadway location's estimated accident rate exceed a set background level, that location could be classified as a potentially hazardous location and receive corresponding geometric 3R/4R treatment. This probabilistic method differs both quantitatively and qualitatively from traditional linear modeling and does not suffer the same problems with timeframe that may affect regression modeling. Details on the Bayesian approach to accident rate modeling are given in Appendix B.

2.3.2 PROBLEMS WITH EXPERIMENTAL CONTROL

As previously mentioned, certain non-geometric related factors such as driver behaviour or environmental conditions, can directly or indirectly influence accidents in an interactive and unpredictable manner. Because these factors often cannot be attributed to any single quantifiable geometric factor, it is difficult to conduct controlled experiments in this area.

2.3.3 DIFFERENCES IN DATA COLLECTION AND RECORDING METHODOLOGY

Methods of reporting non-fatal (and in some cases, fatal) accidents frequently differ between state, province, and federal agency. Essentially, accident data developed from one jurisdiction often cannot be used with any degree of confidence for another region because the accident rates developed from the accident data will vary depending on the level and type of accident reporting done. For example, if data from province A contained accidents with a property damage only (PDO) reporting threshold of $1000, and data from province B used a lower $500 reporting threshold, rates derived from data A would differ from estimated rates of data B.

2.3.4 HIGHWAY SAFETY SYSTEM CHANGES

Because vehicle characteristics and performance changes over time, accident rates developed and data compiled over long periods of time may also lose statistical meaning. For example, recent developments in airbag and anti-lock braking system technologies have not only reduced the severity of accidents in certain cases, but have changed the nature of vehicle and accident severity relationships previously developed. To an extent, other factors (such as driver behaviour or demographics) that have changed with time can also affect these relationships developed from historical data.

Despite these issues in quantifying safety-geometric relationships, the need for such relationships exist, nonetheless. Relationships between accident rates and geometric factors offer a clear and definitive means of quantifying the impact of geometric improvements on highway safety.

2.4 GEOMETRIC FACTORS AFFECTING HIGHWAY SAFETY

Eight key geometric design features were identified to contribute significantly to the frequency and severity of highway accidents by a TRB study committee drafting recommendations into 3R/4R design (***TRB, 1987***). They are:

1. Bridge width
2. Horizontal alignment
3. Sight distances
4. Intersections
5. Pavement surface conditions
6. Pavement edge drops
7. Lane and shoulder width, and shoulder type
8. Roadside geometrics, clear zones and sideslopes

Based on the committee's expert opinion (note that this has not been necessarily quantified), improvements to these design attributes were felt to have a greater and more measurable effect on safety over any other geometric design factor. As a result, other notable geometric features such as vertical alignment and cross slopes, which may affect safety, were not included. One major factor that was not included in this list was traffic volume. This omission does not deny the relationship between traffic volume and accidents, but merely notes that traffic volume is a geometric related factor, and not a geometric design factor.

For each attribute, three issues were addressed: (1) whether a relationship actually exists between the geometric factor and safety, (2) the direction of the relationship (does a change in the factor increase or decrease safety?) and (3) the magnitude of impact of each improvement to the overall 3R safety improvement. The results of this research into the effects of geometric is summarized below.

2.4.1 NON-CROSS-SECTIONAL RELATED GEOMETRIC FACTORS

Bridge Width

Bridge hazards can play a significant role in accidents, not only due to constricted roadways at the entrances to a bridge, but because the physical terrain constraints often dictate that approaches are on a down grade (resulting in unintentionally higher vehicle operating speeds) with horizontal curves as well (***TRB, 1987***). Details into the relationship between accidents and bridge width can be found in Appendix C.

Horizontal Alignment

Due to the physiological and psychological conditions imposed on a driver negotiating a curve (as well as the physical effects on a vehicle), horizontal curves contribute to a large share of accidents on highways. Further details can be found in Appendix C of this thesis.

Sight Distance

Sight Distance is defined as the length of road ahead visible to the driver (***TRB, 1987***). Because sight distances are so important in allowing drivers time to stop in avoiding hazards, increasing stopping sight distances can noticeably increase safety.

Intersections

Rural highway intersections have the third highest concentration of accidents, after horizontal curve locations and bridges (***Brinkman and Smith, 1984***). Details into these accident types can be found in Appendix C.

Pavement Surface Conditions

A new pavement surface will often increase pavement skid resistance (and consequently reduce stopping distances and increase driver avoidance abilities), but can also can lead to higher operating speeds as a result of changes in driver behaviour. Further details can be found in Appendix C.

Pavement Edge Drops

Investigation into this area has quantitatively determined that these drops pose a significant hazard since they can readily turn a minor run-off-road encroachment into a serious roadside collision.

2.4.2 CROSS-SECTIONAL RELATED GEOMETRIC FACTORS

Lane And Shoulder Width, And Shoulder Type

Wide lanes and shoulders provide drivers with increased opportunity to maneuver and recover safely from a run-off-the-road (ROR) accident. They also increase lateral separation to reduce sideswiping (the two largest types of accidents on 2-lane highways). Research has indicated that accident rates decrease as lane and shoulder widths increase (***Zegeer and Deacon, 1987***). As well, it appears that lane widening is more cost effective with highway safety than shoulder widening (***Zegeer et al., 1987***) A generalized linear model was developed (log-linear) to represent the relationship between accident rates and roadway widths and other factors.

$$AR = 0.0019\,(ADT)^{0.882}\,(0.879)^{W}\,(0.919)^{PA}\,(0.932)^{UP}\,(1.236)^{H}\,(0.882)^{TER1}\,(1.322)^{TER2}$$

(EQN 2.2)

where, AR = predicted accident rate based on the following factors
ADT = average daily traffic volume
W = paved lane width
PA = paved shoulder width
UP = unpaved shoulder width
H = median roadside hazard rating (1 to 7 - most hazardous)
TER1= 1 for flat terrain, 0 otherwise
TER2= 1 for mountainous terrain, 0 otherwise

This model was selected through curve fitting of model to data obtained from over 4,000 sites throughout the United States. These sites consisted of over 5,000 miles of 2-lane primary highways from seven states (Alabama, Michigan, Montana, North Carolina, Utah, Washington, and West Virginia) and were selected through a stratified random sampling process (***Zegeer et al., 1987***). The resulting analysis curve represented the greatest correlation with actual data. Further detail into these factors is presented later in this chapter.

Roadsides And Sideslopes

By definition, encroachment of a vehicle occurs when a vehicle travels beyond the outer edge of the highway shoulder. This departure from the roadway surface into the roadside environment is usually uncontrolled and unintentional. Roadside encroachments become extremely hazardous to the vehicle and operator if roadside objects are located within the vehicle trajectory. The magnitude of these types of accidents is significant. An estimated 30% of all highway accidents are run-off-the-road (ROR), or encroachment accidents (***Graham and Harwood, 1982***). Safety researchers have generally agreed that at vehicle operating speeds of 90 km/hr, the lateral distance from the pavement edge that is free of objects should have a sideslope of no greater than 6:1 and extend at least 10 m out from the edge of the roadway. This lateral distance is referred to as a roadside clear zone and constitutes an area that increases the likelihood of a driver recovering from an ROR incident without more severe consequences. When the pavement edge is flat, encroachments rarely approach 10 m - hence the use of a 10 m clear zone. However, much of this work has remained qualitative (***TRB, 1987***).

Roadside encroachment models have been developed to examine safety effects of special roadside hazards (culverts and abutments, for example). These models account for the size and shape of a roadside feature, distance from travel lanes, and probability that a collision will result in physical human injury or fatality. The form that these models generally take is a linearized regression model with conditional probability equations used to represent the likelihood of a collision in the event of a run-off-road accident.

2.5 CROSS-SECTIONAL GEOMETRIC ACCIDENT RATE MODELS

The key issue with developing a working relationship between safety and geometric design is the sheer volume of factors that play a role in accidents. Although eight roadway and roadside geometric factors were found to have a significant effect on accident rates and highway safety (***TRB, 1987***), it does not discount the fact that a multitude of other geometric factors can also affect highway safety, albeit in a secondary manner. As well, non-geometric factors do play a role in accidents, as explained previously. One study estimated that at least 50 factors (from categories including roadway, roadside, driver behavioral and psychological condition, vehicle condition, traffic and environmental conditions) can play a critical role in an accident (***Banks, Brown and Beatty, 1978***).

Three roadway geometric elements - lane width, shoulder width, and shoulder type were singled out as having the greatest potential impact on highway accidents directly attributable to roadway geometrics (***Zegeer and Deacon, 1987***). These elements and their effects on highway safety were examined in depth in that report, which is summarized below.

2.5.1 GEOMETRIC ACCIDENT RATE MODEL STRUCTURE

The format that was selected by Zegeer and Deacon was a log-linear model with parameters derived from regression (***Zegeer and Deacon, 1987***). Nine different accident studies from 1970 to 1982 conducted over nine different states (Louisiana, North Carolina, Ohio, Idaho, Washington, California, Maryland, Kentucky and Texas) were critically assessed for information and data pertaining to the effects of cross-sectional factors on accident experience. From the analysis of the nine accident studies, Zegeer and Deacon concluded that run-off-the-road (ROR) and opposite direction (OD) incidents were the most significant accident types involving lane width, shoulder width and type (***Zegeer and Deacon, 1987***). Based on this conclusion, the rates at which ROR and OD accidents occurred were selected as the dependent variables for the regression model. Although three key factors (road and shoulder width, shoulder type) were already selected as independent variables, Zegeer and Deacon noted that their extensive literature review yielded evidence that other variables such as roadside characteristics, roadway curvatures, traffic volume and intersections also interacted with the three factors to create complex accident relationships.

In light of the fact that these additional factors, though influential, would add additional levels of complexity to the safety-geometric relationships being developed, Zegeer and Deacon decided to limit the scope of the model to predict only overall levels of accidents. This limitation was done by simplifying the number of independent variables to the three key geometric factors in the scope of the research. The potential effects of non-geometric factors on the accident rate model were not analyzed by the researchers.

Data from four of the original nine studies was selected for use in developing the safety model. The remaining five studies were discarded due to one or more of the following reasons: (1) failure to analyze lane and shoulder width combinations, or (2) lack of detailed analysis into the effects of highway geometrics with respect to different accident types. The resulting database of accident statistics contained rural, 2-lane highway geometric data covering the states of Kentucky (***Zegeer, Mayes and Dean, 1981***), Ohio, California, and Texas (***Rogness, Fambro and Turner, 1982***). The study conducted by Rogness, Fambro and Turner involved controlled before-and-after studies undertaken to model the relationship between accident rates and corresponding geometric factors, and to predict changes in accident rates as a result of incremental lane and shoulder improvements. This concept of predicting "after" accident rates on the basis of "before" observational geometric and accident data is one of the key uses of these accident rate models. The statistical reliability of these predictions and the model itself, strongly depends on the integrity of the input variables and underscores the importance of measurability, significance, and causality in the observed input data.

To account for the interrelated variables that were not specifically being modeled (such as ADT volumes), Zegeer and Deacon adjusted each predicted accident rate based on a plot of unadjusted accident rates versus volume levels developed in prior research (***Zegeer, Mayes and Dean, 1981***). The mathematical form of this regression was expressed as follows:

$$AR = (C_1)(C_2)^{L}(C_3)^{S}(C_4)^{LS}(C_5)^{P}(C_6)^{LP} \qquad \textbf{(EQN 2.3)}$$

$$\ln(AR) = \ln(C_1) + \ln(C_2)L + \ln(C_3)S + \ln(C_4)LS + \ln(C_5)P + \ln(C_6)LP \qquad \textbf{(EQN 2.4)}$$

where, AR = the predicted number of ROR and OD accidents per million vehicle miles
L = the lane width (in feet)
S = shoulder width (in feet) including both stabilized and non-stabilized portions
P = width of stabilized portion of shoulder (in feet)
P = 0 for unstabilized shoulders
P = 1 for full-width stabilization of shoulders
C_i's = regression coefficients
LS = interactive effect of lane and shoulder width
LP = interactive effect of lane width and paved shoulder portion

The model form used in Equation 2.3 is called a log-linear model, and represents the possibility of non-linear effects of the various cross-sectional geometric factors on accident rate. Zegeer and Deacon's reasoning behind selecting the geometric factors as the exponential power in the model revolves around the model transformation. This equation can be linearized by taking the natural log of both sides of the equation to yield a log-linear relationship as seen in Equation 2.4. However, if the geometric factors were not used as exponents, the log-linear transformation of the equation would yield a log-log relationship with accident rate.

The regression coefficients were developed using a least-squares fit of the equation to the data. Calibration of the model yielded the following form:

$$AR = 40.290\,(0.7329)^{L}\,(0.8497)^{S}\,(1.0132)^{LS}\,(0.7727)^{P}\,(1.0213)^{LP} \quad \textbf{(EQN 2.5)}$$

Corrections to the model were required due to the fact that the model reflected not only the effects of the three factors on safety, but all other potentially confounding variables as well. These corrections are documented in the subsequent section of this chapter. In addition, the predicted accident rate overestimates the effect of lane and shoulder improvements alone, since roadways with deficient cross-sectional attributes probably also have deficiencies in their general geometric design. Essentially, while this version of the model may predict a greater reduction in AR due to lane and shoulder improvements, other inferior geometrics (such as horizontal and vertical curvatures, and sight distances) will to some extent, reduce the effectiveness of those improvements. This additional issue required the use of accident reduction factors to eliminate the confounding effects of the confounding secondary variables.

Two permutations were made to the model incorporating an accident reduction factor. The first involved removing the external elements (those parameters that were outside of the scope of the model) based on Zegeer's key hypothesis- that as long as the predicted difference between before-and-after accident rates was small, the effects of all other "confounding" variables was also likely to be small (***Zegeer and Deacon, 1987***).

Conversely, when the difference between the two became pronounced (large safety gains), then the effects of the external elements also grew in magnitude. To account for this hypothesized effect, the following relationship was used to derive a correction factor to apply to the safety model:

$$ARF_a = (ARF_m)^C \qquad \textbf{(EQN 2.6)}$$

where, ARF_a = an estimate of the accident reduction factor that can actually be achieved by lane and shoulder improvements (adjusted ARF)
ARF_m = the accident reduction factor resulting from the application of the mathematical safety relationship model (unadjusted ARF)
C = the calibration constant used to modify the mathematical safety relationship

C, which was determined to be 0.4293 by a least-squares fit of the adjusted versus unadjusted ARFs was then applied to the mathematical safety relationship.

Because the previous adjustment yielded an equation that appeared to Zegeer and Deacon to underestimate actual improvement effects (in one case, a 3% reduction in accidents was predicted whereas actual improvements may have reduced accidents by up to 55%) , three further adjustments or assumptions were made, largely based on intuition and judgment. They were:

1. The assumption of a 20% decrease in accident rates when shoulder widths are increased from zero feet to eight feet of stabilized shoulders. This required adjustment of the model coefficients to reflect a 20% observed decrease in accident rate from the Kentucky data.

2. The differences in accident rates between paved and unpaved shoulders for mid-range roadway width were assumed to be true. This assumption was generally supported by the observed differences in accident rates experienced with paved and unpaved shoulders.

3. The equation was adjusted to match observed accident rates experienced for nine foot (2.74 m) lane widths. Because the majority of surveyed highway sections had nine foot (2.74 m) lane cross-sections, Zegeer and Deacon felt that the final model should reflect the significance of those sections.

The final model form was:

$$AR = 4.1501\,(0.8907)^{L}\,(0.9562)^{S}\,(1.0026)^{LS}\,(0.9403)^{P}\,(1.0040)^{LP}$$

(EQN 2.7)

However, the authors document certain limitations to this model. They are: (1) the model is applicable to lane widths between 7 and 12 feet (2.13 m and 3.66 m) and shoulder widths between 0 and 10 feet (3.05 m) only, (2) the relationships developed in the model are applicable for 2-lane, two-way primary or secondary roadways only, and (3) that the relationships cannot account for the effect of intersections on accident rates. Data used was from rural, homogeneous roadway sections.

2.6 MODIFIED GEOMETRIC ACCIDENT RATE MODEL

The final model form that Zegeer and Deacon developed (Equation 2.7) can be modified to account for sectional terrain type, as well as traffic volumes on the roadway (ADT) and roadside hazard severity (***TRB, 1987***). This model is different in that it includes a number of factors that were initially considered outside of the scope proposed by Zegeer and Deacon. Among the factors included is the roadside hazard severity- a measure of the relative safety of a highway cross-section pertaining to the roadside (side-slopes, ditches, obstructions at the roadside, etc.) and traffic volume (ADT). A report on the Safety Effects of Cross-Section Design for Two-Lane Roads includes a modified model based on Zegeer and Deacon's earlier work (***Zegeer et al., 1987***).

The database compiled for the model developed by Zegeer et al. was sizable, consisting of 62,676 accidents occurring over 4,785 miles (7,700 km) of rural 2-lane highway from seven states. Detailed quantifiable geometric data in this database included lane and shoulder widths, ADT, pavement type, horizontal and vertical curvature, and the locations of intersections, bridges, and driveways along each highway sub-section. Other geometric data collected and tabulated included terrain type, speed limit, sideslope ratios, and ditch types. This geometric data enabled the researchers to isolate specific types of accidents occurring with specific geometric conditions, and to incorporate this detailed knowledge of geometric contributing conditions into their accident rate model.

The following accident characteristics were compiled for each highway section of the geometric database: (1) the number of years of accident data (mainly five years), (2) total number of accidents, (3) a detailed breakdown of accidents by severity, (4) light conditions, (5) pavement conditions, (6) accident type, and (7) number of accidents involving fixed object collisions (***Zegeer et al., 1987***) From this, detailed accident rates were determined for use in their regression modeling.

Aggregate accident data by accident type were compiled to determine significant variables for use in the accident rate model. In addition, the design objectives of the model required that the following variables be included into the analysis: (1) highway traffic volume expressed as ADT, (2) lane and shoulder width, (3) pavement type, and (4) one or more quantifiable roadside conditions expressed as a roadside hazard rating (***Zegeer et al., 1987***). These quantifiable roadside conditions were represented by such factors as sideslope ratios, and roadside clear zones.

2.6.1 DETERMINATION OF SIGNIFICANT VARIABLES

Chi-square analyses were performed by Zegeer et al. on the various accident types to determine which accident types were the most correlated with the designed objective factors (lane width, shoulder width, shoulder type, roadside hazard rating, and traffic volume). Chi-square tests are used to determine the goodness of fit between observed, empirical data, and expected values based on a theoretical distribution (such as Poisson or Gaussian). The Chi-square test is defined as follows:

$$\chi^2 = \sum_{j=1}^{k} \frac{(X_j - np_j)^2}{np_j} \quad \textbf{(EQN 2.8)}$$

where, X_j = accident occurring corresponding to the specific geometric variable (observed)
n = number of accident cases, by type of accident
p_j = probability of an accident occurring within each highway sub-section, using the calculated accident rate (expected)

The Chi-square value determines differences between expected and observed accident rates in which case, a $\chi^2 = 0$ implies perfect correlation between observed accident occurrence for each accident type, and the corresponding geometric variable presumed to be significant in the accident. As the χ^2 value increases, the difference between observed and predicted values also increases. This implies that large χ^2 values are the result of a poor fit of predicted values with observed values and implies a poor correlation between the two.

Accident types that were tested for correlation are detailed in Table 2-1:

Table 2-1 Accident Types Tested for Correlation (Zegeer et al., 1987)

Accident Type	Description
1	Total accidents occurring per sub-section
2	Fixed-object collisions
3	Run-off-road (ROR) with a vehicle rollover
4	Other ROR
5	Head-on collisions
6	Opposing direction sideswipes
7	Same direction sideswipes
8	Rear ending
9	Striking parked vehicles
10	Accidents involving pedestrians, bicycles, or motorcycles
11	Angle or turning accidents
12	Train collisions
13	Animal collisions
14	Other and unknown categories
15	All single vehicle accidents
16	Selected multi-vehicle accidents

The χ^2 values for each accident type were found to be less than $\chi^2_{0.05}$ (tested at the 0.05 level of significance) in most of the test cases, implying an extremely strong correlation between the accident types and the corresponding geometric factors. However, χ^2 tests of significance are generally looked upon with suspicion when observed values agree so closely to expected values, since the implication of this small difference between predicted and observed accident rates is that there is a near perfect correlation between the two. Because perfect correlation between observed and predicted values are extremely rare, especially in accident rate samples, the small χ^2 values can be misleading.

A key element in determining the validity of the χ^2 value is the number of degrees of freedom, ν, for the observed sample. Sampling theory states that an observed sampling distribution (such as a sample of different accident types) will more closely approximate the χ^2 distribution for larger ν values, as detailed in Equation 2.9.

$$\nu = k - 1 \qquad \textbf{(EQN 2.9)}$$

where, ν = number of degrees of freedom
k = observed sample size

Because it is rare for observed sample distributions to closely match predicted sample distributions at high degrees of freedom, large samples (with large corresponding ν's) that also exhibit very low χ^2 values are looked upon with suspicion and are generally discounted (***Speigel, 1980***).

The apparent strong correlation between each accident type and the key geometric variables in Zegeer's study was felt to be a direct result of the large accident type sample sizes and judged to be unreliable. As a result, a matrix of contingency coefficients was used to measure the degree of association between the accident types and each geometric variable. The contingency coefficient is defined as:

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}} \qquad \textbf{(EQN 2.10)}$$

where, C = Contingency coefficient
χ^2 = Chi-square observed value for each accident type
n = Observed number of cases

(*M. Speigel, 1980*)

As the value of C grows larger, a greater correlation between accident type and each key geometric variable can be implied. The use of the contingency coefficient arises from situations where low χ^2 values are likely observed as a result of large samples sizes and not necessarily due to an valid relationship between the observed and expected values. Because of this issue, sample size, n, is introduced through the contingency coefficient to account for the role that it may play in the χ^2 values.

Accident types with contingency coefficients greater than 0.220 were considered by Zegeer et al. to be significant in context of the geometric variables being analyzed. The contingency coefficient value of 0.220 corresponded to the upper 33 percent of the contingency coefficients in the analysis, and was selected by Zegeer et al. because it represented the greatest one third of the correlation between accident type and the geometric variables. The accident types found by Zegeer et al. to have the greatest correlation with roadway features were those involving collisions with fixed objects, ROR accidents, and OD accidents involving head-on collisions and OD sideswipes.

Based on this analysis, the developed model was limited to predicting the following types of accidents only:

1. Single-vehicle (fixed object collision, or roll-over, or run-off-road)
2. Related multi-vehicle (head-on, sideswipe in both directions)
3. Total of the two previous types

To conserve model economy of variables, non-primary geometric variables were eliminated on the basis of the accident data review, the review of previous literature concerning accident rate modeling, and engineering judgment of the quality of each geometric variable.

A non-linear regression of each individual geometric variable was performed in relation to the accident types selected as being significant from the previous Chi-square analyses. Accident rates based on accident type were used as the dependent variable, and each geometric variable was the independent variable for each regression. Six variables were found to play a significant role in accident rates: (1) shoulder width, (2) median roadside hazard rating, (3) lane width, (4) median sideslope ratio, (5) percent of section with ≥ 2.5 degree or horizontal curvature, and (6) median recovery area distance. Because traffic volume (expressed as ADT) was a component of each dependent variable, it was not included in this regression analysis, but was later added on the basis of its strong correlation with accident rates observed in the study conducted by Smith et al. into the identification and quantification of 2-lane rural highway safety problems (***Smith et al., 1983***).

Again, Chi-square tests for contingency coefficients were performed on all available geometric variables to confirm their significance in relation to the key accident types. Variables identified as being significant included ADT, terrain, lane and shoulder widths, type of surrounding development, vertical alignment, horizontal curvature, roadside recovery distance, median roadside hazard ratings and the number of driveways per mile of sub-section.

2.6.2 MODEL FORMAT (ROADSIDE FACTORS INCLUDED)

On the basis of the tests of variable significance, and on Zegeer's past work, four general model forms were selected for evaluation. They were:

$$A/M/Y = C_0\,(ADT)^{C1}\,(C_2)^{W}\,(C_3)^{PA}\,(C_4)^{UP}\,(C_5)^{H} \qquad (1) \qquad \textbf{(EQN 2.11)}$$

$$A/M/Y = C_0\,(C_1)^{ADT}\,(C_2)^{W}\,(C_3)^{PA}\,(C_4)^{UP}\,(C_5)^{H} \qquad (2) \qquad \textbf{(EQN 2.12)}$$

$$A/M/Y = C_0 + C_1\,(ADT) + C_2W + C_3PA + C_4UP + C_5H \qquad (3) \qquad \textbf{(EQN 2.13)}$$

$$A/M/Y = C_0\,(ADT)^{C1}\,(W)^{C2}\,(PA)^{C3}\,(UP)^{C4}\,(H)^{C5} \qquad (4) \qquad \textbf{(EQN 2.14)}$$

where, A/M/Y =accidents per mile per year

$$= \frac{A}{L \times T}$$

A = Number of accidents
L = Highway sub-section length, in feet
T = Number of years of accident data
ADT = Bi-directional average daily traffic volume
W = Lane width, in feet
PA = Average paved shoulder width, in feet
UP = Average unpaved/unstabilized shoulder width, in feet
H = Median roadside hazard rating (quantifiable)

Three accident types were represented in each of the four models: (1) all single vehicle accidents (AS) involving fixed object collisions, and run-off-road incidents, (2) all single vehicle accidents, opposite direction head-on collisions, and all sideswipes (AO), and (3) total recorded accidents.

The effects of interactive terms (such as lane widths x shoulder widths) were also tested for model significance, but none of these terms were found by Zegeer et al. to significantly improve the model fit the observed data.

Again, as seen in the previous section, the location of the geometric coefficients were modeled in different positions to represent both a log-log relationship with accident rate, a log-linear relationship, as well as a normal linear relationship. Equation 2.11 and 2.12 represent variations of a log-linear regression model, while Equation 2.14 represents the log-log relationship. All three models can be linearized through a logarithmic transformation of the equation (***Rilett, 1988***). Equation 2.13 was included in the analysis to provide a linear relationship comparison to the other three logarithmic relationship models.

Models 1 and 2 were found to more accurately fit observed accident data better than models 3 and 4. Despite the fact that model 2 exhibited a higher R^2 value than model 1 (indicating a better fit of the model to observed accident data), model 1 was selected as the representative geometric accident rate model for two reasons: (1) R^2 values between model 1 and model 2 was small, and (2) the relative effects of lane widths, shoulder widths, and shoulder types on accident rates directly contradicted Zegeer et al.'s findings in their literature review. As a result, model 1 was selected as the best model.

The final model format included both median roadside hazard ratings (H) and terrain factors (TER1, TER2) to account for their intuitive effects on accident rates.

$$AO/M/Y = 0.0019\,(ADT)^{0.882}\,(0.879)^{W}\,(0.919)^{PA}\,(0.932)^{UP}\,(1.236)^{H}\,(0.882)^{TER1}\,(1.322)^{TER2}$$

(EQN 2.15)

where,

AO/M/YR =	Number of ROR, OD and same direction (SD) accidents per mile per year
ADT=	bi-directional average daily traffic volume
W =	lane width, in feet
PA =	paved shoulder width, in feet
UP =	unpaved shoulder width, in feet
H =	median roadside hazard rating for the highway section (this is a subjective measure in terms of 1 = least hazardous to 7 = most hazardous)
TER1=	1 for flat terrain, 0 otherwise,
TER2=	1 for mountainous terrain, 0 otherwise.

The R^2 value for this model (using AO as the significant accident type) was 0.456, and was within 5 percent of the highest R^2 value obtained (using total accidents, AT, as an accident rate measure, $R^2 = 0.463$) in Zegeer et al.'s analysis of the three types of accident categories. This model was chosen for the following reasons: (1) it included those accidents intuitively judged to be affected by roadway geometrics, (2) the R^2 value and coefficients appeared to agree with existing literature on geometric accident rate relationships, and (3) terrain effects could be included in the model.

2.6.3 MODEL VALIDATION

Zegeer et al. validated this model by randomly taking 75 percent of the observed database and deriving a log-linear equation from the regression. The geometric data from the remaining 25 percent of the database was entered into the final model form to obtain predicted accident rate values. By comparison of these two rates (observed and predicted) for the smaller dataset (25 percent sample), the standard deviation of 0.3607 was found to be under half of the observed mean accident rate of 0.73 accidents per mile per year. Plotted residuals yielded large deviations only at the extreme accident rates (near zero and extremely high rates) and was considered to be within parameter tolerance.

Because of the many confounding and possibly secondary variables involved in accidents, the issue of multicollinearity was also addressed. Multi-collinearity was defined as the difficulty in separating the effects of two independent variables due to an extremely high correlation between the two (***Zegeer et al., 1987***). A correlation matrix was set up between each of the variables used in the model. The results of this matrix identified no high correlation between any two factors, with the highest observed correlation being paved shoulder width (PA) and unpaved shoulder width (UP) at -0.513. As a result, multicollinearity was not considered to be significant in this analysis.

2.7 SUMMARY

This chapter explores both the advantages and short-comings involved in developing safety-geometric relationships. The need for quantifiable relationships between accident rate and highway geometric design factors is a result of the escalating costs associated with highway construction and 3R/4R projects. By developing models that predict the safety benefits (primarily measured in terms of accident rate reduction) achievable through selective and incremental improvements to roadway and roadside geometrics, an accurate picture of the cost of highway safety can be realized. As a result, potential savings in the design and implementation of 3R/4R projects may be realized by allocating project dollars towards a tangible safety gain. Among the most important safety-geometric relationships is the relationship between accident rates and cross-sectional geometric factors.

However, because highway geometrics are only one element of general highway safety, obstacles such as unquantifiable and non-geometric related factors can affect the identification of a safety-geometric relationship. Nonetheless, several key geometric factors have been found by prior research to affect highway accidents. They are: roadway width, roadside geometrics, bridge widths, horizontal alignment, stopping sight distances, intersections, pavement surface condition, and pavement edge drops. These factors appear to influence accident rates and safety in a greater manner than any other geometric factors.

Because of the large volume of factors capable of affecting highway safety, the scope of this research project is limited to examining the effects of the cross-sectional geometric factors on safety and accident rates. As explained previously, research into these geometric factors can be considered independent of the other factors, since the effects of intersections, bridges, horizontal, and vertical alignment on accident rate do not overlap with the effects of cross-section, in general.

While this chapter identifies in detail, prior research into the effects of cross-sectional geometric design elements on accident rates and safety in general, it does not effectively address the current state of accident-geometric modeling, the application of such models to engineering design and evaluation, and the impact that the models presented in the chapter have on 3R/4R geometric design. Chapter 3 of this thesis addresses these issues, and attempts to identify prevalent models and methodologies used by North American state and provincial transportation agencies to analyze accidents and geometrics. It is hoped that the identification of the most widely used safety-geometric models or methodologies by transportation agencies will aid in identifying suitable models to calibrate within Alberta.

Chapter 3

A Survey of Principal North American 3R/4R Safety-Geometric Design Practices

3.0 INTRODUCTION

A feasibility study on the development of a quantitative relationship between roadway geometric factors and collision rates was undertaken by the University of Alberta and Alberta Transportation and Utilities in order to identify potential 3R/4R (Restoration, Rehabilitation, Resurfacing, and Reconstruction) geometric design guidelines for use in Alberta (***Alberta Transportation and Utilities, 1995***). As part of the project, a survey of all state and provincial transportation agencies within the United States and Canada was conducted to obtain information on procedures for incorporating safety considerations within the 3R/4R geometric design process. Of the 64 agencies surveyed 45 responded with completed forms. The survey consisted of questions regarding the use of accident analysis data and safety design concepts in 3R/4R design guidelines and was divided into four separate sections: 1) general information on 3R/4R design standards (if any), 2) methods of accident data acquisition, 3) accident data analysis, and 4) the use of safety-geometric models to identify and quantify the relationship between safety and roadway geometric features.

This chapter focuses on key geometric design elements of the survey results, primarily by highlighting the similarities and differences of regional design practices within North America. The various agencies that currently undertake 3R/4R project work and any associated safety considerations for doing so were identified from the survey results. Lastly, agencies that use safety-geometric models in their safety assessment process were identified and their safety-geometric collision modeling practices examined in detail. The identification of these models were a key component in the selection of suitable safety-geometric models for use in Alberta.

3.1 OVERVIEW OF SURVEY

Stage two of the feasibility study consisted of a mail-back survey of all state and provincial transportation agencies in Canada and the United States. The survey was tailored to identify the state of 3R/4R practice among these jurisdictions, their methods of accident (or collision) data collection and application to safety concerns, and the state of their own usage of geometric collision modeling. This information was necessary to help identify current practices of other agencies that could possibly be applicable to Alberta Transportation and Utilities safety-geometric design process.

3.1.1 Survey Methodology

The survey described above was administered over an approximate one month period in September, 1994 and was composed of four main parts. Part A outlined questions concerning the availability and usage of 3R/4R geometric design standards within the particular jurisdiction. These questions were designed to identify agencies that currently use 3R/4R design procedures and any standards associated with these procedures.

Part B of the survey focused on the area of accident data and its method of acquisition. Because highway accident data is integral to statistical safety analysis, it is important to accurately record as many factors as possible that contribute to an accident. Questions from this section included a survey of accident factors that each agency records, methods of recording collision severity, and the availability of a Geographical Information System database for localizing accidents and integrating geometrics into an analysis.

The applications to which the data was used was the focus of Part C of the survey. In particular, this section identified which agencies conduct an analysis of highway accidents at various levels of detail. Part D then focused on a detailed examination of advanced techniques used in safety analysis by the agencies that replied to Part C. Specifically, Part D was designed to identify agencies that use a type of safety-geometric collision model and the type or method that they use.

3.1.2 Survey Response

Of the 64 transportation agencies that received surveys 45 responded. This translates into a return rate of approximately 70%, and together with the fact that surveys were received from all the major geographical locations in North America, the results were determined to be an adequate sample of 3R/4R design practice within North America. Figures 3-1 and 3-2 on the following pages identify those states and provinces within the geographical regions of North America that responded to the survey. Return rates from various geographical regions in North America ranged from a low of 55% in the U.S. South region, to 100% in the U.S. Southwest and the Canadian Central regions. The return rate by region is listed in Table 3-1.

TABLE 3-1 Survey Return Rates by Geographical Region in North America

Geographic Region	Number of Responses	Response Rate
U.S. Northeast	12/16	75
U.S. South	6/11	55
U.S. Central	7/9	78
U.S. Northwest	5/6	83
U.S. Southwest	6/6	100
Canadian Atlantic	3/4	75
Canadian Central	2/2	100
Canadian Prairie	1/2	50
Canadian Pacific	1/1	100

The survey was distributed in August, 1994, and respondents were given until September, 1994 to reply. Due to the use of pre-paid mail-back envelopes for the United States and Canada, a high number of responses was expected and achieved. Contact was initiated at the senior management level for each transportation agency to facilitate

Figure 3-1 Distribution of Survey Results - Canada

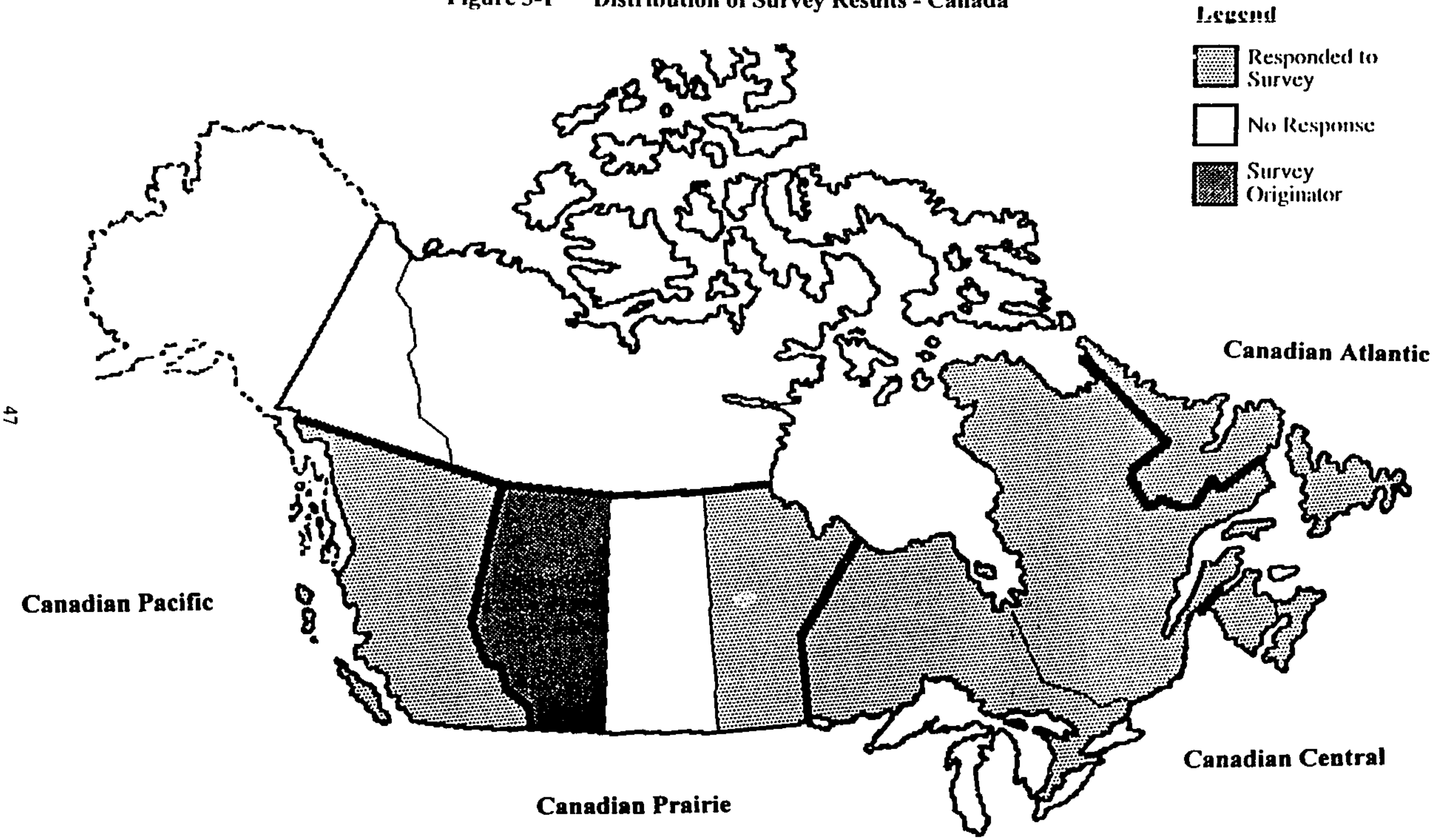
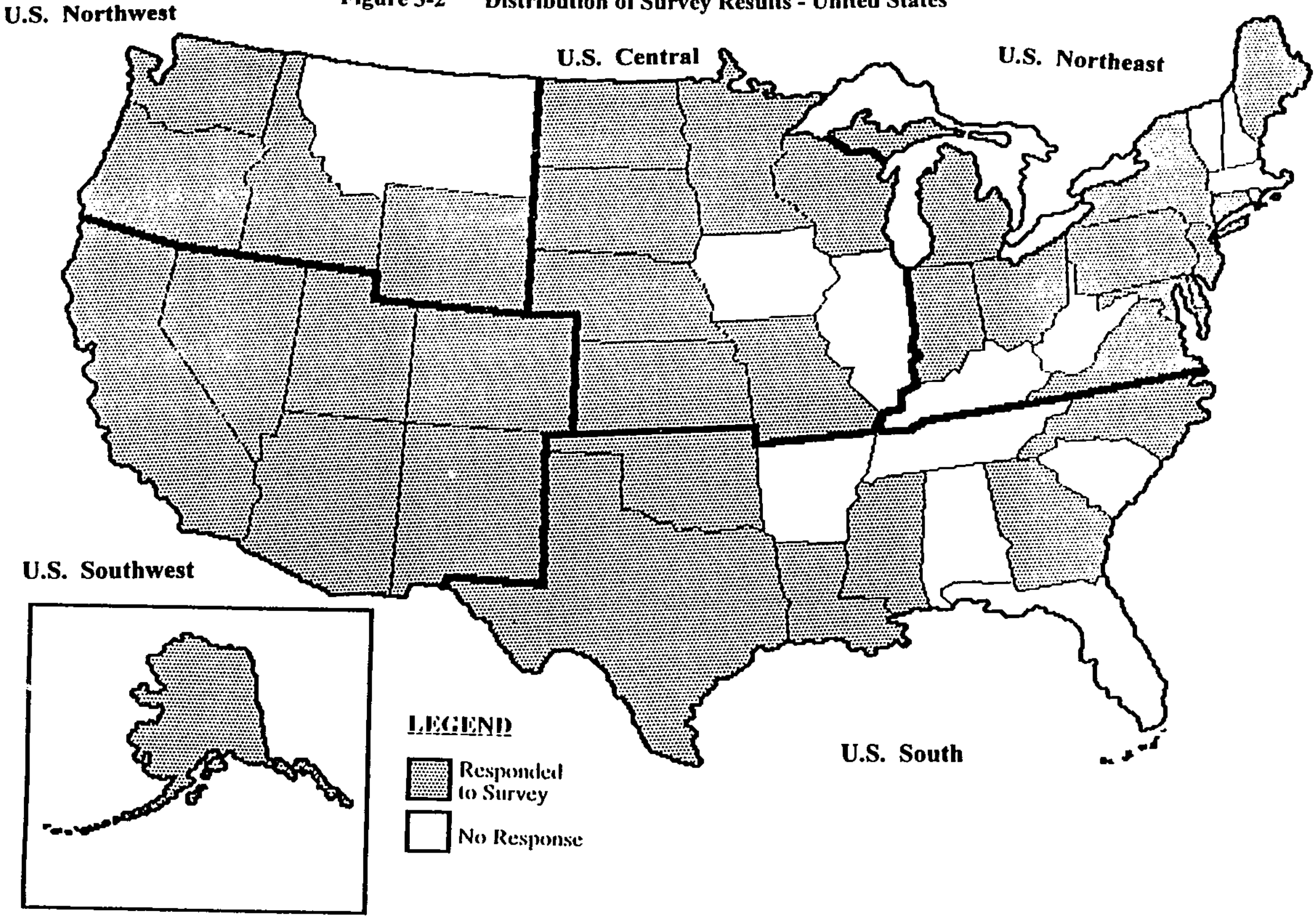


Figure 3-2 Distribution of Survey Results - United States



distribution to the proper staff in each agency. Information on the staff who replied was collected in order to facilitate future contact with each agency. However, it should be noted that these findings reflect only the views of those surveyed.

3.2 SURVEY FINDINGS

The key findings of the survey are documented in the following sections. These findings are summarized into three main areas: 1) use of 3R/4R design procedures in North America, 2) Accident data analysis within 3R/4R design, and 3) Safety-geometric accident model usage in North America.

3.2.1 USE OF 3R/4R DESIGN PROCEDURES IN NORTH AMERICA

This section examines current practice with respect to 3R/4R design practices in North America. The results of the first three questions of the survey are presented and discussed. The results of the first question of the survey - *Are 3R/4R geometric design standards currently being used within your jurisdiction?* - are presented in Table 3-2.

TABLE 3-2 3R/4R Geometric Design Standard Usage

Geographic Region	Number of Positive Responses	Response Rate (%)
Uses 3R/4R guidelines (United States)	36 / 36	100
TRB or National guidelines	26 / 36	72
State originated guidelines	28 / 36	78
combination of guidelines	12 / 36	33
FHWA guidelines	4 / 36	12
AASHTO guidelines	6 / 36	18
Uses 3R/4R guidelines (Canada)	1/7	14
TRB or National Guideline	1/7	14
Uses 3R/4R guidelines (Other)	2/2	100
TRB or National Guideline	1/2	50

Table 3-2 shows that 39 of 45 respondents confirmed that they use some type of 3R/4R geometric design standard within their agency. Every American state surveyed reported that 3R/4R guidelines were being used. The majority of these standards used were based either on a nationally developed guideline: Transportation Research Board (TRB), Federal Highway Administration (FHWA), American Association of State Highway and Transportation Officials (AASHTO), or a combination of the national guidelines and state developed standards. Within Canada, every province except for Nova Scotia reported the absence of any 3R/4R design standard or guideline.

The widespread use of 3R/4R design standards and guidelines throughout the United States is dramatically different from standard practice in Canada. This may be due to the lack of large-scale Canadian federal funding of 3R/4R programs. The use of federally mandated 3R/4R standards and guidelines was found to be uniform across all of the U.S. geographical regions. It should be noted that many states incorporate state developed guidelines into their 3R/4R design policy.

The second question of the survey -- *What geometric road features are incorporated within the 3R/4R design standards? What factors (Agency, User, Safety, Other costs) are considered for each design?* -- sought to identify specific geometric features that each agency applied to their 3R/4R design. Geometric design features surveyed from each Department of Transportation's design standards included: horizontal and vertical curvatures, road and shoulder widths, roadway cross-sections, sideslope ratios, roadside object location, and pavement edge drops. Figure 3-3 identifies the number of agencies (of the 36 previously identified as using 3R/4R guidelines) that incorporate each specific geometric feature into their standards.

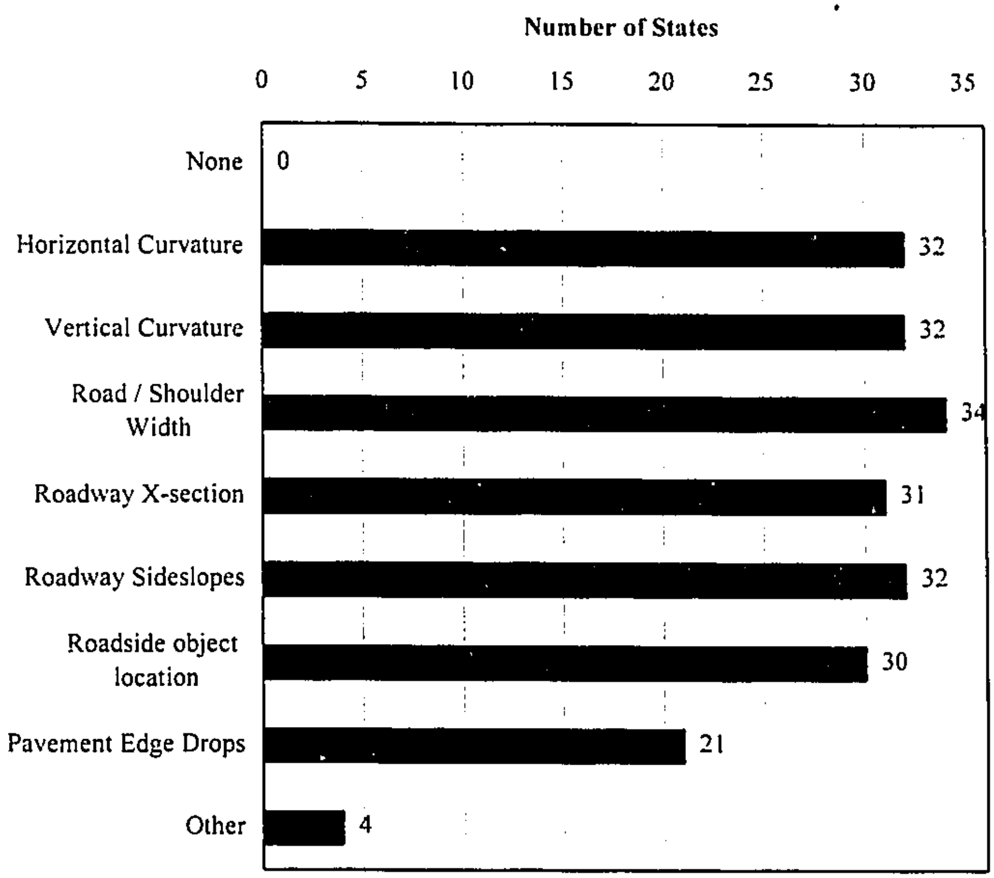


Figure 3-3. Geometric Features Incorporated in 3R/4R Design - United States

It was found that all regions in the United States incorporate a majority of the roadway design features identified above into their 3R/4R programs. This was as expected given the well know relationships between these factors and collision rates and severity. A significant number of states, however, do not use Pavement edge drops in their 3R/4R design process (15 of 36 or 42%). This statistic is interesting because pavement edge discontinuities have been identified as being potentially hazardous (***TRB SR214, 1987***).

Of the different types of costs associated with each feature, 27 of 36 U.S. state transportation departments were identified that consider agency costs, or the direct cost to the agency for implementing a particular geometric feature, as playing a dominant role in 3R/4R design. The results are summarized in Figure 3-4.

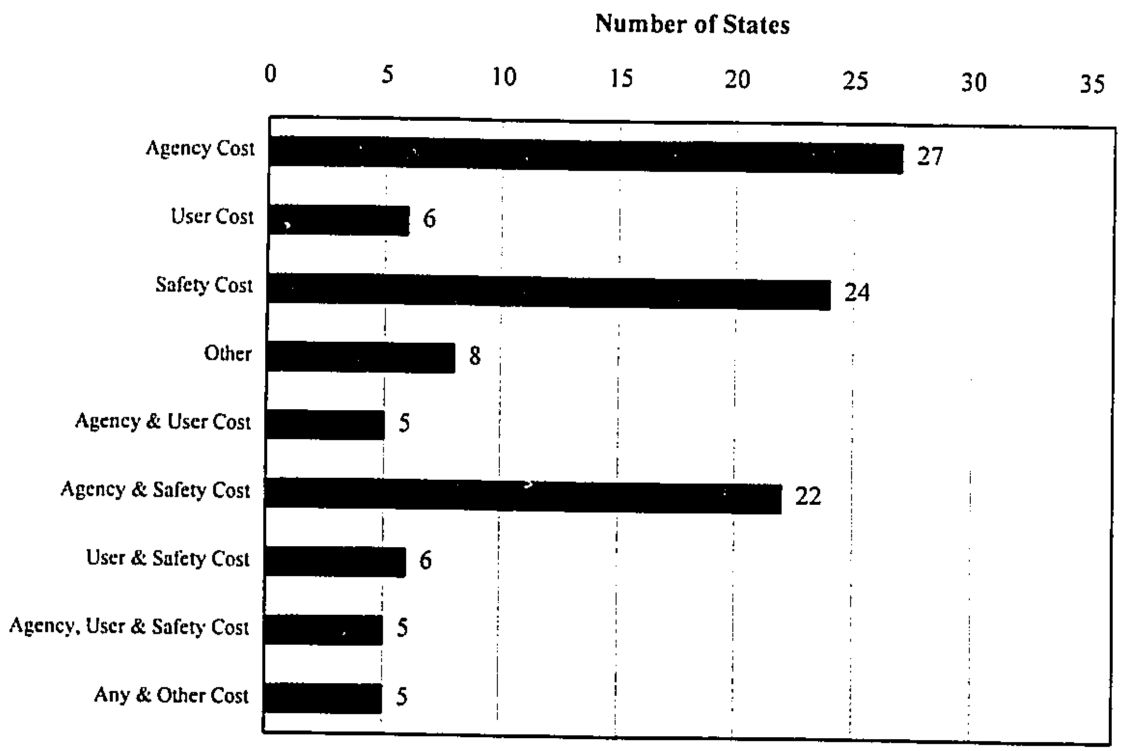


Figure 3-4. Cost Factors Considered in 3R/4R Geometric Design - U.S.

Safety costs reflect the "expense" of not implementing a design in terms of roadway safety. These "expenses", such as increased accident frequencies and associated deaths and injuries, are usually converted to a dollar equivalent using predetermined factors. It was found that 24 of the 36 respondents considered safety costs in their design, and 22 of the 36 agencies used both agency and safety costs in evaluating their 3R/4R geometric design. It was interesting to note that only 6 of the 36 agencies considered user costs in their designs. Users are generally classified as the driving public and the user costs reflect the direct expense to this population associated with a particular design. Again, any non-monetary externalities are typically converted to a dollar equivalent through the use of predetermined factors. It should be pointed out that those agencies that did consider user cost in their design, used it in conjunction with either agency, safety costs, or both.

While 3R/4R design practice in the United States generally encompasses agency and safety costs, practice can vary between regions. The predominant practice in the U.S. Northwest region is to use both agency and safety costs in their design process. Only Alaska and Oregon considers user cost in their design. In the U.S. Southwest region, while the use of safety costs is widespread, agency cost is only considered in California and Nevada. The U.S. central region is similar to the U.S. Northwest, except that safety costs are not a key factor in 3R/4R design in Wisconsin and Missouri. Practice in the U.S. Northeast and South also resembles the U.S. Northwest region closely, with agency and safety cost being predominant. Only Indiana, Maryland, and the District of Columbia consider user cost, in addition to agency and safety costs.

The final question in this section -- *Question 3: What type of analysis is done to select the most suitable 3R/4R projects?* -- attempted to identify the techniques used by various agencies for identifying their 3R/4R investment program. The responses to this question are summarized in Figure 3-5 which shows the type of analysis method as a function of geometric region.

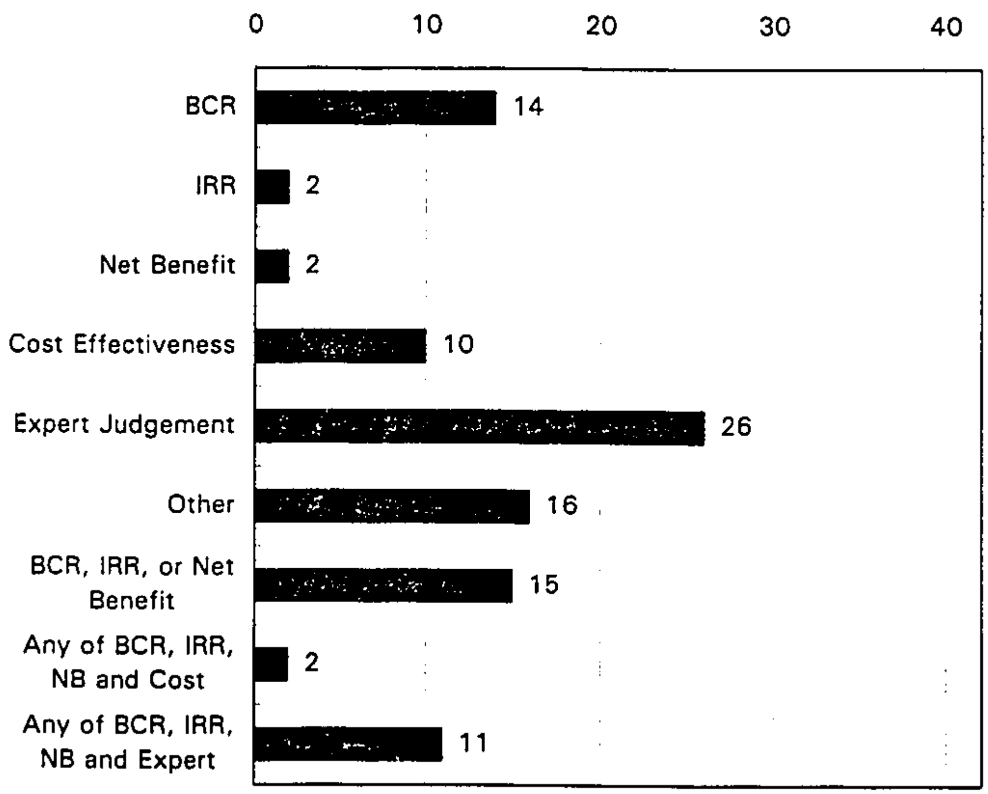


Figure 3-5. Analysis Method Considered in 3R/4R Geometric Design

By far the most popular method of selection is the reliance on expert judgment, with the more traditional engineering economic analysis methods (Benefit Cost Ratio or

BCR, Internal Rate of Return or IRR, Net Benefit methods) placing second. Cost effectiveness of a project was used by only a few jurisdictions as a criteria in the selection of projects. It is important to note that most agencies use a combination of the standard engineering economic methods and expert judgment to select projects. Only Nova Scotia, Georgia, and Ohio rely solely on expert judgment to determine 3R/4R project ranking. These results indicate the selection of 3R/4R projects continue to rely in some part, on the use of qualitative measurements (judgment) in project prioritization. This is probably related to the fact that many of the inputs to the analysis (i.e. level of safety increase as a result of a geometric improvement) and their associated monetary value are often not well defined. Consequently, it would be expected that the engineers would attempt to keep some level of engineering judgment in the final project selection phase. Although not surveyed, the presence of political and interest groups in the decision-making process offers a fourth method of project selection. Decisions based on political or interest group criteria often may be critical in determining a 3R/4R project's priority and funding. However, this choice was not included since it does not involve some type of engineering analysis.

3.2.2 ACCIDENT DATA ANALYSIS WITHIN 3R/4R DESIGN

The second stage of the analysis attempted to document the various state and provincial methods of accident data collection and the extent of computerization of their respective accident databases. In addition, the use to which this data was used within 3R/4R design practices was also explored.

Figure 3-6 identifies the results of Question 8 of the survey -- *Is a GIS (Geographical Information System) database available/in use? Are recorded accidents localized on the GIS database?* -- which identified the availability and usage of a GIS database with safety-geometric design features in the United States.

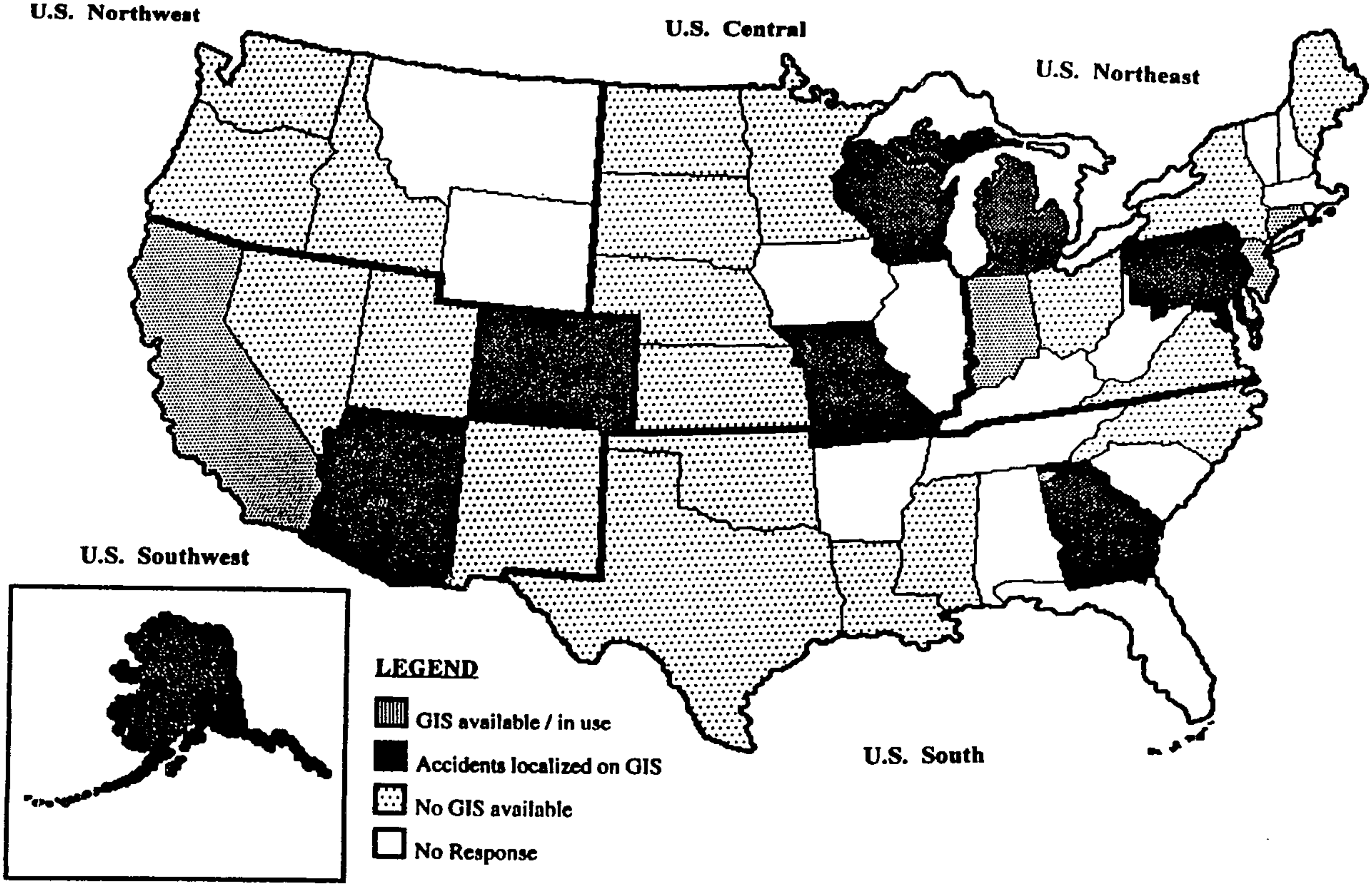
Within Canada, only Manitoba and Newfoundland have indicated the availability of a GIS database for use in roadway design. In addition, Newfoundland localizes accidents on their database. It was found that U.S. states have a much wider proliferation of GIS database usage in highway geometric design, with 13 of 34 states reporting the use of GIS. Of these, the majority (9 of 13) also use the GIS to localize accidents on the database. While states that exhibit a large and developed transportation roadway network such as Ohio, California, and Michigan have a GIS database, other notable states such as New York and Texas do not operate a GIS database for 3R/4R design and accident analysis. It is expected that North America-wide, the use of a GIS database will continue to grow and will be an important tool in 3R/4R design and safety analysis.

Table 3-3 tabulates the results of question 9 -- *Are accident data currently used for 3R/4R geometric assessment?* -- which attempts to identify the use of accident data within the 3R/4R design process.

TABLE 3-3 Accident Data Usage In 3R/4R design

Geographic Region	Number of Responses	Response Rate (%)
Uses accident data (U.S.)	32 / 36	88
Uses accident data (Canada)	2 / 7	29
Uses accident data (Other)	1 / 2	50

Figure 3-6 GIS Database Usage - United States



It was found that 35 of 45 respondents use collected accident data in some type of 3R/4R geometric assessment or design. This assessment techniques that are used included ranking accident location risk by i) number of collisions, ii) collision rate, iii) factors. As can be seen in Table 3-3, 32 U.S. states use accident data in some form of geometric assessment. The only exceptions to this are were Connecticut, Indiana, Nebraska, and the District of Columbia. Because of the limited usage of 3R/4R guidelines in Canada, few provincial jurisdictions use accident data for 3R/4R assessment and design. It is anticipated that this difference in usage may change when and if Canadian provinces implement 3R/4R guidelines into their respective roadway programs.

It was found that 24 of the 35 agencies that utilize accident data in their respective 3R/4R geometric and design use some type of risk categorization to rank roadway locations. The breakdown by method is shown in Figure 3-7.

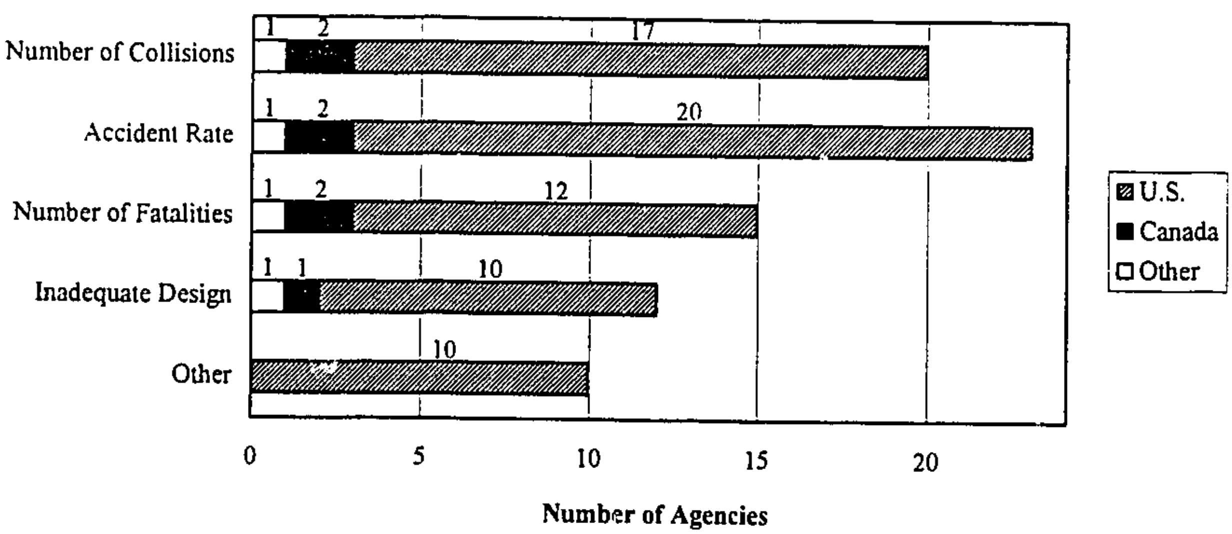


Figure 3-7. Risk Ranking Factors in 3R/4R Geometric Design

It may be seen that the most popular method of ranking is to use accident rates as a gauge of an individual roadway location's hazard level. The use of the numbers of collisions (similar to accident rates) is nearly as widespread as the former. Less used as gauges of the risk (or hazard) that a roadway section exhibits are numbers of fatalities, and design inadequacy.

It should be noted that the common practice is to use a combination of the above metrics in identifying hazardous locations. This would be expected given that no ranking system is universally accepted and each has a number of advantages and disadvantages. The results were cross-classified to indicate which combination of techniques were utilized most often and the results are presented in Figure 3-8.

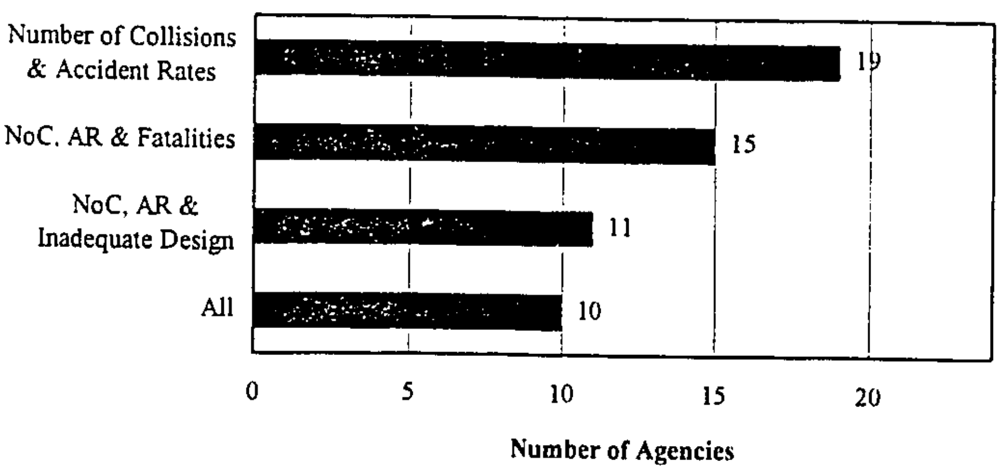


Figure 3-8. Risk Ranking Factors - North America

The use of both numbers of collisions and accident rates was identified as the predominant method for identifying hazardous locations in that 19 of the 24 agencies used both. Over one third of all agencies reporting some type of ranking system used all methods represented to a certain extent.

3.3 Safety-Geometric Collision Model Usage in North America.

It was found that 12 of the 43 respondents use a safety-geometric collision model in their safety and 3R/4R design analysis. As well, the state of Virginia is currently assessing the need for future modeling in this area. While differences do exist between states that use such a model, there are a number of similarities among the types of model that have been adopted by the different agencies. An overview of the different safety geometric models is presented in the next section.

1. California ***(Unpublished data. California Department of Transportation, 1994)***

While the California Department of Transportation (Caltrans) indicated that some type of safety-geometric accident model was being used, they failed to identify in detail the type of model used. Based on their description, it appears Caltrans uses a type of rate quality control model such as a regression model. Their model identifies accident prone locations at a state-wide level, as well as an individual project level. Caltrans relates a proposed 3R/4R project location's accident history with similar control locations to determine a relative magnitude of accident history for that site. The location information, primary accident event descriptions, impact severity, light, road physical, environmental, and surface conditions are then used within the model to predict accident rates and severity. A three year block of accident data is used in assembling a database for their safety analysis with the thought being that this time frame allows designers to account for the statistically random nature of accidents (and statistical regression-to-the-mean effects). No adjustments are made to the database such as the removal of accidents caused by animal strikes, alcohol related accidents or suicides).

The advantages of using a rate-quality control method are the relative ease of use and the ability to develop quick results without intensive numerical methods. This method yields a relatively reliable estimation of hazardous locations. However, disadvantages of this method may be in the inherent limitations of the model structure itself. Even one accident at a location may trigger the quality control model (over a 1:1,000,000 background accident rate, for example) even if no deficiency or even change in roadway geometrics exist. This occurs simply because of the extremely infrequent nature of accidents. As well, non-geometric related accidents can trigger a 3R/4R response to the roadway location, since the majority of roadway accidents will involve some non-geometric related factor. As a result, in order to obtain even a reasonably predictive result, a sizable accident database (usually five or more years) may be required, even to facilitate the removal of certain non-geometric related accidents.

2. Colorado ***(Unpublished data, Colorado Department of Transportation, 1994)***

The Colorado DOT uses both a rate quality control technique and a Poisson regression model to identify accident prone locations. From the following factors used in the models -- location information, primary accident event, impact severity, speed, types of vehicles, driver action and condition, lighting conditions, physical, environmental, and surface conditions of the road, -- the number, rate and severity of accidents are predicted for a given location. As with California, a three year block of accident data is used in their models, with no adjustments made to the database to account for certain types of accidents.

3. Idaho ***(Unpublished data, Idaho Department of Transportation, 1994)***

Idaho uses the ROADSIDE modeling program and a safety/cost benefit analysis to determine accident prone locations. Using location information, primary accident events, impact severity, driver action, lighting condition, and physical and surface conditions of the road, an accident reduction rate is predicted. Data used in the model is assembled over a three to five year time frame, with suicides removed from the analysis. As well, intersection and related accidents are only used when dealing with intersection design. The advantage to removing accidents such as suicides is that since they do not appear to be geometric related, they are not considered for a geometric design solution. However, disadvantages to this include the fact that the removal of these accidents may often yield too small of a database to draw significant conclusions. As well, geometrics may contribute in some part to the accident (animal strikes may be caused by poor stopping sight distances, for example).

4. Maine *(Unpublished data, Maine Department of Transportation, 1994)*

Maine typifies American state methodology of accident geometric modeling by using a rate quality control method to identify accident prone locations. MDOT is able to determine a predicted number and rate of accidents. Data used in the model is assembled from a three year block of accident data that includes such details as location information, primary accident event type, severity, type of vehicles involved, driver condition, light conditions, physical, surface, and environmental conditions of the road, vehicle mechanical condition, as well as both a diagram and written descriptions. As with Idaho, suicides are the only accident type removed from the analysis.

5. Maryland *(Unpublished data, Maryland Department of Transportation, 1994)*

While similar to the other states using a rate quality control model to identify accident-prone locations (black spot analysis), the Maryland DOT also uses generalized linear and Poisson regression models to calculate parameters for their existing models. The use of regression models seem to indicate that the Maryland DOT model parameters can be continually refined to fit accident data as it changes. Data used in black spot analysis is selected from a three year block of accident data, and has all intersection accidents removed for the purpose of studying high accident sections only. Maryland's model utilizes the same types of input data and produces the same results format (identification of number and rate of accidents) as Maine.

6. Michigan ***(Unpublished data, Michigan Department of Transportation, 1994)***

Michigan adopts a somewhat different approach to geometric collision modeling. Primarily using a cost/benefit analysis of improvement vs. crash reduction predicted (supplemented with an IRR/time of return project ranking) method, they also identify accident prone locations through a rate-quality control approach. Using among the most detailed accident data, these models can predict not only accident rates and severity, but crash types as well. Data is used from a five year time frame, with no specific types of accidents removed before use in the model.

In addition, research is currently being conducted into tight diamond interchange sight restrictions, "wye" intersection designs, center lane left-turn crash reduction strategies, and median width designs.

7. Minnesota ***(Unpublished data, Minnesota Department of Transportation, 1994)***

The Minnesota DOT uses the ROADSIDE modeling program to identify accident prone locations. Their methods appear to be similar to those used by Idaho, with the primary difference being that the model results are in the form of a predicted number of accidents and accident severity.

8. New Jersey *(Unpublished data, New Jersey Department of Transportation, 1994)*

The New Jersey DOT (NJDOT) also uses the ROADSIDE modeling program to identify accident prone locations. However, NJDOT does not evaluate highways state-wide. Rather, they focus on design and analysis on a case-by-case, location specific basis. The FHWA Utility Pole Accident Countermeasures Evaluation (UPACE) Program is also being used to model roadside hazards. Further research into the use of 3-D geometrics for safety design is also being considered at this time.

9. North Carolina *(Unpublished data, N.C. Department of Transportation, 1994)*

The North Carolina DOT specifically uses a cost/benefit analysis package to determine the suitability of sideslope improvements over the installation of guard-rails. Their method identifies high accident prone locations through a predicted number and rate of accidents. A three to five year block of accident data is used and no accident types are removed from the database before modeling.

10. Ohio *(Unpublished data, Ohio Department of Transportation, 1994)*

The Ohio DOT plots and analyses accidents to locate potentially hazardous locations and their causes (likely a rate quality control method). Their methodology is similar to states mentioned previously. Accident rates can be predicted from this type of black-spot analysis. As with several other states, intersection accidents are removed from the analysis and treated separately. Accident data is used from the prior three years.

11. Virginia *(Unpublished data, Virginia Department of Transportation, 1994)*

While the Virginia DOT does not currently use any type of safety-geometric accident model, there is currently a proposal to conduct before and after highway studies. No further details were provided in the survey response.

12. Washington *(Unpublished data, Washington DOT, 1994)*

Possibly among the most advanced of the states surveyed in accident modeling, the Washington DOT is currently testing a Poisson regression and a negative binomial regression approach to geometric collision modeling. These models currently are being developed and calibrated for state-wide use. This research is being conducted in conjunction with the University of Washington's research into "Accident Prediction Models Based on Highway Geometrics", and "Accident Prevention at Intersections". Accident data is used from the prior two years, and has animal strikes, pedestrian accidents, and suicides removed from the database before use. Location information, primary accident event type, and the estimated traveling speed of the vehicle(s) data are used in the model.

13. Nova Scotia *(Unpublished data, Nova Scotia DOT&C, 1994)*

The Nova Scotia Department of Transportation uses a rate quality control model to identify high accident prone locations. This was, by far, the predominant method of analysis for the agencies surveyed. The model utilizes location data, impact severity and AADT data collected by their Traffic Engineering department, and predicts an accident rate for a given location. Further research is also being conducted by their Traffic Engineering department on comparing predicted and actual accident rates province-wide.

3.4 CONCLUSIONS

While it should be stressed again that these findings reflect the views of those surveyed only, several key points were identified from the survey results:

1. Much of the research and work conducted in the area of 3R/4R geometric assessment and design originates from the United States. Most standardization of guidelines is done at the national level through American federal transportation agencies. Canadian provinces generally lag behind their American counterparts in the development and use of 3R/4R design guidelines.
2. The use of GIS and computerized accident databases within North America is currently not widespread. Only 13 of 34 reported states and only two Canadian provinces currently use GIS. However, those agencies that do operate a GIS system for 3R/4R work generally also use their database for accident analysis. The level and method of use within Canada is comparable to the U.S.
3. Most state agencies conduct accident analysis in their 3R/4R geometric design process. The method of analysis varies by jurisdiction; however, a sizable majority of states use some type of risk ranking factor by location (generally through accident rates) in their 3R/4R geometric assessment of roadway sections.

4. The level of research into safety-geometric collision modeling varies widely between agencies. While state Departments of Transportation (DOTs) such as California, Colorado, and Michigan conduct or sponsor research in this area, other agencies such as Maine DOT and Washington DOT have also been identified as pursuing model research. This may indicate that a continent-wide survey of agencies may be the best method of locating most agencies conducting this type of research. Few Canadian provinces have been identified as conducting research in this field. The current trend in safety-geometric collision modeling is in the area of generalized linear modeling, primarily with rate-quality control models or some type of other regression model. Growing demand in this area for better modeling techniques may be reflected in the number of agencies identified that are undertaking research into various types of regression modeling, as well as an Empirical Bayesian approach to accident rate modeling.

Further research into the development and validation of generalized linear models for safety-geometric relationships is enhanced by the results of the survey and prior research. The model identified from the feasibility study as being potentially applicable within Alberta was a rate-quality control model, also known as a generalized linear regression methodology. Chapter 4 explains the theory behind regression analysis, concepts in generalized linear modeling, and the reasoning behind the selection of the model to be used in later chapters.

Chapter 4

Generalized Linear Models and Regression Analysis for Cross-Section Geometrics

4.0 INTRODUCTION

Generalized linear models can be effective tools in establishing the relationships between highway safety (measured in terms of accident rate) and geometric roadway features. Linear models are used to define mathematical functions (i.e. linear functions) that represent relationships arising in correlated statistical data. It should be noted that any data used in generalized linear modeling must either be linearly related or transformed into a linear form. Models in the form of log-linear or other multiplicative functions are intrinsically non-linear, but can be transformed into linear functions by taking the natural logarithm of both sides to simplify curve-fitting (***Rilett, 1988***). Based on prior research, cross-sectional geometric design elements appear to have a generalized linear relationship with safety and accident rates (***TRB, 1987***).

The advantage of linear modeling over methods such as an Empirical Bayesian approach to modeling, is that deriving the linear function from relationships in the data is often less numerically intensive than other approaches. However, this numerical approach to mathematical modeling has its limitations, as introduced in Chapter 2, and expanded upon in Appendix B. This appendix introduces the concept of Empirical

Bayesian modeling and explains possible limitations of the generalized linear model in quantifying a safety-geometric relationship.

4.1 GENERAL MODEL THEORY AND STRUCTURE

The form that generalized linear models take is shown in equation 4.1. This equation relates a dependent variable, Y, in terms of a series of functions involving independent variables, X, and a random error component, E.

$$Y = a + bX_1 + cX_2 + ... + zX_n + E_1 + E_2 + ... + E_n \qquad \textbf{(EQN 4.1)}$$

where, $Y =$ dependent variable of the relationship,
$a, b, c =$ coefficients that describe each unique linear function, and
$X_1, X_2, X_n =$ dependent variables or parameters of the relationship
$E_n =$ random error component (for each X_n)

Determining the mathematical relationship between safety and highway geometrics in a modeling context can be found through the statistical process called regression. Regression is the estimation of the dependent variable from a series of independent variable(s) or parameters. One of the most common methods of regression used is the method of least squares estimation. The method of least squares is based on the principle that a "best-fit" line or relationship is one that minimizes the squared sum of all distances between any data point in the set and the line itself. This method is also useful when dealing with multiple regression. Rather than a regression line, a regression plane (two factors), a regression space (three factors), or a regression continuum (more than three factors) is the solution form.

Regression modeling can be used with Equation 4.1 to identify the solutions for the dependent variable, Y, given a series of parameters, X, and a set of coefficients (a, b, c) pre-determined from regression analysis. In essence, it is possible to determine likely values of Y (or accident rates, in the case of this thesis) given a set of conditions X (or quantified geometric factors).

Four assumptions concerning the properties of a generalized linear model are made in order for the regression analysis to be meaningful. They are:

1. The probability distribution of the random error components, E, each have mean values of zero over an infinitely long sampling period. This assumption states that the values of all positive and negative sampling errors for each independent factor effectively negate each other over a long period of time. Essentially, by sampling over an extended period, it is assumed that any bias due to random error is negated and insignificant to the model.
2. The variance of the probability distribution of E is constant over all values of the independent variables, X. This assumption is important since it implies that no bias exists between varying independent factors, X (for example, that the random error in X_n is equal to the random error in X_{n+1}). If this assumption is not valid for a regression, then the bias can be accounted for by weighting each independent term in proportion to it's bias in error.

3. The probability distribution of the random error components, E, are normally distributed. This distribution implies that the frequency of positive error occurrence in each independent variable is equal to the frequency of negative error occurrence and that no bias will exist due to positive or negative error.

4. Random errors, E, associated with each independent variable, X, are themselves, independent. The importance of this assumption is to prevent errors of one variable from influencing the errors of another variable and introducing potentially confounding effects into the regression analysis.

In practice, these four assumptions do not need to be satisfied completely for the method of least squares regression analysis and further test statistics to be valid (***Mendenhall, 1988***). This issue is of importance, as the structure and nature of accident and geometric data will probably not conform exactly with all of the four assumptions in whole.

4.1.1 ECONOMY OF VARIABLES

Several key issues arise from the model structure. First, in creating an effective model, an economy of variables is required. Linear modeling is particularly effective if the number of variables used in modeling the safety relationship is sufficient enough to model the actual relationship, but not duplicate it (***Meyer, 1975***). Too many variables used may, in fact, yield a better fit to actual data, but results in a model that is often unmanageable in terms of the number of input variables required to obtain an estimation. As well, collection of data required for too many variables is not only cumbersome, but unrealistic as well. Any benefits gained from a closer fit to the data is outweighed by the advantage of having a simple and concise mathematical relationship. Thus, one key element of any linear model is that there is a distinct trade-off between the accuracy of the model (in comparison to actual data) and it's simplicity of use. In essence, a "good" model uses only the minimum number of variables required to model a situation without losing any information relevant to the parameters.

4.1.2 MEASURABILITY, SIGNIFICANCE AND CORRELATION

General linear models should also have variables that are measurable, significant and exhibit some degree of correlation. Derived generalized linear relationships between variables that do not exhibit one or more of the three characteristics may be of doubtful usefulness.

Measurability is the most important attribute that a variable must have, since the lack of it results in the inability to quantify a relationship. For example, if an independent (and input) cross-sectional geometric factor is unmeasurable, then it is not possible to determine it's effect on the dependent variable, accident rate. In order for the dependent variable of a relationship to be quantifiable, all independent variables of the relationship must also be measurable.

Significance plays a role in determining an economy of variables in linear modeling. Statistical significance can be defined as the existence of a generally linear relationship between the dependent and independent variables. For example, an independent variable that exhibits a linear relationship with the dependent variable is considered statistically significant. Variables that are deemed to be statistically insignificant through tests of significance should not be included in the analysis, as their presence only serves to add complexity to the modeling problem. Conversely, significant variables should be included in the model, as their omission can lead to a model that fails to identify key relationships. However, it should also be noted that the development of a linear relationship also requires that statistically significant variables are not correlated with each other. Should this occur between two or more input variables, they would no longer be considered independent of each other.

Correlation is an indication of a relationship between the model outcome and it's input variables. This variable characteristic is important to the linear modeling process since a lack of correlation between the dependent and independent variables implies that the manipulation of the independent variable will not result in any change in the dependent variable. In essence, both variables would be considered independent, and exhibit no relationship between each other. Correlation determines whether a generalized linear model actually produces useful results, and should not be confused with causality.

Causality is the theoretical cause-and-effect relationship between the dependent variable and it's independent factors. While this characteristic may be important in order to understand the reasoning behind the behaviour of the dependent variable under manipulation, it is not essential in developing a generalized linear model. Because the purpose of the generalized linear model is to merely identify the trend in behaviour of the dependent variable without having to identify the reasoning behind it, causality is not as important as establishing correlation (***Meyer, 1975***).

While these three characteristics may appear to be intrinsic and trivial to linear modeling, they are clearly important since they establish the fundamental fact that a relationship does exist between the independent (or input) and dependent (or output) variables in a generalized linear model.

4.1.3 SCOPE

A third element of linear models deals with scope, which is closely related to the complexity of a model. The scope of a model can be defined as limiting the selection of variables that exhibit measurability, significance and causality to those factors that directly contribute to the relationship being determined. A model with an "good" scope will include enough variables to adequately represent the data being modeled, without including variables that may play, at best, a secondary and minimal role in the relationship's outcome. A complex model with many defined parameters based on one set of data has limited use in that further data relating to highway safety and geometrics may not provide a good "fit" for the model developed from earlier data. This can be characterized as a problem of calibration and verification or parameters. Essentially, the scope of a model must be broad enough to include all parameters vital to the safety-geometric relationship, yet concise enough to omit those that play, at best, a minimal role.

4.2 LINEAR TRANSFORMATIONS OF NON-LINEAR FUNCTIONS

The nature of certain types of data preclude the direct use of a linear function to model their relationships. For example, prior research into the effects of cross-sectional geometric elements on accident rates have been found to exhibit a distinctly non-linear relationship (***Zegeer et al., 1987***). However, regression modeling techniques can still be applied to non-linear functions through the use of linear transformations.

A more generalized form of the linear equation seen in Equation 4.1 is given below:

$$Y = a * bX_1^{(m)} * cX_2^{(m+1)} * \ldots * zX_n^{(m+n)} * E_1 * \ldots * E_n \qquad \textbf{(EQN 4.2)}$$

where, Y = dependent variable of the relationship,
a, b, c = coefficients that describe each unique linear function, and
X_1, X_2, X_n = dependent variables or parameters of the relationship
m, n = power coefficients
E_n = random error component (for each X_n)

This equation is clearly non-linear with respect to the dependent variable, Y, due to the power coefficients m and n. As well, the other significant difference between this equation and Equation 4.1 is the presence of interaction between each independent variable, X, that represents a more generalized form of modeling the independent variables. It would appear that the use of regression modeling in this case would not be applicable. However, the application of a natural logarithmic function transform to both sides of the equation yields the following "linearized" equation:

$$\ln[Y] = \ln[a] + \ln[b] + \ldots + (m)\ln[X_1] + (m+1)\ln[X_2] + \ldots + \ln[E_1] + \ln[E_2] + \ldots$$

(EQN 4.3)

Note now that the equation is linear in terms of $\ln[Y]$ and $\ln[X_n]$, and not Y and X_n. Regression analysis such as the method of least squares can then be applied to this transformed function, referred to as a log-log function.

A further subset of non-linear models is a specialized function referred to as a log-linear function. Equation 4.4 is an example of this form of function:

$$Y = a * bm^{X_1} * c(m+1)^{X_2} * \ldots * z(m+n)^{X_n} * E_1 * \ldots * E_n$$ **(EQN 4.4)**

This form of the non-linear model differs from Equation 4.2 in that the interaction between Y and X_n is no longer linear even when the natural logarithm is applied to both sides. However, the application of the natural log to both sides does yield a linear relationship in terms of $\ln[Y]$ and X_n. The transformed relationship in Equation 4.5 can be used in linear regression modeling.

$$\ln[Y] = \ln[a] + \ln[b] + \ldots + \ln[m](X_1) + \ln[m+1](X_2) + \ldots + \ln[E_1] + \ln[E_2] + \ldots$$

(EQN 4.5)

This relationship is referred to as a log-linear relationship due to the interaction between the dependent logarithmic variable and the independent linear variables. The resulting analysis of this function is referred to as log-linear regression modeling. The advantage of this log-linear model over the log-log model is that the non-linear interactions of the geometric factors (independent variables) can be modeled in a linear fashion with respect to ln[Y]. This model is simpler than the log-log model since the influence of the independent variables are linear, and not logarithmic.

4.3 LOG-LINEAR ACCIDENT RATE-GEOMETRIC MODEL

Because of the prevalence of log-linear functions in the relationships between accident rate and cross-sectional geometric design factors, it was necessary to select a log-linear model to analyze. As seen previously in Chapter 2, prior accident rate-geometric models used a predominantly log-linear relationship to model geometric effects. One such model was examined in Chapter 2 and is presented in Equation 4.6:

$$AR = 0.0019\,(ADT)^{0.882}\,(0.879)^{W}\,(0.919)^{PA}\,(0.932)^{UP}\,(1.236)^{H}\,(0.882)^{TER1}\,(1.322)^{TER2}$$

(EQN 4.6)

where,
- AR = predicted accident rate based on the following factors
- ADT = average daily traffic volume
- W = paved lane width
- PA = paved shoulder width
- UP = unpaved shoulder width
- H = median roadside hazard rating (1 to 7 - most hazardous)
- TER1= 1 for flat terrain, 0 otherwise
- TER2= 1 for mountainous terrain, 0 otherwise

This model was selected over a log-log model due to the simpler model form and the lack of significant differences in results between a log-linear and a log-log function. For these reasons, a similar model format was selected for modeling the effects of Alberta cross-sectional highway geometric factors on Alberta accident rates.

4.4 SUMMARY

This chapter presents an overview of generalized linear models and the regression methodologies to analyze them. In order for these types of models to be statistically valid, several underlying assumptions or premises should be valid for the relationship.

Chapter 5 of this thesis presents an overview of the Alberta accident and geometric data, as well as an analysis of the two sets of data in preparation for the application of the modeling methodology presented in this chapter.

Chapter 5

Methodology

5.0 INTRODUCTION

Alberta Transportation and Utilities currently uses an accident rate prediction model developed for the Transportation Research Board (***Zegeer et al., 1987***) for use with two lane rural highways. This model estimates an accident rate per mile per year based on geometric variables found to be statistically significant in run-off-the-road (ROR) and opposite direction (OD) type accidents. The basis of this research is to calibrate and validate the use of this type of model under Alberta geometric conditions. The selection of regression as the analysis tool was partially dictated by Alberta Transportation's use of the generalized log-linear regression model developed by Zegeer for the TRB (***TRB, 1987***). Further evidence of the widespread use of regression modeling is apparent from the results of the survey conducted across North America of state and provincial transportation agencies and documented in Chapter 3.

The TRB accident rate model (hereafter referred to as the TRB model) is:

$$AR = 0.0019\,(ADT)^{0.882}\,(0.879)^{W}\,(0.919)^{PA}\,(0.932)^{UP}\,(1.236)^{H}\,(0.882)^{TER1}\,(1.322)^{TER2} \qquad \textbf{(EQN 5.1)}$$

where, AR = Number of ROR, OD and same direction (SD) accidents /mile/year
ADT = Bi-directional average daily traffic volume
W = Lane width, in feet
PA = Paved shoulder width, in feet
UP = Unpaved shoulder width, in feet
H = Median roadside hazard rating for the highway section (a subjective measure in terms of 1 = least hazardous to 7 = most hazardous)
TER1,2 = 1 for flat terrain, 0 otherwise

Calibration of this model to Alberta conditions required quantitative data from two databases: a highway geometric database, and an accident database.

5.1 ALBERTA MODEL STRUCTURE

The calibrated version of the TRB model (referred to as the Alberta Accident Model) is in a form similar to the TRB model. The median roadside hazard rating variable and the terrain factors were not included in the calibration of the TRB model due to the unavailability of those data variables from the Alberta geometric database. As a result, the median roadside hazard rating and both terrain factors were considered to be zero and are therefore assumed to be neutral for this calibration only. The effects of terrain factors on accident rates can be incorporated into the model in the future, when the appropriate data becomes available. The revised model to be calibrated is of the following form:

$$AR = C_0\,(AADT)^{C1}\,(C2)^{W}\,(C3)^{Sh}\,(C4)^{H} \qquad \textbf{(EQN 5.2)}$$

Where, AR = Number of ROR, OD, and SD accidents /mile/year
AADT = Bi-directional average annual daily traffic volume (veh./day)
W = Lane width, in feet
Sh = Paved shoulder width, in feet
H = Roadside hazard rating

AADT (values were used in replacement of the ADT variable from the TRB model due to the unavailability of ADT values in the Alberta Transportation highway geometric database. It was not made clear in Zegeer's research into accident rate relationships (***Zegeer et al., 1987***) why ADT values were selected over AADT values to represent observed highway traffic volumes. Based on the literature search documented in Chapter 3, it appears that Zegeer averaged ADT values for each section of the surveyed highways to determine the effect of traffic volumes on accident rates. This ADT volume, in effect, is similar to the AADT volume defined in the Highway Capacity Manual (***TRB, 1985***) and used in Alberta. Although it is recognized that there may be differences between AADT volumes and ADT volumes used by Zegeer, this variance between the two variables was judged not to be significant in the analysis of the TRB model. As a result, it was assumed that the ADT values mentioned by Zegeer were AADT values instead.

The unpaved shoulder width factor (UP) was also omitted since no primary highways in Alberta remain with unpaved shoulders. Of the 752 rural two-lane primary highway sub-sections in Alberta, only 38 do not have shoulders at all (and consequently no shoulder type). All primary highways in Alberta, whether two-lane or multi-lane, have asphalt concrete pavement type of shoulders. These types of shoulders are considered to be stabilized (PA) in the TRB model (***TRB, 1987***) and precludes the use of the unpaved shoulder width factor in the Alberta Accident Model. Factors were presented in U.S. standard units since the original model was developed using feet and miles as the units of measurement. However, models were also derived in SI units as well.

5.2 ALBERTA GEOMETRIC DATABASE STRUCTURE

Data used in the modeling and calibration of the Alberta Accident Model was obtained from two separate databases. Geometric and other relevant highway data were contained within an Alberta highway geometric database containing all Alberta primary highways as of 1991. Data was presented in a sectional format, listing each primary highway by highway number (HWYNO), master control section number (MCS), safety sub-section number (SUB), and subsection length (KMPOST) in kilometres. Additional data of importance to this analysis is shown in Table 5-1:

Table 5-1 Alberta Primary Highway Geometric Database - 1991

Variable Name	Description
UR • U • R • F	Locational setting of the highway safety sub-section • Urban Highway • Rural Highway • Federal Park Highway
KMPOST	Beginning and end of each safety sub-section
SUBLENGTH	Total sub-section length, in kilometres
HWYTYP • 1 • 2 • 3 • 4 • 5 • 6	Primary highway classification • Gravel • 2-lane undivided • 4-lane undivided • 4-lane divided (expressway) • 4-lane divided (freeway • 6 or more lanes divided
SURFTYP • 1 • 2 • 3 • 4 • 5 • 6 • 7	Primary highway surface type • None, or unknown • Gravel • Double surface treatment • Oiled / dust control • Seal coat (chip seal) • Asphalt concrete pavement • Concrete
SURFWD	Road surface width in metres
SHTYPE	Shoulder type (same as surface type)
SHIN, SHOUT	Inside and outside shoulder width in metres
MEDWD	Median width (if applicable)
AADTYR	Year of recorded AADT volume
AADT	Average Annual Daily Traffic volume
SPDAY, SPNITE	Daytime and Nighttime posted speed limit

For the purposes of this analysis, only roadways with a rural classification (UR = R) were used, as they were the predominant category of primary highways in Alberta. In addition, the analysis was divided into two separate sections: one for two-lane roadways, and the second for all other multi-lane roadways.

A more detailed statistical breakdown comparing rural, two-lane, two-way, paved highways in Alberta to total primary highways is shown in Table 5-2. Approximately 11,900 kilometres of paved, rural, two-lane roadway sections were available for use in the calibration analysis. As seen in Table 5-2, two-lane, paved roadways comprise the bulk of Alberta's primary highway network, accounting for approximately 85% of the total length of paved roadways.

Table 5-2 General Statistics, Alberta Primary Highway Database - 1991

	2-lane, rural, paved primary highways	Total primary highway network
Number of Cases	752	1031
Total Length	11,897.235 km	14,037.191 km
Mean SUB Length	15.821 km	13.615 km
Median	12.864 km	8.976 km
Standard Deviation	14.202 km	13.924 km
Minimum	0.080 km	0.080 km
Maximum	76.331 km	76.331 km

Of the 11,897 total kilometres of two-lane, rural highway length in Alberta, the predominant daytime posted speed limit was 100 km/hr. Table 5-3 summarizes the number of highway sub-sections and corresponding total kilometre lengths for each speed limit group. No difference was discerned between day and night posted speed limits for two-lane primary highways in Alberta.

Table 5-3 Speed Limit, Two-Lane, Rural, Primary Highways - 1991

Posted Speed Limit (Daytime)	Number of Sub-sections	Total Length (km)	Percent of Total
Missing	1	11.26	.1
60 km/hr	6	65.47	.8
70 km/hr	1	0.81	.1
80 km/hr	85	1,624.37	11.3
90 km/hr	59	1,061.04	7.8
100 km/hr	600	9,134.29	79.8
Total	752	11897.24	100.0

In comparison with the data used for the TRB model (*Zegeer et al., 1987*), the distribution of Alberta speed limit data was remarkably similar. Whereas 86 percent of the TRB data had speed limits of 55 mph (approximately 88 km/hr), the Alberta database contained nearly 80 percent of the sub-sections with speed limits of 100 km/hr. The lower American speed limit was a direct result of American federal highway policy, which lowered speed limits on federal primary highways as a result of the 1973 oil embargo. As with the TRB model analysis and development, the predominance of 100 km/hr posted speed limits within Alberta precluded a more detailed analysis of the relationship between speed limits and accident rates.

5.3 ALBERTA HIGHWAY ACCIDENT DATABASE STRUCTURE

Four years of primary highway accident data were contained within four files titled PHWY91, PHWY92, PHWY93 and PHWY94. These ASCII datafiles contained 33,582 accidents of all types recorded over the four year period and were categorized using case numbers for each individual accident event. The PHWY datafiles consist of fifty five accident variables recorded by the accident parties, the highway law enforcement agency, and tabulated by Alberta Transportation's Motor Transport Services division. Certain variables of interest among the fifty five listed variables included the following:

Table 5-4 Alberta Primary Highway Accident Databases - 1991-94

Variable Name	Description
CASENO	Accident case number
OBJNO	Number of vehicles or objects involved in accident
YEAR	Year of event
MON	Month of event
DAY	Date of event
HOUR	Hour of event
INJURED	Total number injured
FATALS	Total fatalities involved
HR	Hit and Run accident
RADIUS	Curve radius, if applicable
HWY	Primary highway number
MCS	Master Control Section number
SUB	Safety Sub-section number
INTTYPE	Intersection type (if applicable)
PE	Primary event
COLLSEV	Collision severity
OBJTYP	Object type
OBJID	Object identification
LIGHTA	Natural light conditions occurring
LIGHTB	Artificial lighting conditions
RDCOND	Contributing physical road condition (if applicable)
DRVA	Primary driver action
DRVC	Primary driver condition
DEVICE	Presence of a traffic control device
SPEED	Possibility of unsafe operating speeds contributing to event
ECOND	Contributing environmental conditions (if applicable)
SCOND	Contributing road surface conditions (if applicable)

These variables were believed to be key non-geometric contributing factors to highway accidents in varying degrees. Isolation and elimination of events that occurred as a result of certain contributing factors such as alcohol, wildlife, or environmental conditions were performed based on the data variables available from the PHWY databases. As a result, elimination of non-geometric variables from the analysis would minimize the confounding effects of these secondary correlative variables.

5.4 AGGREGATION OF GEOMETRIC AND ACCIDENT DATAFILES

The first stage in consolidating the geometric and accident datafiles involved the removal of extraneous formatting and variables from the datafiles using custom programs coded in Lahey FORTRAN77™ and detailed in Appendix A of this thesis. The four primary highway accident databases were consolidated into one datafile and the geometric database was then imported into the consolidated primary highway accident file through a statistical software package shell. SPSS® for Windows™, version 6.1 was used as the statistical analysis tool on a Microsoft Windows™ based 486DX/33 computer with 32 MB RAM. SPSS® was selected for the analysis on the basis of cost, ease of use, and the computational power available for this research work.

Each individual accident event was linked to its corresponding highway geometric data through the use of three keying variables analogous within the two datafiles: HWYNO, MCS, and SUB. The resulting datafile contained accident events with not only basic highway section location information, but with more detailed cross-section and highway volume information. As a result, this combined database contained the essential data required to analyze the relationship between the primary highway accident rate and various geometric design factors.

5.5 ISOLATION OF GEOMETRIC RELATED ACCIDENTS

The identification, isolation and removal of non-geometric related accident events was performed as the next step. Based on previous research conducted into significant variables involved with primary geometric improvement measures (lane width, shoulder width, shoulder type, and sideslope rating), several categories of accidents were identified and removed from the analysis (***Zegeer et al., 1987***). 33,582 accidents were analyzed from the four years of accident data spanning 1991 through 1994. A breakdown of accident cases by numbers of vehicles or objects involved in each accident is listed in Table 5-5:

Table 5-5 Magnitude of Accidents Occurring in Alberta from 1991-94

Number of Vehicles or Objects	Number of Accidents	Percent of the Total Number of Accidents
0 *	1	0.0
1	6,741	20.1
2	26,004	77.4
3	719	2.1
4	94	0.3
5 or more	24	0.1
Total	33,583	100.0

One case with the object number reading zero indicated a hit-and-run incident where an identification on the vehicle was not recorded. Because hit-and-run cases are eventually removed from the analysis (and due to the minuscule effect that this case plays in the analysis), this single case does not affect the analysis.

The first category removed from the analysis encompassed recorded accidents involving animals. This category of accidents were specifically removed since Zegeer's research conducted into two-lane rural roadway accident models concluded that their effects on a geometric accident model were secondary in nature (***Zegeer et al., 1987***). The conclusion that Zegeer reached from his Chi-square analysis of accident data was that accident types involving animals was not highly correlated with primary geometric roadway variables being modeled. As a result, 13,910 accidents over the four year period involving animals were removed from the analysis. 19,672 accident cases in the four year sample remained that did not involve an animal.

Recorded accidents that involved operator impairment by alcohol, other chemical use, driver fatigue, or a driver medical defect were also removed from the working database. Because the effects of operator impairment on accident rate may mask the effects of cross-sectional geometrics on accident rates, this confounding category of accidents were omitted from the analysis. Likewise, impairment due to other chemical use, as well as driver impairment due to fatigue or a medical defect (a heart attack, for example) were also felt to be of a sufficiently confounding nature in the accident and design geometrics relationship. As a result of this confounding data, any accident involving operator impairment was not considered for the analysis. A breakdown of the remaining accidents (after removal of animal-type accidents) is shown in Table 5-6.

Table 5-6 Accidents Involving Driver Impairment, 1991-94

Type of Impairment	Number of Accidents	Percent of Total
None	12,346	62.8
Alcohol	1,387	7.1
Drug	15	0.0
Fatigue or Asleep	862	4.3
Medical Defect	73	0.4
Missing, or Unknown	4,989	25.4
Total	19,672	100.0

While the percentage of accidents involving alcohol is approximately 7% of the total sample, impaired accidents (those accidents involving operator impairment primarily due to alcohol) frequently account for a fair greater proportion of more severe accident types. For example, the percentage of all accidents occurring on Alberta's primary highway that involved alcohol was 9.2%, while the percentage of fatal accidents that involved alcohol for the same sample of accidents was 27.9% (Alberta Traffic Collision Statistics, 1992). While the effects of alcohol on alcohol appear to significantly influence the severity of accidents, it remains that the scope of this research work did not focus on severity, but rather the rate of occurrence. In this manner, alcohol related accidents were not included in the accident rate analysis.

A significant number of recorded cases (4,989) from the 1991-94 database were either missing the driver condition variable (DRVC) or else the driver condition was unknown. While this number is sizable in terms of the total database size, the exclusion of these cases might have resulted in an insufficient accident database to derive accident-geometric relationships from. As a result, 17,335 accidents were retained for further use in the analysis.

Likewise, accidents involving operator hit-and-run were removed since the key geometric data elements required for the accident rate analysis was frequently not available for recording due to the lack of driver information. As well, the percentage of accidents involving hit-and-run was not found to be significant, accounting for approximately 3.4% of the remaining accidents (after removing animals and impaired accidents) on two-lane primary highways from 1991 to 1994. 16,743 accident cases remained in the analysis.

Table 5-7 Hit-and-Run accidents, Alberta Two-Lane Primary Highways 1991-94

Accident Type	Frequency	Percent of Total	Percent Valid
Hit and Run	592	3.4	3.9
None	14,610	84.3	96.1
Unknown	2,133	12.3	-
Total	17,335	100.0	100.0

Table 5-8 Object Type, Alberta Two-Lane Primary Highway Accidents 1991-94

Object Type Involved	Frequency	Percent of Total
Single Vehicle	16,514	98.6
Pedestrian	5	0.0
Motorcyclist	178	1.1
Bicyclist	5	0.0
Fixed Object	15	0.1
Missing or Unknown	26	0.2
Total	16,743	100.0

Table 5-8 contains detailed data of the types of objects involved in each accident. As clearly seen, single vehicle drivers comprise the bulk of the object types in the database. These drivers operated one of the following types of vehicles: (1) standard, four wheel passenger car, (2) light pickup truck (Gross Vehicle Tonnage less than 4,500 kg), (3) minivan, (4) truck with GVT greater than 4,500 kg, (5) truck tractor, (6) bus, (7) motorhome, (8) construction, maintenance, or emergency equipment, or (9) farm equipment. Because the primary research focused on vehicular accidents, those accidents involving pedestrians, motorcyclists, bicyclists, or fixed objects were removed from the analysis. It should also be noted that these other vehicle modes do not comprise a significant quantity of cases, as seen in Table 5-8. The remaining database contained 16,540 valid accidents for analysis.

Accidents within the Alberta primary highway database attributable to highway intersection geometrics were also removed since intersection geometrics affect accidents in a separate manner than that of cross-sectional geometrics (*TRB, 1987*).

Table 5-9 Traffic Devices - 1991-94

Signal Device Involved	Frequency	Percent of Total
None	13,568	82.0
Signal Light	327	2.0
Stop Sign	1,024	6.2
Yield Sign	76	0.5
Merge Sign	48	0.3
Ped. X-walk	3	0.0
School Bus	15	0.1
Lane Control	28	0.2
RR X-ing	43	0.3
Unknown	1,266	7.6
Missing	142	0.9
Total	16,540	100.0

Table 5-9 breaks down accident types involving traffic control devices into detailed groupings of device types. The variable DEVICE in the PHWY accident database was used as a keying variable to remove intersection related accidents. Because of the model parameters outlined previously, accidents involving traffic control devices were removed. The presence of a traffic control device was assumed to indicate the presence of an intersection. Highway subsections containing unknown and missing values for this variable were assumed to be non-intersection type sections and were included in the analysis. The removal of intersection type accidents yielded 14,976 datasets for use in the analysis.

Finally, only those accidents occurring on rural, two-lane, two-way, and paved primary highways were retained for use in the analysis. This exclusion of multi-lane, urban, and divided highways was a direct consequence of a model parameter outlined by the TRB accident rate model (*TRB, 1987*). This parameter specified that the model applied only to the following conditions:

1. Two-lane, two-way paved rural roads on primary and secondary systems
2. Homogeneous roadway sections, without additional accidents expected or occurring at highway intersections.

Table 5-10 illustrates the remaining valid accident cases by highway type:

Table 5-10 Accident Cases by Highway Type - 1991-94

Highway Type	Frequency	Percent of Total
Gravel	205	1.4
2 Lane Undivided	9,530	63.6
4 Lane Undivided	128	0.9
4 Lane Divided (Expressway)	3,108	20.8
4 Lane Divided (Freeway)	1,531	10.2
6 or more Lanes	469	3.1
Unknown or Missing	5	0.0
Total	14,976	100.0

As seen from Table 5-10, the number of accident cases remaining for analysis in the calibration of the two-lane, rural, undivided highway accident rate model is 9,535 (two-lane undivided and Unknown or missing cases). Closer inspection of the accident database revealed that the 5 accident cases with unknown highway types actually occurred on two-lane, undivided highway sections, when visually compared against an

Alberta Highway Inventory Map. These five cases were added to the two-lane, undivided highway group.

5.6 SUMMARY OF ALBERTA GENERAL ACCIDENT CHARACTERISTICS

The combination of all vehicular, non-intersection type accidents yielded a valid database of 9,535 cases on Alberta two-lane, rural, undivided, primary highways between the four year period spanning from 1991 to 1994. Nearly 70 percent of all accidents recorded in this time period for the given conditions were property damage only accidents (6,557 cases). 2,704 injury accidents and 274 fatal accidents occurred in this period, as seen in Table 5-11. An accident rate per kilometre of two-lane, rural, primary highway per year was also calculated to compare to summary statistics used in the TRB model (***Zegeer et al., 1987***). This rate was found by the following formula:

$$\text{Accidents/km/yr} = \frac{\text{Event Frequency}}{(\text{2-lane hwy length x No. of sample years})} \qquad \textbf{(EQN 5.3)}$$

where, 2-lane highway length = 11,897.24 km
No. of sample years = 4

Table 5-11 Collision Severity, Two-Lane, Rural Primary Highways - 1991-94

Event Severity	Frequency	Percent of Total	Accidents/km/yr
Fatality Involved	274	2.9	0.00576
Injury Only	2,704	28.4	0.0568
Property Damage	6,557	68.9	0.137
Total	9,535	100.0	0.200

A comparison between Alberta primary highway accident experience and the experience documented in the TRB research study is presented in Chapter 6.

As well, it is interesting to note that the predominant forms of vehicular accidents occurring on two-lane primary highways in Alberta during this four year period involved single and two vehicle cases. Table 5-12 presents the breakdown of these accident cases by the number of vehicles involved.

Table 5-12 Number of Vehicles Involved in Each Accident - 1991-94

Number of Vehicles Involved in Accident	Frequency	Percent of Total
1	5,376	56.4
2	3,866	40.5
3 or more	277	2.9
Unknown	16	0.2
Total	9,535	100.0

These values are significant since they seem to indicate that the predominant forms of accidents observed on two-lane rural roadways are either single vehicle Run-off-the-road type incidents, or else are some type of collision (head-on, sideswipe, or read-end) between two vehicles. Accident "chaining", where multiple vehicles are involved with several collision events, appear to occur infrequently (less than 3 percent) on Alberta's primary highways. This appears to agree with empirical observations of Alberta highway incident information.

A breakdown of accident cases by event type is detailed in Table 5-13. As seen from the table, the majority of accident events that occurred in the four year period were collision events involving stationary objects. This event category may mask certain secondary events occurring in the accident, such as a run-off-the-road accident, or a sideswipe. However, because each these events appear to be geometrically related, the entire dataset was retained for use in the analysis.

Table 5-13 Primary Event - 1991-94

Event	Frequency	Percent of Total	Accidents/km/yr
Struck Fixed Object	446	4.7	0.00937
ROR Left	1,820	19.1	0.0382
Right Angle	446	4.7	0.00937
Pass Left Turn	474	5.0	0.00996
Left Turn Apr	173	1.8	0.00364
Left Sideswipe	408	4.3	0.00857
Other	501	5.3	0.0105
Rear End	1,365	14.3	0.0287
ROR Right	3,047	32.0	0.0640
Head On	343	3.6	0.00721
Pass Right Turn	53	0.6	0.00111
Right Sideswipe	412	4.3	0.00866
Backing	46	0.5	0.00097
Unknown	1	0.0	0.0
Total	9,535	100.0	(mean accident rate) 0.200

(note: ROR is Run-Off-The-Road)

Accident rates were also calculated using Equation 5.3 and are presented in Table 5-13. A comparison and discussion of similarities and differences between these rates is presented in Chapter 6. Note that these accident rates are divided over the *entire* Alberta primary highway inventory length, and are not indicative of experienced accident rates at individual sub-sections. A more detailed breakdown of observed accident rates by sub-section is also discussed in the analysis and discussion of two-lane rural highway accident and geometric data in Chapter 6.

A series of Chi-square tests were performed on the event types to confirm the correlation of specific event types with the key geometric variables to be modeled (lane width, shoulder width, shoulder type, daytime posted speed, mean and extreme values for sideslope ratio, ditch width, and backslope ratio). This analysis was conducted to confirm the applicability of selecting certain categories of accident events for modeling and calibration (***Zegeer et al., 1987***). Results of the Chi-square tests are also documented in Chapter 6.

5.7 AGGREGATION OF ACCIDENTS BY HIGHWAY SECTIONS

The first step in creating the log-linear regression model involved the aggregation of accident cases by primary highway sub-sections. Aggregation was performed to identify sub-sections without any recorded accidents over the four year test period and to determine mean accident rates for subsections and by highway length. Because of the presence of a number of primary highway sub-sections without a recorded accident over the four year test period, these zero accident sub-sections were obviously not represented in the AT&U accident database. To account for this, the zero accident sections were added to the revised working database (referred to hereafter as the University of Alberta, or UofA database) as statistical control sections and to ensure that the dataset to be modeled included all relevant sub-sections. The result of aggregation and the addition of zero accident sub-sections is that each primary highway sub-section has a corresponding accident frequency and rate (expressed in number per year, per kilometre per year, and per 100 million vehicle kilometres per year). As well, accident rates were calculated in terms of miles to establish a direct comparison with the U.S. TRB model.

5.8 DETERMINATION OF SIGNIFICANT GEOMETRIC VARIABLES

Aided by an understanding of the TRB model's variable selection process (***Zegeer and Deacon, 1987***), the following key geometric variables were selected for use in the model: (1) AADT, (2) Lane Width, (3) Shoulder Width, (4) Mean Sideslope Ratio, (5) Mean Ditch Width, and (6) Mean Backslope Ratio. Because the majority of these variables were initially selected for use within the TRB model parameters, validation of their significance for Alberta accident and geometric conditions was also required. The Chi-square test for goodness of fit and variable significance performed on the accident database to verify certain accident-geometric relationships was used to also validate the use of the geometric variables mentioned above. Chapter 6 details this analysis of geometric factor significance.

As well, a cross-correlation coefficient matrix of the geometric variables was calculated to check for multicollinearity effects between the all geometric related variables. This effect occurs when two or more independent variables are highly correlated and makes separate identification of the effects of each variable difficult (***Zegeer et al., 1987***). Results of the cross-correlation coefficient matrix are also presented in Chapter 6.

5.9 REGRESSION ANALYSIS

Two types of regression analysis were performed on the database. They were: (1) single variable regression, and (2) multi-variable log-linear regression.

5.9.1 SINGLE VARIABLE REGRESSION

Single variable regression was performed to determine the direct correlative effects, if any, between accident rate and each key geometric variable. The following variables were studied in relation to accident rate: (1) AADT, (2) Lane width, and (3) Shoulder width. In addition, regression of shoulder width was performed using grouped shoulder width data to determine generalized effects of incremental shoulder width improvements on accident rate. Unfortunately, due to a significant lack of sub-sections in Alberta with lane widths greater or less than 3.70 m, it was not possible to analyze the effects of incremental changes in lane width to accident rates.

Shoulder widths were grouped in the following manner:

(1) shoulder width = 0.00 metres
(2) 0.00 m < shoulder width < 1.20 m
(3) 1.30 m < shoulder width < 1.80 m
(4) 1.95 m < shoulder width < 2.70 m
(5) shoulder width > 2.70 metres

The values 1.20 metres, 1.80 metres, 2.20 metres, and 2.70 metres were chosen because they corresponded to the U.S. standard shoulder width groupings of 4 feet, 6 feet, and 8 feet (***Zegeer and Deacon, 1987***).

The relationships between accident rate, expressed in terms of a rate per million vehicle miles per year, and a rate per 100 million vehicle kilometres per year and the three key geometric variables (AADT, lane width, and shoulder width) were then plotted and graphed. The use of U.S. imperial units are required to perform a direct comparison between the Alberta accident rate model results and those obtained from the TRB model. Units per 100 million vehicle kilometres are included in this thesis since they are the predominant units of measure within Alberta Transportation and Utilities for accident rates. The regression analysis in Chapter 6 includes these plots and graphs.

5.9.2 MULTI-VARIABLE LOG-LINEAR REGRESSION

Next, the log-linear function representing the modified TRB model was modeled and analyzed. This model differed from the theoretical model in that the units representing accident rate were in terms of vehicle miles or kilometres. Final comparisons between this model and the TRB model required a conversion factor for the accident rate.

$$AR = C_0 (AADT)^{C_1} (C_2)^{W} (C_3)^{Sh} (C_4)^{H} \quad \textbf{(EQN 5.4)}$$

where, AR = Accident rate in either /MVM or /100 million veh. km
C's = Model coefficients
AADT = Bi-directional average annual daily traffic volume
W = Lane width, in units applicable to AR
Sh = Paved shoulder width, in units applicable to AR
H = Estimated Roadside Hazard Rating

5.9.3 COMPARISON OF TRB MODEL vs. ALBERTA MODEL RESULTS

After being developed from the log-linear regression analysis, the Alberta model was compared to the TRB model for similarities. Actual geometric variables from the Alberta database were applied to the TRB model to determine hypothetical accident rates for each actual primary highway sub-section. These predicted accident rates were then compared to those accident rates developed from application of geometric variables to the Alberta developed model. The resulting comparison would indicate the applicability of the TRB model to Alberta conditions, and whether correction factors needed to be applied to account for any variance.

5.10 SUMMARY

This chapter provides a detailed methodology on the application of Alberta primary highway accident and geometric data to the TRB accident rate model described in Chapter 3. The selection of this model to be calibrated as an Alberta accident rate model was based primarily on the prevalent use of it through North American state and provincial highway and transportation agencies. Chapter 6 documents the application of this methodology and illustrates several key relationships between accident rates and individual cross-sectional geometric factors. As well, Chapter 6 also presents the comparison between predicted accident rates derived from both the TRB accident rate model, and the Alberta accident rate model. Differences between the two models will form the basis of a calibration or correction coefficient to be applied to the current TRB model.

Chapter 6

Analysis of Two-Lane, Rural, Primary Highways

6.0 INTRODUCTION

The development of an Alberta specific model from relationships between observed accident rates and corresponding roadway and roadside geometric factors followed the methodology detailed in Chapter 5. As outlined in the previous chapter, three separate databases were integrated for the model development and calibration. The Alberta primary highway geometric database from 1991 and the secondary roadside geometric database were initially merged to form one geometric database. This database, in turn, was merged with the Alberta primary highway accident database (1991-1994) to form a database encompassing all primary highway vehicular accident events and their related geometric attributes at the event locations. From this data, the relationship between accident rates and highway geometric factors was developed.

6.1 SUMMARY OF GENERAL ACCIDENT STATISTICS

The recorded Alberta accident database was grouped by the primary accident event type and aggregated to facilitate the Chi-square analysis for correlation with key geometric variables. Aside from the primary accident event types, additional event type groups were created for the analysis and included: (1) total accidents occurring on two-lane rural roads, (2) Total ROR accidents, (3) all angle and turning accidents, (4) all single vehicle accidents, and (5) selected multi-vehicle and single vehicle accidents. Table 6-1 illustrates preliminary accident rates established for each accident type to aid in determining the quality of the accident data being used for modeling. Two accident rates were calculated: an accident rate per 100 million vehicle kilometres, and an accident rate per kilometre per year.

Total sub-section length and AADT were aggregated for each accident type to account for the effects of vehicle exposure to the highway sub-section on accident rates. By using these aggregate measures, the two accident rates were then determined using the following equations:

$$\text{AR} \ / \ 100 \text{ million veh. km} = \frac{N \times 100{,}000{,}000}{4 \times SUBLN \times AADT} \qquad \textbf{(EQN 6.1)}$$

where, AR = Accident rate per 100 million vehicle kilometres
N = Number of Accidents occurring over four year sample period
SUBLN = Total length of all sub-sections with an accident corresponding to each event
AADT = Total AADT of all sub-sections with an accident corresponding to each event

$$AR \,/\, km \,/\, yr = \frac{N}{4 \times SUBLN} \qquad \textbf{(EQN 6.2)}$$

where, AR/km/yr = Accident rate per kilometre per year
N = Number of Accidents occurring over four year sample period
SUBLN = Total length of all sub-sections with an accident corresponding to each event

As seen in Equation 6.1 and Equation 6.2, differences will occur between the two accident rates, due to the inclusion of AADT traffic volumes in determining one of the accident rates. Equation 6.1 can be used as a measure of accident rates in relation to the volume of traffic exposure on a given sub-section of primary highway. Equation 6.2

merely determines an accident rate in terms of the length of primary highway exposed to the accidents, without accounting for vehicular traffic (AADT). The accident rate that accounts for exposure to accidents due to both highway section length and vehicle volumes may be a better measure since this rate incorporates the dependence of accident rate on traffic volume (*TRB, 1987*).

Table 6-1 Summary of Alberta Accident Statistics by Accident Type - 1991-94

Event	Number of Accidents	Number of Sub-sections Event Occurred	Total Sub-Section Length	Total AADT	ACC/100 million veh. km	ACC/km/yr
1. Struck Fixed Object	446	249	6,117.74	561,760	3.244	0.0182
2. ROR Left	1,820	466	10,445.43	919,700	4.736	0.0436
3. Right Angle	446	235	5,118.35	548,090	3.975	0.0218
4. Pass Left Turn	474	266	6,218.64	560,820	3.398	0.0191
5. Left Turn Apr.	173	117	2,502.17	338,620	5.105	0.0173
6. Left Sideswipe	408	229	5,350.05	576,180	3.309	0.0191
7. Rear End	1,365	401	8,966.73	876,780	4.341	0.0381
8. ROR Right	3,047	505	11,004.85	992,210	6.976	0.0692
9. Head On	343	194	4,360.61	515,470	3.815	0.0197
10. Pass Rt. Turn	53	51	1,138.43	137,210	8.843	0.0166
11. Rt Sideswipe	412	225	5,095.22	570,620	3.543	0.0202
12. Backing	46	45	1,007.47	128,750	8.866	0.0114
13. Other	501	258	6,199.86	599,840	3.368	0.0202
14. Unknown	1	1	1.69	4,270	3,464.8	0.1479
Total ROR (2,8)	4,867	528	11,225.20	1,028,740	10.537	0.1084
Angle + Turn (3,4,5,10)	1,146	379	8,281.40	837,790	4.129	0.0346
Single Vehicle (1,2,8)	5,313	532	11,277.07	1,034,340	11.387	0.1178
Single + Selected Multi-Vehicle (1,2,4,5,6,8,9)	6,711	547	11,369.27	1,073,290	13.749	0.1476
Acc. Severity						
Total Accidents	9,535	561	11,461.21	1,099,980	18.908	0.2080
PDO Accidents Only	6,454	548	11,379.85	1,076,570	13.170	0.1418
Injury Accidents Only	2,807	494	10.701.63	995,670	6.586	0.0656
Fatal Accidents	274	217	5,349.33	538,250	2.379	0.0128

Table 6-1 shows 14 individual accident event types and their corresponding accident rate statistics. Several of these event types were aggregated into an additional four event group categories. The first group, Total ROR, consisted of run-off-the-road accidents occurring both left and right (ROR Left and ROR Right). The angle and turn total contained all right angle accidents, as well as passing accidents involving both left (Pass Left Turn, Left Turn Approaches) and right turns (Pass Right Turn). The single vehicle accident type included all ROR accidents as well as the struck object category, while the multi-vehicle accident type included left turn passing accidents, left turn opposing direction approach accidents, left sideswipes (opposing direction), and head on collisions.

Aggregate accident rates were also calculated by accident severity, using the categories of PDO (property damage only), injury only involved in an accident event, and fatality involved in an accident event as severity categories. A total accident exposure rate was also calculated for the entire accident sample size.

A comparison was drawn between the aggregate accident exposure rates developed from Alberta accident data, and those developed by Zegeer et al (***Zegeer et al., 1987***). Table 6-2 summarizes aggregate accident rates calculated by Zegeer et al. for the TRB research program into safety effects of cross-sectional design elements. The accident rate values were obtained from 4,785 total miles (approximately 7,700 km) of rural two-lane highways throughout seven U.S. states (Alabama, Michigan, Montana, North Carolina, Utah, Washington, and West Virginia). Accident rate values were converted into metric units to facilitate a direct comparison between the TRB values and those developed from Alberta accident data.

Table 6-2 Summary of TRB Accident Statistics by Accident Type

Event	Number of Accidents	ACC/100 million veh. km	ACC/km/yr
1. Struck Fixed Object	12,091	34.270	0.3730
6. Left Sideswipe	2,997	8.135	0.1119
7. Rear End	12,420	24.792	0.5158
9. Head On	2,113	5.221	0.0746
11. Rt Sideswipe	2,288	6.395	0.0994
13. Other	1,883	6.060	0.0684
ROR Rollover	4,245	15.227	0.1119
ROR Other	2,840	9.410	0.0994
Angle + Turn	14,730	35.003	0.6339
Accident Severity			
Total Accidents	62,676	165.538	2.2933
PDO Accidents Only	38,857	104.089	1.4730
Injury Accidents Only	22,944	58.794	0.7955
Fatal Accidents	875	2.660	0.0249

Source- Zegeer et al., 1987

The comparison of accident rates of various accident types and severity between the two accident databases resulted in significant differences between accident rates expressed in 100 million vehicle kilometres, and in kilometres per year. Because of this sizable discrepancy between accident rates obtained using TRB and Alberta accident data, Table 6-3 was developed to identify differences between each individual accident type as a percentage of the entire accident database. By comparing the proportion of each accident type and severity as a percentage of the entire database, it was hoped that differences in proportions could help to explain the sizable differences seen in aggregate accident rates. It should be noted that several event categories from Table 6-1 are missing from Table 6-2 and Table 6-3, as a result of differences between the Alberta database and the TRB data in recording accident types. For example, the TRB database in Table 6-2 included a more detailed breakdown of ROR accident events (ROR Rollover and Other) than those used from the Alberta database. Those event types appearing in Table 6-3 are categories that were the best attempt to mirror accident groupings used in the TRB analysis.

Table 6-3 Comparison of TRB and Alberta Accident Rates (by percent of total accidents)

Event	Number of Accidents (Alberta)	Percent of Total Accidents (Alberta)	Number of Accidents (TRB)	Percent of Total Accidents (TRB)
1. Struck Fixed Object	446	4.7	12,091	19.3
6. Left Sideswipe	408	4.3	2,997	4.8
7. Rear End	1365	14.3	12,420	19.8
9. Head On	343	3.6	2,113	3.4
11. Rt Sideswipe	412	4.3	2,288	3.7
13. Other	501	5.3	1,883	3.0
ROR - Rollover	-	-	4,245	6.8
Angle + Turn	1,146	12.0	14,730	23.5
Accident Severity				
Total Accidents	9,535	100.0	62,676	100.0
PDO Accidents Only	6,454	67.7	38,857	62.0
Injury Accidents Only	2,807	29.4	22,944	36.6
Fatal Accidents	274	2.9	875	1.4

Source- Zegeer et al., 1987

6.2 DIFFERENCES BETWEEN ALBERTA AND TRB ACCIDENT RATES

From an inspection of Table 6-3, it is immediately noted that the frequencies of accidents occurring from the TRB database is over six times greater than that seen in the Alberta database (62,676 cases versus 9,535 cases). As well, sizable differences exist for Struck Fixed Object event and Angle + Turning event percentages between the two databases. Several possible explanations exist for these discrepancies and differences between the two databases and their corresponding accident exposure rates.

1. Small Alberta Accident Sample Sizes

The smaller Alberta accident sample size between the two databases may be the first reason for discrepancies between the accident rates of the two accident databases. While the two databases are not noticeably different in sampling duration (the Alberta database was sampled over a continuous four year period and the TRB sample was obtained over non-continuous four to five year periods, depending on the state), the TRB accident data was drawn from a survey of seven states, compared to the Alberta data only. As a result, the total accident experience of the TRB database could possibly be up to seven times greater than that of the Alberta database, due to seven times the number of jurisdictions examined for accident data.

2. Differences In Roadway Network Lengths

The total surveyed roadway length of the TRB database was approximately 33% smaller than Alberta's two-lane, rural primary highway network (7,700 km versus approximately 12,000 km in Alberta). As seen in Equations 6.1 and 6.2, accident rates were calculated as a function of total two-lane, rural primary highway network length. Because the total TRB accident frequency occurs over a smaller total roadway network length, the aggregate accident rate for the TRB database could be as much as three times higher than that of the Alberta database as seen in Equation 6.3, assuming no other factors influence the differences in accident rates between the two accident samples. However, since this assumption is not a very realistic assessment of accidents and their contributing factors, the effect of this problem is probably much less than previously mentioned. Nonetheless, this effect cannot be neglected in identifying reasons for the differences in accident rates.

$$\Delta AR = \frac{1}{1 - \frac{L_{TRB}}{L_{ALB}}} \qquad \textbf{(EQN 6.3)}$$

where, ΔAR = Difference in accident rate between TRB and Alberta database
L_{TRB} = Total length of TRB two-lane, rural primary highway network (in km)
L_{ALB} = Total length of Alberta two-lane, rural primary highway network (in km)

3. Differences In Traffic Volumes

Significant differences in traffic volumes, measured in terms of AADT in the Alberta database and ADT in the TRB database, were found between the two. It should be noted that the rationale behind using ADT in the TRB accident database was not made clear by Zegeer et al. in their analysis, since the form of the ADT values used more closely resembled AADT values as opposed to ADT values. Because no mention was made of the variation in daily traffic volume between each day of a year (as is the case using ADT as a measure of volume), it was assumed for the purpose of this thesis that ADT referred to the traffic volume experienced on a roadway section averaged over one year (AADT as a measure of volume). As such, the differences between the Alberta sample and the TRB sample were primarily in the distribution of volumes observed on individual roadway sections in each database. Figure 6-1 illustrates the distribution of traffic volumes observed for the two accident databases. Traffic volumes for both cases were grouped on the basis of TRB groupings for convenience. Each traffic volume value indicates the upper limit of the volume range (i.e. volume of 1,000 vehicles = ADT or AADT range from 751 to 1,000 vehicles).

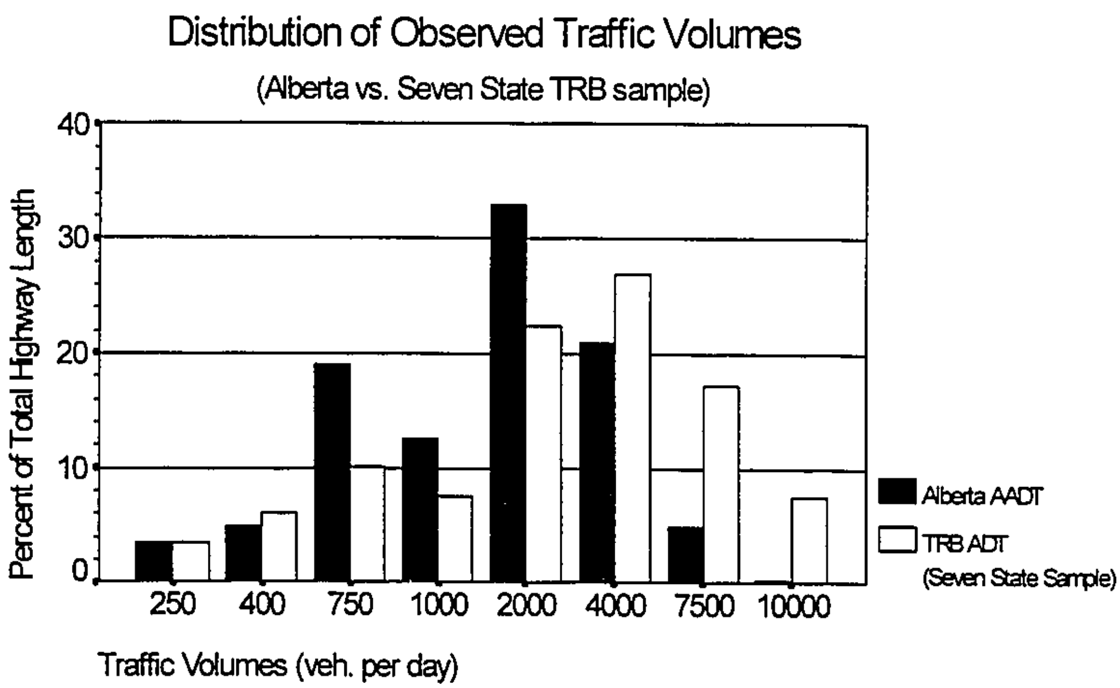


Figure 6-1. Distribution of Observed Traffic Volumes

Figure 6-2 below illustrates the same traffic values cumulatively distributed. While differences in distribution can be seen in Figure 6-1 (i.e. for the volume group of 4,000 to 7,500 vehicles per day, the difference in percentages between Alberta and the TRB data is about two and a half times), the cumulative distribution of volume groups illustrates more clearly the skew seen in the TRB samples towards larger ADT values for roadway sections.

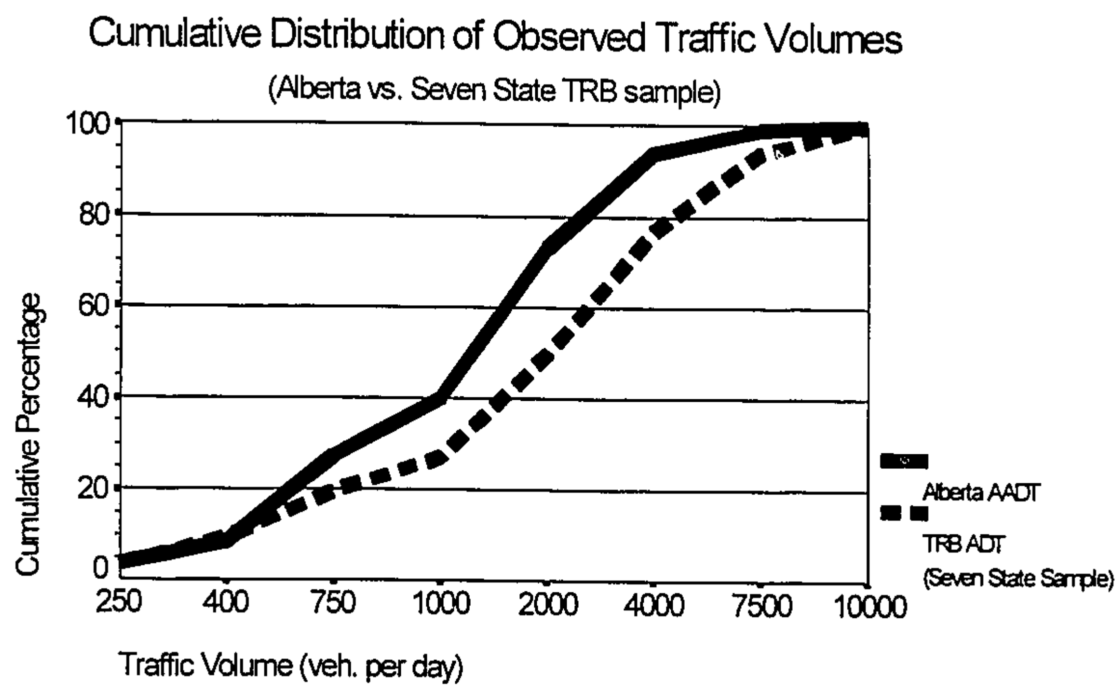


Figure 6-2. Cumulative Distribution of Observed Traffic Volumes (ALB vs. TRB) (TRB Sample source- Zegeer et al., 1987)

These two figures clearly show that there is a noticeable difference in proportion of ADT/AADT groups, especially at the higher volume end. Whereas nearly 80% of the Alberta database contains AADT values of less than 2,000 vehicles/day, only 50% of the TRB ADT values fall in that category. Because a sizable proportion (approximately

25%) of the TRB database consists of ADT values greater than 4,000 vehicles/day, it can be concluded that the sample of U.S. roadways contained a greater percentage of high volume roads than the corresponding Alberta database.

This conclusion may have a direct impact on Alberta accident rates, both in terms of AADT and by highway length alone. This difference in the distribution of roadway volumes (with the U.S. TRB sample containing far more high-volume roadways than the Alberta sample) may explain why accident rates appear to be lower in Alberta since lower volumes imply fewer potential vehicle interactions and conflicts.

6.3 TRB CHI-SQUARE ANALYSIS

A Chi-square contingency table was produced by Zegeer et al. (***Zegeer et al., 1987***) to confirm the significance of accident types modeled in their evaluation of safety effects of geometric cross-section design. Zegeer et al. identified seven types of accidents occurring on rural two-lane roadways that appeared to be significantly influenced by the following cross-sectional design variables: (1) lane width, (2) shoulder width, (3) shoulder type, (4) median sideslope code, (5) median roadside rating, and (6) median recovery distance. Of the seven geometric variables involved in the analysis, five were found by Zegeer to be consistently correlated with certain accident types. They were: lane and shoulder width, shoulder type, median roadside hazard rating, and roadside recovery distance measures. Two of the geometric variables appeared to be closely related in his findings- those involving the roadside hazard rating (a subjective measure of the degree of hazardness of the roadside), and the roadside recovery distance (also related to the roadside hazardness measure).

Accident types found to be significantly correlated with the previous geometric features were found by Zegeer et al. to be: (1) total accidents, (2) fixed object collisions, (3) ROR rollover and other, (4) opposing direction sideswipes, (5) single vehicle accidents, and (6) selected multi-vehicle, opposing direction accidents. The Chi-square analysis yielded no contingency coefficients in excess of 0.363, yielding an upper one-third coefficient range between 0.220 to 0.363. Of those categories found to consistently exhibit high levels of significance with respect to the relationship with geometric factors being analyzed, three categories were found to be appropriate and logical for use in the final model development: (1) single vehicle accidents, (2) single and selected multi-vehicle accidents, and (3) total accidents. The focus of the Alberta Chi-square analysis was to verify that these accident groupings were valid for use in an Alberta accident rate model.

6.4 ALBERTA CHI-SQUARE ANALYSIS

The results from the Chi-square analysis performed on the Alberta accident data is shown in Table 6-4. Included for each accident type and geometric factor pairing is the Chi-square value obtained from the test, as well as the level of significance that the null hypothesis is valid over. The null hypothesis in this analysis is the statistical hypothesis that the accident type and the geometric factor in question are independent of each other.

Table 6-4 Chi-square Values for Accident Types (Alberta primary, rural, 2-lane highways 1991-94)

Type	LANEWD	SHOUT	SHTYP	SPDAY	MSS	MDT	MBS	ESS	EDT	EBS
1. Struck Object	7.571 (0.751)	58.807 (0.373)	1.570 (0.210)	0.867 (0.833)	47.841 (0.058)	26.283 (0.069)	33.355 (0.264)	3.020 (0.555)	3.919 (0.561)	6.456 (0.264)
2. ROR Left	11.286 (0.420)	87.095 (0.005)	1.185 (0.276)	26.034 (0.000)	41.324 (0.181)	17.440 (0.425)	28.634 (0.484)	11.675 (0.020)	10.234 (0.069)	7.423 (0.191)
3. Rt. Angle	34.050 (0.000)	147.383 (0.000)	0.981 (0.322)	6.071 (0.108)	53.540 (0.018)	20.869 (0.232)	40.138 (0.082)	10.822 (0.029)	3.223 (0.666)	2.622 (0.758)
4. Pass Left Turn	24.485 (0.011)	124.159 (0.000)	1.786 (0.181)	10.703 (0.013)	71.822 (0.000)	49.654 (0.000)	59.220 (0.001)	11.608 (0.021)	6.125 (0.294)	10.999 (0.052)
5. Left Turn Approach	8.816 (0.639)	53.203 (0.581)	1.677 (0.195)	14.882 (0.002)	32.295 (0.551)	16.178 (0.511)	30.835 (0.373)	6.182 (0.186)	8.034 (0.154)	6.835 (0.233)
6. Left Sideswipe	13.628 (0.254)	75.990 (0.039)	2.211 (0.137)	4.263 (0.234)	47.367 (0.064)	32.256 (0.014)	44.104 (0.036)	27.599 (0.000)	26.952 (0.000)	28.452 (0.000)
7. Other	4.089 (0.967)	55.158 (0.507)	0.853 (0.356)	2.218 (0.528)	39.458 (0.239)	21.112 (0.221)	33.037 (0.276)	1.347 (0.853)	1.388 (0.926)	9.706 (0.084)
8. Rear End	3.657 (0.979)	95.357 (0.001)	0.040 (0.841)	10.063 (0.018)	62.196 (0.002)	32.648 (0.012)	53.522 (0.004)	6.276 (0.179)	22.103 (0.001)	8.754 (0.119)
9. ROR Right	18.749 (0.066)	98.442 (0.000)	0.572 (0.449)	1.653 (0.647)	69.974 (0.000)	43.217 (0.000)	70.265 (0.000)	24.713 (0.000)	25.360 (0.000)	22.653 (0.000)
10. Head On	15.103 (0.178)	56.352 (0.462)	1.328 (0.249)	12.146 (0.007)	53.629 (0.017)	37.192 (0.003)	58.182 (0.001)	27.461 (0.000)	31.011 (0.000)	27.644 (0.000)
11. Pass Right Turn	4.861 (0.938)	44.383 (0.869)	1.954 (0.162)	3.374 (0.338)	15.781 (0.997)	9.482 (0.924)	16.849 (0.964)	1.975 (0.740)	4.313 (0.505)	2.216 (0.818)
12. Right Sideswipe	29.795 (0.002)	84.484 (0.008)	0.014 (0.906)	3.738 (0.291)	43.044 (0.138)	10.453 (0.884)	42.398 (0.052)	8.136 (0.087)	10.855 (0.054)	9.288 (0.098)
13. Backing	12.587 (0.321)	36.413 (0.980)	0.089 (0.765)	0.909 (0.823)	35.189 (0.412)	21.115 (0.221)	27.197 (0.561)	5.605 (0.231)	3.317 (0.651)	8.073 (0.152)
14. Total ROR (2,9)	26.347 (0.006)	161.849 (0.000)	2.437 (0.119)	24.246 (0.000)	95.425 (0.000)	53.065 (0.000)	100.609 (0.000)	40.193 (0.000)	39.865 (0.000)	40.979 (0.000)
15. Angle + Turn (3,4,5,11)	37.339 (0.000)	127.740 (0.000)	0.361 (0.548)	26.068 (0.000)	45.611 (0.088)	32.812 (0.012)	33.841 (0.245)	6.580 (0.160)	2.851 (0.723)	5.630 (0.344)
16. Single Vehicle (1,2,9)	23.900 (0.013)	170.645 (0.000)	1.078 (0.299)	27.522 (0.000)	85.174 (0.000)	50.827 (0.000)	92.546 (0.000)	36.562 (0.000)	37.891 (0.000)	39.442 (0.000)
17. Single+Multi-Vehicle (1,2,4,5,6,9,10)	19.274 (0.056)	120.952 (0.000)	1.435 (0.231)	16.929 (0.001)	61.049 (0.003)	34.918 (0.006)	64.349 (0.000)	16.410 (0.003)	18.370 (0.003)	17.310 (0.004)

By proving the null hypothesis false, the opposite of the null hypothesis (that the accident type and the geometric factor are dependent) then holds true. This alternative hypothesis (that a relationship does exist between accidents and geometric factors) is the ultimate goal of this analysis.

Using a 0.05 level of significance, those cells that have significance of less than 0.05 are in bold-type to indicate the rejection of the null hypothesis and the acceptance of the alternative hypothesis. These cells appear to exhibit some degree of relationship between the accident type and the corresponding geometric factor in question. The 0.05 level of significance was selected on the basis of commonly accepted statistical practice as well as matching the confidence limit used by Zegeer et al. (***Zegeer et al., 1987***).

Because of differences in recording accident types between the TRB accident database and the Alberta accident database, a exact comparison between the Alberta and TRB accident data was not possible. However, certain accident types were grouped together in the Alberta database to form the following groupings that mirror those analyzed in the TRB research. These additional groups are included at the bottom of Table 6-4 and are: (1) Total ROR, (2) Angle and Turning Accidents, (3) Single Vehicle Accidents, (4) Single and Selected Multi-vehicle Accidents, and (5) Total Accidents.

The following cross-sectional geometric factors were selected for the chi-square analysis illustrated in Table 6-4:

1. lane width (LANEWD) - lane width, shoulder width, and shoulder type were selected for the analysis based on their presence, not only in the TRB model, but in most bodies of research in this area as well (***TRB, 1987***)

2. outside shoulder width (SHOUT)

3. shoulder pavement type (SHTYP)

4. daytime posted speed (SPDAY) - this geometric related factor was included in the chi-square analysis to provide additional verification of the significance of the other geometric factors. Despite it's somewhat qualitative relationship with the other primary cross-sectional geometric factors, posted speed limits are established on a sufficiently quantifiable design speed basis as to make their relationship with each accident event type valid. Regardless, the availability of this data for analysis resulted in it's inclusion in the chi-square analysis.

5. mean sideslope ratio (MSS) - the following mean and extreme values represented the value of sideslope ratio, ditch width, or backslope ratio observed for each sub-section length. Because of differences between Primary Highway Inventory sections and Safety Sub-sections as detailed in chapter 5, two separate values were calculated for each roadside geometric factor: (1) a mean value, representing the mean value of each factor over the entire sub-section length, and (2) an extreme value, representing the maximum ditch width, or minimum slope ratio observed at any point in the sub-section length. While these values were not directly analyzed in the TRB

research, they are represented by the use of median roadside hazard indices and median roadside recovery distances in Zegeer et al.'s research (***Zegeer et al., 1987***).

6. mean ditch width (MDT) - see mean sideslope ratio
7. mean backslope ratio (MBS) - see mean sideslope ratio
8. extreme sideslope ratio (ESS) - see mean sideslope ratio
9. extreme ditch width (EDT) -see mean sideslope ratio
10. extreme backslope ratio (EBS) - see mean sideslope ratio

6.5 RESULTS OF ALBERTA CHI-SQUARE ANALYSIS

From the tests of significance shown in Table 6-4, the following individual and grouped accident types are significantly related to the majority of the primary cross-sectional geometric factors being examined: (1) pass left turn accidents (type 4), (2) total ROR accidents (type 14), (3) single vehicle accidents (type 16), (4) single and selected multi-vehicle accidents (type 17), and (5) total accidents. Because of the predominance of literature and research pointing to the relationships between accidents and lane and shoulder widths, only those accident event types exhibiting a high degree of relationship with those two geometric factors at a minimum were considered for incorporation into the Alberta Accident Model. The five significant accident types identified from Table 6-4 also exhibited high significance in five or more of the other geometric factors.

6.5.1 Pass Left Turn (Accident Event Type 4)

This category exhibited high levels of significance for relationships between the following geometric factors: lane width, shoulder width, daytime posted speed, mean sideslope ratio, mean ditch width, mean backslope ratio, as well as extreme sideslope ratio. As expected from expert judgment (***Zegeer et al., 1987***), both lane and shoulder width appear to play highly significant roles in this type of accident by dictating the separation between the turning vehicle, and the opposing direction vehicle.

6.5.2 Total ROR (Accident Event Type 14)

This event type appeared to exhibit one of the highest degrees of significance with respect to each geometric factor's relationship. All factors except for shoulder type were found to be significant within the 0.05 level of significance, with shoulder width, mean sideslope, mean ditch width, and mean backslope ratio especially so (below 0.001 significance level). These results also appear to agree with expert opinion that both shoulder factors and roadside factors are highly significant in run-off-the-road types of accidents (***TRB, 1987***).

6.5.3 Single Vehicle (Accident Event Type 16)

This category of accidents included all ROR accidents (Event Types 2 and 9), as well as those involving the vehicle striking a stationary roadside object (Event Type 1). Because the majority of the accidents in this group involved ROR Left and Right incidents, this group's Chi-square and significance results were very similar to those exhibited for total ROR accidents (Event Type 14). Again, this group appeared to be most strongly related to shoulder width, mean sideslope ratio, mean ditch width, and mean backslope ratio.

6.5.4 Single and Selected Multi-Vehicle (Accident Event Type 17)

This category of accident types was the only one of the five significant groups with a low level of significance for the lane width-accident type relationship. One possible explanation for this may be due to the confounding effects of other factors arising from vehicle interaction in the accident. Despite this, the other roadside geometric factors were found to be highly significant for this group of accident types-similar to those groups previously mentioned.

6.5.5 Total (Accident Event Type 18)

This category was analyzed to compare similar Chi-square results obtained by Zegeer et al., as well as to provide a measure of the effect of the key roadway geometric factors and their relationship with accident occurrence (***Zegeer et al., 1987***). The Chi-square and significance levels obtained from the analysis were consistent with those obtained by Zegeer et al., not only in the types of geometric factors found to be significant, but in their strength of relationship as well. The level of significance recorded for lane width, shoulder width, daytime posted speed limit, as well as all mean and extreme roadside values (sideslope ratio, ditch width, and backslope ratio) were all below 0.05.

6.6 ALBERTA CONTINGENCY COEFFICIENT ANALYSIS

Contingency coefficient analysis of the Alberta accident data yielded results similar to those obtained by Zegeer et al. Based on the same one-third range method used by Zegeer et al. to screen out less significant relationships (documented in Section 3.3.1), the following accident types were found to be significantly correlated with lane and shoulder width, daytime posted speed, mean sideslope ratio, mean ditch width, and mean backslope ratio:

1. Left turn passing accidents (Event Type 4)
2. All ROR accidents (Event Type 14)
3. Single vehicle accidents (Event Type 16)
4. Total accidents (Event Type 18).

Table 6-5 Contingency Coefficient Values for Accident Types (Alberta primary, rural, 2-lane highways 1991-94)

Type	LANEWD	SHOUT	SHTYP	SPDAY	MSS	MDT	MBS	ESS	EDT	EBS
1. Struck Object	0.0282	0.0783	0.0128	0.0095	0.0722	0.0535	0.0600	0.0182	0.0207	0.0266
2. ROR Left	0.0344	0.0952	0.0112	0.0522	0.0671	0.0436	0.0559	0.0357	0.0334	0.0285
3. Rt. Angle	0.0597	0.1230	0.0101	0.0252	0.0763	0.0477	0.0661	0.0344	0.0189	0.0169
4. Pass Left Turn	0.0506	0.1134	0.0137	0.0335	0.0883	0.0735	0.0800	0.0356	0.0259	0.0347
5. Left Turn Approach	0.0304	0.0745	0.0133	0.0395	0.0593	0.0420	0.0580	0.0260	0.0296	0.0273
6. Left Sideswipe	0.0378	0.0889	0.0152	0.0212	0.0718	0.0593	0.0693	0.0549	0.0542	0.0557
7. Other	0.0207	0.0759	0.0095	0.0153	0.0656	0.0480	0.0600	0.0121	0.0123	0.0326
8. Rear End	0.0196	0.0995	0.0021	0.0325	0.0822	0.0597	0.0763	0.0262	0.0491	0.0310
9. ROR Right	0.0443	0.1011	0.0078	0.0132	0.0872	0.0686	0.0873	0.0519	0.0526	0.0497
10. Head On	0.0398	0.0767	0.0118	0.0357	0.0764	0.0637	0.0795	0.0547	0.0581	0.0549
11. Pass Right Turn	0.0226	0.0681	0.0143	0.0188	0.0415	0.0322	0.0429	0.0147	0.0217	0.0156
12. Right Sideswipe	0.0558	0.0937	0.0012	0.0198	0.0685	0.0338	0.0679	0.0298	0.0344	0.0319
13. Backing	0.0363	0.0617	0.0031	0.0098	0.0619	0.0480	0.0545	0.0248	0.0190	0.0297
14. Total ROR (2,9)	0.0525	0.1292	0.0160	0.0504	0.1016	0.0760	0.1043	0.0662	0.0659	0.0668
15. Angle + Turn (3,4,5,11)	0.0625	0.1150	0.0062	0.0522	0.0705	0.0598	0.0607	0.0268	0.0177	0.0248
16. Single Vehicle (1,2,9)	0.0500	0.1326	0.0106	0.0537	0.0961	0.0744	0.1001	0.0631	0.0643	0.0655
17. Single+Multi-Vehicle (1,2,4,5,6,9,10)	0.0449	0.1120	0.0123	0.0421	0.0815	0.0617	0.0836	0.0423	0.0448	0.0435

Table 6-5 presents contingency coefficients for the various accident type categories. The upper one-third range of coefficients found was between 0.060 and 0.133. This range of coefficients did not include those obtained for the total accident group. When the one-third range of contingency coefficients used was compared to the coefficient range used by Zegeer et al., smaller values were found in general for the Alberta database. However, because the distribution of upper one-third values of coefficients throughout the Alberta contingency table were remarkably consistent with the TRB contingency table, differences in coefficient magnitude were seen to be primarily due to differences in the two accident databases' sizes, and not in the underlying accident type distribution of relationship with each geometric factor. In essence, the results of the Alberta contingency coefficient table strongly resemble those of the TRB contingency tables, with respect to both the accident type and geometric factors found to be highly correlated and significant.

Contingency coefficient results compared with those obtained by Zegeer et al. yielded several accident types for potential use in developing the Alberta Accident Model. Of significance in this analysis was that the Alberta Chi-square analysis appeared to closely parallel results obtained from Zegeer et al.'s study that indicated single vehicle accidents involving ROR accidents and fixed object collisions, related multi-vehicle and single vehicle accidents, as well as total accidents occurring on two-lane rural roadways were applicable for use in modeling.

The Alberta contingency coefficient analysis consistently placed single vehicle accidents and total accidents in the upper one-third of contingency values for all three roadway geometric variables and three roadside variables being analyzed. In addition, both left turn passing accident and total ROR accident categories were also found to be correlated with the key geometric factors. These results appeared to validate the use of the accident types selected for use in the Alberta Accident Model, and yielded a valuable comparison of accident types with the Zegeer et al. research and the TRB models.

This agreement between Chi-square analysis results between two relatively disparate accident samples seems to indicate that the use of either single vehicle accident types (Event Type 16) or total accidents (Event Type 18) for developing an observed accident rate experience would be suitable for the Alberta Accident Model. In light of this assumption, total accidents were selected as the accident group to be analyzed in the regression analysis due to their consistently high significance and contingency coefficient scores in relation to the key roadway and roadside geometric factors.

6.7 CROSS-CORRELATION BETWEEN GEOMETRIC FACTORS

One potential issue identified from prior literature in the area of safety-geometric accident modeling involved collinearity or cross-correlative effects between geometric factors. As addressed in Chapter 5, the identification and resolution of this problem is critical to a robust regression analysis due to the potentially confounding nature of collinear factors on any regression analysis. For example, should two geometric factors be correlated not only with the accident occurrence but also highly correlated with each other, it would not be statistically possible to identify which of the two factors (or combination of the two) is statistically responsible for the accident-geometric relationship.

In this cross-correlation analysis, Pearson correlation coefficients were calculated for each pair of geometric variables to be modeled, with absolute values ranging from 0.0 to 1.0. The Pearson correlation coefficient is a commonly used measure to quantify the strength of linear association between two variables (*Norusis, 1992*). Based on this coefficient, any pair of factors exhibiting a high coefficient value signifies a strong cross-correlative effect between the two, and also a potential problem for a regression model. Table 6-6 displays the Pearson correlation coefficients for the following geometric variables to be potentially included in the Alberta Accident Model: (1) AADT, (2) Lane width, (3) Shoulder width, (4) Mean sideslope ratio, (5) Mean ditch width, and (6) Mean backslope ratio. Those coefficients with values of 0.9 or greater are indicated in bold-print to signify potentially strong collinear effects.

Table 6-6 Cross-Correlation Matrix for Significant Alberta Geometric Factors (Pearson Correlation Coefficients)

	AADT	Lane Width	Shoulder Width	Mean Sideslope Ratio	Mean Ditch Width	Mean Backslope Ratio
AADT	-	0.0769	0.3618	0.3550	0.3212	0.3678
Lane Width	0.0769	-	0.1028	0.1308	0.1439	0.0835
Shoulder Width	0.3618	0.1028	-	0.7437	0.7185	0.7352
Mean Side-slope Ratio	0.3550	0.1308	0.7437	-	**0.9065**	**0.9826**
Mean Ditch Width	0.3212	0.1439	0.7185	**0.9065**	-	**0.8950**
Mean Back-slope Ratio	0.3678	0.0835	0.7352	**0.9826**	**0.8950**	-

As seen in Table 6-6, all roadway geometric factors exhibit a certain amount of collinearity with each other. This is as expected, since Alberta roadway geometric design standards tend to emphasize coherent design practices (for example, high AADT = wide lanes and shoulders) (***Kenny, 1995***). It is also unsurprising that AADT is more correlated with shoulder widths and roadside geometric factors than with lane widths. This is primarily due to the rigid standardization of Alberta's primary highway design guideline of 3.70 m for each traveled lane width. Because of the distinct lack of data for lane widths deviating significantly from the 3.70 m guideline, it is difficult to draw a correlation between AADT and lane width. This correlation between AADT and shoulder widths and roadside factors implies a distinct design guideline of milder secondary geometric features for higher AADT values. No combination of roadway and roadside geometric factors appear to be of great enough collinearity to be of concern for modeling.

This however, cannot be said for the roadside geometric factor pairings. As seen in Table 6-6, paired sideslope ratios, ditch widths, and backslope ratios yielded correlation coefficients ranging from 0.8950 to 0.9826. These high values seem to indicate a strong relationship between the three variables. Empirically, these strong relationships seem to have merit, as roadsides with ample width would intuitively have gentle sideslope ratios, wide ditches or drainage channels, and gentle backslopes. Conversely, roadsides with restricted clear widths may have more predominantly steeper sideslope grades and narrow ditches. As a result of this collinearity, the inclusion of each roadside variable as a separate entity in the regression analysis was not recommended as possible collinearity effects would mask the identification of the dominant contributing factor (of the two collinear geometric factors) to an accident-geometric model.

Instead, as seen in the Zegeer et al. research, a combinative factor was substituted for the three roadside geometric factors. Zegeer et al. examined two similar types of roadside factors: a median roadside hazard rating based on roadside grades and clear zones, and a median roadside recovery distance based on lateral distance required for a vehicular recovery from a roadside encroachment. By using guidelines established by Alberta Transportation and Utilities on clear zone distances for sideslopes, ditches, and backslopes, a roadside hazard rating was estimated for each combination of the three roadside factors (***Zegeer et al., 1987***). This corresponding hazard rating was then incorporated into the various regression models developed.

6.7.1 ESTIMATING ROADSIDE HAZARD RATINGS

Roadside hazard ratings were estimated for each sub-section of Alberta's two-lane, rural, primary highways by first establishing a roadside recovery distance value for each combination of sideslope ratio, ditch width, and backslope ratio. Next, using the conversion table between roadside recovery distances and hazard ratings developed by Zegeer et al., an approximate value for the roadside hazard rating can be determined. Table 6-7 shows Zegeer et al's hazard rating equivalencies for given roadside recovery distances.

Table 6-7 Roadside Hazard Rating Equivalencies for Roadside Recovery Distances

Average Roadside Recovery Distance	Corresponding Range of Roadside Hazard Rating	Most Likely Roadside Hazard Rating
0 to 5 ft.	4 to 7	5
6 to 10	3 to 7	4 or 5
11 to 15	1 to 5	4
16 to 20	1 to 5	3 or 4
≥ 20 ft.	1 to 4	3

Source- Zegeer et al., 1987

Despite Zegeer et al. effort to quantify the relationship between recovery distances and the hazard rating, the assignment of those ratings still involved much subjective judgment (***Zegeer et al., 1987***). As a result, the assignment of roadside hazard ratings for Alberta roadside conditions also remains, to a large extent, qualitative.

Roadside recovery distances for the Alberta Accident Model were estimated using the following formula:

$$R_{Rec} = (1.5\text{ m} \times \text{MSS}) + \text{MDT} \qquad \textbf{(EQN 6.4)}$$

where,
R_{Rec} = Roadside Recovery Distance, in metres
MSS = Mean Sideslope Ratio (expressed in terms of x:1)
MDT = Mean ditch width, in metres

(Source- Kenny, 1996)

It is significant to note here that the mean backslope ratio was not included in the roadside recovery distance calculations. This was a result of two reasons: (1) Zegeer et al. found that the clear zone dictating the roadside recovery distance usually ended at the hinge between the ditch and the reverse slope (backslope) (***Zegeer et al., 1987***), and (2) the resulting combination of the horizontal sideslope component and the ditch width component of the Alberta roadside recovery distance exceeded 20 feet in most cases.

Once R_{Rec} was calculated and converted into feet, the corresponding most likely value for the hazard rating was used. In cases where two hazard ratings were equally likely to occur, a mean value of the two ratings was used as the Alberta hazard rating. As mentioned previously, the majority (717 of 725 valid cases) of Alberta's roadside hazard ratings were 3.0, as expected. A noticeable proportion of sub-sections (31 cases) were missing roadside inventory information, and were subsequently omitted where roadside data was required for analysis. These sub-sections are located within National Park boundaries, and while are included within the accident database, they are not considered a part of the Alberta highway geometric inventory database.

6.8 SINGLE VARIABLE REGRESSION

One key aspect in the determination of a relationship between accident rate and geometric design is identifying any direct relationship between accident rates and individual geometric factors. Using single variable regression, the following geometric factors were examined individually for general effects on accident rate: (1) lane width, (2) shoulder width, (3) daytime posted speed limit, (4) mean sideslope ratio, (5) mean ditch width, (6) mean backslope ratio, and (7) estimated roadside hazard rating.

Each geometric factor was analyzed over an AADT range to account for the influence of traffic volumes on accident rates. By analyzing individual geometric factors over AADT, gains in safety could potentially be identified due to incremental improvements in geometrics over the broad spectrum of AADT values.

6.8.1 AADT, LANE WIDTH AND ACCIDENT RATES

In Figure 6-3, accident rates for each two-lane, rural, primary highway sub-section in Alberta were plotted against AADT to identify underlying general trends in the accident-geometric relationship. Because of the predominance of uniform lane widths standardized at 3.70 m within Alberta, it was not possible to perform a robust analysis of the effects of incremental lane widening on accident exposure rates. Of a total of 752 two-lane, rural, primary highway sub-sections in Alberta, 706 sections have a traveled lane width of 3.70 m. Because this group of lane widths accounts for nearly 95% of the entire sub-section population, and due to the lack of sub-sections with lane widths exceeding 3.70 m (5 cases) or lane widths less than 3.70 m (41 cases), an analysis of incremental changes in lane widths was not performed.

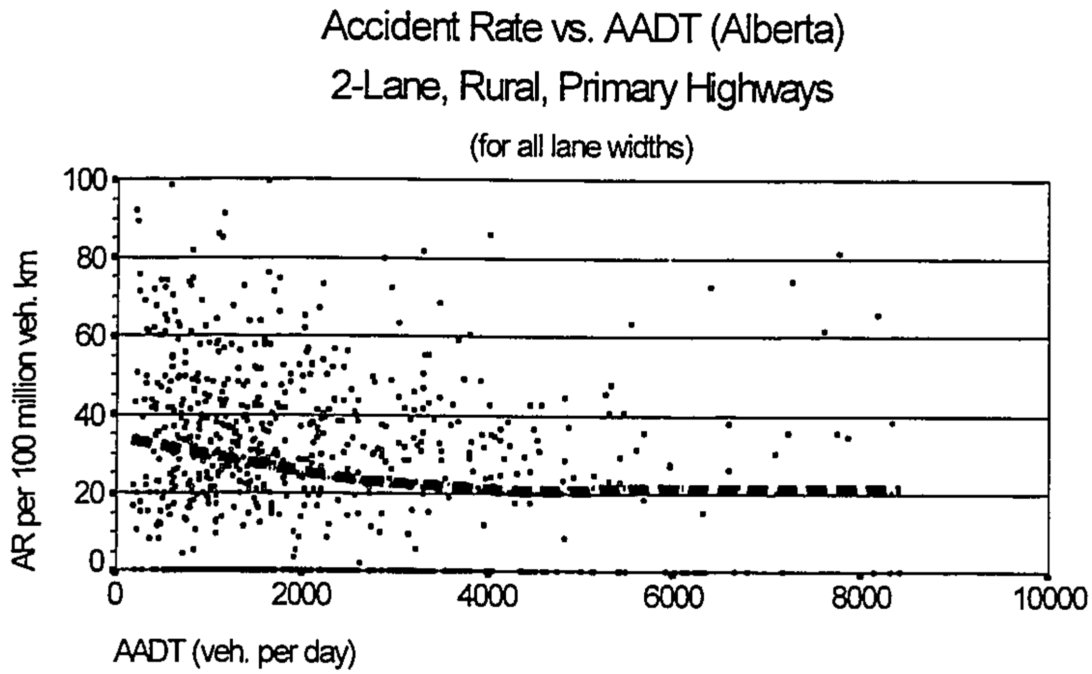


Figure 6-3. Observed Accident Rate Experience Over AADT (for all lane widths) (source- Alberta highway accident database, 1991-1994)

The regression line seen in Figure 6-3 represents a regression using the method of least squares estimate and locally weighted by variance about each AADT value to smooth the curve. The resulting regression line is essentially plotted through the mean values of accident rates for all AADT values. The R^2 value for this function was found to be 0.032, indicating an extremely large degree of dispersion over a 95% confidence interval.

As seen in Figure 6-3, there is a noticeable declining trend in observed accident rates as AADT increases for a given sub-section in Alberta. At extremely low AADT values (less than 100 vehicles per day), the accident rate, expressed in terms per million vehicles miles traveled, is approximately 35 per 100 million veh. km. This rate declines as AADT increases, reaching a low value of about 0.3 accidents per million veh. miles. It should be noted that there is also a slight but significant increase in accident rates when AADT exceeds 5,000 vehicles/day.

Because the accident rate is measured in terms of a vehicle exposure rate on the sub-section (per million vehicle miles traveled) the initial decline in accident rate can be directly attributable to the increase in AADT volumes. If accident occurrence frequency is equal for two sub-sections of unequal traffic volumes, the sub-section with a higher AADT volume will experience a lower accident rate in terms of vehicle distance traveled. Conversely, low volume sub-sections experiencing a given accident occurrence frequency will appear to exhibit an unusually high accident rate per vehicle distance traveled on that sub-section.

The noticeable increase in accident rates when traffic volumes exceed 5,000 vehicles/day can be attributed to increasing friction between operating vehicles on the road, and consequently, increased probability of collision or accident. These two conclusions appear to be supported by prior independent research conducted into accident rates by Alberta Transportation and Utilities for accident data from 1986 to 1990 (***Kenny, 1995***).

As expected, the data in Figure 6-3 shows a great deal of dispersion, with a very low R^2 value of 0.0136 from linear regression. Because of factors such as extremely low occurrence frequency and other potentially confounding non-geometric contributors to an accident, as explained in chapters 2 and 3, such dispersion and scatter was expected. To present the relationship between accident rate and AADT more clearly, mean values of accident rates per million vehicle miles were calculated over groups of AADT values. The AADT groups used in this a subsequent single variable analysis were based on ADT groupings used by Zegeer et al. (*Zegeer et al., 1987*). They are: (1) 0-400 vehicles/day, (2) 401-750 vehicles/day, (3) 751-1000 vehicles/day, (4) 1001-2000 vehicles/day, (5) 2001-4000 vehicles/day, (6) 4001-7500 vehicles/day, and (7) greater than 7501 vehicles/day. Figure 6-4 presents the same data as Figure 6-3, with mean values representing groups of accident rate values for given AADT groups.

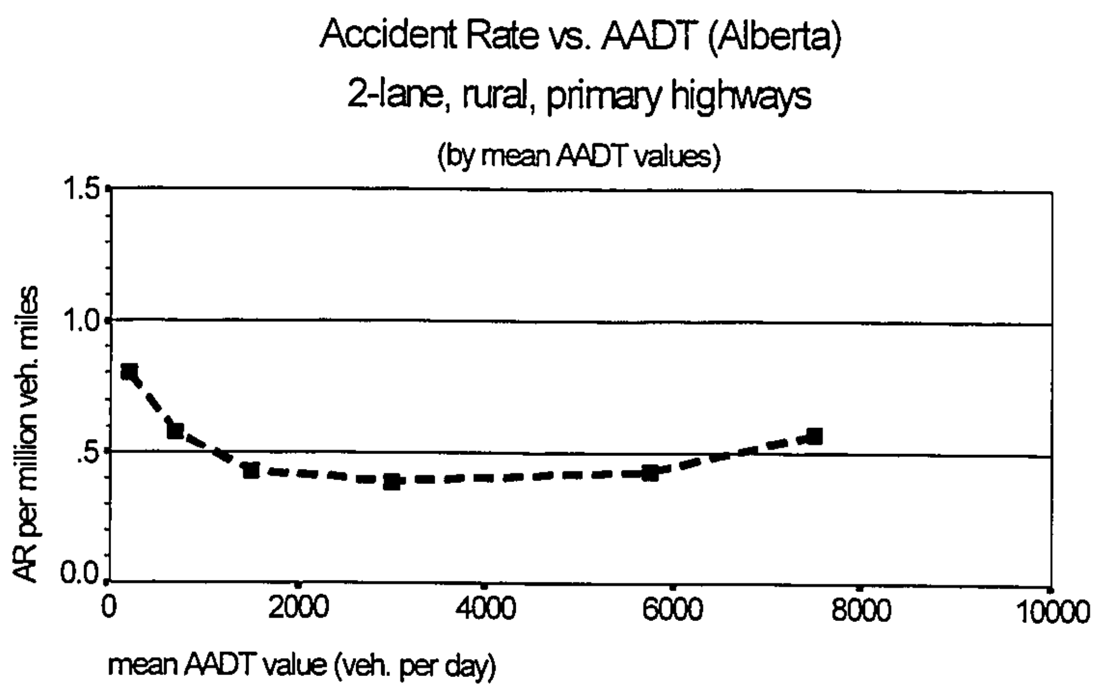


Figure 6-4. Observed Mean Accident Rate Experience Over AADT (source- Alberta highway accident database, 1991-1994)

Figure 6-5 presents the same information as Figure 6-4 using the standard metric units used by Alberta Transportation and Utilities within Alberta. It is standard practice in

Alberta to express accident exposure rate in terms per 100 million vehicle kilometres traveled.

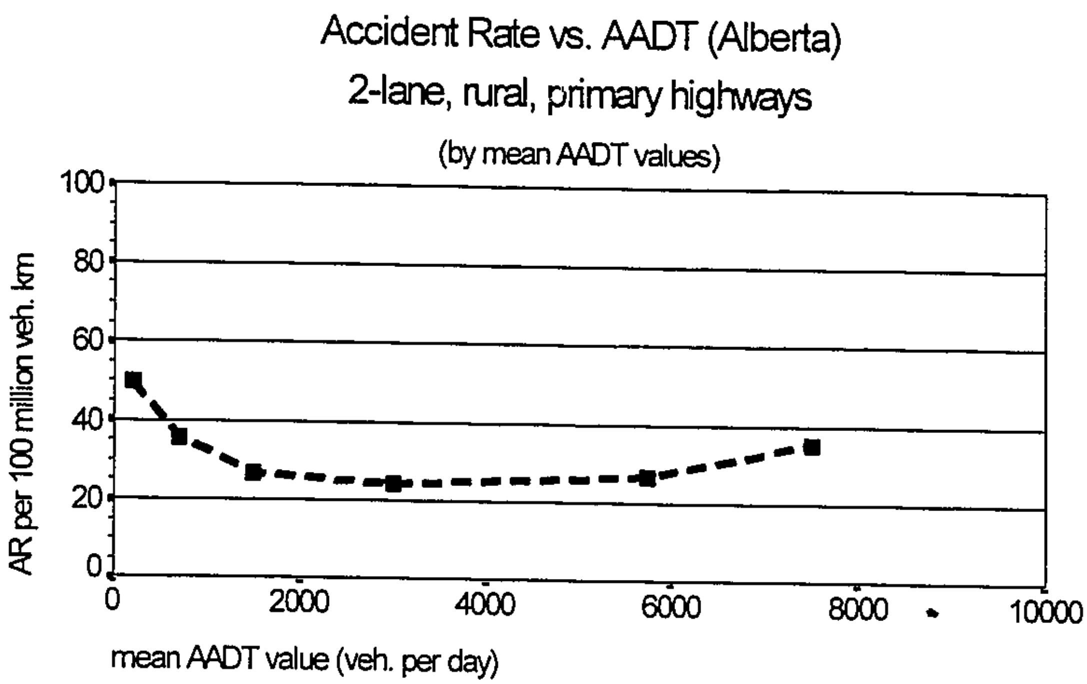


Figure 6-5. Observed Accident Rate Experience Over AADT (for all lane widths) (source- Alberta highway accident database, 1991-1994)

The data points used in these figures were derived from weighted mean accident values over each corresponding AADT value. As mentioned previously, there is a decline in accident rate as AADT volumes increase for a given roadway section, until approximately 5,000 vehicles/day are using the road. At this point, due to increasing vehicle conflicts and increases in collision probability, accident rates begin to increase again.

While the trend exhibited by Figures 6-3, 6-4, and 6-5 are all similar to that in the 3R/4R Geometric Design Guidelines by Alberta Transportation and Utilities, the actual values of accident rates is significantly lower, in some cases, as much as 50%. This is primarily due to the size of the final accident database used in this analysis. While only 9,535 of the original 33,582 accidents were used as a result of data suitability, the accident rates developed for the 1995 3R/4R Geometric Design Guidelines included all accident types. As such, it should be stressed that the resulting accident rates developed from this analysis are observed "geometric-attributable" accident rates, and not total accident rates.

Alberta's observed accident rates were also compared with rates developed by Zegeer and Deacon at the University of Kentucky (***Zegeer and Deacon, 1987***). Figure 6-6 compares accident rates observed for all traveled lane widths in Alberta (but primarily 3.70 m or 12.2 ft wide) and 11 to 12 feet wide traveled lane widths in Kentucky (***see Zegeer, Mayes and Dean, 1981***). AADT values were plotted on a log-scale to linearize the resulting accident relationship. Accident rates between the two plots are also slightly different, in that the Alberta accident dataset contained all single and multiple vehicle geometric related accidents, while those from the TRB dataset contained only ROR accidents. The effects of this are not felt to be significant as the majority of the Alberta accidents are of the ROR type (5,313 of 9,535 total accidents).

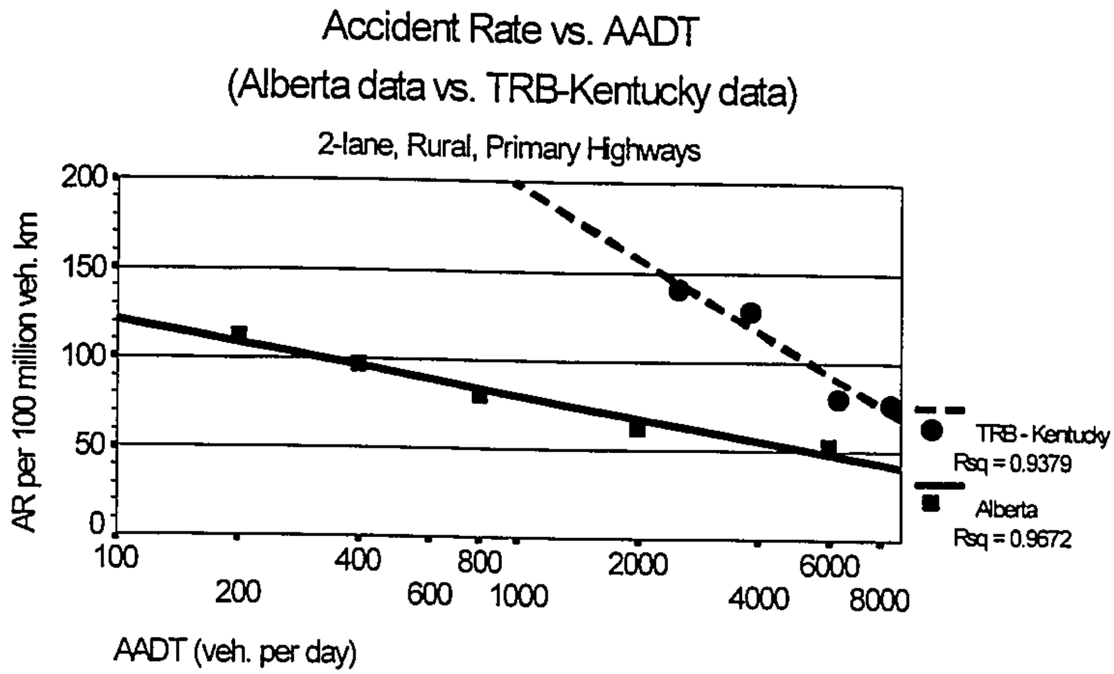


Figure 6-6. Observed Accident Rate Experience Over AADT (log-scaled) (source- Zegeer, Mayes and Dean, 1981)

As seen in the figure, both observed accident rates decline noticeably as AADT increases (note that the increase in the Alberta accident rate at higher AADT values is not noticeable when plotted on a log scale). The differences in accident rates between the two samples for a given AADT may be due to the narrower lane widths seen in the TRB sample, as well as a smaller length of highway than used for Alberta in determining total vehicle miles traveled (***Zegeer, Mayes and Dean, 1981***). Zegeer et al. also noticed a significant decline in accident rates due to incremental increases (from 9 and 10 feet to 11 and 12 feet lane widths) in lane widths. Although this was also seen in results developed by Alberta Transportation (3R/4R Geometric Design Guidelines, 1995) for various lane and shoulder combinations, these conclusions could not be substantiated in this analysis, due to the high degree of lane width standardization in Alberta and resulting lack of variance in lane width values. However, on the basis of those conclusions, it would be expected that the Alberta accident rate would be lower than those rates from Kentucky, due to differences in lane widths.

6.8.2 AADT, SHOULDER WIDTH AND ACCIDENT RATES

While design standardization prevented the analysis of the effects of incremental changes in lane widths on accident rates, it was not an issue for shoulder widths.

Table 6-8 Mean Accident Rates in Alberta by Shoulder Width and AADT, 1991-94

Shoulder Width (m)	AADT (vehicles/day)	Mean Accident Rate (AR / 100 million veh. km)	Number of Sub-sections
0 - 1.20 m	0 - 400	47.32	20
	401 - 1000	39.67	66
	1001 - 2000	30.50	67
	2001 - 4000	26.56	41
	4001 - 7500	15.81	8
	7501 -	0	0
1.30 - 1.80 m	0 - 400	53.75	11
	401 - 1000	34.22	49
	1001 - 2000	26.82	84
	2001 - 4000	24.07	46
	4001 - 7500	30.12	18
	7501 -	0	0
1.95 - 2.70 m	0 - 400	51.05	3
	401 - 1000	28.90	31
	1001 - 2000	22.87	70
	2001 - 4000	23.57	78
	4001 - 7500	20.87	29
	7501 -	27.06	3
> 2.70 m	0 - 400	0	0
	401 - 1000	38.87	14
	1001 - 2000	26.88	24
	2001 - 4000	23.32	54
	4001 - 7500	33.13	30
	7501 -	39.29	6

Table 6-8 displays mean (aggregated) observed accident rates by shoulder width and sub-section AADT. Because of the level of dispersion seen in the Alberta accident database and exhibited in the plot of accident rate vs. AADT for lane widths (Figure 6-3), Table 6-8 was used to identify mean accident rates occurring over large numbers of sub-sections in Alberta. The threshold value for number of sub-sections was initially set at 5% of the entire population (38 sub-sections) and lowered to 2% (15 sub-sections). This removal of low occurrence mean accident rates (for a specific shoulder width and AADT combination) served to clarify the relationship between shoulder width, AADT, and accident rate by eliminating accident rates that may have been biased due to low population sizes. The accident rates for combinations of shoulder widths and AADTs that are applicable for at least 2% of the total sub-section population are in boldface type in Table 6-8. Based on Table 6-8, accident rates were plotted against AADT for various shoulder width groups, as seen in Figure 6-7.

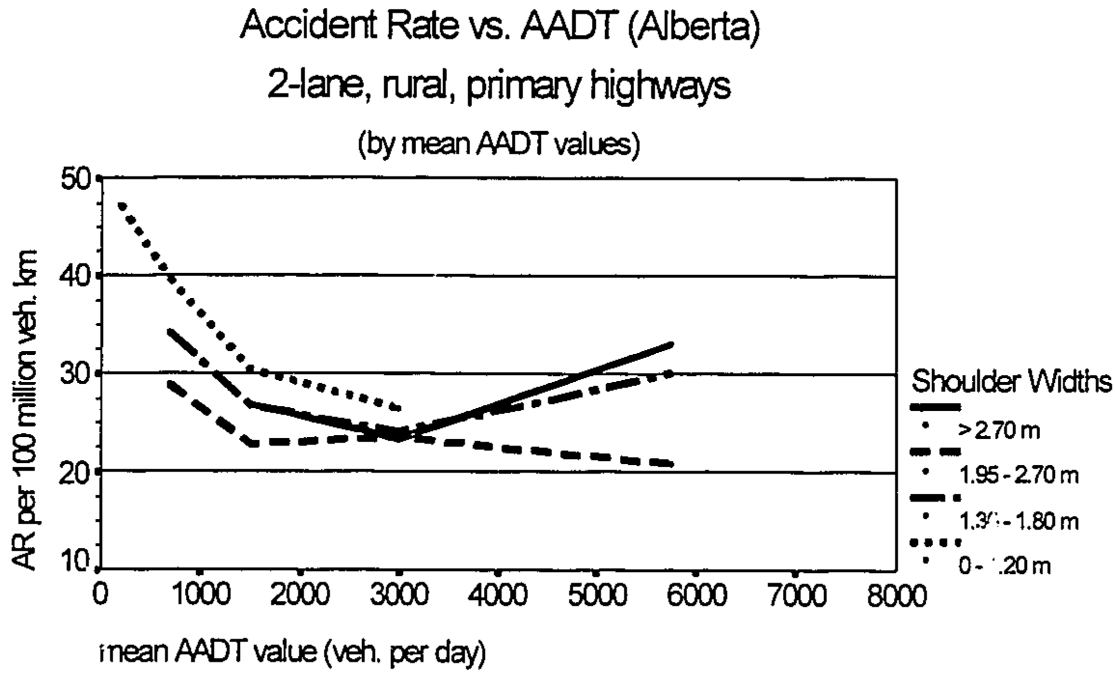


Figure 6-7. Mean Accident Rate Experience Over AADT (for shoulder widths) (source- Alberta highway accident database, 1991-1994)

Again, the mean accident rate values were determined over AADT, and plotted. Regression lines were established using the method of least squares. In Figure 6-7, two items should be noted. First, the concave (declining and increasing over AADT) trend first seen in Figure 6-3 for lane widths is also seen in general here. It is interesting to note that while three of the four shoulder width groups both appear to experience minimum observed accident rates at approximately 3,000 vehicles/day (AADT), the fourth group (shoulder widths between 0 and 1.20 m) could also have a minimum accident rate occurring at 3,000 vehicles/day if it's curve was extrapolated into the upper AADT regions. This appears to indicate that, for shoulder widths alone, an AADT value of 3,000 vehicles/day may be the threshold where incremental gains achieved through shoulder width improvements may be outstripped by rising AADT (and corresponding increases in vehicle friction and collision probability).

The second point of interest is that while there appears to be a declining trend in accident rate as shoulder width increases, the decrease in accident rate over incremental increases in shoulder widths appears to be less pronounced than the TRB accident rate-shoulder width relationship. Figure 6-8 illustrates the differences in accident rate for given shoulder widths between the observed Alberta accident data and the TRB developed rate more clearly.

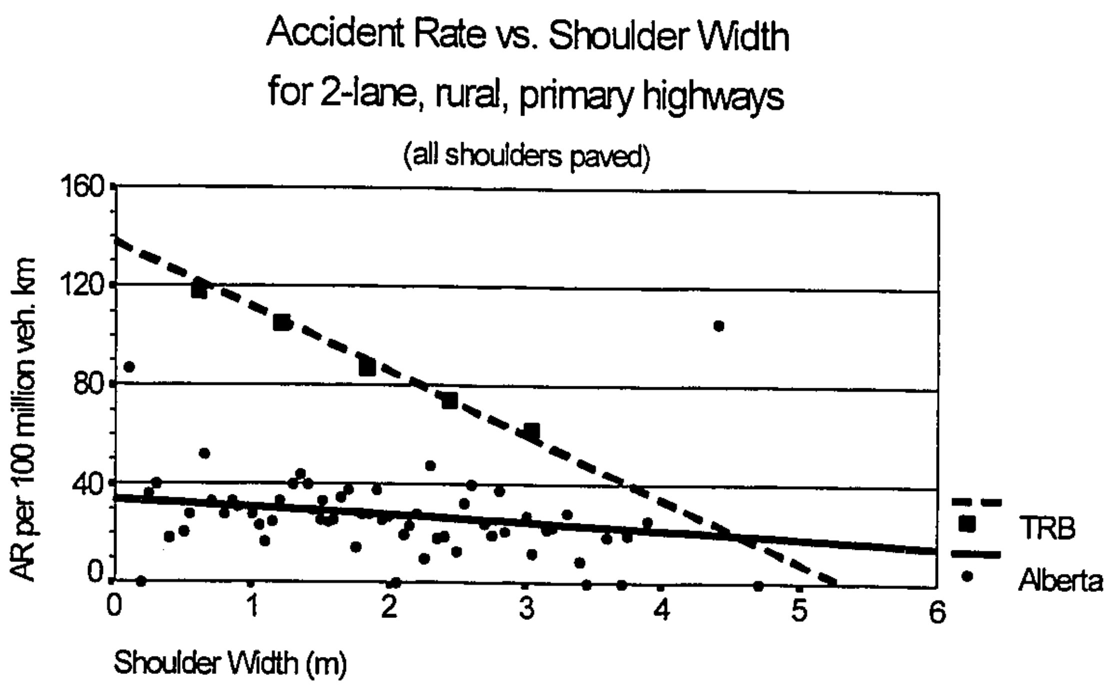


Figure 6-8. Accident Rate Experience Over Shoulder Widths (Alberta source- Alberta highway accident database, 1991-1994 TRB source- TRB, 1987)

TRB accident rates in Figure 6-8 were derived from accident data for a standardized lane width of 12 feet (***TRB, 1987***). As mentioned previously, the majority (706 of 752 sub-sections) of Alberta's primary highway lane widths have been standardized at 3.70 m. Because this value corresponds to the U.S. 12 foot lane width value, the comparison between the TRB and Alberta results for lane widths of approximately 3.70 m wide should be valid.

In order to display the degree of dispersion over various shoulder widths, mean values were not used to normalize the Alberta accident data plot. The resulting linear regression line for the Alberta data in Figure 6-8 appears to exhibit very little decrease in accident rate over incremental increases in shoulder width (from approximately 35 accidents per 100 million vehicle km to 15 accidents/100 million veh. km between 0 and 5 m shoulder widths). While it is possible that Alberta primary highways exhibit a lower accident rate than those surveyed for the various TRB studies, it is counter-intuitive that there is very little change in accident rate as shoulder widths increase. This result may possibly be due to an insufficiently large enough Alberta accident sample size, as well as an insufficient geometric database to determine a relationship.

6.8.3 AADT, DAYTIME POSTED SPEED AND ACCIDENT RATES

The relationship between daytime posted speed limits and accident rates was examined to verify possible design relationships between posted speed limits and corresponding geometric factors such as lane width, shoulder width, and roadside geometrics through accident rates. Figure 6-9 below display the relationship between daytime posted speed limits on Alberta's two-lane, rural, primary highways, corresponding AADT values, and observed accident rates.

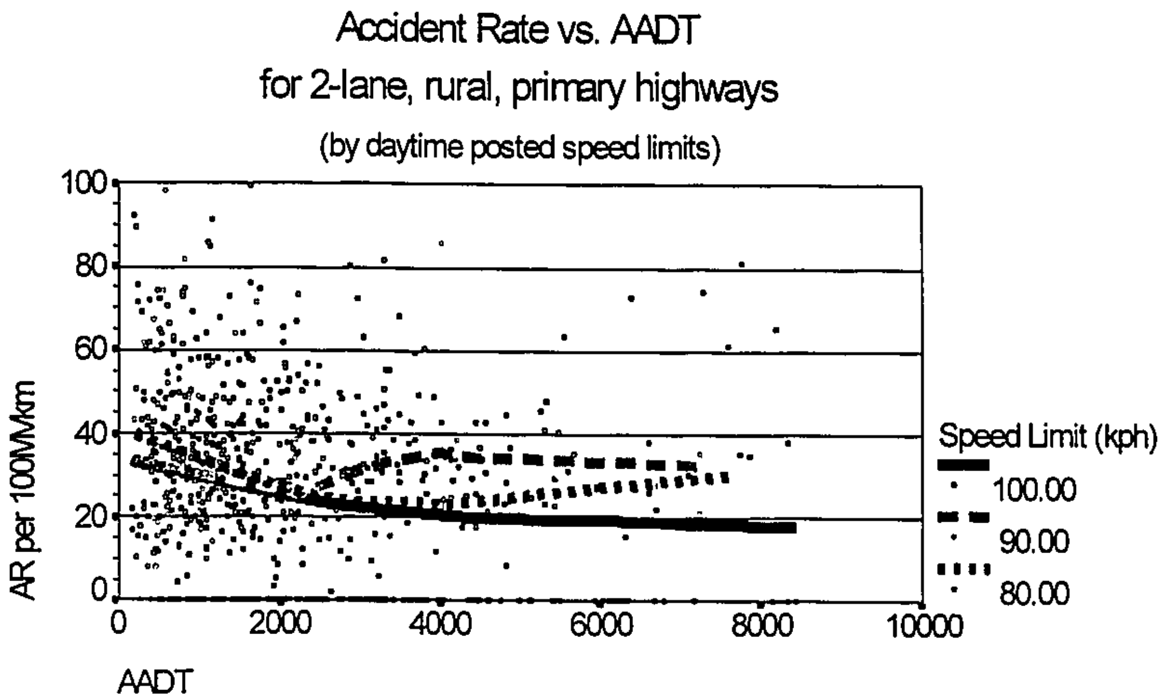


Figure 6-9. Mean Accident Rate Experience Over AADT (for Speed Limits) (source- Alberta highway accident database, 1991-1994)

As seen in Table 5-3 in Chapter 5, the predominant posted speed limit on Alberta's two-lane primary highways is 100 km/hr. Because this group comprises nearly 80 percent of the total sub-sections, it's corresponding relationship with AADT and observed accident rate should be the most reliable. As seen in Figure 6-9, there appears to be a steady and noticeable decline in the 100 km/hr accident rate from extremely low AADT values (under 50 vehicles per day) to approximately 3,000 vehicles/day. There appears to be very little change (a very slight decrease) in accident rate above 3,000 vehicles/day for the 100 km/hr sub-sections. In the case of the 90 km/hr and 80 km/hr sub-sections, a noticeable rise occurs in accident rates above 2,000 vehicles/day and 3,000 vehicles/day respectively. While this result appears to correspond to accident rate increases seen with regards to lane widths and shoulder widths (Figures 6-5 and 6-7), the rise is probably due to a lack of sub-sections with posted speed limits under 100 km/hr and high AADT (above 3,000 vehicles/day) volumes as well. As seen in Table 5-3 of Chapter 5, no more than 85 and 59 of 752 roadway sub-sections exhibit posted speed limits of 80 km/hr and 90 km/hr respectively. Fewer than one percent of the total number of sub-sections have speed limits below 80 km/hr. As a result, the reliability of the 80 km/hr and 90 km/hr relationships with AADT and accident rate is questionable. It should also be noted that the discontinuity seen for the 90 km/hr function at the 2,000 vehicles/day AADT value is probably due to insufficient data, and not due to any underlying relationship between accident rate and AADT for 90 km/hr speed limits.

Because lane widths in Alberta have been predominantly standardized at 3.70 m wide, improvements in accident rates for various posted speed limits were attributable to other factors instead. The lack of a noticeable increase in accident rates for 100 km/hr posted speed limit sub-sections with AADT volumes above 3,000 vehicles/day may be a direct result of incremental improvements to shoulder widths and other roadway and roadside factors once a section of highway approaches a certain vehicle volume.

6.8.4 SIDESLOPE RATIO AND ACCIDENT RATES

Chapter 5 outlined the methodology and reasoning behind using mean values of sideslope ratios for each sub-section length. The initial analysis of the geometric database determined that approximately 30 percent (222 of 752 sub-sections) of the two-lane, rural, primary subsections did not have homogeneous sideslope ratios throughout the entire length of the sub-section. Table 6-9 shows the distribution of mean sideslope ratio values for the geometric database. Note that 31 cases corresponding to the National Park System primary highways within Alberta provincial boundaries were not included in the AT&U primary highway geometric inventory survey, and as a result, were missing all roadside geometric factors.

Table 6-9 Mean Sideslope Ratios
Alberta primary highways, 1991-94

Mean Sideslope Ratio Value	Number of Sub-sections	Percent of Total Sub-sections
<3 : 1	11	1
3 : 1	30	4
3 : 1 to 4 : 1	23	3
4 : 1	245	32
4 : 1 to 5 : 1	98	13
5 : 1	102	13
5 : 1 to 6 : 1	60	8
6 : 1	156	21
missing/unknown	31	4

A plot of a corresponding observed accident rate for each sub-section versus the sub-section's mean sideslope ratio was then performed and a weighted mean best fit curve was plotted through the points. Because certain sideslope ratios such as the 4:1 ratio occurred more frequently in the Alberta geometric database, each mean sideslope ratio value was weighted according to it's frequency of occurrence. The resulting plot is shown in Figure 6-10.

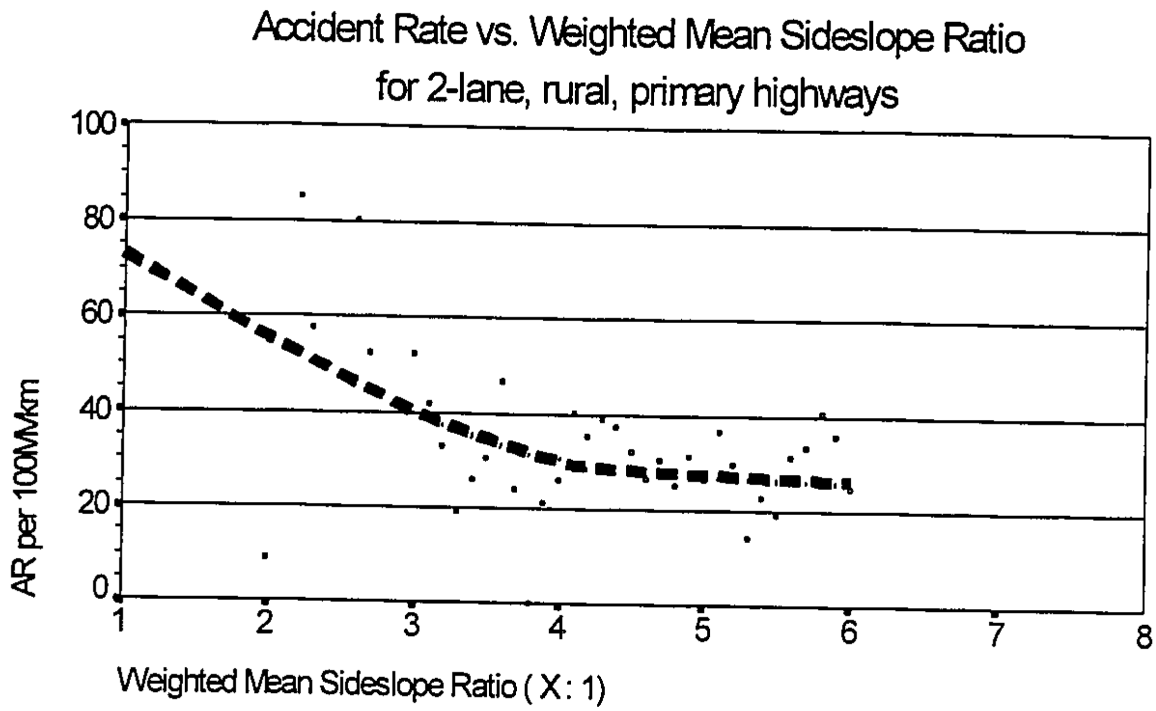


Figure 6-10. Accident Rate Experience Over Weighted Mean Sideslope Ratio (source- Alberta highway accident database, 1991-1994)

This figure clearly shows a substantial decline in observed accident rates when mean sideslope ratios are increased from 1:1 to approximately 3.5:1. Accident rates are observed to decrease by over 50 percent over the 1:1 - 3:1 ratio range. However, the decline in accident rates is much less noticeable for incremental sideslope ratio increases from 4:1 and onwards. In effect, this figure implies that the safety effects of incremental sideslope ratio flattening are more pronounced for ratios less than 4:1. For sideslope improvements exceeding the 4:1 ratio, the benefits of a improved safety through lowered accident rates may not outweigh the cost of the improvement itself. However, as seen in Table 6-9, further data may be required for sideslopes of 3:1 and steeper, as the limited data (41 sub-sections) may be skewing the plot results.

6.8.5 DITCH WIDTH AND ACCIDENT RATES

As with the sideslope ratio, roadside ditch widths obtained from the secondary Alberta geometric database were also non-homogeneous over most sub-sections. As a result, mean ditch width values were also calculated and plotted against individual sub-

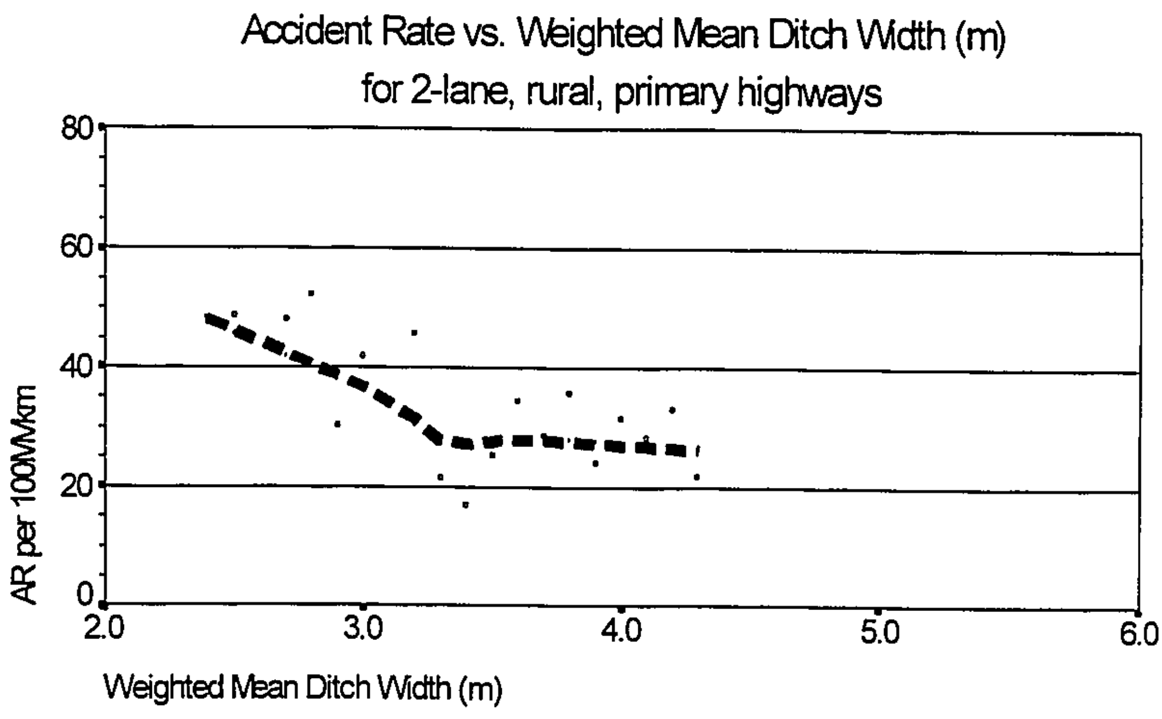


Figure 6-11. Accident Rate Experience Over Weighted Mean Ditch Width (source- Alberta highway accident database, 1991-1994)

section observed accident rates. Figure 6-11 displays the effect of mean ditch width values on accident rates.

Again, a possible insufficiency in accident data for ditch widths between 3.0 m and 3.5 m may have caused the discontinuity in the function seen in Figure 6-11. Because the function represents a smoothed best-fit regression line through mean values, the lack of sufficient accident rate values in the region mentioned may skew the results of the regression line.

Mean ditch width values for each sub-section were weighted by frequency of ditch width occurrence in order to find the best-fit regression line. Again, two trends should be noticed. As seen in Figure 6-11, accident rates decline as ditch widths increase in general, much the same as the trend seen for mean sideslope ratios (Figure 6-10). Second, accident rates appear to decline much faster between 2.4 m and 3.4 m mean ditch widths than over ditch widths of greater than 3.4 m. Again, this general trend was also exhibited in the plot of accident rates against mean sideslope ratios (Figure 6-10) and may be a direct result of the cross-correlation seen between the two factors (see Table 6-6).

The predominant number of sub-sections exhibited mean ditch widths of between 3.5 m and 4.3 m wide (664 of 752 sub-sections). Consequently, the lack of a significant number of sub-sections with mean ditch widths of less than 3.5 m may skew the relationship between accident rate and narrow ditch widths (less than 3.5 m). Despite this limitation, there appears to be two conclusions that may be drawn from Figure 6-11.

First, accident rates are generally seen to decline as ditch widths increase. While this may be a result of collinearity between the ditch width factor and the sideslope ratio, the decline in accident rates is nonetheless apparent and significant (approximately 50 percent decline from widths of 2.3 m to 4.3 m). Second, incremental improvements to ditch widths (primarily through widening of existing ditches) may not be cost-effective for existing ditches with widths of 3.0 m or wider. While there may be measurable safety advantages to widening ditches with widths of less than 3.0 m, there appears to be little, if any reduction in observable accident rate for ditch widths greater than 3.0 m.

6.8.6 ROADSIDE HAZARD RATING AND ACCIDENT RATES

Development of a relationship between estimated roadside hazard ratings for Alberta's two-lane, rural primary highways and accident rates was not possible due to the predominance of sub-sections with a hazard rating of 3 (717 of 725 valid cases). Based on the estimated median roadside recovery distances calculated from sideslope ratio and ditch width, and the corresponding estimated hazard rating, only one sub-section had a hazard rating greater than 3.5. As well, no sub-sections had estimated hazard ratings of less than 3. As a result, no single variable regression analysis was done between estimated median roadside hazard rating and corresponding accident rates.

Despite this fact, the median roadside hazard rating was still included in the generalized log-linear regression modeling and calibration for two reasons: (1) possible effects on other geometric factors when combined together that may not be apparent in single variable regression, and (2) direct comparison and calibration of the TRB model for Alberta use would not be as accurate without considering some measure of roadside geometrics, namely the median roadside hazard rating used in the TRB model.

6.9 GENERALIZED LOG-LINEAR REGRESSION MODELING

The combinative effects of the six geometric factors (AADT, lane width, shoulder width, mean sideslope ratio, mean ditch width, and estimated roadside hazard rating) on accident rates was examined next. As explained in the methodology outlined in Chapter 5, the form of the generalized log-linear model to be used for observed Alberta data was that of the TRB model:

$$AR = C_0\,(AADT)^{C_1}\,(C_2)^{W}\,(C_3)^{Sh}\,(C_4)^{H} \qquad \textbf{(EQN 6.5)}$$

where, AR = Accident rate in number of accidents per mile per year
C's = Model coefficients
AADT = Bi-directional average annual daily traffic volume, in vehicles per day
W = Lane width, in feet
Sh = Paved shoulder width, in feet
H = Estimated roadside hazard rating

Note that the accident rate used in this log-linear model is not the same accident rate used to develop the accident and geometric factor relationships seen previously in this chapter. As previously seen, it is common practice to identify accident rates occurring on a sub-section or at a location in relation to traffic volume experienced. For example, relationships developed in sections 6.8.1 to 6.8.5 all refer to accident rate in terms of number of accidents expected to be experienced over 100 million vehicle-kilometres traveled on the sub-section, or in terms of the U.S. units of million vehicle-miles traveled. These accident rates are deemed to be useful primarily because they account for the significant role that traffic volume (in terms of AADT) plays in highway vehicle accidents. However, because traffic volume is accounted for in the highway vehicle accident model in Equation 6.5, it was not necessary to include it again as a separate element of the accident rate.

The general regression function in SPSS® for Windows™ was used to generate the model coefficient parameters for Equation 6.5. SPSS® uses a sequential quadratic programming technique to identify optimum solutions to a generalized regression equation. This programming technique searches for an optimal solution using a double iterative step searching algorithm. The first (or major) iterative step involves searching for an optimal direction to apply the second (or minor) iterative step. This minor step searches along the identified vector for a local regression optimum, and then compares the new regression line's sum of squares (SS) residual value with the previous sum of squares residual to determine if the path choice is correct. If the new SS is smaller than the previous SS, the iteration continues from the new path.

6.9.1 RESULTS OF LOG-LINEAR REGRESSION MODELING

Initially, the coefficient parameters were unconstrained in the sequential quadratic programming to determine a base Alberta model to calibrate other model variations against. Initial estimates of coefficient parameters were as follows: C0 = 0.0, C1, C2, C3, and C4 = 1.0. The first model and coefficient parameters developed from the unconstrained log-linear regression is presented in Equation 6.6:

$$AR = 0.0006\,(AADT)^{2.716}\,(0.870)^{W}\,(0.963)^{Sh}\,(1.289)^{H} \qquad \textbf{(EQN 6.6)}$$

This equation was transformed into a log-linear format by taking the natural log of each side. This resulting transformed equation was then used to determine the significance of each coefficient parameter and geometric factor pair, using statistical tests of significance.

$$\ln[AR] = -7.379 + 0.999\ln[AADT] - 0.139W - 0.038Sh + 0.254H \qquad \textbf{(EQN 6.7)}$$

The corresponding version of Equation 6.6 in SI units is shown in Equation 6.8 and expresses accident rate in terms of number of accidents per kilometre per year.

$$AR = 0.0006\,(AADT)^{2.716}\,(0.959)^{W}\,(0.988)^{Sh}\,(1.080)^{H} \qquad \textbf{(EQN 6.8)}$$

This regression yielded a sum of squares residual value of 155.95 and an R^2 value of 0.656. This R^2 value is higher than the corresponding R^2 value found by Zegeer et al. (R^2 = 0.456) from Section 2.6.2, and indicates that the relationship developed appears to have a moderately high degree of correlation between accident rate and the various geometric factors. Figure 6-12 illustrates the distribution of residuals for both observed and expected accident rate values using a P-P (Expected Probability-Observed Probability) plot. The close fit of the residual line in comparison to the 1:1 slope indicates that the distribution of residual error from the geometric factors is normally and equally distributed over the analysis range. This result is important in meeting the regression assumptions criteria documented in Chapter 4 and validating the developed relationship.

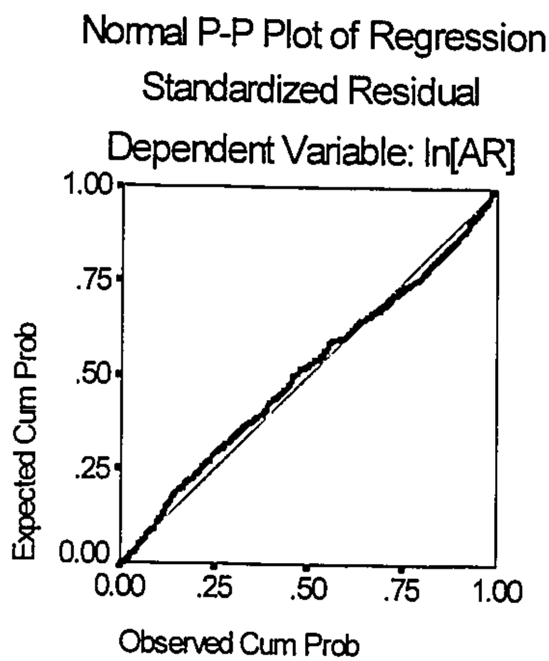


Figure 6-12. Residual P-P Plot of Observed and Predicted Accident Rates (for Equation 6.8)

Of note in this equation is that while all coefficients solved for in this model appear similar to those seen in the TRB model, some differences do exist. It appears that AADT may have a greater effect on accident rate in Alberta than for those states using in the TRB model, due to the larger exponential coefficient in Equation 6.6 (2.716 versus 0.882). Significantly, the effects of the other three geometric factors (lane width, shoulder width, and hazard rating) all appear to influence accident rate in much the same manner and magnitude as seen in the TRB accident rate model. A comparison of the coefficients and their effects on the geometric factors is given in Table 6-10.

Table 6-10 Comparison of Alberta and TRB Model Coefficients

Geometric Variable	Alberta Accident Rate Model Coefficient Value	TRB Accident Rate Model Coefficient Value
Constant	0.0006	0.0020
AADT	2.716	0.882
Lane Width	0.870	0.879
Shoulder Width	0.963	0.932
Hazard Rating	1.289	1.236

Aside from the gross differences seen in AADT coefficient values, the other differences between geometric coefficients were not significant. More importantly, the direction of influence found for each factor confirmed prior TRB and Alberta Transportation and Utilities' findings. The model in Equations 6.6, 6.7, and 6.8 shows that incremental increases in lane and shoulder widths result in decreases in accident rate, and that the increase in hazard rating also increases predicted accident rates.

Despite the fit of the model in relation to the TRB accident rate model, the issue of variable significance required changes to the final model form. Standard T-tests of significance were performed on the coefficients for each geometric variable to determine their significance in the equation. The results of these tests of significance are:

Table 6-11 Tests of Significance for Equation 6.6

Geometric Variable	T-Statistic Value	Significance
AADT	30.536	0.000
Lane Width	-2.023	0.044
Shoulder Width	-4.299	0.000
Hazard Rating	0.441	0.660
Constant	-3.494	0.000

It is evident from the level of significance in Table 6-11 that neither the hazard rating nor lane width geometric factor contains enough variability to be statistically significant in the analysis. Because of the absolute lack of variability in the Hazard Rating, an attempt was made to substitute an alternative measure of roadside safety using the estimated roadside recovery distance (RRec). However, the results of this factor were found to be equally insignificant with hazard rating.

The alternative model required removal of both the lane width factor and the hazard rating from the regression analysis. The resulting equation below yielded a sum of squares residual of 170.57 and a similar R^2 value of 0.654. All factors were found to be significant within the 95% confidence interval.

$$AR = 0.0002\,(AADT)^{2.748}\,(0.989)^{Sh} \qquad \textbf{(EQN 6.9)}$$

where, AR = Accident rate in terms of /km/yr
AADT= Bi-directional average annual daily traffic volume, in vehicles per day
Sh = Paved shoulder width, in metres

The probability plot of regression residuals for Equation 6.9 was very similar to Figure 6-12. Both P-P plots appear to indicate that the variance of the model is normally and equally distributed.

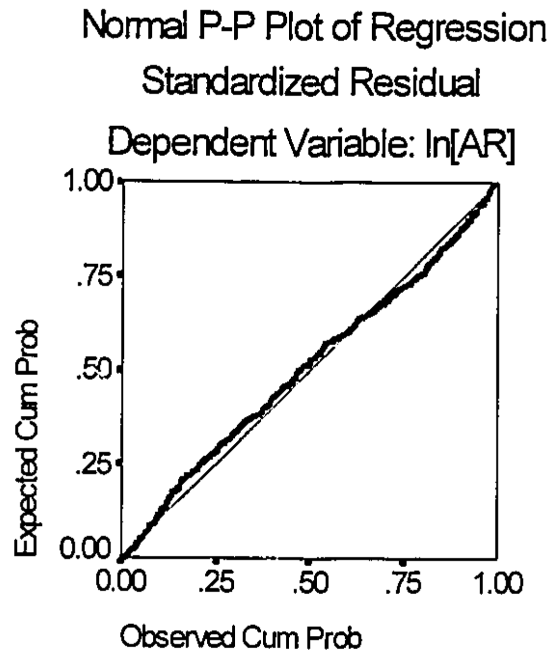


Figure 6-13. Residual P-P Plot of Observed and Predicted Accident Rates (for Equation 6.9)

Several further models were developed by using coefficient constraints in an attempt to force the quadratic search for an optimum solution down other major vectors. However, none of the constrained models exhibited relationships of greater validity than the models in Equation 6.8 and Equation 6.9.

A further set of three multi-variable regressions were performed on the model to determine if slightly different initial starting values for the five coefficients (C0, C1, C2, C3, and C4) might yield a better curve fitting and algorithmic search. Three sets of starting parameters were used in comparison: (1) all coefficients initially set to 0, (2) all coefficients initially set to 1, and (3) coefficients initially set to the final coefficients used in the TRB model. Conclusive results from these three regressions indicated that the starting values selected appeared to make no difference in the final model outcome and form, since each of the three subsequent regressions resulted in approximately the same sum of squares residual values.

6.9.2 SELECTION OF ALBERTA LOG-LINEAR REGRESSION MODEL

Because coefficient constraints were not required to force the model into a form similar to observed findings, the unconstrained version was selected as representative of observed Alberta accident and geometric conditions. However, because of a lack of data variability in both the lane width and hazard rating factor, the unconstrained model was modified to include only those parameters that were sufficiently robust for regression analysis. This final Alberta model (in terms of SI units) is:

$$AR = 0.0002\,(AADT)^{2.748}\,(0.989)^{Sh} \qquad \textbf{(EQN 6.10)}$$

where, AR = Accident rate in terms of /km/yr
AADT= Bi-directional average annual daily traffic volume, in vehicles per day
Sh = Paved shoulder width, in metres

6.10 COMPARISON OF OBSERVED AND PREDICTED ACCIDENT RATES

Once the Alberta Accident Model was developed, the next stage in the analysis was to compare various observed and predicted accident rates with each other to determine possible calibration factors to be applied to the Alberta Accident Model and the TRB Accident Model. The first pair of accident rates to be compared against each other was observed Alberta accident rates from actual accident data versus accident rates predicted by the Alberta Accident Model through geometric factors. Figure 6-14 displays a plot of observed versus predicted Alberta accident rates by sub-section, as well as a best-fit linear regression line through the data points.

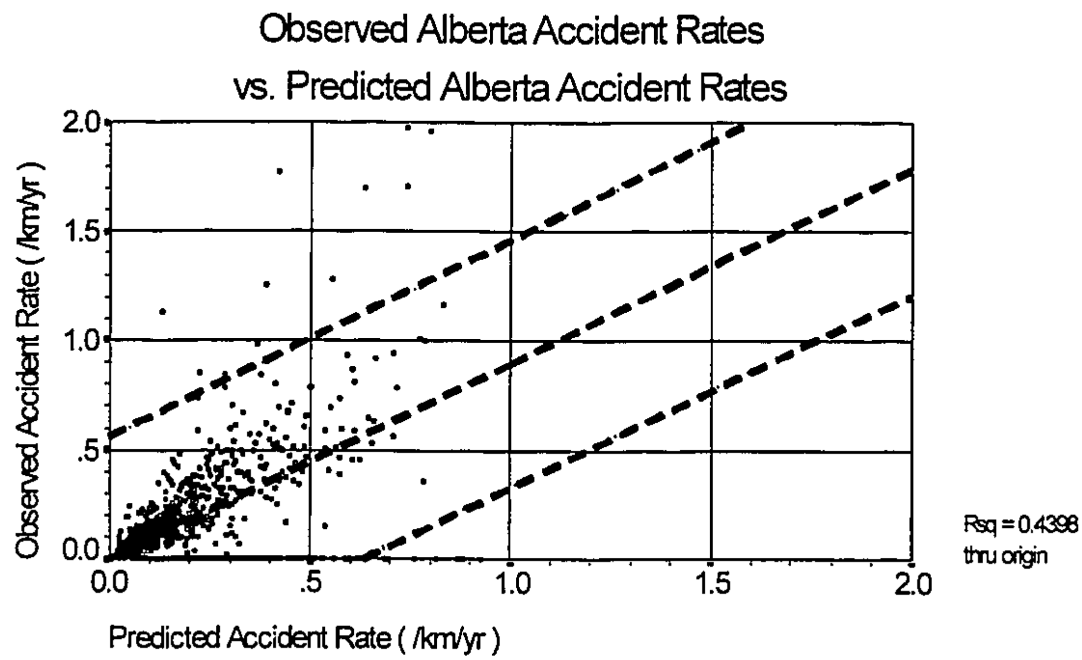


Figure 6-14. Observed Alberta Accident Rate vs. Predicted Alberta Accident Rate

Although there appears to be a significant degree of dispersion in the data, this is as expected due to the nature of highway accidents previously discussed in Chapter 2. Several points should be noted from this linear regression plot.

First, from visual inspection, the majority of both observed and predicted accident rates occur at rates less than 1.0 accidents per kilometre per year (corresponding to the lower left quadrant of Figure 6-14). Because of the lack of accident data existing for rates exceeding 1.0 accidents per kilometre per year, the linear relationship seen in Figure 6-12 may not be valid for predicted rates in excess of 1.0 accidents per kilometre per year.

However, this evidence does not invalidate the linear relationship between observed accident rates, and those rates predicted by the Alberta Accident Model. Because of the best-fit approximation of the various geometric factors and accident rates used to derive the Alberta Accident Model in section 6.9, it was expected that the linear regression between observed and predicted accident rates would yield a correction coefficient of approximately 1.0, as exhibited in the slope of the regression line in Figure 6-14. The actual correction coefficient obtained (the slope) was 0.893, with an R^2 value of 0.440 and a sum of squares residual of 61.31. This implies that the best-fit approximation used in section 6.9 to obtain the Alberta Accident Model yielded a model that produced results quite similar to the observed accident data, despite moderate dispersion (as evidenced by the R^2 value < 1.0). It should also be noted that the 95% confidence intervals shown in the figure also give an indication to the significant degree of dispersion in the accident data.

Figure 6-15 illustrates the comparison of observed Alberta accident rates versus accident rates predicted by the unmodified TRB Accident Model using Alberta primary highway geometric data.

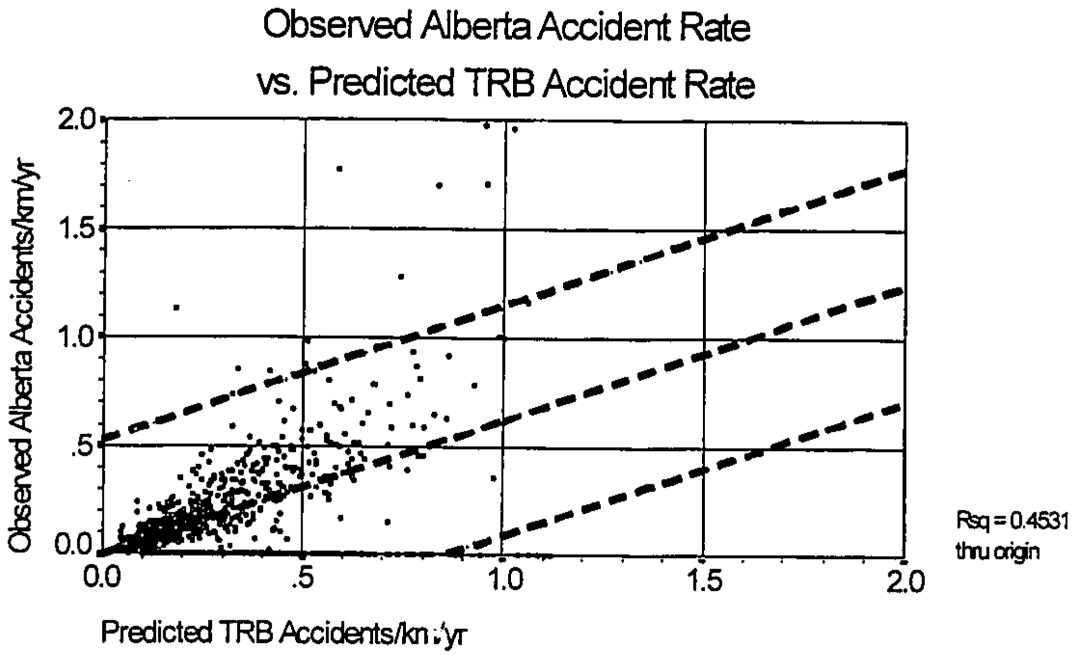


Figure 6-15. Observed Alberta Accident Rate vs. Predicted TRB Accident Rate

The findings from this relationship are quite similar to those seen in Figure 6-14, namely in the location of the majority of the data points, the dispersion of the data points (as evidenced by the respective, and similar R^2 values for each figure), as well as the linear relationship between observed Alberta accident rates and predicted TRB accident rates. What differs in Figure 6-15 is that the linear relationship between observed and predicted does not have a linear coefficient of approximately 1.0. In this case, the correction coefficient apparently needed to adjust predicted rates from the TRB model to meet Alberta observational analogues is 0.620. Again, it should be noted that the 95% confidence intervals shown on Figure 6-15 indicate a significant degree of dispersion in the data. The linear equation approximating the relationship between observed Alberta accident rate and the TRB predicted accident rate is given in Equation 6.10:

$OBS_{ALB} = 0.620\ PRED_{TRB}$ **(EQN 6.11)**

where, OBS_{ALB} = Observed Alberta accident rate (in Accidents/km /yr)
$PRED_{TRB}$ = Predicted TRB accident rate (in Accidents/km/yr)

This relationship can best be defined as a calibration coefficient that can adjust the predictive model to more accurately represent actually observed results.

Finally, a relationship was developed between accident rates predicted by the Alberta Accident Model and those rates predicted by the TRB Accident Model, using Alberta primary highway geometric data to determine accident rates for both models. Figure 6-16 represents this relationship:

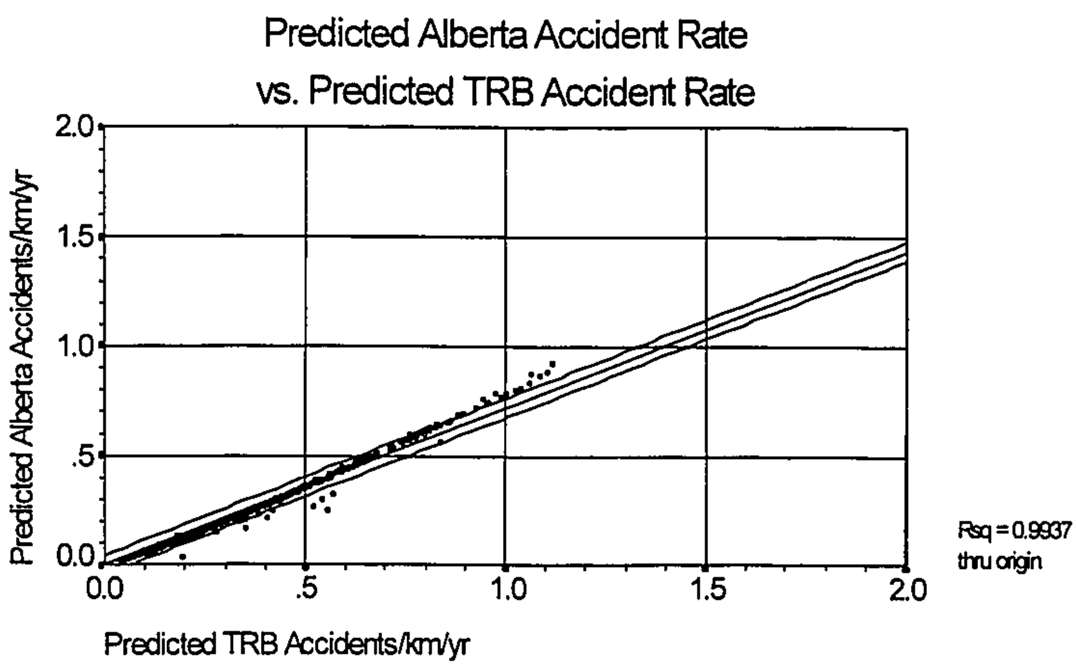


Figure 6-16. Predicted Alberta Accident Rate vs. Predicted TRB Accident Rate

Because of the similarities in both models' log-linear structure as well as the common source of geometric data used to develop the accident rates, the level of dispersion was very low, as expected. The R^2 value of 0.994 as well as the extremely narrow 95% confidence interval of this figure indicates that the two models appear to be very closely correlated with each other with very few data outliers. The sum of squares residual found for this figure was 0.366, and the T-statistic was significant at 338.51.

Several best-fit regression lines were considered for Figure 6-16. The linear regression line was selected to represent the relationship between the two predicted accident rates for the following reasons: (1) the linear relationship is the simplest relationship among higher order relationships, and (2) the improvements in R^2 values between the first order and subsequently higher order relationships were not noticeably significant. As a result of this, the following equation was used to represent the relationship between predicted accidents of the Alberta Accident Model and the TRB Accident Model:

$$PRED_{ALB} = 0.719\ PRED_{TRB} \qquad \textbf{(EQN 6.12)}$$

where, $PRED_{ALB}$ = Predicted Alberta accident rate (in Accidents/km/yr)
$PRED_{TRB}$ = Predicted TRB accident rate (in Accidents/km/yr)

6.11 SUMMARY OF FINDINGS

Based on the analysis of Alberta primary highway accident and geometric data over a four year span, seven key conclusions can be reached from the work presented in this chapter.

1. While the Alberta accident database and the TRB accident database exhibit certain differences in magnitude and background traffic volume factors, the two databases are similar enough that a comparison between accident rates for both models is feasible.

2. Roadway and roadside geometric factors to be modeled were found to be significant for similar accident types and groups of accidents for both the Alberta accident database and the TRB accident database. Among the accident event types found to be most highly significant with respect to geometric factors were aggregate single vehicle accidents and total vehicular accidents. To spare further reduction of the Alberta accident database and possible insufficiency of data to develop a model, the total vehicular accident event type was used to develop observed accident rates.

3. Because of strong cross-correlation between the three roadside geometric factors available for modeling (mean sideslope ratio, mean ditch width, and mean backslope ratio), a combinative factor was used to represent the effect of mean roadside geometrics on accident rates. This factor (HAZARD) was a direct analogue of the median roadside hazard rating used in the TRB Accident Model (***TRB, 1987***). It was also found that the vast majority of Alberta's two-lane, rural, primary highways have moderately low hazard ratings (3.0 - 3.5), in accordance to Alberta 3R/4R geometric design guidelines (***Kenny, 1995***).

4. The effect of traffic volume on accident rates in Alberta is also consistent with prior research in the area of safety-geometric modeling. Accident rates were generally found to decline as traffic volumes (measured in terms of AADT) increased. A

noticeable increase in accident rates occur at moderate traffic volumes, and continues to climb, as a direct result of increasing traffic friction and probability of incidence from higher traffic volumes. These findings on AADT and accident rates are consistent with those of the TRB (***Zegeer and Deacon, 1987***) and Alberta Transportation and Utilities' own internal findings (***Kenny, 1995***).

5. Both lane and shoulder widths were found to affect accident rates in a manner similar to that seen in prior TRB funded research. However, due to the widespread standardization of lane widths of Alberta's primary highways, it was not possible to determine the effects of incremental lane width increases on accident rates. The future use of secondary highways, with there associated lower levels of geometric standardization may introduce enough variation in lane widths to perform an analysis.

6. Individually, both mean sideslope ratio and mean ditch width appeared to have a well defined relationship with accident rates. The predominant effect of these roadside geometric factors occur at the lower extreme end of their values. Safety improvements (through lower accident rates) are greatest for incremental improvements in the steepest sideslope ratios and the narrowest ditches. Further increases in sideslope ratio past 3.5:1 or ditch widths past 3.5 m do not return comparable increases in safety. These findings may be important in refining benefit-cost assessments of further incremental improvements to roadside geometrics.

7. The log-linear regression model developed from Alberta geometric and accident data appears to be in a similar form as the TRB Accident Model. However, the key difference between the two models lie in terms of the magnitude of accident rates predicted. As a result, a correction coefficient has been developed to adjust those accident rates obtained from the TRB Accident Model to more closely represent Alberta accident experience. This difference in accident rates may be a direct result of the differences seen in geometrics and accidents explained previously. Nonetheless, based on these findings, the Alberta Accident Model not only bears a striking resemblance to the TRB Accident Model, but also validates the general use of the TRB Accident Model within Alberta.

It should be noted that similar to the findings of the TRB research (***Zegeer et al., 1987 and TRB, 1987***) into safety-geometric relationships, these findings apply to a specific class of roadways only. These findings and the Alberta Accident Model have thus far, been validated for use with two-lane, rural, primary highways.

An additional interest of this research program concerned the possible application of the Alberta Accident Model to Alberta's four-lane and multi-lane, rural, primary highways. Chapter 7 of this thesis details the methodology, analysis and findings of this further research.

Chapter 7

Analysis of Multi-Lane, Rural, Primary Highways

7.0 INTRODUCTION

The goal of the analysis presented in this chapter is to develop a relationship between accident rates and cross-sectional geometric factors for multi-lane, rural, primary highways in Alberta. As illustrated in Chapter 3 and Chapter 6 of this thesis, prior research into accident rate and cross-sectional geometric relationships concentrated primarily on two-lane highways, with the resulting safety-geometric accident rate model applicable for the two-lane roadway scope only. Despite this, the development of a multi-lane accident rate model using the same methodology was conducted in this chapter for the following reasons: (1) to provide some sort of comparison to the results obtained for the two-lane accident rate model, (2) to develop relationships between accident rate and multi-lane specific geometric factors (such as median width) at the minimum, and (3) determine if the TRB model is in fact, not limited to two-lane roadway geometric design only.

7.1 MULTI-LANE GEOMETRIC FEATURES

Within Alberta, the multi-lane, rural primary highway category consists primarily of four-lane divided rural freeways. Aside from the uni-directional traffic flow associated with divided roadways, the roadway and roadside geometrics for this class of highways only differ slightly from two-lane, rural, primary highways. The key geometric differences between four-lane and two-lane roadways in Alberta lies in additional inside shoulder width and median width data associated with four-lane, rural, divided highways. Unlike two-lane, rural, primary highways where shoulder widths for each side are identical, multi-lane roadways in Alberta have differing pavement widths for the inside and the outside shoulders. In general, the inside shoulder is more narrow than the outside shoulder, with typical design widths shown in Chapter 1.

The most significant difference between two-lane and multi-lane, rural primary highways in Alberta concerns standardization of geometric design.

7.1.1 ROADWAY GEOMETRIC FACTORS

The relatively broad degree of variation with geometric factors seen in Chapter 5 and Chapter 6 for two-lane roadways is not apparent for multi-lane roadway geometrics in Alberta. The lack of sizable variation in roadway and roadside geometric factors is a direct result of the high degree of standardization applied to multi-lane highway design in Alberta. Within Alberta, 255 highway sub-sections are classified as multi-lane rural divided freeways. Tables 7-1 and 7-2 demonstrate the lack of significant variation in geometric factors.

Table 7-1 Summary of Alberta Multi-lane Highway Sub-sections, by Lane Width

Lane Width	Number of Sub-sections	% of Total Sub-sections
3.20 m	2	0.8
3.60 m	8	3.1
3.70 m	245	96.1
Total	255	100.0

Table 7-2 Summary of Alberta Multi-lane Highway Sub-sections, by Outside Shoulder Width

Outside Shoulder Width	Number of Sub-sections	% of Total Sub-sections
1.50 m	2	0.8
2.40 m	11	4.3
2.70 m	1	0.4
3.00 m	237	92.9
3.30 m	1	0.4
3.75 m	1	0.4
4.00 m	2	0.8
Total	255	100.0

Tables 7-1 and 7-2 clearly show that the predominant roadway geometric features of multi-lane rural highways in Alberta are lane widths of 3.70 m and outside shoulder widths of 3.00 m.

The third roadway geometric factor of importance in this analysis is inside shoulder width. Unlike the previous two roadway factors, there was a significant degree of variation in these factor values. Table 7-3 illustrates the distribution of inside shoulder width values.

Table 7-3 Summary of Alberta Multi-lane Highway Sub-sections, by Inside Shoulder Width

Inside Shoulder Width	Number of Sub-sections	% of Total Sub-sections
0.0 - 1.0 m	16	6.3
1.1 - 1.5 m	75	29.4
1.6 - 2.0 m	41	16.1
2.1 - 2.5 m	81	31.8
2.6 + m	32	12.5
Missing	10	3.9
Total	255	100.0

7.1.2 ROADSIDE GEOMETRIC FACTORS

Multi-lane, rural primary highway roadside geometrics exhibited a much greater degree of variability in factor values than roadway factors. As with two-lane primary highways, two outside roadside factors were examined. They are: (1) mean sideslope ratio, and (2) mean ditch width. Median width, a third roadside factor that was not applicable to two-lane, rural highways, was also examined. Tables 7-4 through 7-6 illustrates the distribution of values for each roadside factor.

Table 7-4 Summary of Alberta Multi-lane Highway Sub-sections, by Mean Sideslope Ratio

Mean Sideslope Ratio	Number of Sub-Sections	% of Total Sub-sections
4:1 - 4.5:1	25	9.8
4.5:1 - 5.5:1	97	38.0
5.5:1 +	112	43.9
Missing	21	8.3
Total	255	100.0

Table 7-5 **Summary of Alberta Multi-lane Highway Sub-sections, by Mean Ditch Width**

Mean Ditch Width	Number of Sub-sections	% of Total Sub-sections
≤ 3.80 m	64	25.1
3.90 m	28	11.0
4.00 m	27	10.6
4.10 m	37	14.5
4.20 m	12	4.7
4.30 m	66	25.9
Missing	21	8.2
Total	255	100.0

Table 7-6 **Summary of Alberta Multi-lane Highway Sub-sections, by Median Width**

Median Width	Number of Sub-sections	% of Total Sub-sections
0.0 - 5.0 m	25	9.8
5.1 - 10.0 m	36	14.1
10.1 - 15.0 m	7	2.8
15.1 + m	187	73.3
Total	255	100.0

Whereas outside roadside factors such as sideslope ratio and ditch width exhibit many non-standard values, the majority (73.3%) of median widths in Alberta are greater than 15.0 m. This fact may be significant in any analysis of multi-lane, divided highways since it indicates that the separation between the two traveled directions may be sufficiently large enough to analyze each direction as a separate entity. Accidents occurring on a section of divided highway can be considered not to be affected by traffic and factors occurring on the opposing highway direction. In essence, the interaction of accident and geometric factors between each direction of travel is not significant, due to the prevalence of extremely wide median widths that prevent opposing direction interactions from occurring.

7.2 SINGLE VARIABLE REGRESSION

Chapter 6 documented the application of single variable regression to illustrate the relationships between accident rates and individual cross-sectional geometric factors for two-lane, rural primary highways. On the assumption that the methodology used in Chapter 6 was valid for multi-lane, divided, rural primary highways, the same regression methodology was applied to accidents occurring on multi-lane highway sections and multi-lane geometric factors. Testing this assumption by analyzing the relationships between accident rates and cross-sectional geometric factors of multi-lane, rural, primary highways forms the basis of this chapter. As in Chapter 6, an important step in developing the log-linear relationship between accident rate and cross-sectional geometric factors is determining the relationship between accident rate and individual factors in question.

The distribution of accident event types that occurred on divided primary highway sections in Alberta between 1991 and 1994 was remarkably similar to that seen on two-lane, undivided primary highways (see Table 5-13). Table 7-7 illustrates the distribution of accidents by event type for those that occurred on multi-lane, divided primary highways and the complementary frequency distribution of two-lane accident event types.

Table 7-7 Primary Event, Multi-lane, Rural, Primary Highways - 1991-94

Event	Frequency	% of Total Multi-lane Event Frequency	% of Total 2-lane Event Frequency
1. Struck Fixed Object	315	6.0	4.7
2. ROR Left	1,229	23.5	19.1
3. Right Angle	284	5.4	4.7
4. Pass Left Turn	42	0.8	5.0
5. Left Turn Approach	74	1.4	1.8
6. Left Sideswipe	54	1.0	4.3
7. Other	277	5.3	5.3
8. Rear End	746	14.2	14.3
9. ROR Right	1,644	31.4	32.0
10. Head On	59	1.1	3.6
11. Pass Right Turn	2	0.0	0.6
12. Right Sideswipe	496	9.5	4.3
13. Backing	13	0.2	0.5
14. Missing	1	0.0	0.0
Total	5,236	100.0	100.0

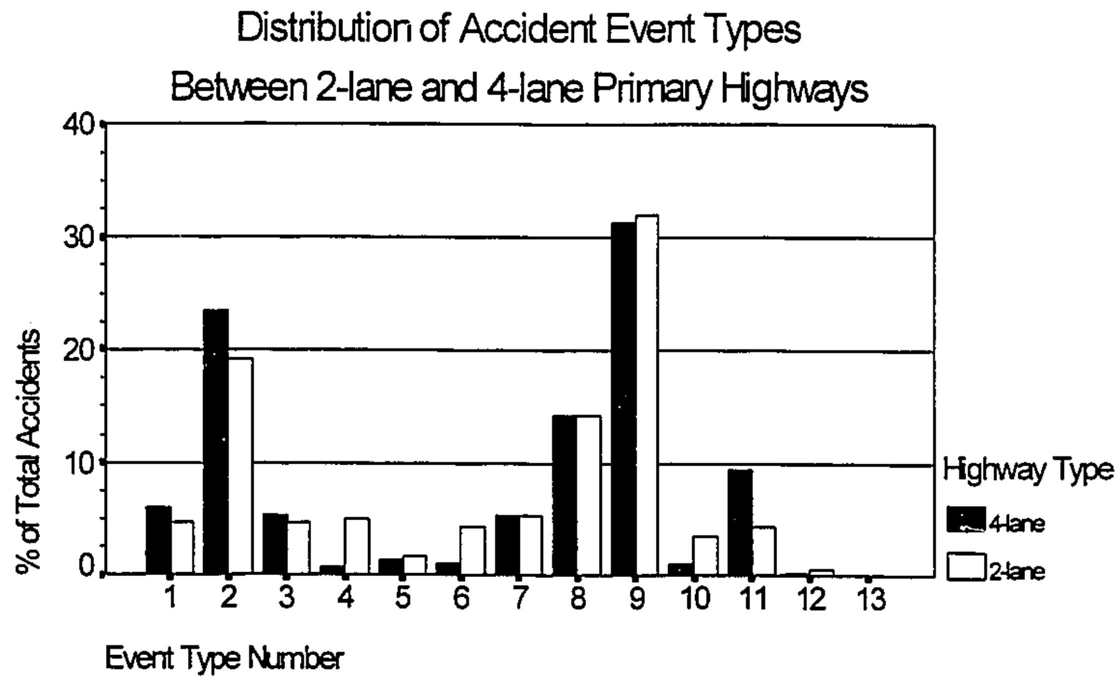


Figure 7-1. Distribution of Accident Event Types (2-lane vs. 4-lane)

The comparison between accident event types by percentage distribution between two-lane and four-lane primary highways is illustrated in Figure 7-1. Each event number in Figure 7-1 corresponds to the event number and event type indicated in Table 7-7.

As clearly seen in Figure 7-1, both accident samples are distributed in a similar manner between event types. The event categories that exhibit differences between the two accident populations are: (1) pass left (type 4) and right (type 11) turns, (2) left turn approach collisions (type 5), (3) head on collisions (type 10), and (4) ROR left accidents (type 2).

As expected, the percentage frequency occurrence of accident events involving opposing direction movement (such as head on and left turn approach collisions) is much greater for the two-lane population than the four-lane accident database. The percentage frequency of those accident types for the four-lane accident population was extremely low, due to the generally wide separation between opposing direction traffic flow as explained previously in this chapter.

Another noted difference between the two accident populations concerned ROR left accidents. These accidents appear in a greater percentage for four-lane accidents over two-lane accidents possibly due to geometric differences between the two highway types. It is possible that ROR left accidents may occur more frequently on uni-directional, divided highways since a higher proportion of vehicles may be traveling in the left lane. Increased flow in the inner (left) lane may result in a greater probability of vehicle encroachment on the inside shoulder and roadside. This encroachment usually results in an ROR left type accident event.

In general, differences between the two accident databases were noted for event types involving opposing direction vehicle interactions.

7.2.1 AADT AND ACCIDENT RATES FOR FOUR-LANE HIGHWAYS

The high degree of lane width standardization seen in Table 7-1 did not yield an adequate analysis of the effects of lane width on accident rates. A sizable variation in AADT volumes for the 255 multi-lane, rural highway sub-sections in Alberta did however, enable a relationship to be developed between accident rates and AADT. This relationship is presented in Figure 7-2.

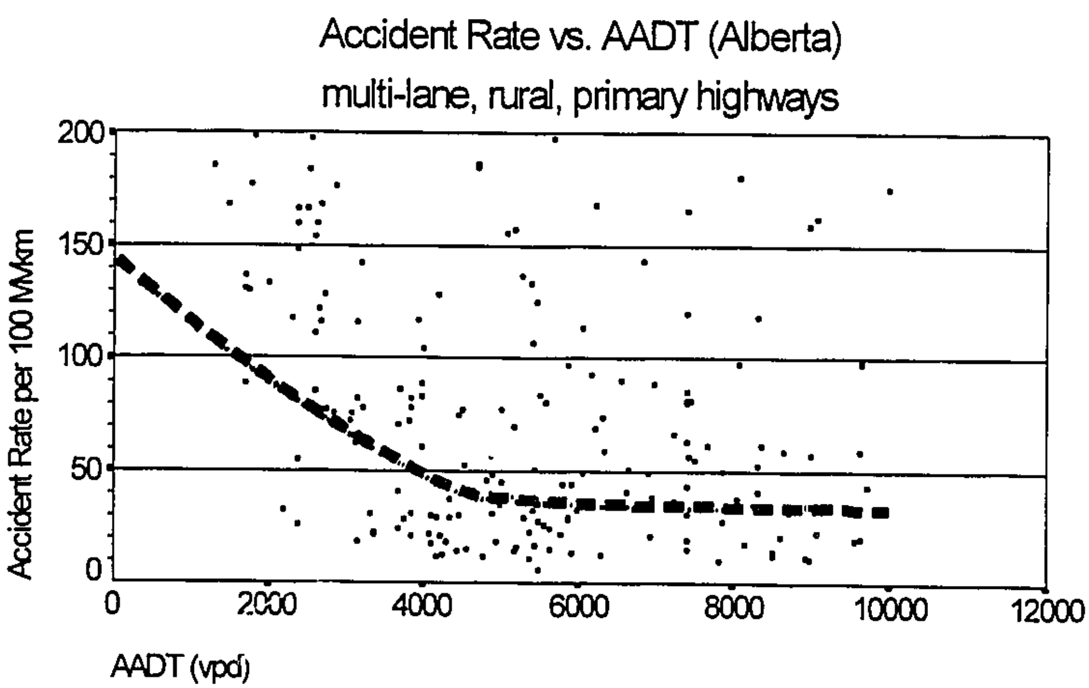


Figure 7-2. Alberta Accident Rate Experience Over AADT - Multi-lane Hwys.

The most noticeable element of this figure is that there appears to be two distinct regions of AADT influence on accident rates. The first region occurs for multi-lane, rural primary highways with traffic volumes up to approximately 5,000 vehicles per day. From the lowest AADT values of approximately 1,000 vehicles per day, there is a sharp and very noticeable decline in accident rates as AADT increases. This decrease in accident rates may be a direct result of incremental improvements in the overall geometrics of the divided highway as AADT increases tax the highway system. Another likely contributor to the sharp decrease in accident rates over AADT may be the fact that the accident rate is also measured in terms of a background traffic volume exposure (in terms of 100 million vehicle kilometres traveled on the sub-section). For example, a lower AADT value coupled with identical accident experiences for two separate sub-sections will result in a higher accident rate (per 100 million vehicle km) than that section with the higher AADT volume.

The second significant finding from Figure 7-2 is that no significant safety gains (through accident rate reduction) appear once AADT increases above 5,000 vehicles per day. This may signify that the geometric design standards for multi-lane highways in Alberta are sufficient to lower the accident rate to that of a background random accident rate. Further improvements in cross-sectional geometrics may not achieve safety gains desired.

7.2.2 INSIDE SHOULDER WIDTHS AND ACCIDENT RATES

Insufficient variation in outside shoulder width values prevented a regression analysis of the effects of outside shoulder widths on accident rates. However, as seen in Table 7-3, a significant variability in inside shoulder widths permitted an analysis of the effects of that shoulder width on accident rates. The results of this analysis are presented in Figure 7-3.

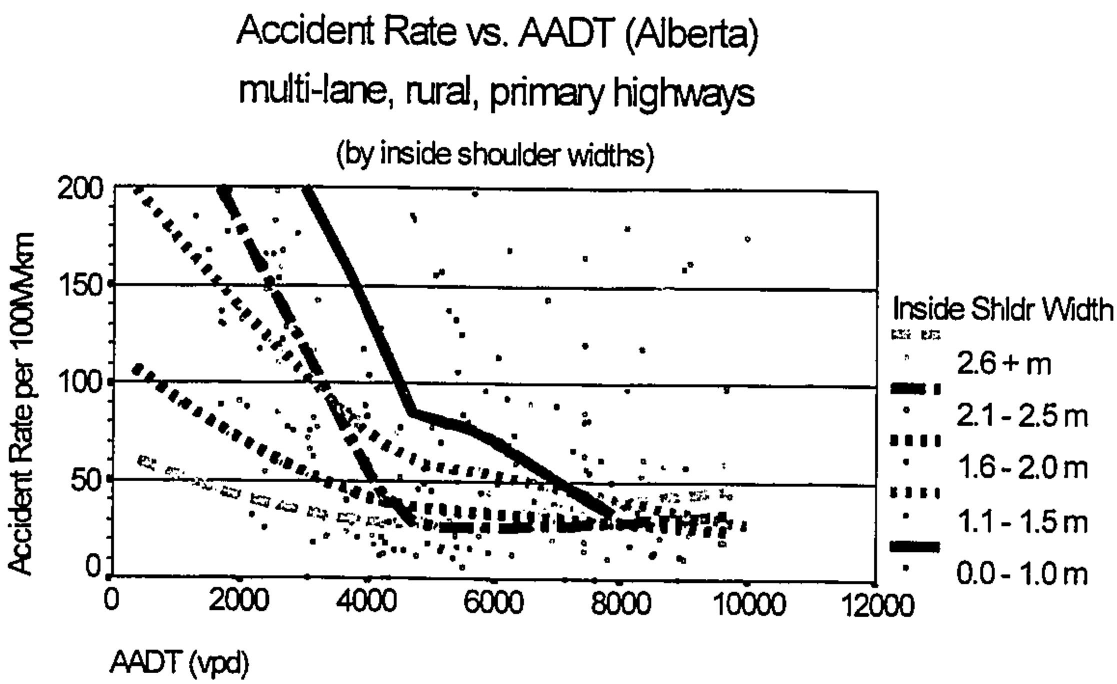


Figure 7-3. Observed Accident Rate Experience Over AADT (for inside shoulder widths)

Best fit regression lines were drawn for each set of inside shoulder width groups. It is clear from the figure that accident rates appear to decline consistently as AADT increases in volume. More importantly, the declining trend over AADT is also matched in the decline in accident rates as inside shoulder widths increase as well. Note that for relatively low AADT values (approximately 2,000 vehicles per day on a multi-lane, divided highway) a significant difference (from over 200 accidents per 100 million veh. km to below 100 accidents per 100 million veh. km) appears in accident rates between inside shoulder widths of 0.0 - 1.0 m, and those exceeding 2.5 m. Incremental increases in inside shoulder widths also appear to have a significant impact on the decline of accident rates as well. As seen in Figure 7-3, this influence of inside shoulder width on accident rate disappears above AADT volumes of approximately 5,000 vehicles per day. This may signify that the incremental improvements to inside shoulder widths may only be effective for roadways experiencing up to 5,000 vehicles per day. Conversely, the cross-sectional geometrics at higher volumes of traffic may currently be of sufficient quality that further improvement may not be necessary. This issue could be potentially resolved with further research involving highway before-and-after studies. This topic is briefly discussed in Appendix B of this thesis.

7.2.3 MEDIAN WIDTHS AND ACCIDENT RATES

The results of the regression between median width and accident rates are remarkably similar to those seen in section 7.2.2. Figure 7-4 illustrates the relationship established between median width and accident rates over varying AADT volumes. Median widths were grouped into 5.0 m increments to aid in determining the effects, if any, incremental changes in median widths had on accident rates.

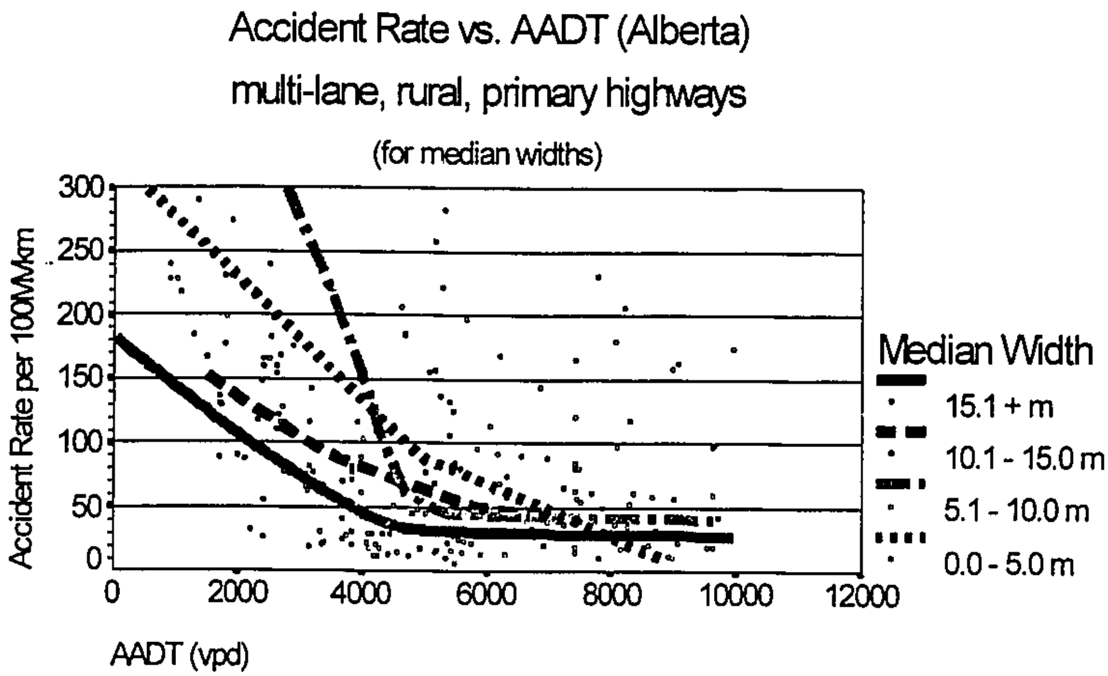


Figure 7-4. Observed Accident Rate Experience Over AADT (for median widths)

In Figure 7-4, accident rates are seen to decline as AADT increases up to an approximate AADT value of 5,000 vehicles per day. At this point, the accident rate for each median width group appears to remain constant despite further AADT increases. Accident rates are also seen to generally decline as median widths increase. Despite the fact that a predominant number of sub-sections in Alberta have median widths greater than 15.0 m (all but eliminating possible opposing direction vehicle conflicts), the remaining sub-sections have median widths sufficiently narrow that the opposing direction traffic may increase the probability of an accident occurring. Again, this conjecture may require additional data of highways exhibiting narrow medians to further improve the relationship.

7.2.4 SIDESLOPE RATIOS AND ACCIDENT RATES

The relationship between mean sideslope ratio and accident rates was not as clear as those seen for either shoulder widths, or median widths (sections 7.2.2 and 7.2.3). A plot of accident rates against AADT using means sideslope ratios to develop best-fit regression lines for the data is illustrated in Figure 7-5.

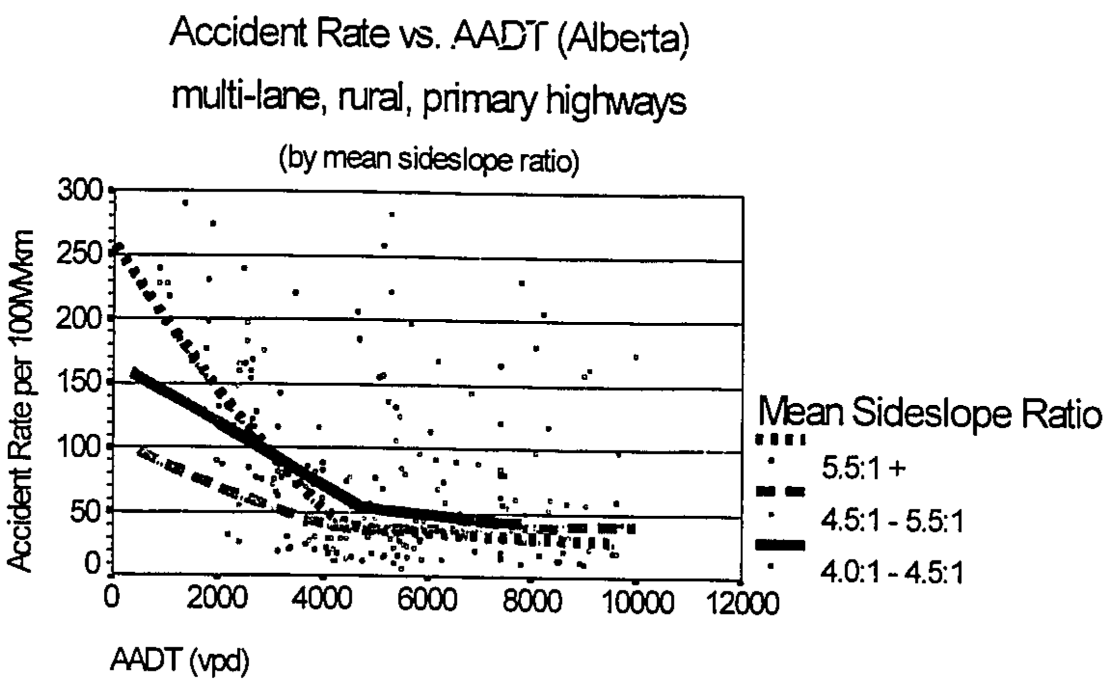


Figure 7-5. Observed Accident Rate Experience Over AADT (for mean sideslope ratios)

This figure differs from previous figures in that the decline seen in accident rates as a result of incremental changes in sideslope ratio are counter-intuitive. While prior research into two-lane highway cross-sectional geometrics seem to indicate that accident rates decline as sideslope ratio increases or flattens (*TRB, 1987*), what is exhibited in

Figure 7-5 is the opposite, with accident rates greatest for the gentlest sideslopes. At upper AADT values however, accident rates are both constant (approximately 30-40 per 100 million vehicle km) and relatively similar despite incremental changes in sideslope ratio. The discrepancies between Figure 7-5 and expected sideslope-accident relationships for lower AADT values may be a result of insufficient robustness of sideslope ratio data on lower volume (below 5,000 vehicles per day) multi-lane, rural, divided primary highways in Alberta.

7.2.5 DITCH WIDTHS AND ACCIDENT RATES

The developed relationship between mean ditch widths and accident rates over AADT were no clearer than those seen in the previous section concerning mean sideslope ratios. Figure 7-6 illustrates the results of the application of best-fit regression lines for each set of mean ditch widths. Note that for the purposes of this analysis, and on the basis of the results presented in Table 7-5, the ditch width values of 4.0 m and 4.2 m were not included in the analysis due to insufficient sample sizes.

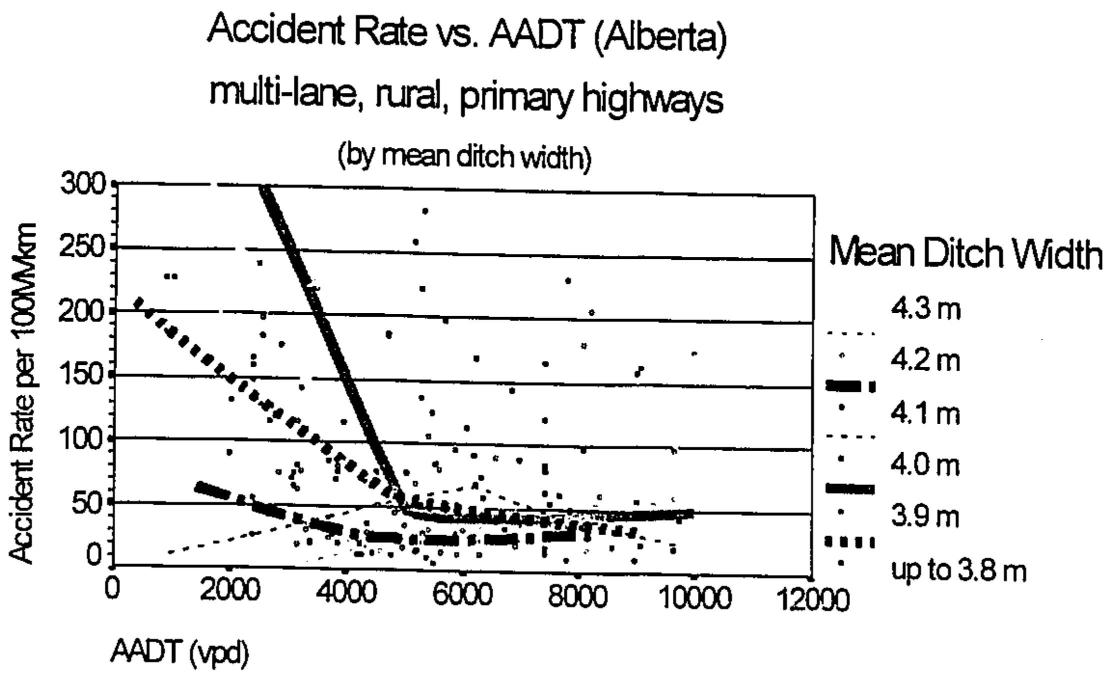


Figure 7-6. Observed Accident Rate Experience Over AADT (for mean ditch widths)

As expected, the general accident rate trend seen in Figure 7-6 decline as AADT increases; up to an approximate value of 5,000 vehicles per day. From this point, no noticeable changes occur in accident rates, regardless of mean ditch width. Discrepancies with expected trends (***TRB, 1987***) are seen when analyzing incremental changes in mean ditch width, and their effects on accident rates.

As expected, both narrow width categories (3.9 m and less than 3.8 m) exhibit higher general accident rates below 5,000 vehicles per day of AADT. Surprisingly however, accident rates for the widest ditch width category (4.3 m) are significantly higher than all other width categories, except for 3.9 m. Despite the fact that the 4.3 m and the sub-3.8 m width categories both contain the largest sample size (approximately

25% each) of the entire multi-lane, primary highway sub-section population, the counter-intuitive results of accident rates for wide ditches cannot be easily explained. One explanation for this result may be the distinct possibility that other geometric factors may be influencing the relationship in a combinative manner. Another possibility may lie in the small sample size (255 sub-sections compared to over 750 sections for the two-lane analysis) being used to develop these relationships. Unfortunately, the lack of both quantity of accident and geometric data, and the lack of diversity in geometric factor values, currently hinders performing further analysis into the root causes of these results.

7.2.6 DAYTIME POSTED SPEED LIMITS AND ACCIDENT RATES

Although daytime posted speed limits are not considered a cross-sectional geometric design factor in the traditional sense, determining possible relationships between speed and accident rates may identify possible global effects of geometric design factors on highway safety. This assumption is drawn from the fact that operational and design speed limits are directly determined by the geometric design parameters of a given roadway section (***RTAC, 1986***). As a result, determining the direction of influence that speed limits have on accident rates may indirectly verify the relationships of individual geometric factors with accident rates. The distribution of daytime operating speed limits, by number of sub-sections is shown in Table 7-8. An analysis of night-time operating speed limits was not performed due to the uniformity of data presented (244 of 255 sub-sections have speed limits of 100 km/hr at night).

Table 7-8 **Daytime Posted Speed Limit for Multi-lane, Rural, Primary Highways - 1991-94**

Daytime Posted Speed Limit	Number of Multi-lane Sub-sections	% of Total Multi-lane Sub-sections
< 90 km/hr	1	0.4
90 km/hr	10	3.9
100 km/hr	134	52.6
110 km/hr	110	43.1
Total	255	100.0

As seen in the table, insufficient data for speed limits below 90 km/hr on multi-lane, rural, primary highways in Alberta prevented the development of any relationships with accident rate. Of the remaining three speed limit categories, the 90 km/hr speed limit category may not be of sufficient enough size to produce a significant relationship with accident rate.

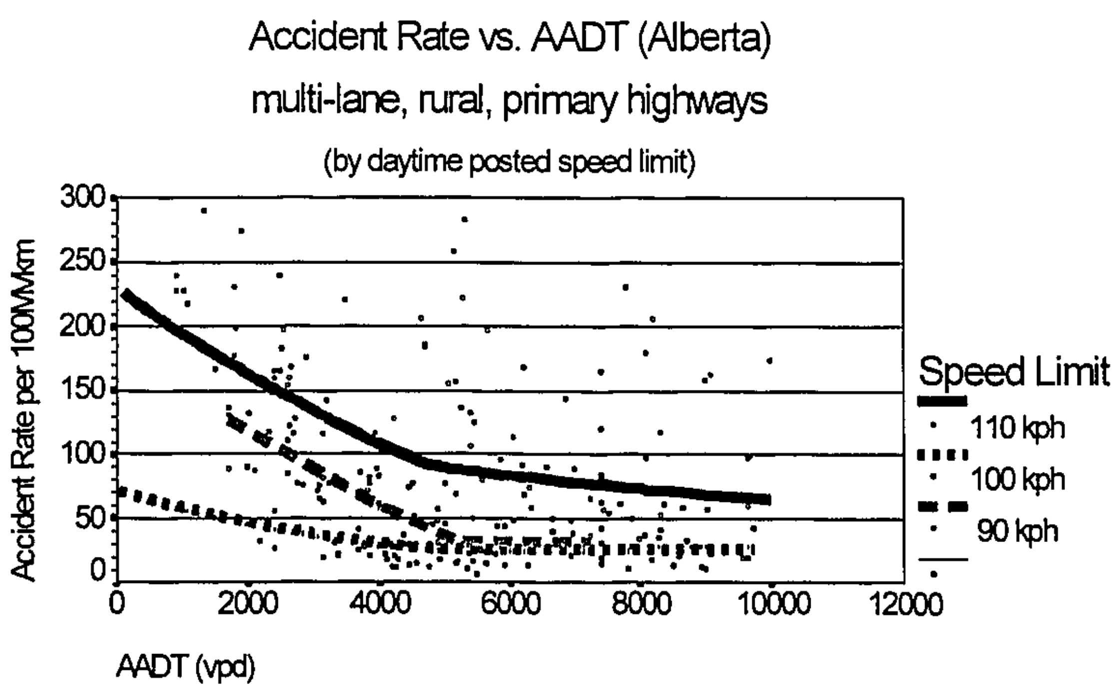


Figure 7-7. Observed Accident Rate Experience Over AADT (for daytime speed limits)

Figure 7-7 illustrates the best-fit regression lines developed for three distinct operational daytime speed limits: (1) 90 km/hr, (2) 100 km/hr, and (3) 110 km/hr. The noticeable lack of change in accident rates seen as AADT increases above 5,000 vehicles per day is once again seen here. This appears to confirm relationships between posted speed limits and various cross-sectional geometric factors. Without assigning causality to any of the relationships illustrated in the figure, it is also clear that sub-sections with operating speed limits of 110 km/hr experience significantly greater accident rates than those of 100 km/hr or 90 km/hr (more than twice the lower speed limit's exposure rate, in some cases). While it is not within the scope of this research project to examine causal relationships between accident rate, speed limits, and geometric factors, further research into additional geometric design factors, such as roadway alignment, may clarify this issue.

Nonetheless, possible explanations for the behaviour of the accident rate relationship for 110 km/hr speed limits may lie in a combination of driver operating speeds, driver perception and reaction, greater vehicle interaction, and potential deficiencies in geometrics. Unfortunately, while this section identifies a correlation between posted speed limit and accident rates, it is not possible to examine the causal relationships in this thesis.

7.3 LOG-LINEAR REGRESSION MODELING

The methodology used to analyze two-lane, rural primary highways in Chapter 5 and Chapter 6 was also applied to multi-lane, rural, primary highways in this chapter. Although it is recognized that differences exist in geometric factors and accident types between two-lane and multi-lane highways, this analysis is an attempt to quantify the interrelated effects of geometric factors on accident rates for multi-lane highways. A log-linear regression methodology was used to analyze multi-lane highways enabling a modicum of comparison between the developed Alberta two-lane accident rate model, and a multi-lane model.

The log-linear regression model used in this analysis is in the following form:

$$AR = C0(AADT)^{C1}(C2)^{W}(C3)^{SHOUT}(C4)^{SHIN}(C5)^{MEDWD}(C6)^{RRec} \qquad \textbf{(EQN 7.1)}$$

where,

	AR =	Accident rate, in accidents per mile per year
	AADT =	Annual Average Daily Traffic volume
	W =	Lane width, in feet
	SHOUT =	Outside shoulder width, in feet
	SHIN =	Inside shoulder width, in feet
	MEDWD =	Median width, in feet
	RRec =	Clear roadside recovery distance, in feet

Several issues concerning the geometric factors used in this regression should be noted. First, the use of an estimated roadside hazard rating (HAZARD from Chapter 6) was not viable in this analysis since the equivalent hazard rating for every multi-lane, primary highway subsection in Alberta was 1.0 (least hazardous). This uniform and low hazard rating is due entirely to extremely broad roadside recovery zones that varied between 9.5 m wide, and in excess of 13.5 m wide. As a result, two separate sets of roadside geometric factors were used in modeling the relationship between geometrics and accidents: (1) individual roadside geometric factors, and (2) a combined roadside geometric factor in the form of a roadside recovery zone. The roadside recovery zone was calculated using Equation 6.4 in Section 6.7.1 of Chapter 6.

The second issue concerning developing the relationship between geometrics and accident rates involved the effects of uniform geometric factors on the regression model. As detailed previously in Section 7.7.1, two key geometric factors in the two-lane, rural, primary highway model (lane width and outside shoulder width), did not contain sufficient variability to develop a single variable regression model. As a result of these data limitations, three models were developed for the multi-lane, rural, primary highway regression: (1) an unconstrained model, (2) a constrained model, and (3) a constrained model with the uniform geometric factors removed from the analysis.

7.3.1 UNCONSTRAINED LOG-LINEAR MODEL

Log-linear regression was performed on the geometric variables to determine a best-fit curve through the data, as well as to estimate the coefficients in Equation 7.1. The regression yielded the following equation (with coefficients):

$$AR=954.275(AADT)^{0.083}(0.545)^{W}(1.236)^{SHOUT}(0.888)^{SHIN}(0.999)^{MEDWD}(0.993)^{RRec}$$

(EQN 7.2)

The residual sum-of-squares value for this model was 2661.730, and the R^2 value was 0.063, indicating a large degree of observed data dispersion from the predicted model and a poor fit of data to the model.

This model has two items of interest. First, the various cross-sectional geometric factors all appear to affect accident rates in the proper direction (i.e. increasing lane width decreases predicted accident rate) except for outside shoulder width. This counter-intuitive result may be caused by the lack of diversity in data values for outside shoulder width. Without a broad spectrum of data values to model, the model results can potentially be misleading in their relationships.

The second item of interest concerns the estimation of coefficient values. As seen in Equation 7.2, the initial coefficient C0, and the coefficient for AADT, C1, differ in magnitude from those seen in the TRB model presented in Chapter 6. While the end result of combining the two coefficients appears to be similar to that seen in Chapter 6, the coefficients were constrained to yield a model much closer to the TRB and the Alberta two-lane Accident Model.

7.3.2 CONSTRAINED LOG-LINEAR MODEL

In this model, constraints were applied to the coefficients in Equation 7.1 in an attempt to force the log-linear regression solution to fit the format of the Alberta two-lane Accident Model. The following constraints were used:

1) $0.0 \leq C0 \leq 1.0$ - This constraint attempted to correct the magnitude differences between the C0 and C1 coefficients of the multi-lane model and the two-lane model. See Section 7.3.1 for detail.

2) $C1...C6 \geq 0.0$ - These constraints prevented the iterative search process from assigning negative, and unrealistic, values to the coefficients.

The resulting model developed using constraints was as follows:

$$AR = 1.000\,(AADT)^{0.083}(0.896)^{W}(1.333)^{SHOUT}(0.886)^{SHIN}(0.998)^{MEDWD}(0.993)^{RRec}$$

(EQN 7.3)

The fist item to note is that the coefficient C0 = 1.0, the upper boundary constraint for this model. This local optimal solution found for the parameters and constraints exhibits several of the characteristics of the previous model (Equation 7.2) including a counter-intuitive outside shoulder width coefficient value. Because of the upper constraint value being reached for the coefficient C0, further refinements to the geometric factors used in the model were required.

7.3.3 MODIFIED CONSTRAINED LOG-LINEAR MODEL

The key change in this model was the removal of lane and outside shoulder width variables as contributing factors to the log-linear regression. Because of a lack of robustness in the data for the two geometric factors, their effects on the accident rate were assumed to be unquantifiable, and therefore, neutral to the regression modeling. The removal of the two geometric factors resulted in the following model:

$$AR = (AADT)^{0.149} (0.893)^{SHIN} (1.004)^{MEDWD} (1.008)^{RRcc} \qquad \textbf{(EQN 7.4)}$$

The log-linear regression of this model yielded a residual sum of squares value of 2722.064, and an R^2 value of 0.041. This model appears to indicate that both median width and roadside clear zone values do not greatly influence the resulting accident rate, as evidenced by their coefficient values of approximately 1.0. As well, assuming that the effects of lane width or outside shoulder width are neutral to the accident rate and the model, the only geometric factor that appears to affect accident rate is the inside shoulder width.

7.4. DISCUSSION AND CONCLUDING REMARKS

These findings from the three models are not surprising, due to the rigid standardization of design geometrics present in multi-lane, primary highways in Alberta. The lack of variation in design factors will result in a relatively poorly developed relationship between geometrics and accident rates, as seen in Equation 7.4. Although both lane and outside shoulder widths are widely assumed to affect accident rates (***TRB, 1987***), modeling this relationship is difficult without robust and diverse geometric data.

However, despite these limitations in developing an accident-geometric model for multi-lane, primary highways, several conclusive findings were observed from the relationships between accident rates and individual geometric variables. They are:

1. The types of vehicle accidents occurring on multi-lane, rural, primary highways are distributed in a similar manner as those occurring on two-lane, rural, primary highways. The accident types that differ significantly between two-lane and multi-lane highways are: accidents involving opposing direction vehicle interactions (head on collisions, for example), and ROR left accidents. The first event type occurrence on divided highways is obviously negligible, since the separation between the two opposing directions of flow is in excess of 9.5 m (median width). The second event type occurrence is a direct result of the nature of multi-lane highways. Because a significantly higher frequency of vehicles use the inner-most (left) lane on multi-lane highways (as opposed to it's use as a passing lane on two-lane highways), left side accidents such as ROR events are assumed to increase proportionately as well.
2. Due to the high degree of geometric design standardization, a limited number of relationships between accident rate and geometrics were possible. However, a definitive relationship between AADT and accident rate was identified. Accident rates were found to decline as AADT increased, up to an approximate value of 5,000 vehicles per day. The fact that no additional decreases in accident rate as AADT increases past 5,000 vehicles per day may indicate that Alberta's cross-

sectional geometric design standards are sufficient to reduce the accident rate to that approximating a background, random accident event occurrence. Additionally, the AADT value of 5,000 vehicles per day appeared to be the general threshold value for the effectiveness of incremental improvements of cross-sectional geometrics on accident rates.

3. Incremental increases to the inside shoulder width was found to effectively decrease the accident rate up to a traffic volume of approximately 5,000 vehicles per day. 0.5 m incremental increases to the inside shoulder width were found to have significant impact on the reduction of accident rates for highways with AADT volumes below 5,000 vehicles per day.

4. Daytime operating speed limits were found to have a correlative relationship with accident rates. The increase in the posted speed limit from 100 km/hr to 110 km/hr appeared to raise accident rates two-fold or greater. In no manner does this conclusion imply a causal relationship between speed and accident rates. While this causality may be valid, it was not possible to determine this on the basis of the data available and the scope of this thesis.

5. The distinct lack of variability for most of the cross-sectional geometric factors yielded a log-linear regression model that mirrored the problem. The unconstrained model, which best represented the observed accident and geometric data, appears to be the optimal solution for multi-lane, rural highway accident rate modeling. However, the problems identified by the constrained models (namely the lack of data variability) should not be forgotten when interpreting the unconstrained model.

In general, this analysis of multi-lane, rural, primary highways in Alberta yielded results that were expected due to their similarity to previous results documented in Chapter 6 of this thesis. However, determining the significance of the cross-sectional geometric differences between two-lane and multi-lane, rural highway sections in Alberta is not possible without the availability of non-standard multi-lane highway designs. This analysis presented in Chapter 7 of this thesis is limited not in scope, but in the lack of proper data to analyze the relationship between accidents and cross-sectional geometric factors.

Chapter 8

Conclusions and Recommendations

8.0 OVERVIEW

This thesis is an attempt to quantify relationships between accident rates and cross-sectional geometric factors in Alberta. In order to place these Alberta-based accident-geometric relationships in a general geometric design perspective, a comparison between the Alberta relationships and those developed for the TRB Special Report on Design Safer Roads (***TRB, 1987***) was performed. The resulting relationships developed from this thesis and validated against prior research may prove to be useful in further quantifying highway safety in 3R/4R geometric design in Alberta.

Chapter 2 of this thesis accomplishes two purposes: to provide an overview of the multitude of geometric and non-geometric factors influencing highway safety and corresponding accident rates, and to examination in detail, the effect of key cross-sectional geometric factors on accident rates. These key cross-section factors are: lane width, shoulder width, shoulder type, roadside hazard ratings, and terrain considerations.

The objective of Chapter 3 is to present an overview of 3R/4R geometric design practices in North America, as well as identifying the methodologies used to model accident-geometric relationships throughout North America. This survey of methodology answers many questions on the utility and prevalence of differing approaches to the accident-geometric relationship issue in 3R/4R geometric design.

Chapter 4 presents a detailed explanation behind the selection of generalized linear regression as the methodology used to develop the accident-geometric relationship. An overview of the statistics involved in generalized linear modeling and regression analysis are included in this chapter, along with key assumptions and characteristics of regression modeling.

Chapter 5 presents a methodology to develop an Alberta-based model relating accident rate with various cross-sectional geometric factors identified as significant in Chapter 2. This methodology is applied to the analysis of two-lane, rural, primary highways in Chapter 6, and the analysis of multi-lane, primary highways in Chapter 7.

Both Chapter 6 and Chapter 7 present relationships identified between accident rate and individual cross-sectional geometric factors, as well as the interaction between the geometric factors and their effects on accident rate. These relationships are defined in terms of mathematical models presented in each chapter.

8.1 CONCLUSIONS

1. Two-Lane, Log-Linear Regression Findings

The result of the log-linear regression of accident and cross-sectional geometric data yielded models in the form of the following equations:

$$AR = 0.0006\,(AADT)^{0.999}\,(0.959)^{W}\,(0.988)^{Sh}\,(1.080)^{H} \qquad \textbf{(EQN 8.1)}$$

$$AR = 0.0002\,(AADT)^{1.011}\,(0.989)^{Sh} \qquad \textbf{(EQN 8.2)}$$

where, AR = Accident rate in number of accidents per kilometre per year
AADT = Bi-directional average annual daily traffic volume, in vehicles per day
W = Lane width, in metres
Sh = Paved shoulder width, in metres
H = Estimated roadside hazard rating

These equations closely resembles the model developed for the TRB and currently used by a majority of North American transportation agencies in identifying the effects of cross-sectional geometric elements on highway safety. Although limitations in data quantity and robustness were encountered throughout the analysis of two-lane, rural, primary highways, the results of the analysis were sufficiently similar to the TRB Accident Model that a comparison between the two could be performed. Equation 8.2 was used in the comparison analysis with the TRB accident rate model as the lane width and hazard rating factors in Equation 8.1 were not sufficiently robust to warrant their inclusion into the Alberta accident rate model.

The results of this comparison between the two models yielded a correction coefficient that can be applied to prior accident rates developed using the TRB model to correspond to Alberta geometric conditions. The correction (or calibration) coefficient derived from the comparison is:

$$OBS_{ALB} = 0.719\,PRED_{TRB} \qquad \textbf{(EQN 8.3)}$$

where, OBS_{ALB} = Observed Alberta accident rate (in Acc/mile/year)
$PRED_{TRB}$ = Predicted TRB accident rate (in Acc/mile/year)

This coefficient is derived on the basis of accident and cross-sectional geometric data that have limitations, and as such, should be used as a correction coefficient with discretion and engineering judgment only. Additional input of accident data or the availability and use of more detailed and accurate cross-sectional geometric data can potentially improve both the Alberta Accident Model and the resulting correction coefficient.

2. Multi-Lane, Log-Linear Regression Findings

The analysis of the relationship between accident rates and cross-sectional geometrics for multi-lane, rural, primary highway sections was not as clear as that of the two-lane highway analysis. The extremely high degree of standardization involved in the cross-sectional geometric design of multi-lane, divided highways in Alberta did not yield a sufficiently robust geometric data sample to determine accident-geometric relationships for many geometric factors. The data sample for several key factors examined in the two-lane analysis in Chapter 6 but insufficient for the multi-lane analysis included both lane width, and outside shoulder width (or simply shoulder width for two-lane samples). As a result, the model developed to reflect the relationship between accident rate and cross-sectional geometric factors does not reflect the potential and probable influence that lane and outside shoulder width have on accident rates. The final multi-lane, rural, primary highway accident rate model is given in Equation 8.4:

$$AR = (AADT)^{0.149}\,(0.893)^{SHIN}\,(1.004)^{MEDWD}\,(1.008)^{RRcc} \quad \textbf{(EQN 8.4)}$$

where, AR = Accident rate in number of accidents per mile per year
AADT = Bi-directional average annual daily traffic volume, in vehicles per day
SHIN = Paved inside shoulder width, in feet
MEDWD = Median width, in feet
RRcc = Outside clear roadside recovery distance or zone, in feet

The same equation in SI units is given in Equation 8.5:

$$AR = (AADT)^{0.149}\,(0.966)^{SHIN}\,(1.001)^{MEDWD}\,(1.002)^{RRcc} \quad \textbf{(EQN 8.5)}$$

This model is not likely to be of great use as a predictive tool unless highway sections with non-standard or sub-standard cross-sectional geometrics become available to incorporate into the analysis. As well, the extremely low R^2 value determined for this model also limits the usefulness and possibly the accuracy of it. The lack of key cross-sectional geometric factors from the relationship limits the scope of the model and consequently, it's usefulness. Overcoming this deficiency in data may require the use of non-traditional methods of analysis, such as the Empirical Bayesian methodology and the before-and-after techniques of data analysis presented in Appendix B of this thesis, since it would not be practical to under-design highways.

8.2 RECOMMENDATIONS

Five recommendations should be considered to further improve on the methodology in determining the relationship between accident rates and cross-sectional highway geometrics in Alberta. They are:

1. If possible, improve the diversity and accuracy of values within each cross-sectional geometric factor. By increasing the heterogeneity of values in each geometric factor, stronger regression relationships between accidents and cross-sectional geometrics can be developed. Because the use of mean roadside factors introduces an element of uncertainty in the results, improving the accuracy of each geometric factor will directly improve the model as well.
2. Expand the scope of the research to include the Alberta secondary highway system. The addition of these highways, with a variety of design standards and geometric conditions may be an effective method of introducing geometric factor diversity into the regression models. The inclusion of secondary highways in the model may also yield relationships between accident rates and pavement conditions (paved vs. unpaved) as well.
3. Expand sample timeframe beyond a four year accident data sample. The inclusion of more accident data over a longer period of time may improve the resulting relationships between accidents and geometrics. However, care should be taken to ensure that the sampling period does not overlap gross changes in vehicle

performance characteristics, new vehicle and highway safety improvements, multiple highway rehabilitation, changes in driver behaviour, and other factors that may undermine or cloud causal relationships between accidents and geometrics. Increasing the sample period from four years to six or eight years may be sufficient to produce additional data points without introducing additional confounding factors mentioned above.

4. Incorporate changing patterns of traffic volumes, roadway and roadside geometric improvements into the model to more accurately model the accident-geometric relationship. The assumption that traffic volumes used in the analysis remain at 1991 levels is inaccurate and introduces further error into the model. As well, the additional assumption that roadway and roadside geometric conditions do not change over the four year span are also inaccurate. The use of techniques such as an Empirical Bayesian methodology may aid in incorporating these ongoing network changes into the accident rate model. However, merely combining accident data with geometric data on a yearly basis (i.e. 1996 primary highway accident statistics with the 1996 primary highway cross-sectional geometric inventory) will improve on the accuracy and validity of the accident rate model.

5. A continuous calibration of the Alberta Accident Model is advisable, as new accident and geometric data becomes available. To ensure that the data used in the model is not obsolete, the earliest year of both accident and geometric data can be removed from the analysis and the most recent accident and geometric data incorporated.

References

Alberta Transportation and Utilities. (1995). Feasibility Study on the Modeling of Relationships Between Highway Geometric Design Features and Roadway Geometrics. Edmonton, Alberta.

Alberta Transportation and Utilities. (1992). Alberta Traffic Collision Statistics - 1992. Edmonton, Alberta.

Al-Masaeid, Hashem R., K. C. Sinha, and T. Kuczek. (1993). Evaluation of Safety Impact of Highway Projects. Transportation Research Record 1401. Washington, D. C.: Transportation Research Board.

Banks, Joseph F., Brown, D. B., and Beatty, R. L. (1978). Cost and Safety Effectiveness of Highway Design Elements. NCHRP Report 197. Washington, D.C.: Transportation Research Board.

Brinkman, C. P. and S. A. Smith. (1984). Two-Lane Rural Highway Safety. Public Roads, Vol. 48, No. 2.

Chan, P. L. P. (1994). Use of Delay in Accident Prediction at Intersections Without Traffic Signals. Unpublished M.Sc. Thesis, University of Alberta, Edmonton, Alberta.

Cleveland, D. (1987). Effects of Resurfacing on Highway Safety. TRB State of the Art Report 6. Washington, D.C.: Transportation Research Board.

Glennon, J. C., Neuman, T. R., and Leisch, J. E. (1986). Safety and Operational Considerations for Geometric Design of Rural Highway Curves. Report FHWA-RD-86/035. Washington D.C.: U.S. Department of Transportation.

Graham, J. L., and Harwood, D. W. (1982). Effectiveness of Clear Recovery Zones. NCHRP Report 247. Washington, D.C.: Transportation Research Board.

Hauer, E. (1986). On the Estimation of the Expected Number of Accidents. Accident Analysis and Prevention, Vol. 18, No. 1.

Higle, Julia L., and Witkowski, J. M. (1988). Bayesian Identification of Hazardous Locations. Transportation Research Record 1185. Washington, D.C.: Transportation Research Board.

Kenny, B. (1995). 3R/4R Geometric Design Guidelines For Rural Paved Highways in Alberta. Edmonton, Alberta: Alberta Transportation and Utilities.

Kenny, B. (1996, July). Conversation concerning roadside recovery distance calculation.

Lin, Feng-Bor. (1990). Flattening of Horizontal Curves on Rural Two-Lane Highways. Journal of Transportation Engineering, Vol. 116, No. 2, New York: American Society of Civil Engineers.

Mendenhall, W. (1988). Statistics for the Engineering and Computer Sciences. San Francisco: Dellen Publishing Company.

Meyer, S. L. (1975). Data Analysis for Scientists and Engineers. New York: John Wiley and Sons.

Norusis, Marija J. (1992). SPSS Systems User's Guide. Chicago: SPSS Inc.

Olsen, P. L. (1984). Parameters Affecting Stopping Sight Distance. NCHRP Report 270. Washington, D.C.: Transportation Research Board.

Paluri, A. (1987). Relationships Between Design Parameters, Operating Speeds, and Accident Rates for Future Highway Design Consideration on Two-Lane Rural Curved Road Sections. Unpublished MSc. Thesis, Clarkson University, New York.

Persaud, B. (1986). Relating the Effect of Safety Measures to Expected Number of Accidents. Accident Analysis and Prevention, Vol. 18, No. 1.

Rilett, L. R. (1988). Development, Analysis and Application of Load Equivalency Factors from Canroad Test Data. Unpublished M.A.Sc. Thesis, University of Waterloo, Ontario.

Rogness, R. O., Fambro, D. B., and Turner, D. S. (1982). Before-After Accident Analysis for Two Shoulder Upgrading Alternatives. Transportation Research Record 855. Washington, D.C.: Transportation Research Board.

RTAC. (1986). Manual of Geometric Design Standards for Canadian Roads. Ottawa: Roads and Transportation Association of Canada.

Smith, S. A., Purdy, J., McGee, H. W., Harwood, D. W., St. John, A. D., and Glennon, J. C. (1983). Identification, Quantification, and Structuring of Two-Lane Rural Highway Safety Problems and Solutions. Report FHWA-RD-83/021. Washington D.C.: U.S. Department of Transportation.

Speigel, Murray R. (1975). Probability and Statistics. New York: McGraw-Hill.

TRB. (1985). Special Report 209: Highway Capacity Manual. Washington, D.C.: Transportation Research Board.

TRB. (1987). Special Report 214: Designing Safer Roads. Practices for Resurfacing, Restoration, and Rehabilitation. Washington, D.C.: Transportation Research Board.

Turner, Don S. (1984). Prediction of Bridge Accident Rates. Journal of Transportation Engineering, Vol. 110, No. 1, New York: American Society of Civil Engineers.

Zegeer, Charles V. and Deacon, J. A. (1987). Effect of Lane Width, Shoulder Width, and Shoulder Type on Highway Safety. TRB State of the Art Report 6. Washington, D.C.: Transportation Research Board.

Zegeer, C. V., Hummer, J., Reinfurt, D., Herf, L., and Hunter, W. (1987). Safety Effects of Cross-Section Design for Two-Lane Roads. Report FHWA-RD-87/008. Washington, D.C.: U.S. Department of Transportation.

Zegeer, C. V., Mayes, J. G., and Dean, R. C. (1981). Cost-Effectiveness of Lane and Shoulder Widening of Rural Two-Lane Roads in Kentucky. Transportation Research Record 806. Washington, D.C.: Transportation Research Board.

Glossary of Terms

3R/4R -	Restoration, Rehabilitation, Repaving, and Reconstruction highway infrastructure improvement projects.
AADT -	Annual Average Daily Traffic volume. The total vehicle volume recorded over a section of highway in one year divided by 365 (days).
AASHTO -	American Association of State Highway Transportation Officials
Accident -	A common term used to describe a vehicle or driver failure event on a highway. The fault for the incident is not attributable in this usage.
AT&U -	Alberta Transportation and Utilities
Clear Zone -	The horizontal cross-sectional distance measured from the outside edge of the roadway shoulder to either the hinge point where any sideslope or backslope ratio becomes steeper than 4:1, or where any irregular surface or discontinuity (rocks, treeline, or sheer face) begins. This relatively flat, unobstructed and smooth distance is also referred to as the Roadside Recovery Distance.
Collision -	A specific type of accident involving a vehicle or vehicles striking either a stationary or moving object.

Cross-Sectional - Attributes of a highway involving the width of the section. These factors include lane and shoulder widths, roadway pavement types, as well as roadside slope and width geometrics.

Encroachment - The action occurring when a vehicle travels beyond the outer edge of the highway shoulder and enters the roadside zone. This departure from the roadway surface into the roadside environment is usually uncontrolled and/or unintentional

FHWA - Federal Highway Administration

Design Geometrics - Design factors attributable to roadway geometric design

Hazard Rating - A scale between 1 and 7 (1 being least hazardous and 7 being most hazardous) that quantifies the probability of accident occurrence and the severity of an accident due to the various roadside factors.

Reconstruction - Geometric improvement involving the complete rebuilding of a roadway section, usually including all aspects of design geometrics. This improvement is the most costly of the four types.

Recovery Distance - See clear zone

Regression-to-mean - This effect states that if the value of an event for any given time period exceeds the prior mean value of the event, then the likelihood of a future event value being either below or at the mean event value increases. A corollary of this is that if the event value is below the historical mean value of the event, then future events are more likely to be above the historical mean.

Rehabilitation -	A roadway improvement designed to improve existing geometrics to a higher standard.
Repaving -	A roadway improvement incorporating the restoration of the pavement surface condition to the design standard or better.
Restoration -	Geometric improvement involving the rebuilding of a geometric element to it's design condition or specifications. These changes are not as cost intensive as Reconstruction improvements.
Roadside -	Any feature of the highway attributed to the geometrics not on the roadway.
Roadway -	The portion of the highway cross-section that includes the traveled lanes and both paved and/or unpaved shoulders.
ROR -	Run-Off-the-Road. A type of accident involving a vehicle departing the paved edge of the roadway and entering the roadside.
TRB -	Transportation Research Board

Appendix A

FORTRAN Code

A.1 ALBERTA GEOMETRIC DATABASE DATA STRIPPER

```
c-----------------------------------------------------------------
c  File:   ABGEO.FOR
c  Type:   Header/Footer stripper
c  Desc:   This file strips extraneous characters from an ASCII data
c          file.  Output file format is in the form of tab-delineated
c          data columns.
c  By:     R. Dai
c  Coding Date:    July 4, 1995
c  Revision Date:  July 5, 1995
c-----------------------------------------------------------------

c-----------------------------------------------------------------
c  Define variables
c    * note: all variables must be char$ since headers for each
c            page of AT&U's datafile are char$.
c-----------------------------------------------------------------

      integer count
      character*1 coll, hwytyp, srftyp, shtyp
      character*3 ctl, sub, spday, spnite
      character*4 hwyno, shwdo, aadtyr, aadt
      character*5 shwdi, medwd
      character*6 srfwd

c-----------------------------------------------------------------
c  Open Files for input / output
c-----------------------------------------------------------------

      open(unit=5,file='e:\ray\ABGEO91.DAT')
      open(unit=6,file='e:\ray\ABGEO91.OUT')

      count = 1

c-----------------------------------------------------------------
c  Read the first row of the data file
c    * note: coll will be used for error-checking only (no output)
c-----------------------------------------------------------------

  10      read(5,51) coll, hwyno, ctl, sub, hwytyp, srftyp, srfwd, shtyp,
     +      shwdi, shwdo, medwd, aadtyr, aadt, spday, spnite
  51      format(A1,T3,A4,T8,A3,T12,A3,T70,A1,T75,A1,T77,A6,T87,A1,
     +      T90,A5,T97,A4,T104,A5,T111,A4,T118,A5,T125,A3,T130,A3)

c-----------------------------------------------------------------
c  Check if the row is a valid data set
c    * note: if coll = 0, data set is valid and follows
c-----------------------------------------------------------------

      if(coll.eq."0") then

         write(6,60) hwyno, ctl, sub, hwytyp, srftyp, srfwd, shtyp,
     +      shwdi, shwdo, medwd, aadtyr, aadt, spday, spnite

  60      format(A4,1X,A3,1X,A3,1X,A1,1X,A1,1X,A6,1X,A1,1X,A5,
     +      1X,A4,1X,A5,1X,A4,1X,A5,1X,A3,1X,A3)
```

```
         count = count + 1
         write(*,*) 'count =  ',count,'  col1 = ',col1

c-----------------------------------------------------------------
c  Check if EOF is reached
c   * note: EOF reached when col1 = 9
c-----------------------------------------------------------------

      else
         if (col1.eq."9") then
            goto 100
         end if

      end if

      goto 10

c-----------------------------------------------------------------
c  Close all open external files
c-----------------------------------------------------------------

  100 close(unit=5,status='keep')
      close(unit=6,status='keep')

      stop
      end
```

A.2 ALBERTA ACCIDENT DATABASE DATA STRIPPER

```
c-----------------------------------------------------------------
c  File:   PHWY.FOR
c  Type:   Editor for PHWY*.DAT accident files
c  Desc:   This file strips extraneous characters from an ASCII data
c          file.  Output file format is in the form of tab-delineated
c          data columns.
c  By:     R. Dai
c  Coding Date:    July 4, 1995
c  Revision Date:  July 5, 1995
c-----------------------------------------------------------------

c-----------------------------------------------------------------
c  Define variables
c   * note: all variables must be char$ since headers for each
c             page of AT&U's datafile are char$.
c-----------------------------------------------------------------

integer count
      character*1 col1, hwytyp, srftyp, shtyp
      character*3 ctl, sub, spday, spnite
      character*4 hwyno, shwdo, aadtyr, aadt
      character*5 shwdi, medwd
      character*6 srfwd

c-----------------------------------------------------------------
c  Open Files for input / output
c-----------------------------------------------------------------

      open(unit=5,file='e:\ray\ABGE091.DAT')
      open(unit=6,file='e:\ray\ABGE091.OUT')
```

```
      count = 1

c-------------------------------------------------------------------
c  Read the first row of the data file
c   * note: col1 will be used for error-checking only (no output)
c-------------------------------------------------------------------

 10     read(5,51) col1, hwyno, ctl, sub, hwytyp, srftyp, srfwd, shtyp,
    +     shwdi, shwdo, medwd, aadtyr, aadt, spday, spnite
 51     format(A1,T3,A4,T8,A3,T12,A3,T70,A1,T75,A1,T77,A6,T87,A1,
    +     T90,A5,T97,A4,T104,A5,T111,A4,T118,A5,T125,A3,T130,A3)

c-------------------------------------------------------------------
c  Check if the row is a valid data set
c   * note: if col1 = 0, data set is valid and follows
c-------------------------------------------------------------------

      if(col1.eq."0") then

        write(6,60) hwyno, ctl, sub, hwytyp, srftyp, srfwd, shtyp,
    +     shwdi, shwdo, medwd, aadtyr, aadt, spday, spnite

 60     format(A4,1X,A3,1X,A3,1X,A1,1X,A1,1X,A6,1X,A1,1X,A5,
    +     1X,A4,1X,A5,1X,A4,1X,A5,1X,A3,1X,A3)

        count = count + 1
        write(*,*) 'count =  ',count,'  col1 = ',col1

c-------------------------------------------------------------------
c  Check if EOF is reached
c   * note: EOF reached when col1 = 9
c-------------------------------------------------------------------

      else
        if (col1.eq."9") then
          goto 100
        end if

      end if

      goto 10

c-------------------------------------------------------------------
c  Close all open external files
c-------------------------------------------------------------------

 100  close(unit=5,status='keep')
      close(unit=6,status='keep')

      stop
      end
```

Appendix B

Identification of Accident Locations:

An Empirical Bayesian Methodology

B.0 INTRODUCTION

As seen in the prior chapters, generalized linear models can be used to identify relationships between accident rates (a common measure of accident experience and highway safety) and contributing cross-sectional geometric factors. These models relate an accident rate, Y, to a measurable, and presumably contributory geometric factor X, multiplied by some adjustment coefficient.

As one of the key reasons for their widespread usage, these models are not usually mathematically intensive or difficult to use. From the accident rate estimated from geometric input parameters, a highway section can be identified as deficient in safety if it's rate exceeds a background rate established as a safety standard. The background accident rate may be defined as an average accident rate occurring throughout the province based on historical accident data, plus a multiple of the standard deviation of accident rates in smaller regions throughout the province. The background rate, in effect, establishes the average rate of accidents experienced historically in the region, plus a certain level of confidence with the statistical mean.

Therefore, generalized linear modeling can be used to first, identify whether a certain geometric design treatment is effective in reducing accident rates, and next, identifying whether sites with the treatment in question are experiencing greater than average accident rates (hazardous location). In essence, generalized linear models are a useful tool in aiding engineers in quantifying safety effects of their geometric designs.

B.1 PROBLEMS WITH GENERALIZED LINEAR MODELS

While useful for the simplicity of modeling data, generalized linear models have three key inherent limitations. These limitations, or problems are:

1. Errors of estimation caused by potential regression-to-mean effects.
2. Possibility of false-negative results from hazardous location identification.
3. Reactive nature of linear modeling in accident prevention.

B.1.1 REGRESSION-TO-MEAN EFFECTS

The method that general linear models accurately predict the safety-geometric relationship rests in the theoretical modeling of the prior means of all data to be modeled. In essence, all points that define the best-fit regression line must be the prior (or historical) mean of the accident rates and geometric factors that they represent. The practical implication of this assumption is that if a data point is not based on the mean incidence rate at that location (i.e. follows a non-gaussian distribution), the result of the regression may be artificially skewed by the regression-to-mean effect (***Persaud, 1986***).

This effect states that at any one location, if an accident rate for any given year exceeds the prior mean accident rate value, then the likelihood of a future rate being either below or at the mean rate increases (***Hauer, 1986***). In effect, data not representing the prior mean rate of a location may not accurately represent the "true" accident rate of the location, and may introduce inaccuracy into the general linear regression model.

B.1.2 FALSE-NEGATIVES IN HAZARDOUS LOCATION IDENTIFICATION

As a result of the regression-to-mean effect and the inherent randomness of accident data, the possibility arises that false-negative identification of hazardous sites may occur. These can result from a lower than mean rate of accidents experienced at a particular site which cannot be directly attributable to any quantifiable geometric factor (***Higle and Witkowski, 1988***). Due to the extremely infrequent nature of accidents (most highway sections rarely experience any accidents in a given year), the possibility of a highway section experiencing zero accidents for a period of time does exist, even if the highway section is inherently hazardous or geometrically unsafe. As a result, although the site may be experiencing the regression-to-mean effect, and is in fact hazardous, the site is nevertheless assumed as being non-hazardous (***Al-Masaeid et al., 1993***).

Mistyping of locations can often lead to spending budget-constrained project dollars on sites that actually do not require safety improvements (false positive results) or completely missing locations that actually require improvements (false negative results). These flaws in the prediction model can lead to flaws in safety management and decision-making. While false-positive tests are not desirable, since they lead to extraneous spending, false negative tests are clearly the main concern of 3R/4R and safety design engineers as they may lead to potentially unsafe 3R/4R highway geometrics.

B.1.3 THE NATURE OF LINEAR MODELING IN ACCIDENT PREVENTION

Linear modeling is by nature, a reactive approach to safety engineering. When predicting hazardous locations using the traditional linear regression method, only highway sections that actually exceed the background level will be identified. In effect, in order for a section to be identified as hazardous and to qualify for safety improvements, the accident must occur before, and not after the improvement is implemented. This dichotomy is indicative of the reactive nature of linear models- these models cannot estimated the likelihood of further accidents occurring until an accident (with resulting human and socio-economic losses incurred from the accident) has occurred.

B.2 EMPIRICAL BAYES APPROACH TO HAZARD IDENTIFICATION

With these difficulties in accurately identifying a historical rate of accidents for each location (and the practicalities of such a statistical database), identification of hazardous locations using an Empirical Bayes analytical approach can be used. Empirical Bayesian analysis enables prior (or historic) accident rates at specific locations to be integrated into a regional historic background accident rate such that random elements present in all roadway accidents can be mathematically defined and isolated (***Higle and Witkowski, 1988***). Furthermore, the use of Bayesian statistics allows one to adjust accident rates by incorporating the most current historical data (as it becomes available). This is a distinct advantage over the traditional linear modeled rate quality control method in that each new data element (recent incident or accident rate) cannot be integrated into the linear model, since the model represents only mean events. In essence, Empirical Bayesian techniques may be more flexible in modeling accident rates than traditional generalized linear models.

By eliminating the problems associated with the regression-to-mean phenomenon, the possibility of testing out false-negatives in site identification may also be eliminated. False-negatives are less likely to occur when using the Empirical Bayesian methodology since prolonged zero accident rates (a highway section appearing "normal", yet indicative of a potential false-negative due to inherent geometric deficiencies) will be incorporated into the Bayesian model as a matter of course, and can change the calculated probability of an accident occurring, but not the actual identification process.

In addition, an Empirical Bayes approach to modeling accident rates allows one to identify hazardous sites due to the probability of that site's accident rate exceeding a background level (as opposed to the reactive approach of linear modeling where the site rate must actually exceed the background level before that site can be identified as hazardous). This probabilistic identification method differs both quantitatively and qualitatively from the traditional linear modeling (and rate quality control) method (***Higle and Witkowski, 1988***).

B.3 BAYESIAN METHODOLOGY

Rather than ignoring the fact that accident rates for specific highway sections are to a large extent, randomly driven (which is what a traditional linear model approach would entail), the Empirical Bayesian methodology treats local accident occurrence as random, and uses regional accident rates and the location's accident history as a basis for estimating the probability of hazard (***Higle and Witkowski, 1988***).

The Empirical Bayesian methodology can involve a two-step procedure. First, regional accident data is aggregated to determine the probability distribution of accident rates throughout the region. This probability distribution is then combined with a particular location's accident history to determine a more precise estimate of the accident probability distribution associated with that site. From this analysis (and those of other sites in question), local sites can then be identified as being hazardous from their accident probabilities. The following notation is used in this methodology:

$\lambda_i =$	Accident rate at a highway section, i (λ_i is Poisson-randomly distributed)
$N_i =$	Number of accidents at highway section, i during a time period
$V_i =$	Vehicle volume through highway section, i during a time period
$f_i(\lambda_i \mid N_i, V_i) =$	Probability density function associated with the accident rate at highway section, i, given that N_i and V_i are observed.
$f_R(\lambda_i) =$	Probability density function associated with the regional accident rate.

Therefore, f_R is the function associated with the regional rate, and f_i is the function associated with the individual (and more refined) estimate of site i's accident rate. As well, the cumulative distribution function (CDF) associated with the accident rate λ_i is:

$$P(\lambda_i \leq \bar{\lambda}) = \int_0^{\bar{\lambda}} f_i(\lambda | N_i, V_i) d\lambda \qquad \textbf{(EQN B1)}$$

Two key assumptions must be made for this approach to work. They are:

1. Regardless of the accident rate at any given location, the actual number of accidents experienced follows a Poisson distribution with expected value λV_i (mean).

$$P\{N_i = n \mid \bar{\lambda}_i = \lambda, V_i\} = \frac{(\lambda Vi)^n}{n!} e^{-\lambda V_i} \qquad \textbf{(EQN B2)}$$

   This assumption is based on the explicit treatment of actual accident numbers at individual sites as random, as mentioned previously.

2. The probability distribution of the regional accident rate, $f_R(\lambda)$, is gamma distributed. Expected numbers of accidents in the region is also a random variable and is gamma distributed. These assumptions imply that:

$$f_R(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda} \qquad \textbf{(EQN B3)}$$

   for some a and b (***Al-Masaeid, et al., 1993***).

To determine $f_R(\lambda)$, a and b must first be determined. Most commonly, the method of moments estimates (MME) is used, where a and b are chosen such that the mean and variance of the gamma distributed function is the same as the mean and variance of the sample. Another method used involves selecting a and b such that they represent values that are most likely to yield the observed data (method of maximum likelihood estimates - MLE). As well, other methods exist to select a and b (***Higle and Witkowski, 1988***). These methods are not discussed in this paper. Once $f_R(\lambda)$ is found (by determining the parameters a and b), Bayes' theorem can be applied to find the site-specific accident probability distribution functions.

$$f_i(\lambda | N_i, V_i) \propto f(N_i | \lambda, V_i) \times f_R(\lambda) \qquad \textbf{(EQN B4)}$$

Because of assumption 2 made previously, $f_i(\lambda \mid N_i, V_i)$ is as a result, also gamma distributed with parameters $a_i = a + N_i$, and $b_i = b + V_i$, based on initial parameters chosen. Therefore, the probability density function of expected accident rates associated with a particular location i, can be given as the following:

$$f_i(\lambda | N_i, V_i) = \frac{\beta_i^{\alpha_i}}{\Gamma(\alpha_i)} \lambda^{\alpha_i - 1} e^{-\beta_i \lambda} \qquad \textbf{(EQN B5)}$$

It is interesting to note that as N_i and V_i increase, the significance of the initial choices for a and b decreases (i.e.. $a_i \rightarrow N_i$, and $b_i \rightarrow V_i$).

Once these probability density functions are determined, it is then possible to evaluate each site for hazard significance by comparing each site with some background accident rate criteria (mean λ), subject to some tolerance level d.

$$P(\lambda_i > \overline{\lambda} \mid N_i, V_i) > \delta \qquad \textbf{(EQN B6)}$$

B.4 APPLICATIONS OF BAYESIAN ANALYSIS

As mentioned, the Empirical Bayesian methodology can be applied to highway accident data associated with or caused by certain geometric factors. The resulting analysis may then be used to identify the probability that other sites exhibiting similar characteristics will also experience probable accident rates similar to prior situations. This technique has the added advantage that levels of statistical confidence can be changed without requiring numerical recomputation of the entire function. As well, the significance of being able to compute a probability that the actual site accident rate may or may not exceed the regional accident rate can be important to 3R/4R highway design engineers. Knowing the probability of hazard associated with a section of roadway enables an engineer to determine trade-offs between implementing safety improvements in a section, staging improvements, or doing nothing. This may not seem significant, but being conscious of degrees of hazard (through probabilities) is preferable to the traditional approach of having to fix all hazards, regardless of it's impact on accident rates and safety in general.

B.5 SUMMARY

While this overview of Empirical Bayesian identification techniques is by no means exhaustive, nor an indictment on the capabilities and advantages of traditional accident models (rate-quality control in a general linear format), it does raise several issues concerning the suitability of universally applying a rate control model to all accident data. Careful analysis and consideration of underlying assumptions need to be made if the results of the model are to accurately reflect actual and possible future situations.

What is also clear is that the Bayesian methodology excels in situations where accident data is sparse or discontinuous through the years. Gaps in accident data can lead to bias creep in traditional statistical models. But because historical gaps may not contribute significantly to differences in a regional mean accident rate, as well as having no effect on the application of prior data to the Bayes model (one requires only what data is available), this problem does not affect Bayesian modeling. However, it should be noted that as the sample size of accident records increases, the differences seen between a Bayesian approach and a traditional linear model disappear. It is thus reasonable to state that the Bayesian methodology works best under conditions where historical data is sparse, yet recent data is readily available - the state that many transportation agency accident and geometric databases are in today.

Appendix C

Non Cross-Sectional Related Geometric Factors

Affecting Highway Safety

C.0 INTRODUCTION

Chapter 2 of this thesis examined the effects of cross-sectional highway geometric factors on accident rates. Other geometric related factors that may contribute to highway accidents, but lie outside of the scope of this thesis are: (1) bridge widths, (2) horizontal alignment, (3) sight distances, (4) intersections, (5) pavement surface conditions, and (6) pavement edge drops.

C.1 BRIDGE WIDTH

Bridge hazards may be potentially significant contributors to highway accidents due to constricted roadways at entrances, and the common physical terrain constraints featured with most bridges and their approaches. Other factors, such as the premature icing of bridge decks, or rain on the deck, also play key roles in bridge safety. A 1984 study conducted into the geometric relationship of bridge widths and accident rates resulted in the following accident rate estimation model:

$$AR = 0.50 - 0.061\,(RW) + 0.0022\,(RW)^2 \qquad \textbf{(EQN C.1)}$$

where, AR = accident rate per million vehicles
RW = relative bridge deck width, in feet ($0 < RW \leq 14$ ft)

This model is not applicable to bridges with approach lanes wider than the total clear deck width since the drastic reduction in lane width at the approaches were found to greatly increase the accident rate at the corresponding bridge (***Turner, 1984***). The model specifically applies to bridge situations involving total roadway narrowing (lane and shoulder widths) at the approaches, with lane widths remaining constant throughout the length of the approaches and bridge deck.

From this quantitative relationship developed, the following conclusions were determined: (1) increasing the bridge width with respect to the width of approach lanes can reduce the frequency of accidents by as much as 40%. This gain can result from as small an increase as 0.5 m, and (2) incremental safety gains from the actual widening of bridge lanes decreases as lane width increases. This is additional proof that merely widening bridge decks may not be a cost-effective safety measure. However, the quadratic nature of the model indicates that initial incremental increases in bridge widths do yield potentially large returns in accident rate reduction.

Despite the widespread and intuitive belief that low-cost improvements such as barriers, rails, and special signage benefit roadway safety, their actual incremental effects on bridge safety and accident rates have not been quantified in a reliable manner. This area of bridge safety continues to be guided by expert and qualitative judgment (***TRB, 1987***)

C.2 HORIZONTAL ALIGNMENT

Due to the physiological and psychological conditions imposed on a driver negotiating a curve (as well as the physical effects on a vehicle), horizontal curves contribute a larger share of accidents than straight sections of roadway. Accident rate models seem to indicate that degree of curvature, as well as vehicle exposure (represented in vehicle miles traveled through a curve) are the most significant contributors to curve accidents. It has been estimated that for each degree reduction in curvature, three accidents can be eliminated per 100 million vehicle miles traveled (***TRB, 1987***).

Several variations exist in modeling the relationship between accident rates and the geometry of alignments, but most generally relate accident rates as a function of degree of curvature. Based on a linear regression fit to data obtained in a 1987 New York study (***Paluri, 1987***), a relationship between accident rate (per million vehicle miles) and degree of curvature was found to be (***Lin, 1990***):

$$A = -1.079 + 1.302\, D \qquad \textbf{(EQN C.2)}$$

where, A = accidents per million vehicle miles
D = degree of curvature

This relationship has an r^2 value of 0.43, and a standard error of estimate of 7.8 accidents per million vehicle miles. To account for traffic exposure (i.e. volume of traffic on the section of curvature), Lin determined a new relationship:

$$A = 1.09 + 1.61\, D\, (1 - 1.20\, m + 0.40\, m^2) \qquad \textbf{(EQN C.3)}$$

where, m = traffic exposure of the curve

Based on an analysis of covariance by Kerlinger in 1973, it was concluded that once degree of curvature and traffic exposure were accounted for in the accident rate of horizontal curves, no other factors (such as lane width or curve length) or combination of factors were statistically significant contributors to accident rates.

Prior research into the application of safety in 3R/4R geometric design supports this conclusion (***Glennon, Neuman, and Leisch, 1983***). Glennon et al.'s work affirms the general relationship between degree of curvature and accident rate (that as curvature increases, accident rates increase as well) using a slightly different approach. This approach represents accident rates on a curved section as two components: a base accident rate (i.e. straight section accident rate) regardless of roadway curvature, and an accident rate distinctly attributable to the degree of roadway curvature. Therefore,

$$A = A_S + A_C \qquad \textbf{(EQN C.4)}$$

where, A = combined accident rate
A_S = accident rate on straight sections
A_C = accident rate on curved sections

and, $$A_S = AR_S\,(L_S)\,(V) \qquad \textbf{(EQN C.5)}$$

where, AR_S = accident rate (per million vehicle miles) on straight sections
L_S = length of straight section, in miles
V = traffic volume in millions of vehicles

or, $$A = AR_S\,(L)\,(V) + [\,A_C - AR_S\,(L_C)\,(V)\,] \qquad \textbf{(EQN C.6)}$$

$$A = AR_S\,(L)\,(V) + \Delta A_C \qquad \textbf{(EQN C.7)}$$

where, L_C = length of curved section, in miles
ΔA_C = increase in accidents due to curvature beyond that anticipated for a straight section of road of the same length.

Calibration from Glennon's research data (study of over 3000 sites in four states: Florida, Illinois, Ohio, and Texas) yielded $\Delta A_C = 0.0336(D)(V)$, where D is the degree of curvature. As well, $AR_S = 0.902$ was determined from the data (presumably from a least squares fit of the accident data). This leads to the final equation:

$$A = 0.902\,(L)\,(V) + 0.0336\,(D)\,(V) \quad \text{for } L \geq L_C \qquad \textbf{(EQN C.8)}$$

and for accidents relating to curved segments alone,

$$A_C = 0.902\,(L_C)\,(V) + 0.0336\,(D)\,(V) \qquad \textbf{(EQN C.9)}$$

where, L_C = length of curved section, in miles
V = traffic volume, in millions of vehicles
D = degree of horizontal curvature

The first part of the equation represents the accident potential of the steady-state effect (similar to a straight through drive) of a vehicle in a curve, and the second part represents the transitional effects of entering and exiting the curve and the potential for an accident to occur as a result of those effects.

C.3 SIGHT DISTANCE

The effectiveness of increasing safety as a result of eliminating sight restrictions has been found to be as much as 52% over control sites (***Olsen, 1984***). The development of sight distance guidelines is documented extensively in most road design manuals, and is mainly dependent on vertical gradients on crest curves.

C.4 INTERSECTIONS

Rural highway intersections have the third highest concentration of accidents, after horizontal curve locations and bridges (***Brinkman and Smith, 1984***). The U.S. National Safety Council estimated in 1985 that 32% of all rural highway accidents occurred at an intersection. The frequency and severity of events occurring at intersections has resulted on a great deal of emphasis placed on designing safer intersections.

Improvements to intersections are often conducted individually, since many intersections exhibit unique geometrics. In general, improvements are designed to reduce potential vehicle conflicts, increase driver awareness, and improve driver decision. The application of a general model similar to that used with lane widths, or horizontal curves cannot be readily applied to intersections because of the large number of geometric factors in intersection design that could contribute to an accident. As well, potential inter-related effects between intersection variables may mask any true relationship between specific geometric elements and a corresponding accident rate. As a result, it is difficult to quantify the effects specific intersection geometric features in a linearized accident rate model form.

C.5 PAVEMENT SURFACE CONDITIONS

The traditional nature of 3R projects usually involves some form of pavement rehabilitation. Resurfacing of roadways traditionally protects the road structure and is believed to enhance the driveability and vehicle handling as a result. The quantifiable safety implications of resurfacing are not as clear, however. Two main effects of resurfacing work opposite to each other in regards to safety. While a new surface will often increase pavement skid resistance (and consequently reduce stopping distances and increase driver avoidance abilities), it also can lead to higher operating speeds as a result of driver behaviour. Research has found that dry pavement accidents can increase up to 10% immediately after resurfacing, probably attributable to higher operating speeds as a result of roadway upgrading (***Cleveland, 1987***). As well, 15% decreases in accidents under wet pavement conditions were also observed. These changes in accident rates were found to decrease over time. As a result, it was found that the net effect of resurfacing on safety is positive over the roadway's lifespan, in that benefits derived from increased driver control far outweighs the consequences of higher operating speeds on the roadway.

C.6 PAVEMENT EDGE DROPS

Edge drops are vertical discontinuities located at the edge of pavement surfaces. They can result from wear on a roadway, or misaligned resurfacing of the main pavement area. Investigation into this area has quantitatively determined that these drops pose a significant hazard since they can readily turn a minor run-off-road encroachment into a serious roadside collision.