EMPIRICAL BAYESIAN METHODS FOR THE IDENTIFICATION OF SITES TO BE CONSIDERED FOR SPECIFIC SAFETY TREATMENTS

BY

Craig Allan Lyon

A thesis submitted in conformity with the requirements for the degree of Master of Applied Science
Graduate Department of Civil Engineering
University of Toronto

© Copyright by Craig Allan Lyon 1999
The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L’auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author’s permission.

L’auteur conserve la propriété du droit d’auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.
ABSTRACT

This study develops and illustrates a methodology for identifying sites which have the potential for safety improvements for a specific treatment type. The methodology is applied to two types of vehicular accidents, those occurring due to the presence of horizontal curves and those that are speed related. The road segments investigated are all on two-lane rural undivided highways; data used were obtained from the Ministry of Transportation, Ontario. The modeling utilizes the Empirical Bayesian methodology to estimate the expected frequency of treatable accidents occurring at each site. This estimate is used as a ranking index for treatment. Models are developed for injury (fatal and non-fatal injury) accidents as well as all severities combined. A validation of the procedure shows that it is more efficient and reliable than commonly used blackspot identification methods based on recent accident counts or rates. A useful side benefit is that the procedure provides the basis for the estimation of the safety effect of treatments applied.
ACKNOWLEDGMENT

This study could not have been completed without the generous assistance of Doug Coulter, Paul Treinhalie and Dave Steed, staff of the MTO central and eastern offices. The guidance and financial support of Dr. Bhagwant Persaud is gratefully acknowledged. My thanks goes to Dr. Amer Shalaby for kindly reviewing this work. The support of the Lyon family throughout my years of post-secondary education deserves thanks. Finally I wish to thank Kelly Read who has given me immeasurable personal support during my study.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>ii</td>
</tr>
<tr>
<td>ACKNOWLEDGMENT</td>
<td>iii</td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td>iv</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>vii</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>viii</td>
</tr>
<tr>
<td>GLOSSARY OF NOTATION</td>
<td>ix</td>
</tr>
<tr>
<td>1.0 INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>2.0 LITERATURE REVIEW</td>
<td>4</td>
</tr>
<tr>
<td>2.1 HORIZONTAL CURVE WARNINGS</td>
<td>4</td>
</tr>
<tr>
<td>2.1.1 ACCIDENTS ON HORIZONTAL CURVES</td>
<td>4</td>
</tr>
<tr>
<td>2.1.2 TREATMENTS OF HAZARDOUS CURVES</td>
<td>6</td>
</tr>
<tr>
<td>2.2 SPEED CONTROL MEASURES</td>
<td>9</td>
</tr>
<tr>
<td>2.2.1 SPEED RELATED ACCIDENTS</td>
<td>9</td>
</tr>
<tr>
<td>2.2.2 TREATMENTS OF SECTIONS WITH A SPEEDING</td>
<td>11</td>
</tr>
<tr>
<td>2.3 THEORETICAL FOUNDATIONS</td>
<td>12</td>
</tr>
<tr>
<td>2.3.1 BACKGROUND</td>
<td>13</td>
</tr>
<tr>
<td>2.3.2 ESTIMATES OF $E(m/x)$ and $VAR(m/x)$</td>
<td>16</td>
</tr>
<tr>
<td>2.3.3 SELECTION OF MODEL VARIABLES</td>
<td>18</td>
</tr>
<tr>
<td>2.4 SUMMARY</td>
<td>19</td>
</tr>
<tr>
<td>3.0 DATA</td>
<td>20</td>
</tr>
<tr>
<td>3.1 PHYSICAL DATA</td>
<td>20</td>
</tr>
<tr>
<td>3.1.1 HIMS FILES</td>
<td>20</td>
</tr>
<tr>
<td>3.1.2 ETR RECORDS AND CONTRACT DRAWINGS</td>
<td>22</td>
</tr>
<tr>
<td>3.2 TRAFFIC INFORMATION</td>
<td>23</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 1. Basic Statistics of Datasets ................................................................. 25
Table 2. Parameter Estimates for Horizontal Curve Model ............................ 44
Table 3. Parameter Estimates for Tangent Model ........................................ 48
Table 4. Parameter Estimates for Speed Related Accident Model ............... 50
Table 5. Top 20 Hazardous Curves ................................................................. 55
Table 6. Comparison of Ranking Methods for Horizontal Curves: Total Accidents of All Severities ................................................................. 58
Table 7. Comparison of Ranking Methods for Horizontal Curves: Treatable Accidents Of All Severities ................................................................. 59
Table 8. Top 20 Sections Identified for Speed Control ................................ 62
Table 9. Comparison of Ranking Methods for Sections for Speed Control ...... 65
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1.</td>
<td>Maximum Likelihood Procedure</td>
<td>35</td>
</tr>
<tr>
<td>Figure 2.</td>
<td>Horizontal Curve Model: Relationship With AADT</td>
<td>45</td>
</tr>
<tr>
<td>Figure 3.</td>
<td>Horizontal Curve Model: Relationship With Radius</td>
<td>47</td>
</tr>
<tr>
<td>Figure 4.</td>
<td>Total Speed Accidents: Relationship With AADT And Surface Width</td>
<td>51</td>
</tr>
<tr>
<td>Figure 5.</td>
<td>Fifth Highest Rated Hazardous Curve</td>
<td>56</td>
</tr>
<tr>
<td><strong>GLOSSARY OF NOTATION</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>AADT</strong></td>
<td>Average Annual Daily Traffic</td>
<td></td>
</tr>
<tr>
<td><strong>ADS</strong></td>
<td>Accident Data System</td>
<td></td>
</tr>
<tr>
<td><strong>EB</strong></td>
<td>Empirical Bayesian</td>
<td></td>
</tr>
<tr>
<td><strong>ETR</strong></td>
<td>Engineering Title Record</td>
<td></td>
</tr>
<tr>
<td><strong>FHWA</strong></td>
<td>Federal Highway Administration</td>
<td></td>
</tr>
<tr>
<td><strong>HIMS</strong></td>
<td>Highway Inventory Management System</td>
<td></td>
</tr>
<tr>
<td><strong>LHRS</strong></td>
<td>Linear Highway Referencing System</td>
<td></td>
</tr>
<tr>
<td><strong>MTO</strong></td>
<td>Ministry of Transportation, Ontario</td>
<td></td>
</tr>
<tr>
<td><strong>MUTCD</strong></td>
<td>Manual of Uniform Traffic Control Devices</td>
<td></td>
</tr>
<tr>
<td><strong>PDO</strong></td>
<td>Property Damage Only</td>
<td></td>
</tr>
<tr>
<td><strong>Injury Accident</strong></td>
<td>A fatal or non-fatal injury accident</td>
<td></td>
</tr>
<tr>
<td><strong>RTM</strong></td>
<td>Regression to the Mean</td>
<td></td>
</tr>
<tr>
<td><strong>k</strong></td>
<td>A parameter describing relationship between $E{m}$ and $\text{Var}{m}$</td>
<td></td>
</tr>
<tr>
<td><strong>$\lambda_c$</strong></td>
<td>Regression estimate of annual accident occurrence on horizontal curves</td>
<td></td>
</tr>
<tr>
<td><strong>$\lambda'_c$</strong></td>
<td>Refined EB estimate of annual accident occurrence on horizontal curves</td>
<td></td>
</tr>
<tr>
<td><strong>$\lambda_t$</strong></td>
<td>Regression estimate of annual accident occurrence on a tangent</td>
<td></td>
</tr>
<tr>
<td><strong>$\lambda'_t$</strong></td>
<td>Refined EB estimate of annual accident occurrence on a tangent</td>
<td></td>
</tr>
</tbody>
</table>
\( \lambda_s \) Regression estimate of annual speed related accident occurrence

\( \lambda_s' \) Refined EB estimate of annual speed related accident occurrence

\( E\{m\} \) Estimate of the expected annual accident frequency at an average site

\( E\{m/x\} \) EB estimate of the expected annual accident frequency at a specific site

\( \text{Var}\{m\} \) Variance of the expected accident frequency at an average site

\( \text{Var}\{m/x\} \) Variance of the refined estimate of accident frequency at a specific site

\( x \) recorded number of accidents

\( p(x) \) Probability that \( x \) number of accidents will occur at a site

\( n \) The ratio of the number of years of accident data used to the number of years the EB estimate is for

\( w_1, w_2 \) Weights used in the EB framework

\( \text{s.e.} \) Standard Error

\( E\{\theta\} \) Estimate of the safety effect of treatments
CHAPTER 1
INTRODUCTION

Motor vehicle accidents contribute a large loss to society. Between 1988 to 1993, in the province of Ontario, there were an average of 227 062 accidents per annum, and an average of 45 353 accidents occurred on rural two-lane roads per annum. The magnitude of this loss to society is evident when considering dollar amounts. In 1994, in Ontario the cost of each death was estimated at $831 429, each injury at $20 084 and each PDO (property damage only) accident at $6 136.

Those responsible for safety on the road system must be able to identify sites which hold an unusually high rate of accident occurrence, and to rationally apply limited budgets to safety improvements at the sites which will likely give the largest safety gain for the dollars spent.

The fundamental step in improving safety is determining which sites are truly hazardous. New regression modeling techniques (e.g. Empirical Bayesian) incorporate geometric, traffic and accident history and can thus be used to determine how “safe” a site is by how many accidents are expected to occur. EB estimates may be used to rank sites according to the expected number of treatable accidents and to identify the sites which are likely to respond favourably to engineering treatments. Treatable accidents are those which may be reduced or eliminated through some form of treatment. A study by Elvik (8), on the
effects of automatic speed enforcement in Norway found that by using the EB technique to correct for regression to the mean (RTM), sites which met accident based warrants experienced a much larger decrease in accidents following treatment than those sites which did not. EB estimates are also useful in before-after studies to evaluate the safety effect of treatments (19).

The object of this study is to develop guidelines for identifying hazardous sites that warrant specific treatments. Two types of accidents are investigated. Those that are occurring due to the presence of a horizontal curve are investigated for the possibility of providing some form of curve warning treatment. Accidents that are speed related are investigated for the application of speed control measures. Intersection related accidents are not dealt with in this study.

These guidelines were developed using data from Ontario provincial 2-lane rural roads to calibrate accident prediction models. The EB method is used to develop the methodology for ranking sites by the potential for safety improvement.

This thesis is divided into 7 chapters. Following the introduction chapter 2 summarizes the literature review conducted on the identification of hazardous highway curves and sites having speed related accidents. The chapter also reviews literature on accident prediction models and the EB method. Chapter 3 describes the data that was collected and used in the study. Chapter 4 gives an overview of the two methodologies developed for
identifying hazardous curves and sections warranting speed control measures. Chapter 5 discusses the calibrated models. Chapter 6 reports on the validation of the methodologies developed. Lastly, chapter 7 provides a summary of the study.
An extensive literature review was conducted on horizontal curve and speed related accidents. In this chapter, a summary is provided along with the basis of the modeling theory.

2.1 HORIZONTAL CURVE WARNINGS

2.1.1 ACCIDENTS ON HORIZONTAL CURVES

Researchers such as Zeeger et al (2) have found that on rural roads accidents occur more frequently on the horizontal curve sections than on the tangent sections. In addition, fatal and injury accidents have been found to have a higher percentage on curves than on tangents. Studies cited for an FHWA informational guide on safety improvements for rural roads have quantified this higher accident rate from 1.5 to 4 times that found on similar tangent sections (3). In addition, horizontal curves were found to have particularly high rates of run-off road and head-on collisions, resulting in higher fatality and injury rates than on tangents. Retting and Farmer (1), in their study on the use of pavement markings to reduce speeds on curves, cited the National Highway Traffic Safety Administration, reporting that 80 percent of all fatal crashes in the United States occurred on two-lane rural roads, and of these 40 percent of fatal roadside crashes occurred on curves. A study by Glennon et al. (4) found that 41.5 per cent of accidents on horizontal curves resulted in an injury or fatality. Given that tangent sections make up a larger
fraction of the road system, the increased risk of fatal and injury crashes on curved sections is apparent.

Given that curves have a higher accident potential than tangents, and that this potential decreases with a decrease in operating speeds, Retting and Farmer (1) concluded that low cost curve warning treatments such as signing, marking and delineation have the potential for large savings if applied to hazardous curves that warrant treatment from a benefit-cost analysis. For rural two-lane roads, curve-warning treatments are particularly desirable. Geometric improvements such as curve flattening are extremely costly and to be economically efficient would require a very large reduction in accidents to be associated with the improvement. On rural two-lane roads which typically have small volumes however, it would be odd to expect a large enough reduction in accidents to warrant such a measure. However, the low cost of curve-warning treatments could be justified by a relatively low expected decrease in accidents.

Limited budgets require a rational decision-making framework for allocating funds for site improvements. All sites with small radii curves and high volumes for example are not eligible for treatment. There must be a proven hazard as reflected by the expected number of accidents and a large enough expected reduction in accidents to warrant the costs of curve warning treatment.
2.1.2 TREATMENTS OF HAZARDOUS CURVES

Aside from the geometric re-design of roads, the use of curve-warning signs, perhaps, is the primary method of improving safety on curves. Curve warning signs are intended to advise drivers of a change in the horizontal alignment which will require an adjustment of driving behaviour, most likely a decrease in speed. Warning devices are particularly targeted to drivers unfamiliar with the road. Drivers experienced in a particular road section will drive according to their knowledge of the curve and choose a safe speed. However drivers who are unfamiliar with this particular curve may only rely on their knowledge of other curves and enter the curve at an unsafe speed. If curve warning signs, with or without speed advisory plates, are successful in raising the awareness of the unfamiliar driver so that they drive at a safe speed, then they have served their purpose in promoting safety. However, as stated in a report to the FHWA (5), the large number of curve warning and speed advisory signs has lessened their impact on driver behaviour. This same study further states that curve warning signs seem to be in place wherever horizontal alignment changes, and that speeds posted on advisory signs are much lower than operating speeds. In effect, this results in drivers placing little confidence in the advice the signs are meant to convey. Retting and Farmer\(^1\) provide in the following statement an excellent description of the danger of inconsistent practice in selecting sites for curve warning treatments, “... if several moderately sharp curves are preceded by curve warning signs, it is reasonable for drivers to expect that any additional curve warning signs they encounter downstream would be followed by similar curves. If an

---

unusually sharp downstream curve is not differentiated by some special warning, driver expectancy is violated, and drivers may not slow sufficiently”.

Despite the need for a consistent scientific based warrant for the installation of curve warning treatments, there is little guidance. The Uniform Traffic Control Devices for Canada manual (25) recommends curve warning signs wherever the safe curve speed is 10 km/h less than the posted speed limit. The MUTCD warrant for a curve warning sign is: “where engineering investigations of roadway, geometric and operating conditions show the recommended speed on the curve to be in the range between 30 and 60 mph (48 and 96 km/h respectively) and equal to or less than the speed limit established by law or by regulations for that section of highway, and advisory speed plates when additional protection is desired”. This vague description does not provide for the consistent implementation of curve warning treatments. A report for the FHWA (5) recognized this and called for a review of the criteria for deploying curve warning measures.

An attempt to develop warrants for curve warnings on low volume rural roads was undertaken by Stockton et al. (6). In this work, the required deceleration distance was calculated from the operating speed on the approach and the safe computed curve speed. Sites which did not permit for the required perception-reaction-deceleration distance were labeled as possible hazardous curves. It was stated further that curve signs are warranted in advance of all curves which have intersecting angles of 45 degrees or more on paved roads, and 60 degrees on unpaved roads, unless the speed limit is less than or equal to 55
kph, or the combination of normal approach speed and safe curve speed requires a perception-reaction-deceleration distance of less than 90 metres. Advisory speed plates are reported to be warranted with curve warning signs when the safe curve speed is 8 kph below the maximum speed warranting a curve sign. These warrants were based on engineering judgement and the analysis conducted for the study.

A critical review on criteria for setting advisory speeds on curves was undertaken by Chowdhury (7). In this research, it was found that the two most common methods of setting advisory speeds, the ball bank and nomograph method, were outdated even by their own criteria. The criteria are based on tests done over 50 years ago and are no longer reflective of vehicle performance and attainable friction factors between the pavement and vehicle tires. In fact it was found that 46% of the curves studied were posted at a lower speed than would be suggested by either method using present day conditions.

A major fault of the previous criteria is the omission of a site’s accident history. If a site is truly hazardous, that will be reflected in an abnormally high accident frequency for its geometric and traffic conditions. By using methods that do not include the consideration of accident history, the actual safety performance of sites are ignored. This may lead to incorrectly identifying non-problem sites as requiring curve warning treatments and problem sites as not requiring curve warning treatments. As mentioned earlier, Elvik (8) showed that sites which met crash-based warrants showed a much higher reduction in accidents than those sites which did not meet the crash-based criteria. It is important to
correctly identify the sites that warrant treatment to obtain the maximum safety benefits for the cost of implementation and maintenance of treatment.

Zwahlen (9), in a study on curve warning systems, developed a methodology for the design of curve warning systems on rural two-lane highways based on curve geometry, traffic conditions, accident history, expected accident severity and human factors. The methodology determines the appropriate curve warning treatment, if any, and provides for a consistent method of evaluating curve warning needs at horizontal curves. The drawback of this procedure is that it is very data intensive and therefore impractical for many jurisdictions to apply to all curves. However, this shows promise for the investigation of specific curve sites identified from a less data intensive methodology.

This study aims to provide and demonstrate a methodology for identifying curves for further investigation utilizing the Empirical Bayesian framework.

2.2 SPEED CONTROL MEASURES

2.2.1 SPEED RELATED ACCIDENTS

Speed-related accidents may be defined as accidents which occurred when one or more of the involved vehicles was either traveling too fast for the prevailing conditions, or travelling over the posted speed limit. These accidents are of a particular concern due to the increased accident severity associated with impacts at higher speeds. Speed-related accidents may occur for numerous reasons such as weather conditions, the speed of other
vehicles in relation to the vehicle involved in the accident, passing sight distance or other aspects of horizontal and vertical geometry.

One cause of speed accidents may lie in how roads are designed. For a given section of road the safe speed of travel will change between design elements. For example, the safe speed of travel will differ between a tangent section and a curve section. Speed limits, however, are set by the lowest maximum speed desired for each element in the section. Drivers naturally will recognize that it is safe to drive at a higher speed on the less hazardous sections of the road and consequently may be traveling at too high a speed when entering a section requiring a lower speed, not expecting the required speed decrease.

It has been suggested that accident rates correlate more strongly with the speed variance of traffic than with the mean speed (11). The theory is that most drivers will make an appropriate speed choice for conditions, while only a small fraction will choose to travel at a dangerous speed (10). This is in effect self-enforcement by drivers. Where drivers are choosing speeds very close to each other and the speed variance is low it may be assumed that most vehicles are traveling at a safe speed. Where the speed variance is high, there will be drivers choosing unsafe speeds.

Published research however has failed to separate speed-related accidents, as reported by police agencies or other reporting body, and all accidents that occurred. However, when
examining sites for implementing speed control measures, it is imperative that sites with a speed-related safety problem are identified and not those sites which have a different safety problem.

As with other safety treatments, the effectiveness of speed control measures depends on them being applied to road sections which will potentially have a reduction in speed related accidents. What is required is a methodology that identifies candidate sites for speed related treatments based upon a high number of treatable speed related accidents.

2.2.2 TREATMENTS OF SECTIONS WITH A SPEEDING PROBLEM

Methods of controlling speeds in an attempt to reduce accidents typically include lowering of the speed limit creating speed zones, and/or increased enforcement. These zones are usually identified through public concerns with speeding, police concerns over speeding or a high number of recent accidents in the zone.

A report for the Transportation Association of Canada (TAC) (10) lists the most common factors in the setting of speed limits as 1) 85th percentile speed, 2) design speed, 3) pace speed, 4) legislated limits and 6) accident rates. The MUTCD (21) lists, in addition to the above, roadside development, pavement friction factors and the safe speed for curves and other hazardous sections within the zone.

Relative to the identification of zones for speed control measures are the 85th percentile
speed, pace speed and accident rates applicable for a road in service. The 85th percentile speed is the speed at which 85 percent of vehicles are traveling at or below. The pace speed is the 10 mph (16.09 kph) range which contains the largest percentage of vehicles. The disadvantage in solely using these measures are that although speeds may be over the speed limit or in a desired range there may not in fact be a history of accident occurrence. The use of accident rates (accidents/km.) may result in selecting short zones with a randomly high accident count when no treatable speed safety problem exists. The use of a modified accident rate (accidents/vehicle*distance), is also a disadvantage in that zones with low volumes may have a high accident rate while a zone with high volumes may have a low accident rate despite experiencing more treatable accidents. Selecting the zone with the high accident rate in this situation results in applying treatment to the zone where a smaller reduction in accidents will occur and thus fewer benefits.

2.3 THEORETICAL FOUNDATIONS

Accidents are random in nature and as such must be modeled using statistical techniques to predict accident occurrence. The typical model form is shown in equation 1:

\[ y = f(b) \] ... (1)

where,

- \( y \) is the dependent variable (accidents per unit of time)
- \( b \) is a group of explanatory variables
Which explanatory variables are included in the model is determined in the modeling process. This, unfortunately, is difficult since not only does the inclusion of a variable depend on its true relationship with the dependent variable, but also on whether the dataset is adequately large and varied for this relationship to be discovered and included.

2.3.1 BACKGROUND

This study requires two types of crash estimation models for two-lane rural roads. The first is for the investigation of hazardous horizontal curves, which requires as part of the explanatory variables information related to horizontal curve geometrics. The second is for the investigation into speed related accidents, which may use models for which alignment data is unknown. This will be further discussed in chapter 3.

Fink and Krammes (12) developed a model to predict accidents related to the curve. The dependent variable was the mean accident rate for horizontal curves on rural two-lane highways with a grade less than 5 percent and AADT between 400 and 3500 vehicles per day using a database of 563 curves. A linear regression equation was developed relating total accident rate (million vehicle miles) to the degree of curvature. The effects of approach tangent length and sight distance were not found to be statistically significant but it did appear to the authors that the adverse safety affect of long tangents and short sight distance becomes more pronounced on sharper curves. The inability to come to statistically significant conclusions was likely due to the small database. In a five year period only 235 accidents were recorded, of which 106 resulted in a fatality or injury.
Zeeger et al (13) used a database of 10,900 curves to model total accidents on horizontal curves and develop accident reduction factors for geometric improvements. Predictor variables found to be statistically significant contributors to accident experience included the length of curve, volume of vehicles passing through the curve, degree of curve, roadway width and the presence of spiral transitions. The model form predicted accidents to increase as degree of curve and curve length increases, while decreasing for an increase in roadwidth and presence of a spiral transition.

The previously mentioned studies make the assumption that accidents increase linearly with AADT. This however has been shown to be an erroneous assumption in studies by Mountain et al (14), Hauer (15) and other researchers.

Vogt and Bared (16) developed models for non-intersection accidents which allowed for a non-linear accident relationship with AADT and also included vertical geometric information. Models were calibrated for total accidents on two-lane rural road sections, with curve and tangent alignments included in each section. Statistically significant variables which were included in at least one of the models included AADT, total roadwidth, roadside hazard rating, degree of horizontal curve, crest curve grade rate, absolute grade, driveway density, lane width and the state in which the road existed (Washington or Minnesota). Models for injury only accidents were developed but were found to be less significant.
Models for which no alignment data was available have been developed by many researchers. Hauer (15) developed non-linear models for total non-intersection accidents with the AADT as the only independent variable. Persaud et al. (17) constructed models for severe (fatal and injury) and property damage only (pdo) non-intersection accidents using AADT, lane width and shoulder width as independent variables. Dai et al. (18) developed a model for total non-intersection accidents. Independent variables included AADT, lane width, shoulder width and a roadside hazard rating as independent variables. Although these models do not relate accident experience to horizontal or vertical alignment they are still useful as the purpose is to develop predictive rather than causal models.

Persaud et al. (17) have proposed that sites should be ranked by the difference between the expected accident rate and what is expected for a site with no safety problem. In effect this is a ranking by the potential safety improvement. This does not imply that a site with no safety problem will have zero accident occurrence but in a practical sense is taken to mean that there is a base level of safety which most roads will conform to. The difference between the expected accident frequency and what is normal for a safe site defines the expected number of treatable accidents.

The disaggregate models produced by researchers for intersection accidents by crash types (e.g. rear end, right angle) have not been equaled for two-lane rural roads. To develop
such statistically significant and therefore defendable disaggregate models the data must contain many more accident occurrences than are required to develop models for aggregate data. Disaggregate models for rural two-lane highways have likely been discouraged for this reason.

2.3.2 ESTIMATES OF E\{m/x\} and VAR\{m/x\}

The left hand side of the equation $y = f(b)$ is the estimate of the mean accident rate $m$ for a reference population of homogeneous sites with the same geometric and AADT characteristics that are included in the model. The mean accident rate, $m$, of sites within a reference population does in fact vary due to factors such as weather, the socio-demographics of drivers and countless other factors which could not be included in the model. This distribution of mean accident frequency within the reference population has been postulated to be adequately represented by the gamma probability distribution model (19).

Where the mean accident frequency of a reference population is described by the gamma distribution, and the accident count at a site is described by the Poisson distribution with a known mean, the accident counts within a reference population may be described by the negative binomial error distribution (19).

To obtain an estimate of the mean accident frequency of a specific site, the accident history of the site is considered as well. This is the essence of the Empirical Bayesian
process. Further discussion will be given in chapter 4 and a comprehensive derivation may be found in a text by Hauer (19).

If the estimates of $E\{m\}$, $\text{Var}\{m\}$ and the accident count $x$ is known then refined estimates of $E\{m/x\}$ and $\text{Var}\{m/x\}$ may be easily obtained using equations 2 and 3.

$$E\{m/x,n\} = w_1 x + w_2 E\{m\} \quad \ldots \ (2)$$

where,

$x$ is the number of accidents recorded in the period for which data is being used in the equation

$n$ is the ratio of the time period over which $x$ pertains to the time period that the estimate $E\{m/x,n\}$ is for

$k$ is a parameter estimated in the modeling procedure, describing the relationship between the variance and mean of the estimate

$$\text{Var}\{m/x\} = \frac{(x + k) \cdot E\{m\}^2}{(k + n \cdot E\{m\})^2} \quad \ldots \ (3)$$

where,

$$w_1 = E\{m\}/(k + nE\{m\}) \quad \ldots \ (4)$$

$$w_2 = k/(k + nE\{m\}) \quad \ldots \ (5)$$
For a negative binomial (NB) error structure the relationship between $E\{m\}$ and $\text{Var}\{m\}$ is:

$$k = \frac{E\{m\}^2}{\text{Var}\{m\}}$$  \quad \ldots (6)

### 2.3.3 SELECTION OF MODEL VARIABLES

The quality of the estimates of $E\{m/x\}$ and $\text{Var}\{m/x\}$ relies on the quality of the accident prediction model. The goal in developing this model is to include the roadway, traffic or other variables that explain the different accident experiences of sites. McCullagh and Nelder (20) have pointed out that a model which fits the data well is not necessarily a good model. For example using correlated variables may give counterintuitive results and make the model unstable. They also postulate that by including limitless variables the model may fit the data perfectly, but not explain anything. Therefore there needs to be a systematic process of determining whether or not to include a variable in the model. This may be done through examining the deviance of the model fit and the t-statistics for parameter estimates for each variable.

In keeping with these principles, variables that seemed a priori to belong in the models were collected, and those which met the criteria of improving the model fit and being significant using t-statistics were used.
2.4 SUMMARY

Several insights have been obtained from the literature review.

- Horizontal curves on 2-lane rural roads have a higher accident rate, estimated at 1.5 to 4 times that of similar tangent sections.

- Hazardous curves have the potential for large safety benefits through a decrease in accidents resulting from relatively low cost curve-warning treatments.

- Accident rates increase with increasing speed variance as opposed to simply an increase in mean speed.

- Published models for analysing the effect of speed on safety have failed to separate speed related accidents from total accidents.

- There is no consistently applied methodology for identifying either hazardous curves or sections warranting speed control measures.

- Vehicular accidents are random in nature and therefore must be modeled statistically. The EB approach is a superior method for doing this.

- It has been shown that when countermeasures are implemented the safety improvement is much more effective on sites which met an accident based method than on sites which did not meet these warrants.

- The EB estimates may be used in the evaluation of safety improvements.
CHAPTER 3

DATA

The data used for this study came from the Ministry of Transportation of Ontario (MTO). There were four sources of data. Physical data on road sections was obtained from Highway Inventory Management System (HIMS) electronic files, Engineering Title Records (ETR) and contract drawings. Traffic files are available electronically for the same road sections in the HIMS file. The Accident Data System (ADS) contained extensive records for each accident reported on MTO roads for the years 1988 to 1993.

To create the required databases, suitable reference populations were created using the HIMS, ETR records and contract drawings for horizontal curves. For speed accidents, reference populations were created using only HIMS data. Traffic and accident data were subsequently matched to the sites chosen for the reference populations.

3.1 PHYSICAL DATA

The physical data was obtained from electronic files (HIMS), ETR records and contract drawings. A description of each source and the data collected is given in the following sections.

3.1.1 HIMS FILES

The Highway Information Management System (HIMS) divides all road sections under the jurisdiction of the MTO into sections using a Linear Highway Reference System.
(LHRS). In the LHRS system, roads are divided into sections which typically begin and end at an intersection or administrative boundary. Each section is identified by an LHRS reference number and intermediate points within the section may be identified by an offset distance from the beginning point of the section.

The HIMS files contain information regarding cross-sectional elements and the classification of the road section. The information collected for this study is listed and described where necessary below:

- LHRS number- this is the unique identification number of each road section and is used to link all data
- offset- this along with the LHRS number describes the beginning of each section. In some cases there are multiple sections with the identical LHRS number.
- distance- this is the length of the section
- number of lanes
- passing sight distance- this is the percentage of the section which is marked for passing
- terrain- terrain may be level, rolling or mountainous
- lane width
- surface type- example: gravel, concrete
- shoulder width
• shoulder type- example: gravel, paved or partially paved
• environment- urban, semi-urban or rural

In order to develop models for speed related accidents, the physical information contained in the HIMS files was adequate. To develop models to evaluate hazardous curves it was necessary to obtain alignment information. This information was found in ETR records and contract drawings.

3.1.2 ETR RECORDS AND CONTRACT DRAWINGS

Horizontal and vertical alignment data cannot be found in electronic files. To obtain this data it was necessary to examine paper files of contract drawings and ETR records. Unfortunately, it is not possible to obtain this information for all road sections under MTO control. Therefore the data, obtained for modeling hazardous curves was limited to what was available and by time constraints. The locations identified were matched to LHRS numbers and offsets. Data collected from these drawings for curves is listed below:

• radius
• length of curve
• presence of spiral
• length of spiral
• distance to previous curve
• distance to next curve
• superelevation
• presence of vertical curve

• grades

3.2 TRAFFIC INFORMATION

Traffic flow data may be obtained from electronic files which match the LHRS sections in the HIMS files. For this study the only information used is the Average Annual Daily Traffic (AADT).

3.3 ACCIDENT DATA SYSTEM

The accident data are obtained from 4 separate electronic files. The first file contains basic information about the location and type of each accident recorded. The second contains information about each driver and vehicle involved in the collision. The third file contains involved person information. The final file contains information on each fatality. For this study, only the first and second files were needed.

3.3.1 BASIC INFORMATION

The required fields for this study are listed and briefly described below:

• Microfilm number. This is a unique reference number for each accident which enables matching with the other ADS files.

• LHRS. The LHRS number of the section that the accident occurred in.

• Offset. The position of the accident within the LHRS section.
- Classification. Accidents are classed as fatal, non-fatal injury, property damage only, non-reportable or other. For this study fatal and non-fatal injury accidents are combined to model as severe injury accidents.

- Accident location. This field lists the location type where the accident occurred such as intersection, at railway crossing or parking lot. This study excludes intersection and intersection related accidents on the two-lane rural roads chosen.

This data is adequate for developing models for identifying hazardous curves. However, information regarding the cause of the accident is needed to separate speed related accidents. This information may be obtained from the driver and vehicle information file.

**3.3.2 DRIVER AND VEHICLE INFORMATION**

To identify speed-related accidents, the accidents from the basic information file were matched to the driver and vehicle information file. The required fields from this file are described below:

- Microfilm number. This number is used to match the driver and vehicle information to the accidents in the basic information file.

- Driver action. The driver action is reported by the investigating police officer. It may be recorded as one of: unknown, driving properly, following too close, speed exceed limit, speed too fast for condition, speed too slow, improper turn or disobey traffic controls.
Speed related accidents were identified as those that are reported as speed exceed limit or speed too fast for condition.

### 3.4 BASIC STATISTICS OF DATASETS

Some basic statistics on the three datasets assembled are given in Table 1. As discussed in chapter 4, the identification of curve related accidents requires models for horizontal curves and for tangents. The identification of sites for speed control requires models for speed related accidents.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Horizontal curves</th>
<th>Tangents</th>
<th>Sites for speed control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of sites</td>
<td>585</td>
<td>551</td>
<td>2,912</td>
</tr>
<tr>
<td>Total length (km.)</td>
<td>143.79</td>
<td>206</td>
<td>18,692.4</td>
</tr>
<tr>
<td>AADT range</td>
<td>634 – 15,817</td>
<td>100 – 21,633</td>
<td>42 – 31,075</td>
</tr>
<tr>
<td>Average AADT</td>
<td>3,800</td>
<td>3,412</td>
<td>3,256</td>
</tr>
<tr>
<td>1988-1993 accidents of all severities</td>
<td>1,070</td>
<td>10,301</td>
<td>16,218</td>
</tr>
<tr>
<td>1988-1993 injury accidents</td>
<td>349</td>
<td>3,359</td>
<td>5,457</td>
</tr>
</tbody>
</table>

**TABLE 1: BASIC STATISTICS OF DATASETS**
3.5 **SUMMARY**

The data used for this study came from the Ministry of Transportation of Ontario (MTO). Accident, traffic and some geometric data was attainable in electronic files. Alignment data was collected from Engineering Title Records and contract drawings where available for use in the modeling of hazardous horizontal curves. All information was linked using the Linear Highway Referencing System.
CHAPTER 4
OVERVIEW OF METHODOLOGY

This chapter elaborates on Empirical Bayesian estimates and maps out the procedures for identifying hazardous highway curves and sites warranting speed control measures on two-lane rural highways.

4.1 EMPIRICAL BAYESIAN (EB) ESTIMATES

EB estimates are used as a predictor of accident occurrence. Two sources of information are utilized to estimate the accident frequency of a site. Firstly, the physical and traffic information is used to develop a multivariate statistical model estimate. This estimate is then further refined using the accident history of the site. By using these two sources of information, EB estimates improve the accuracy of predictions and compensate for regression to the mean (RTM) bias (19).

4.1.1 REGRESSION EQUATIONS

Recall equation (1) introduced in chapter 2, \( y = f(b) \). The left hand side of equation (1) is the mean accident rate \( y \) for a reference population of similar sites. The task of the regression equations is to provide estimates for the equation. \( E\{m\} \) is the estimate of \( y \), the expected number of accidents per year. \( \text{Var}\{m\} \) is the variance of this estimate. Fundamentally, regression estimates of accident occurrence are an average accident frequency for a reference population with identical characteristics. These estimates are
adequate for examining general trends in data, for example if the accident frequency is lower for highway sections with wider lanes. However for the purpose of identifying which road sections have a safety problem they are inadequate. The EB approach overcomes this limitation. By using two sources of information, the regression estimates and the actual accident history of sites, the estimates of accident occurrence are now site specific and of improved accuracy. A side benefit is that these same EB estimates may be used in the evaluation of safety improvements, as what would have been expected had no safety improvement been implemented. The evaluation of safety improvements will be discussed in chapter 7.

4.1.1.1 GENSTAT
Genstat is a statistical analysis software package that may be used for non-linear regression modeling. Genstat also allows for the specification of a negative binomial error distribution which researchers are now acknowledging as appropriate for accident modeling (19). Genstat was used to estimate the parameters for the accident prediction models. By using a negative binomial error distribution the modeling procedure must be iterative. This process will be described in section 4.1.5.

4.1.2 ACCIDENT HISTORY
The accident history is used in the EB procedure to refine the estimate obtained from the regression equation. Accident counts are obtained for the desired accident type.
4.1.3 EMPIRICAL BAYESIAN EQUATIONS

The EB equations were introduced in section 2.3.2, they were:

\[
E\{m/x\} = w_1 \cdot x + w_2 \cdot E\{m\} \quad \text{... (2)}
\]

\[
\text{Var}\{m/x\} = \frac{(x + k) \cdot E\{m\}^2}{(k + n \cdot E\{m\})^2} \quad \text{... (3)}
\]

where,

\[
w_1 = E\{m\}/(k + n \cdot E\{m\}) \quad \text{... (4)}
\]

\[
w_2 = k/(k + n \cdot E\{m\}) \quad \text{... (5)}
\]

\[
k = E\{m\}^2/\text{Var}\{m\} \quad \text{... (6)}
\]

and where,

\[
E\{m\} = \text{the expected accidents frequency for an average site in a reference population}
\]

\[
\text{Var}\{m\} = \text{the variance of the accident frequency to occur in the reference population}
\]

\[
x = \text{the observed number of accidents}
\]

\[
n = \text{the ratio of the time period for which accident data is available over the time period for which the estimates are desired.}
\]

\[
k = \text{the parameter which describes the relationship between } E\{m\} \text{ and } \text{Var}\{m\} \text{ and is estimated in the modeling procedure}
\]
\[ \text{E}(m/x) = \text{the refined estimate of the expected number of accidents to occur at a site for a given time period} \]
\[ \text{Var}(m/x) = \text{the variance of the refined estimate of the expected number of accidents to occur at a site for a given time period} \]
\[ w_1, w_2 = \text{weighting factors} \]

By examining the series of equations presented, a few points become apparent. The variance of the estimates, \( \text{Var}(m/x) \), decreases as the number of years of data used increases and as the value of \( k \) increases. Fundamentally this makes sense. As more years of data are used then the more information is known about each site and the estimate of each \( \text{E}(m/x) \) will be better. As the differences between sites are better explained in the model then the variance of estimates, \( \text{Var}(m/x) \), should decrease. As the value of \( k \) increases, \( \text{Var}(m) \) from the regression model estimate decreases. Therefore, as \( k \) increases, the regression model is explaining more of the variation in accident experience across sites and the estimate of \( \text{E}(m/x) \) will also improve.

The effect of \( k \) on the weighting of \( \text{E}(m) \) and \( x \) is also evident. As \( k \) increases the value of \( w_1 \), the weight assigned to the accident count \( x \), decreases and the value of \( w_2 \), the weight assigned to the estimate of \( \text{E}(m) \) increases. Simply put, as the accident prediction model improves and \( k \) increases, more of the accident experience is being accounted for through the model, and therefore the refined estimate \( \text{E}(m/x) \) may rely less upon the accident history of the individual sites to estimate their expected accident frequency.
The relationship between \( k \), \( \text{Var}\{m\} \) and \( \text{E}\{m\} \) is discussed in the next section.

4.1.4 ERROR DISTRIBUTIONS

In the statistical modeling of accidents, it is now commonplace to adapt the negative binomial error distribution. The negative binomial error distribution is derived from the Poisson and gamma distributions. Each of these and their relations to one another are described in the next few sections. A comprehensive discussion of the use of the Poisson, gamma and negative binomial distributions may be found in a text by Hauer (19).

4.1.4.1 POISSON DISTRIBUTION

The Poisson error distribution is an adequate model to describe accident occurrences at a site when the mean is known as it is both a discrete and a non-negative model. The Poisson model is expressed as:

\[
p(x/m,n) = \frac{e^{-mn}(mn)^x}{x!} \quad \ldots (7)
\]

where,

\( p(x/m,n) = \text{the probability of } x \text{ accidents occurring at a site in } n \text{ years} \)

\( x = \text{the number of accident counts per } n \text{ years} \)

\( m = \text{the expected accident frequency at a site per } n \text{ years} \)
For a one year period equation (7) may be rewritten as:

$$p(x/m) = \frac{e^{-m}(m)^x}{x!}$$  \hspace{1cm} ... (8)

The variance of the Poisson distribution is equal to its mean.

$$E(x) = Var(x)$$

As mentioned, the Poisson distribution is useful when the true mean accident frequency is known. However the estimate of $E(m)$ is not the estimate of an individual site's expected accident frequency, but rather an estimate of the expected accident frequency of a reference population with the same physical and traffic characteristics. Despite controlling for all possible factors in a reference population, there are still differences unaccounted for such as weather and the socio-demographics of the driving population. Therefore, the expected accident frequencies of individual sites within a homogeneous reference population vary. Hauer (19) describes this variation as being adequately described by the gamma distribution, to be described in the next subsection.

**4.1.4.2 GAMMA DISTRIBUTION**

The gamma error distribution is described below:

$$f(m) = a^m b^{-1} e^{-am}/\Gamma(b), \hspace{1cm} for \ m>0 \ and \ 0 \ otherwise \hspace{2cm} ... (9)$$
where,
\[ a = \frac{E(m)}{\text{Var}(m)} \]  \hspace{1cm} \ldots (10)
\[ b = \frac{E(m)^2}{\text{Var}(m)} \]  \hspace{1cm} \ldots (11)

When the accident frequency at a site may be represented by the Poisson distribution with a known \( E(m) \) and the distribution of \( m \)'s around the estimate \( E(m) \) is described by the gamma distribution, the expected number of accidents and the variance of this estimate at a site is given by:

\[ E(x) = E(m) \]  \hspace{1cm} \ldots (12)
\[ \text{Var}(x) = E(m) + \text{Var}(m) \]  \hspace{1cm} \ldots (13)
\[ \text{Var}(m) = \frac{E(m)^2}{k} \]  \hspace{1cm} \ldots (6)

where,
\[ E(x) \] = mean accident counts of the reference population
\[ \text{Var}(x) \] = variance of the accident counts of the reference population
\[ \text{Var}(m) = \frac{E(m)^2}{k} \]

\( k \) = a parameter describing the relationship between \( \text{Var}(m) \) and \( E(m) \) estimated in the modeling procedure

In essence, equation (13) shows that the variance in expected accident frequency at a site is due to the variance of the individual site, accounted for by the variance from the Poisson error distribution, and equal to \( E(m) \), and also the variance that accounts for the
differences between sites in the reference population, and equal to \( \text{Var}(m) \).

### 4.1.4.3 NEGATIVE BINOMIAL ERROR DISTRIBUTION

The probability of a site experiencing \( x \) accidents in a year is equal to the product of the Poisson probability of \( x \) accidents given an \( E(m) \) and the gamma probability of the \( E(m) \) being equal to that value. Therefore the probability of a site experiencing \( x \) accidents is the integral of the Poisson probability of \( x \) accidents multiplied by the gamma probability as shown below:

\[
p(x) = \int_0^\infty \frac{e^{-m}m^x}{x!} \frac{\alpha^b m^{b-1}e^{-\alpha m}}{\Gamma(b)} \, dm
\]

\[
= \left( \frac{a^b}{\Gamma(b)} \right) \frac{1}{x!} \int_0^\infty e^{-m(a+1)}m^{x+b-1} \, dm
\]

\[
= \left( \frac{a^b}{\Gamma(b)} \right) \frac{1}{x!} \frac{\Gamma(x+b)}{(a+1)^{x+b}}
\]

\[
p(x) = \left( \frac{a}{a+1} \right)^b \left( \frac{b(b+1)\ldots(b+x-1)}{(a+1)^x \cdot x!} \right)
\]

for \( x = 0, 1, 2, \ldots \infty \)

where,

\[
a = E(m)/\text{Var}(m) \quad \ldots (10)
\]

\[
b = E(m)^2/\text{Var}(m) \quad \ldots (11)
\]
4.1.5 ESTIMATION OF OVERDISPERSION PARAMETER $k$

The overdispersion parameter $k$ describes the relationship between $\text{Var}(m)$ and $E(m)$ as shown below:

$$k = E(m)^2 / \text{Var}(m)$$ \hspace{1cm} \ldots (6)$$

The value of $k$ is determined in the modeling process through a maximum likelihood procedure. This procedure is outlined in Figure 1.

![Diagram](image.png)

**FIGURE 1: MAXIMUM LIKELIHOOD PROCEDURE**

A value of $k$ is assumed for input into the modeling procedure. The parameters are
estimated and the observed and fitted values are printed to an output file. A maximum likelihood procedure uses the observed and fitted values to determine a most likely value of $k$. This new value of $k$ is used as input into the modeling procedure in the subsequent iteration. This procedure is repeated until the input $k$, and most likely $k$, determined by the maximum likelihood program, are identical. This then is the proper value of $k$.

4.2 ESTIMATION OF TREATABLE ACCIDENTS

After developing models to predict accident occurrence, the next step is to refine the methodology to identify the treatable accidents. This section describes the methodology for identifying the two types of accidents investigated in this study, those occurring on horizontal curves and speed related accidents.

4.2.1 HORIZONTAL CURVES

The methodology for horizontal curves seeks to identify those curves which may benefit from some form of curve warning treatment, resulting in a reduction in expected accident frequency. The next step in developing the methodology is to identify the treatable accidents. In this case, these are those accidents which are occurring due to the presence of the curve. By comparing the EB estimate of expected accident frequency to what would be expected on a similar tangent section, an estimate of treatable accidents is obtained. This measure of treatable accidents may be used to rank sites by potential safety improvement. The procedure is illustrated below using hypothetical models.
CURVES

The expected accidents per year may be found as,
\[ \lambda_c = 0.0001 \text{(length)}^{0.6} \text{(AADT)}^{0.8} \]
\[ k = 2.0 \]

TANGENTS

The expected accidents per year may be found as,
\[ \lambda_t = 0.00005 \text{(length)} \text{(AADT)}^{0.9} \]
\[ k = 2.5 \]

step 1

Use the curve model to estimate the expected annual accident rate, \( \lambda_c \) for each curve section.

step 2

Use the EB procedure to refine the estimate \( \lambda_c \). The required equations are:

\[ \lambda'_c = w_1(x) + w_2(\lambda_c) \quad \ldots (2) \]

where,

\[ w_1 = \frac{\lambda_c}{k + n^{\ast} \lambda_c} \quad \ldots (4) \]

\[ w_2 = \frac{k}{k + n^{\ast} \lambda_c} \quad \ldots (5) \]

\[ n = \text{the ratio of the time period for which accident data is being used over the time period to which the estimate applies} \]

step 3

Use the tangent model to estimate the expected annual accident rate, \( \lambda_t \) for an equivalent tangent section.
step 4

The expected number of accidents occurring due to the presence of the curve may be found as:

\[ I = (\lambda_c' - \lambda_t) \] ...

(16)

step 5

The estimate of treatable accidents found in step 4 is used to rank all sites in decreasing order. A suitable cutoff could be chosen to determine which sites would be investigated for treatment. The values of \( \lambda_c' \) may also be used in the evaluation of safety treatments. This will be discussed in chapter 7.

4.2.2 SITES FOR SPEED CONTROL

The methodology for speed control measures seeks to identify sites which have a speed related safety problem. In this case treatable accidents are simply those that are reported as speed related. This procedure is a straightforward application of the EB procedure, applied to speed related accident models. The methodology is outlined below, again using a hypothetical model:

step 1

Calculate \( \lambda_s \), the expected number of speed related accidents per year:

\[ \lambda_s = 0.00002(\text{length})(\text{AADT})^{0.8} \]

step 2

Use the EB procedure to refine the estimate \( \lambda_s \). The required equations are:
\[ \lambda_i = w_1(x) + w_2(\lambda_a) \] ... (2)

where,

\[ w_1 = \lambda_n / (k + n*\lambda_a) \] ... (4)
\[ w_2 = k / (k + n*\lambda_a) \] ... (5)

\[ n = \text{the ratio of the time period for which accident data is being used over the time period to which the estimate applies} \]

**step 3**

All sites are ranked by the estimate of \( \lambda_n / km \). in decreasing order. A cutoff could then be used to select sites for treatment. Again, the EB estimates may be used in the evaluation of safety improvements.

**4.3 SUMMARY**

This chapter has outlined the Empirical Bayesian framework and its application to two methodologies for identifying hazardous sites. The EB method utilizes information from two sources, a regression model including physical and traffic data and the accident history of individual sites. The negative binomial model is used to describe the distribution of accident counts within a reference population. The negative binomial model results from two assumptions. The first, that the distribution of accident counts at a site is described by the Poisson distribution given a known mean, \( m \). The second is that the means, \( m \) of individual sites are distributed according to the gamma distribution within a
reference population. The parameter $k$ which describes the relationship between $\text{Var}(m)$ and $\text{E}(m)$ is determined in the modeling procedure through a maximum likelihood procedure. The methodologies for identifying hazardous sites for treatment uses the EB method to estimate the expected number of treatable accidents occurring annually. These estimates may be used to rank sites by the potential for safety improvement and for use in the estimation of the safety effect of applied treatments.
CHAPTER 5

RESULTS OF MODEL CALIBRATION

This chapter describes the data used for modeling and the final models that were developed. Models were developed using the Genstat statistical analysis software. Models are for non-intersection accidents and calibrated using the counts from 1988 to 1993 and the average AADT for that time period. Prior to discussing the calibrated models, a short discussion on the decision making for selection of the best model is given.

5.1 GOODNESS OF FIT AND SELECTION OF BEST MODEL

Prior to discussing the models developed for use in this study, it should be first discussed how the choice was made between models to select the best model. Goodness of fit measures allow analysts to make this comparison. Which goodness of fit measure to use, however, is not straightforward. Two common methods are the coefficient of determination $R^2$ and the Pearson $\chi^2$ statistic (24). The $R^2$ and a modified version of the $\chi^2$ statistic were both used in a study by Vogt and Bared (16).

The coefficient of determination $R^2$ may be summarized as:

$$R^2 = \frac{\text{explained variation}}{\text{total variation}}$$

The $R^2$ allows for easy comparisons. By definition, it estimates what percent of the variance is being accounted for. However, Miaou et al. (13), state that the $R^2$ measure is a
poor measure for use in accident prediction models for two reasons. Firstly, the $R^2$ measure is derived on the assumption that the random variable is described by the normal distribution, which is not the case (19). Secondly, accident prediction models are typically nonlinear. Hence the $R^2$ measure was not chosen to evaluate the goodness of fit.

The Pearson $\chi^2$ statistic has also been used by researchers, including Bonneson and McCoy (23). The form for accident prediction models is shown below:

$$\chi^2 = \sum_1^n \frac{(x-E\{m\})^2}{\text{Var}\{x\}}$$

However, the use of this measure has also been disputed for accident prediction models (24).

It was decided to use the value of $k$ to compare the fits of models. To review, $k$ describes the relationship between $\text{Var}\{m\}$ and $E\{m\}$ as shown below:

$$k = \frac{E\{m\}^2}{\text{Var}\{m\}} \quad \ldots (6)$$

Equation (6) suggests that models with larger values of $k$ have smaller variances. The value of $k$ has been used previously by researchers to measure the goodness of fit (24, 17).

In addition to examining the $k$ value, as mentioned in chapter 2, the standard errors and $t$-
statistics of the parameters are examined to prove they are significant.

5.2 HORIZONTAL CURVE STUDY

For the identification of hazardous curves, two models were required as discussed in section 4.2.1 to estimate the accidents occurring on horizontal curves due to the presence of the curve itself. These were a model for non-intersection accidents on horizontal curves and a model for non-intersection accidents on tangent sections.

5.2.1 MODELS FOR HORIZONTAL CURVES

The dataset for horizontal curves consisted of 585 sites with a total length of 143.79 km, an AADT range of 634 to 15,817, and an average AADT of 3,800. The total number of non-intersection accidents in the period 1988 to 1993 was 1,070, 349 of which resulted in an injury.

The calibrated models are of the form:

\[
\text{All accidents/year} = a(\text{length})^b(\text{AADT})^c e^{\frac{d}{\text{length}\cdot \text{radius}}} \quad \ldots (17)
\]

\[
\text{Injury accidents/year} = a(\text{length})^b(\text{AADT})^c(\text{radius})^d e^{\frac{e}{\text{length}\cdot \text{radius}}} \quad \ldots (18)
\]

where,

- length is in kilometres
- radius is in metres
The estimated parameters and their standard errors are given in Table 2. All estimated parameters were significant at the 95% confidence interval.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Injury Accidents</th>
<th></th>
<th>All Accidents</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>Standard Error</td>
<td>Estimate</td>
</tr>
<tr>
<td>a</td>
<td>e^{-0.9282}</td>
<td>e^{1.1131}</td>
<td>e^{-5.9127}</td>
</tr>
<tr>
<td>b</td>
<td>0.4144</td>
<td>0.1371</td>
<td>0.6477</td>
</tr>
<tr>
<td>c</td>
<td>0.8254</td>
<td>0.0817</td>
<td>0.7609</td>
</tr>
<tr>
<td>d</td>
<td>0.2510</td>
<td>0.1097</td>
<td>Insignificant</td>
</tr>
<tr>
<td>f</td>
<td>1473</td>
<td>245</td>
<td>789</td>
</tr>
<tr>
<td>k</td>
<td>2.10</td>
<td></td>
<td>2.20</td>
</tr>
</tbody>
</table>

**TABLE 2: PARAMETER ESTIMATES FOR HORIZONTAL CURVE MODEL**

The model gives an estimate of the expected annual accident frequency for an average site within a reference population. This estimate will be further refined in the Empirical Bayesian technique.

Figure 2 graphs the relationship between injury accidents and AADT for a curve of length 200 metres. The value of b indicates that as the traffic volume increases the expected accident frequency increases while the accident rate decreases. This can be seen by examining the lines S_1 and S_2. The accident rate (accidents per volume) at any volume is...
equal to the slope of the line from the xy intercept to the corresponding point on the accident prediction function. It is evident that as the volume increases, the number of expected accidents increases, while the slope of the line connecting to the xy intercept, and thus accident rate decreases. The relationship holds true for all accident severities combined for which the b parameter is less than 1. This supports the hypothesis that accident rates using million-vehicle-km. are a poor representation of safety performance.

The parameter associated with the curve radius, d, in the injury model, at first glance may
The parameter $d$ is estimated to be 0.25. A positive value would indicate that as the radius increases and the sharpness of the curve decreases, the expected accident frequency increases. However the radius variable is also included in the model by the variable $(\text{length}/\text{radius})$. The parameter $f$ associated with this variable is positive showing that as the radius increases, the expected accident frequency decreases. The overall effect is that as the radius increases, the expected accident frequency decreases. Figure 3 illustrates this relationship for a curve of length 200 metres. It is also interesting to note that while increasing the radius always reduces the expected accident frequency, the reduction is very large for radii less than 200 metres and is minimal for radii larger than 400 metres. Models calibrated without the $(\text{length}/\text{radius})$ term do in fact result in a negative value for the parameter $d$. The variable $(\text{length}/\text{radius})$ was included in the model as it improved the goodness of fit.
5.2.2 MODELS FOR TANGENTS

One of the fields in the accident data lists each accident as occurring on a curve or on a straight segment of road. This allowed for the identification of tangent sections. Tangent sections were identified as those road sections which had no curve accidents in the period 1988 to 1993. Based on Poisson probabilities, it is unlikely that there are curves or at
least hazardous curves within these sections. Since the tangent data is being used to model sections where there is no safety problem due to hazardous curves, this assumption is justifiable.

The dataset for tangents consisted of 551 sections with a total length of 206 km., an AADT range of 100 to 21 633 and an average AADT of 3 412. The total number of non-intersection accidents in the period 1988 to 1993 was 10 301, 3359 of which resulted in an injury.

The calibrated model is of the form:

\[ \frac{\text{Accidents/km./year}}{\text{AADT}} = a(AADT)^b \]  \hspace{1cm} \ldots (19)

The estimated parameters and their standard errors are given in Table 3. All estimated parameters are significant at the 95% confidence interval.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Injury Accidents</th>
<th>All Accidents</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>Standard Error</td>
</tr>
<tr>
<td>a</td>
<td>$e^{-9.5917}$</td>
<td>$e^{0.3435}$</td>
</tr>
<tr>
<td>b</td>
<td>1.0084</td>
<td>0.0411</td>
</tr>
<tr>
<td>k</td>
<td>6.70</td>
<td>6.06</td>
</tr>
</tbody>
</table>

**TABLE 3: PARAMETER ESTIMATES FOR TANGENT MODEL**
As with the model for horizontal curves, the parameter $b$, associated with AADT, indicates that as the traffic volume increases the expected accident frequency increases but the accident rate (accidents/veh-km.) decreases.

### 5.3 Speed Related Accidents

The dataset for speed related non-intersection accidents consisted of 2912 sections with a total length of 18692.4 km., an AADT range of 42 to 31 075 and an average of 3 256. In the period 1988 to 1993 there were 16 218 accidents, of which 5457 were injury.

The models were of the form:

$$\text{Injury accidents/km./year} = a(AADT)^b(e^{c \cdot \text{width} + d \cdot \text{terrain}})$$

$$\text{All accidents/km/year} = a(AADT)^b(e^{c \cdot \text{width} + d \cdot \text{terrain} + e \cdot \text{psd}})$$

where,

- width is the sum of the lane and shoulder widths in metres
- terrain is equal to 1 for level terrain
- 2 for rolling terrain
- psd is the percentage of length that is marked for safe passing

The estimated parameters and their standard errors are given in Table 4. All estimated parameters are significant at the 95% confidence interval.
As with the models for curve and tangent accidents, the estimated values of parameter b are less than 1. This indicates that the use of accident rates using million vehicle-km. is misrepresentative. As AADT increases, the accident frequency increases but the accident rate decreases. Figure 4 graphs the relationship between total speed related accident frequency and AADT for various road widths. The estimated values of parameter c indicate that the expected accident frequency decreases with an increase in total width (lanes + shoulders). The estimates of d indicate that the expected accident frequency increases as the terrain increases from flat to rolling terrain. The negative value of the parameter d indicates that as the amount of passing sight distance increases speed related accidents decrease.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Injury Accidents</th>
<th>All Severities</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>Standard Error</td>
</tr>
<tr>
<td>a</td>
<td>$e^{-7.571}$</td>
<td>$e^{-0.164}$</td>
</tr>
<tr>
<td>b</td>
<td>0.734</td>
<td>0.025</td>
</tr>
<tr>
<td>c</td>
<td>-0.1177</td>
<td>0.0160</td>
</tr>
<tr>
<td>d</td>
<td>0.1804</td>
<td>0.0373</td>
</tr>
<tr>
<td>e</td>
<td>insignificant</td>
<td></td>
</tr>
</tbody>
</table>

**TABLE 4: PARAMETER ESTIMATES FOR SPEED RELATED ACCIDENT MODELS**
FIGURE 4: Total Speed Related Accidents: Relationship With AADT And Surface Width
Chapter 4 outlined the procedure for identifying hazardous sites. This chapter illustrates the procedures, and outlines how the methodology for identifying hazardous sites has been validated.

6.1 PROCEDURE FOR TESTING METHODOLOGY

The object of this study is to provide a methodology for identifying sites which have the potential for safety improvements. Firstly, the methodologies were performed in an illustrative application using the models developed. Secondly, the results were examined to evaluate their ability to predict treatable accidents in the future. To test the methodologies developed, their ability to identify sites with treatable accidents in the future was compared to other methods:

1) Rank sites by the recent accident count
2) Rank sites by the recent accident rate (accidents per million-vehicle-kilometres)
3) Rank sites strictly using regression estimates (i.e. eliminating accident counts from the EB method), done for horizontal curves only
4) Rank sites by an EB estimate of accidents of all causes, done for speed related accidents only
Method 3 will be of use when accident data is unavailable or unreliable. Method 4 may be useful when information on which accidents were speed related is not available.

Traffic and accident data for 1988-1993 was divided into two groups, 1988-1990, and 1991-1993. The 1988-1990 data was used to rank sites by the potential for safety improvement (EB method), and the comparison methods mentioned. The 1991 to 1993 data was used to examine the efficiency of the ranking methods. Example calculations and results of the validation exercise are given for the identification of hazardous curves, and for sections warranting speed control measures in the following sections.

6.2 HORIZONTAL CURVES

6.2.1 ILLUSTRATIVE APPLICATION

The procedure followed was as discussed in 4.2.1 as follows:

**step 1**
Use regression equation (17) to estimate the expected annual accident frequency for all severities, $\lambda_c$ for each curve section.

**step 2**
Use the EB procedure to refine the estimate $\lambda_c$ using equations (2), (4) and (5)

**step 3**
Use equation (19) to estimate the expected annual accident rate, $\lambda_t$ for an equivalent tangent section.
step 4
Calculate the expected number of accidents occurring due to the presence of the curve:

\[ I = (\lambda_c' - \lambda_s) \]  \[ \ldots (16) \]

step 5
Use the index found in step 4 to rank sites in descending order. Details of the top 20 sites, which constitute approximately 3 per cent of the sample are given in Table 5.

Note: for method 3 step 2 is skipped and the expected number of accidents occurring due to the presence of the curve is simply:

\[ I = (\lambda_c - \lambda_s) \]
<table>
<thead>
<tr>
<th>radius (m)</th>
<th>length (m)</th>
<th>mindist (m)</th>
<th>avg. aadt</th>
<th>6 yr. total accident count</th>
</tr>
</thead>
<tbody>
<tr>
<td>85.00</td>
<td>169.40</td>
<td>7.03</td>
<td>8758</td>
<td>31</td>
</tr>
<tr>
<td>436.59</td>
<td>731.90</td>
<td>971.63</td>
<td>6183</td>
<td>12</td>
</tr>
<tr>
<td>498.96</td>
<td>597.87</td>
<td>1002.39</td>
<td>7867</td>
<td>21</td>
</tr>
<tr>
<td>160.00</td>
<td>264.80</td>
<td>125.60</td>
<td>8758</td>
<td>12</td>
</tr>
<tr>
<td>155.23</td>
<td>291.24</td>
<td>162.34</td>
<td>3517</td>
<td>12</td>
</tr>
<tr>
<td>582.13</td>
<td>535.86</td>
<td>1302.75</td>
<td>6067</td>
<td>13</td>
</tr>
<tr>
<td>388.08</td>
<td>434.00</td>
<td>145.23</td>
<td>6183</td>
<td>10</td>
</tr>
<tr>
<td>537.35</td>
<td>523.48</td>
<td>932.96</td>
<td>5150</td>
<td>12</td>
</tr>
<tr>
<td>349.28</td>
<td>255.58</td>
<td>72.48</td>
<td>8733</td>
<td>8</td>
</tr>
<tr>
<td>582.13</td>
<td>441.45</td>
<td>62.59</td>
<td>8733</td>
<td>7</td>
</tr>
<tr>
<td>582.13</td>
<td>451.10</td>
<td>27.75</td>
<td>8792</td>
<td>13</td>
</tr>
<tr>
<td>582.13</td>
<td>697.74</td>
<td>105.42</td>
<td>5150</td>
<td>10</td>
</tr>
<tr>
<td>1150.00</td>
<td>154.54</td>
<td>7.03</td>
<td>8758</td>
<td>14</td>
</tr>
<tr>
<td>498.96</td>
<td>416.27</td>
<td>90.85</td>
<td>4383</td>
<td>16</td>
</tr>
<tr>
<td>436.59</td>
<td>336.81</td>
<td>145.37</td>
<td>5167</td>
<td>9</td>
</tr>
<tr>
<td>180.00</td>
<td>326.22</td>
<td>5628.87</td>
<td>9200</td>
<td>9</td>
</tr>
<tr>
<td>436.59</td>
<td>526.92</td>
<td>453.30</td>
<td>5700</td>
<td>7</td>
</tr>
<tr>
<td>1746.38</td>
<td>506.73</td>
<td>625.33</td>
<td>5150</td>
<td>7</td>
</tr>
<tr>
<td>582.13</td>
<td>225.21</td>
<td>28.27</td>
<td>8792</td>
<td>9</td>
</tr>
<tr>
<td>349.28</td>
<td>655.32</td>
<td>140.21</td>
<td>683</td>
<td>4</td>
</tr>
</tbody>
</table>

**TABLE 5: TOP 20 HAZARDOUS CURVES**

mindist = the distance in metres to the nearest adjacent curve

Rated number 1 on the list, is a curve which most engineers would consider as warranting some safety treatment. It has a small radius, is very close to another curve, and has a high number of accidents occurring on a short length. It is interesting to note that the two adjacent curves are also ranked in the top 20, at #4 (125 m. from #1) and #13 (7 m. from #1). It is apparent that the problems on these curves are due in part to their proximity to...
and relationship with the adjacent curves. This seems to indicate that where curves are located close to one another the whole group of curves should be investigated for safety improvements.

For a subgroup of the 585 curves used in this study field visits were conducted and photographs taken. Photos were available for one of the curves identified in the top 20 hazardous curves, shown in Figure 5.

![Figure 5: Fifth Highest Rated Hazardous Curve](image)

The curve rated fifth highest, shown in Figure 5, would be classified as sharp, having a radius of 155 metres. Additional data shows this curve also contains no spirals and provides no illumination. Within the curve are 3 driveways and the curve was given a
roadside hazard rating of D, of a possible A, B, C or D, D being the worst case. The curve is in rolling terrain and has 0% passing sight distance. The rail bridge, which is close to the road offers ample opportunity for collisions with vehicles running off the road. Although protected on the outside of the curve by a guardrail, the opportunity for crashes exists. A curve warning sign did exist, but with 12 accidents in the period 1988-1990, 4 of which were injury, there clearly exists a need for safety improvements. Providing further low-cost curve warning treatments would definitely be more cost-effective than realigning the road.

The ability of the methodology to identify curves such as that in Figure 5 is a validation of the usefulness of the methodology.

**6.2.2 FURTHER VALIDATION**

As the second part of the methodology validation the ability of the EB method to predict accidents of all severities in the future was compared to three other methods:

1) A regression estimate ignoring site specific accident data
2) 1988-1990 accident count
3) 1988-1990 accident rate (accidents/million-vehicle-kilometres)

The regression estimate estimates the safety of a site based on its characteristics alone and without the use of site-specific accident data. The accident count and accident rate are
common methods used to evaluate the safety of sites. By using 3 years of data and calculating an average, this average may (possibly incorrectly) be assumed to be a long-term constant average. This average may then be used to forecast accident experience in the future.

The 585 curves were sorted in descending order of the ranking index for total treatable accidents. Details of the top 20 sites (approximately 3.5% of the sample) are given in Tables 6 and 7. Table 6 lists total accidents, Table 7 lists total treatable accidents. Also listed is the ranking given to those sites by the three methods listed above.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Rank if crash history not used</th>
<th>Rank by accident count</th>
<th>Rank by accident rate</th>
<th>radius (m)</th>
<th>length (m)</th>
<th>86-90 aadt</th>
<th>88-90 total accidents</th>
<th>91-93 total accidents</th>
<th>EB Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>85.00</td>
<td>169.40</td>
<td>8100</td>
<td>21</td>
<td>10</td>
<td>15.337</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>3</td>
<td>102</td>
<td>436.59</td>
<td>731.90</td>
<td>5817</td>
<td>9</td>
<td>3</td>
<td>8.433</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>2</td>
<td>99</td>
<td>498.96</td>
<td>597.87</td>
<td>7517</td>
<td>10</td>
<td>11</td>
<td>8.421</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>8</td>
<td>78</td>
<td>160.00</td>
<td>264.80</td>
<td>8100</td>
<td>6</td>
<td>6</td>
<td>5.462</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>10</td>
<td>24</td>
<td>155.23</td>
<td>291.24</td>
<td>3350</td>
<td>6</td>
<td>6</td>
<td>4.620</td>
</tr>
<tr>
<td>6</td>
<td>27</td>
<td>4</td>
<td>86</td>
<td>582.13</td>
<td>535.86</td>
<td>5750</td>
<td>8</td>
<td>5</td>
<td>5.878</td>
</tr>
<tr>
<td>7</td>
<td>11</td>
<td>7</td>
<td>78</td>
<td>388.08</td>
<td>434.00</td>
<td>5817</td>
<td>7</td>
<td>3</td>
<td>5.365</td>
</tr>
<tr>
<td>8</td>
<td>24</td>
<td>6</td>
<td>79</td>
<td>537.35</td>
<td>523.48</td>
<td>4900</td>
<td>7</td>
<td>5</td>
<td>5.088</td>
</tr>
<tr>
<td>9</td>
<td>43</td>
<td>11</td>
<td>75</td>
<td>349.28</td>
<td>265.58</td>
<td>8367</td>
<td>6</td>
<td>2</td>
<td>4.026</td>
</tr>
<tr>
<td>10</td>
<td>48</td>
<td>13</td>
<td>126</td>
<td>582.13</td>
<td>441.45</td>
<td>8367</td>
<td>6</td>
<td>1</td>
<td>4.789</td>
</tr>
<tr>
<td>11</td>
<td>42</td>
<td>12</td>
<td>131</td>
<td>582.13</td>
<td>451.10</td>
<td>8500</td>
<td>6</td>
<td>7</td>
<td>4.866</td>
</tr>
<tr>
<td>12</td>
<td>9</td>
<td>15</td>
<td>143</td>
<td>582.13</td>
<td>697.74</td>
<td>4600</td>
<td>5</td>
<td>5</td>
<td>4.609</td>
</tr>
<tr>
<td>13</td>
<td>338</td>
<td>5</td>
<td>23</td>
<td>1150.00</td>
<td>154.54</td>
<td>8100</td>
<td>8</td>
<td>6</td>
<td>3.037</td>
</tr>
<tr>
<td>14</td>
<td>53</td>
<td>20</td>
<td>72</td>
<td>498.96</td>
<td>416.27</td>
<td>4187</td>
<td>5</td>
<td>11</td>
<td>3.296</td>
</tr>
<tr>
<td>15</td>
<td>56</td>
<td>16</td>
<td>88</td>
<td>436.59</td>
<td>336.81</td>
<td>4983</td>
<td>5</td>
<td>4</td>
<td>3.205</td>
</tr>
<tr>
<td>16</td>
<td>1</td>
<td>60</td>
<td>174</td>
<td>180.00</td>
<td>326.22</td>
<td>8700</td>
<td>3</td>
<td>6</td>
<td>3.806</td>
</tr>
<tr>
<td>17</td>
<td>8</td>
<td>28</td>
<td>154</td>
<td>436.59</td>
<td>526.92</td>
<td>5700</td>
<td>4</td>
<td>3</td>
<td>3.882</td>
</tr>
<tr>
<td>18</td>
<td>570</td>
<td>9</td>
<td>94</td>
<td>1746.38</td>
<td>506.73</td>
<td>4900</td>
<td>6</td>
<td>1</td>
<td>3.394</td>
</tr>
<tr>
<td>19</td>
<td>144</td>
<td>19</td>
<td>84</td>
<td>582.13</td>
<td>225.21</td>
<td>8500</td>
<td>5</td>
<td>4</td>
<td>2.928</td>
</tr>
<tr>
<td>20</td>
<td>21</td>
<td>54</td>
<td>19</td>
<td>349.28</td>
<td>655.32</td>
<td>650</td>
<td>3</td>
<td>1</td>
<td>2.004</td>
</tr>
</tbody>
</table>

**TABLE 6: COMPARISON OF RANKING METHODS FOR HORIZONTAL CURVES: TOTAL ACCIDENTS OF ALL SEVERITIES**

58
TABLE 7: COMPARISON OF RANKING METHODS FOR HORIZONTAL CURVES: TREATABLE ACCIDENTS OF ALL SEVERITIES

The following observations are related to the ranking methods in Tables 6 and 7:

1. The site ranked #16 has been ranked #60 by the accident count method. The EB method has identified this site as hazardous based upon its sharp radius (180 m.), and relatively high AADT (8 700). The EB method recognized that the accident count of 3 was a random fluctuation below the mean accident frequency and gave it a higher ranking. This proves true as evidenced by the accident count of 6 for the period 1991-

2. The accident rate ranking method identified only 2 of the top 20 sites identified by the EB methodology. Accident rates are subject to select short sites with a low AADT and a randomly high accident count, which results in a high accident rate.

3. The use of regression models without using accident data can provide a reasonable ranking for sites. By using regression models to select a set of sites and then collecting accident data on them, the workload could be reduced where accident data is hard to come by. For example, by selecting the top 27 sites using regression models only, the top 8 sites selected by the EB method could be found. Where accident data is difficult to obtain or is unreliable, ranking sites using only the regression estimate may be a better alternative.

4. The advantage of the EB method over using regression estimates alone is that by including the accident history, a safety problem which is not represented in the regression models is identified. Ten of the top 20 sites identified by the EB method were ranked below 20 by using only the regression models. These sites had accident counts in the 1988-1990 period of 8,7,6,6,6,8,5,5,6,5 and 5,5,2,1,7,6,11,4,4,1 in the 1991-1993 period respectively. Clearly there are safety problems which are not being recognized by ignoring the accident data. An excellent example is curve #13 which was ranked #338 using regression models alone. The low rank was given due to it’s large radii and low AADT. The accident count of 14 in 6 years however clearly shows evidence of a safety problem.

5. The value of EB estimates as predictors of accidents is evident. In the period 1990-
1993 the 585 curves experienced 100 accidents of all severities. This is a drop from 136 accidents in the 1988-1990 period. The drop is most likely due to regression to the mean. The EB estimates, which correct for regression to the mean, estimated the accident count in 1991-1993 to be 102.446, almost equal to 100. Likewise, for treatable accidents the drop between the two time periods was 76.095 to 66.946. The EB estimate of treatable accidents in the second three year period is 69.707. This is further evidence that the EB method as superior to using the accident count for ranking sites and evaluating the effect of safety treatments.

6.3 SECTIONS WARRANTING SPEED CONTROL

6.3.1 ILLUSTRATIVE APPLICATION

Since injury accidents are more likely to be speed related, they are used in the application. The procedure followed was as outlined in 4.2.2 as follows:

step 1

Calculate $\lambda_s$, the expected number of speed related injury accidents per year from equation (20).

step 2

Use the EB procedure to refine the estimate $\lambda_s$ using equations (2), (4) and (5) to obtain $\lambda'_s$.

step 3

Rank all sites by the estimate of $\lambda'_s$/km. in decreasing order.
The details of the top 20 sites identified are given in Table 8:

<table>
<thead>
<tr>
<th>Length (km.)</th>
<th>psd</th>
<th>terrain</th>
<th>lane-width (m.)</th>
<th>shoulder-width (m.)</th>
<th>avg. AADT</th>
<th>6 yr. speed related injury accidents</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>35</td>
<td>2</td>
<td>7.3</td>
<td>3</td>
<td>15967</td>
<td>6</td>
</tr>
<tr>
<td>6.5</td>
<td>39</td>
<td>2</td>
<td>7.3</td>
<td>3</td>
<td>18600</td>
<td>12</td>
</tr>
<tr>
<td>2.2</td>
<td>72</td>
<td>1</td>
<td>7.3</td>
<td>3</td>
<td>18917</td>
<td>4</td>
</tr>
<tr>
<td>9.3</td>
<td>38</td>
<td>2</td>
<td>6.7</td>
<td>2.4</td>
<td>10900</td>
<td>12</td>
</tr>
<tr>
<td>9.3</td>
<td>48</td>
<td>2</td>
<td>8.5</td>
<td>3</td>
<td>14100</td>
<td>18</td>
</tr>
<tr>
<td>6.3</td>
<td>35</td>
<td>2</td>
<td>7.3</td>
<td>3</td>
<td>15967</td>
<td>7</td>
</tr>
<tr>
<td>6.6</td>
<td>100</td>
<td>2</td>
<td>7.5</td>
<td>2.5</td>
<td>8733</td>
<td>10</td>
</tr>
<tr>
<td>0.4</td>
<td>46</td>
<td>2</td>
<td>6.7</td>
<td>3</td>
<td>16817</td>
<td>1</td>
</tr>
<tr>
<td>3.3</td>
<td>17</td>
<td>2</td>
<td>6.7</td>
<td>1.8</td>
<td>8475</td>
<td>4</td>
</tr>
<tr>
<td>0.8</td>
<td>50</td>
<td>1</td>
<td>7.3</td>
<td>2.4</td>
<td>11333</td>
<td>4</td>
</tr>
<tr>
<td>5.1</td>
<td>61</td>
<td>1</td>
<td>7.3</td>
<td>3</td>
<td>15117</td>
<td>7</td>
</tr>
<tr>
<td>2.9</td>
<td>20</td>
<td>2</td>
<td>7.3</td>
<td>2.4</td>
<td>16825</td>
<td>2</td>
</tr>
<tr>
<td>4.7</td>
<td>30</td>
<td>2</td>
<td>6.7</td>
<td>1.8</td>
<td>7267</td>
<td>6</td>
</tr>
<tr>
<td>4.8</td>
<td>21</td>
<td>2</td>
<td>6.7</td>
<td>1.2</td>
<td>10475</td>
<td>7</td>
</tr>
<tr>
<td>3.5</td>
<td>73</td>
<td>1</td>
<td>7.3</td>
<td>3</td>
<td>8758</td>
<td>11</td>
</tr>
<tr>
<td>4</td>
<td>63</td>
<td>1</td>
<td>6.7</td>
<td>1.8</td>
<td>10192</td>
<td>4</td>
</tr>
<tr>
<td>3.5</td>
<td>64</td>
<td>1</td>
<td>6.7</td>
<td>1.8</td>
<td>16833</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>69</td>
<td>1</td>
<td>7.3</td>
<td>2.4</td>
<td>12675</td>
<td>9</td>
</tr>
<tr>
<td>3.1</td>
<td>50</td>
<td>2</td>
<td>7.3</td>
<td>3</td>
<td>10408</td>
<td>9</td>
</tr>
<tr>
<td>3.9</td>
<td>59</td>
<td>1</td>
<td>7.3</td>
<td>3</td>
<td>12033</td>
<td>6</td>
</tr>
</tbody>
</table>

**TABLE 8: TOP 20 SECTIONS IDENTIFIED FOR SPEED CONTROL**

The characteristics of the top 20 sites indicate the following:

1. Firstly, 12 of the top 20 and 8 of the top 9 sites are classified as rolling terrain. This indicates that in areas where the vertical geometry does not consist of small grades, but includes vertical curves and steep grades, the speed related accident frequency is expected to increase.
2. Secondly, and related to the vertical geometry, the passing sight distance (given as the percentage of the road section which has adequate sight distance for passing) of the top 20 sites is 50% or less for 12 of the 20 sections, 11 of these are classified as rolling terrain.

3. The high AADT's indicate that speed related accidents are a problem in areas with high levels of traffic where more conflicts between vehicles arise.

4. In choosing sections for treatment the expected number of treatable accidents would be the expected accident rate per km. multiplied by the section length. The cost of treatments also would be related to the length over which they apply. The final stage in choosing sections for treatment would be to examine the ratio of benefits (expected reduction in accident costs) to the costs of treatment application.

6.3.2 FURTHER VALIDATION

As the second part of the methodology validation the EB method to predict injury speed related accidents in the future was compared to three other methods:

1) 1988-1990 speed related injury accident count

2) 1988-1990 speed related injury accident rate (accidents/million-vehicle-km.)

3) 1988-1990 injury accident EB estimate of all accident types (not only those that were speed related).

Method 3 is used to evaluate the effectiveness of ranking based on all accident types,
where information on which accidents are speed related are unavailable. A model was calibrated using the same sections used for the calibration of speed related accidents. The model is shown below:

\[
\text{injury accidents} = 0.000470(\text{length})(\text{AADT})^{0.849} e^{(0.1260 \times \text{travel} - 0.0491 \times \text{width})}
\]

The 1419 sections which had experienced at least one speed related injury accident in the 1988-1990 period were sorted in descending order of the ranking index. Details of the top 20 sites are given in Table 9. Also listed is the ranking given to those sites by the three methods listed above.
<table>
<thead>
<tr>
<th>Rank</th>
<th>Rank if all accidents used, EB Estimate/ km.</th>
<th>Rank by speed related accident count/km.</th>
<th>Rank by speed related accident rate (mvkm.)</th>
<th>length (km.)</th>
<th>88-90 AADT</th>
<th>88-90 speed related injury accidents</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>16</td>
<td>58</td>
<td>729</td>
<td>3.1</td>
<td>15783</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>12</td>
<td>11</td>
<td>1018</td>
<td>6.5</td>
<td>17917</td>
<td>12</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>97</td>
<td>765</td>
<td>2.2</td>
<td>18983</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>52</td>
<td>3</td>
<td>723</td>
<td>9.3</td>
<td>10483</td>
<td>12</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>1</td>
<td>824</td>
<td>9.3</td>
<td>13717</td>
<td>18</td>
</tr>
<tr>
<td>6</td>
<td>11</td>
<td>30</td>
<td>1019</td>
<td>6.3</td>
<td>15783</td>
<td>7</td>
</tr>
<tr>
<td>7</td>
<td>30</td>
<td>4</td>
<td>512</td>
<td>6.6</td>
<td>8367</td>
<td>10</td>
</tr>
<tr>
<td>8</td>
<td>34</td>
<td>716</td>
<td>479</td>
<td>0.4</td>
<td>16033</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>23</td>
<td>98</td>
<td>507</td>
<td>3.3</td>
<td>8300</td>
<td>4</td>
</tr>
<tr>
<td>10</td>
<td>31</td>
<td>191</td>
<td>213</td>
<td>0.8</td>
<td>11150</td>
<td>4</td>
</tr>
<tr>
<td>11</td>
<td>5</td>
<td>31</td>
<td>791</td>
<td>5.1</td>
<td>12983</td>
<td>7</td>
</tr>
<tr>
<td>12</td>
<td>48</td>
<td>367</td>
<td>1188</td>
<td>2.9</td>
<td>16200</td>
<td>2</td>
</tr>
<tr>
<td>13</td>
<td>111</td>
<td>59</td>
<td>490</td>
<td>4.7</td>
<td>7117</td>
<td>6</td>
</tr>
<tr>
<td>14</td>
<td>45</td>
<td>192</td>
<td>1033</td>
<td>4.8</td>
<td>10717</td>
<td>7</td>
</tr>
<tr>
<td>15</td>
<td>2</td>
<td>12</td>
<td>295</td>
<td>3.5</td>
<td>8100</td>
<td>11</td>
</tr>
<tr>
<td>16</td>
<td>14</td>
<td>99</td>
<td>739</td>
<td>4</td>
<td>9917</td>
<td>4</td>
</tr>
<tr>
<td>17</td>
<td>179</td>
<td>368</td>
<td>1284</td>
<td>3.5</td>
<td>16667</td>
<td>3</td>
</tr>
<tr>
<td>18</td>
<td>6</td>
<td>100</td>
<td>861</td>
<td>4</td>
<td>12233</td>
<td>9</td>
</tr>
<tr>
<td>19</td>
<td>7</td>
<td>101</td>
<td>577</td>
<td>3.1</td>
<td>9967</td>
<td>9</td>
</tr>
<tr>
<td>20</td>
<td>29</td>
<td>60</td>
<td>655</td>
<td>3.9</td>
<td>11333</td>
<td>6</td>
</tr>
</tbody>
</table>

**TABLE 9: COMPARISON OF RANKING METHODS FOR SECTIONS FOR SPEED CONTROL**

The following observations are made relevant to Table 9:

1. The accident rate (accidents/million vehicle-km.) does a poor job of selecting those sections which experience the highest number of speed related accidents. The
tendency is to select relatively short sections, with a low AADT, and which experience a randomly high accident count. For example, the section ranked number one is of length 0.7 km., had an 1988-1990 AADT of 167 and experienced 1 speed related injury accident in 1988-1993.

2. Using an EB estimate based upon all accident types to select sites for speed treatment may not select sites that warrant speed control measures. For example, the sites ranked 13th and 17th by using speed related accidents are ranked 111th and 179th using accidents of all causes. While these sites have relatively few injury accidents over the 6 year period, 0.64 and 0.38 per km. per year respectively, the number of speed related injury accidents is comparatively high, 0.21 and 0.14 per km. per year. This illustrates the importance of selecting sites based on the target accidents, in this case those that are speed related.

6.4 SUMMARY

Chapter 6 illustrates and validates the methodologies for identifying hazardous curves for treatment and sections warranting speed control measures. Several conclusions have been made:

- Ranking sites by their recent accident count or recent accident rate tends to give a higher ranking than deserved to sites with a randomly high recent accident count, and a lower ranking than deserved to sites with a randomly low recent accident count.

- The EB method is capable of recognizing randomly high or randomly low accident
counts and ranking sites according to their estimated long term mean accident frequency. This reduces the number of sites wrongly identified as warranting safety treatments, and the number of sites that warranted treatment but may not have been identified.

- Using only a regression estimate may be useful where accident data is unavailable or difficult to collect.

- It is important to rank sites by the expected frequency of target accidents, not by accidents of all types.

- EB estimates are useful predictors of future accident experience.
CHAPTER 7

SUMMARY

To maximize the safety benefit of treatments to the road network, it is imperative that hazardous sites be chosen based on the potential for safety improvement, that is, a reduction in the expected accident frequency. Thus far there is no consistent methodology being applied amongst jurisdictions which takes this into account. The object of this study was to develop methodologies for identifying sites for the treatment of two accident types on 2-lane rural undivided highways: 1) accidents occurring due to the presence of a horizontal curve and 2) accidents which are speed related.

The methodologies developed utilized the Empirical Bayesian method to rank sites based upon their potential for the reduction of accidents. The EB method utilizes information from 2 sources, the geometric and traffic information and the accident history of the site. EB estimates decrease errors introduced by the regression to the mean (RTM) phenomenon.

Ontario data was used to develop models for non-intersection accidents on horizontal curves, tangents and for speed related accidents. Models were developed for injury (fatal and non-fatal injury) accidents and for accidents of all severities.

For horizontal curves, the methodology identifies the treatable accidents as those
occuring due to the presence of the curve and estimates them as the EB estimate of the curve minus the tangent model estimate of a similar site. Sites were ranked by this measure. For speed related accidents, the treatable accidents were identified as those reported as "speed too fast" or "speed related" by the police. Speed related accidents were ranked simply based on the EB estimate of speed related accidents.

The EB methodology was found to be effective in identifying sites which are experiencing treatable accidents. EB estimates overcome problems associated with common blackspot identifiers, the accident count and accident rate (accidents/vehicle-km.).

It is advantageous that the EB estimates used in the methodologies are also useful in the evaluation of the effectiveness of safety treatments.

Finally, although the methodologies developed outline the procedure for choosing sites for the investigation for potential treatment they do not constitute a warrant. Optimistically, warrants would be developed after using the methodologies to choose sites for treatment and then evaluating the safety effect of the treatments applied.
REFERENCES


