ACCIDENT PREDICTION MODELS FOR TWO-LANE RURAL HIGHWAYS

K. R. Kalokota
Utah Transportation Center
Utah State University

Prianka N. Seneviratne
Utah Transportation Center
Utah State University

May 1994

Acknowledgments

The authors wish to thank Mr. Jeffery Piniati of the Federal Highway Administration for making the HSIS data available. Financial support for this research was provided by the Mountain Plains Consortium.

Disclaimer

TABLE OF CONTENTS

CHAPTER 1. INTRODUCTION 1
1.1 Introduction 1
1.2 Study Objectives and Methodology 3

CHAPTER 2. DATA COLLECTION AND ANALYSIS 7
2.1 Study Area 7
2.2 Database 8
2.3 Compilation of Data Files 9
2.4 Description of Variables in the Data Files 10
2.5 Data Analysis 13

CHAPTER 3. MODEL DEVELOPMENT 15
3.1 Functional Forms of Models 15
3.2 Selection of Model Form and Calibration 15
3.3 Model Verification 17
3.4 Model Validity Over Time and Space 21
3.5 Applying Previous Models to Utah 22

CHAPTER 4. PARAMETER UPDATING AND MODEL TRANSFERABILITY 25
4.1 Rationale 25
4.2 Common Parameter Updating Techniques 26
4.3 Model Development Procedure Selection 28
4.4 Bayesian Updating for Temporal Transfer 29
4.5 Model Verification 30
4.6 Bayesian Updating for Spatial Transfer 31

CHAPTER 5. CONCLUSIONS 35
5.1 Discussion 35
5.2 Conclusion 36

REFERENCES 39

APPENDIX A. Program to Combine Roadway Inventory File and 1985-87 Accident File 43

APPENDIX B. Program to Combine Roadway Inventory File and 1988-89 Accident File 45

APPENDIX C. Program to Combine Roadway Inventory File and 1990 Accident File 47

APPENDIX D. Program to Combine All Accident Files Along with Roadway Inventory Files 49

APPENDIX E. Program to Perform GLM 51

APPENDIX F. Program to Calibrate Corridor Model 53
APPENDIX G. Program to Calibrate Tangent Model 55

APPENDIX H. Program to Calibrate Curve Model 57

APPENDIX I. TABLES 59
Table 1. Distribution of section lengths and accidents on tangents and curves in percentage 61
Table 2. Distribution of curves and accidents in percentage 62
Table 3. Distribution of grade and accidents in percentage 63
Table 4. Distribution of lanes and accidents in percentage 64
Table 5. Distribution of right side shoulder width and accidents in percentage . 65
Table 6. Distribution of AADT and accidents in percentage 66
Table 7. R^2 values to test multi-collinearity 67
Table 8. Geometric variables used in previous models and variables selected for this study 67
Table 9. Accident prediction model forms 68
Table 10. Comparison of Model A predicted values and observed values 69
Table 11. Comparison of Model B predicted values and observed values 70
Table 12. Comparison of Model C predicted values and observed values 71
Table 13. Best accident prediction models for northern Utah 72
Table 14. Comparison of models developed for Utah and previous models . . . 73
Table 15a. Accident prediction models for individual years for corridor 74
Table 15b. Accident prediction models for individual years for tangents 74
Table 15c. Accident prediction models for individual years for curves 75
Table 16a. Bayesian parameter updating of corridor models 76
Table 16 b. Bayesian parameter updating of tangent models 77
Table 16 c. Bayesian parameter updating of curve models 78
Table 17. Bayesian accident prediction models 79
Table 18. Comparison of Bayesian updated model predictions & observed accidents 80
Table 19. Accident prediction models developed for Provo, Utah using 1989 data 81
Table 20a. Bayesian method of parameter updating of southern Utah corridor model 82
Table 20b. Bayesian method of parameter updating of southern Utah tangent model 82
Table 20c. Bayesian method of parameter updating of southern Utah curve model 83
Table 21. Bayesian updated models for southern Utah 83

APPENDIX J. FIGURES 85
Figure 1. Distribution of accidents with section length 87
Figure 2. Distribution of degree of curvature and accidents 88
Figure 3. RSI for degree of curvature 89
Figure 4. Distribution of vertical grade, frequency, and accident rate 90
Figure 5. Distribution of shoulder width and accidents 91
Figure 6. Comparison of accident predictions for corridor 92
Figure 7. Comparison of accident predictions for tangents 93
Figure 8. Comparision of accident predictions for curves 94

Figure 9. Comparison of predicted accidents for corridors 95
Figure 10. Comparison of accident predictions for tangents. 96
Figure 11. Comparison of accident predictions for curves 98

APPENDIX K. RELATED RESEARCH 99

EXECUTIVE SUMMARY

Approximately 2.5 million miles or 63 percent of the highways in the U.S. are classified as two-lane rural highways, and 50 percent of fatalities occur on these highways. Past studies indicate that inconsistencies in highway system characteristic such as geometry, vehicle mix, level of road usage, environment, and driver behavior are the main causes of accidents. Of these, geometry has the potential to directly and indirectly influence the other characteristics.

This report is based on a study aimed at modelling the influence of the geometric design variables on traffic accidents on two-lane rural highways. The objectives of the study were:

1. To review previously developed relationships between accidents and geometric variables, and to identify significant variables.
2. To examine the significance of the above identified variables as well as other geometric variables on accidents in a selected part of highway 89 and 91 in northern Utah.
3. To establish statistically significant relationships between accidents and the geometric variables.
4. To examine the spatial and temporal validity of the above relationships.

It was found that the exposure in terms of distance travelled (length of road section) is the most significant variable in the empirical models developed in the present case. Contrary to previous findings, horizontal curvature and cross-section were found to have rather negligible effects on accident occurrence.

The other main findings of the study were:

1. Disaggregation of data by type of roadway section (curved or tangent) does not improve model predictability.
2. Models developed in the present case are temporally and spatially transferable by updating parameters using Bayesian statistics whereas previously developed models are not.

CHAPTER 1

INTRODUCTION

1.1 Introduction

It is very likely that two-lane highway networks in most countries will be required to play a more active role in the years to come. Many segments of these networks will require upgrading to better meet future traffic volumes, changing vehicle composition, diverse vehicle and driver characteristics, and above all to enhance safety.

One of the key questions that arises out of this scenario relates to the form of upgrading required for safety. More specifically, what, if any, geometric elements of the highways should be upgraded and to what extent.

The answer is very complex. Though road design practices have been refined and simplified over the years, changing driver and vehicle characteristics continue to present engineers with new challenges. For instance, the “Policy on Geometric Design of Highways and Streets," published by the American Association of State Highways and Transportation Officials (AASHTO)(1), has been revised three times since 1949 (i.e., in 1964, 1985, and 1990). Higher standards have emerged each time, but the accident rates on two-lane highways have not declined in proportion to the increased design standards.

Approximately 4 million kilometers or 63 percent of highways in the U.S. are classified as two-lane rural highways, and 50 percent of fatalities occur on these highways (20). They have the highest accident rate of any class of rural highway, with fatal and injury vehicle-mile exposure accident rates (VMER) four to seven times higher than those on rural interstate highways (21). The lower-volume two-lane rural roads exhibit accident experiences quite different from those of higher-volume roads, according to Raff (23), and Khilberg, et al. (24).

Design standards are applied by highway agencies to assure optimal operational and safety performance based on the anticipated use of roads in their system. Ideally, the application of the highest design standards could be expected to maximize safety. However, when operating under financial and time constraints, compromises on standards are inevitable. Therefore, a better understanding of the individualized singular and combined effects of roadway design features on safety is needed to guide decision makers and promote highway safety management efforts.

Previous research (2, 15) has shown that geometric design inconsistencies, operations (traffic mix, volume, and speed), environment, and driver behavior are the common causes of accidents. Environmental conditions and driver behavior can seldom be foreseen. They are case, time, and driver specific. They are also influenced by geometric inconsistencies. Thus, one option, which is easily justifiable and implementable, is to alter the geometry.

Studies that have shown the influence of various geometric design variables on the occurrence of accidents have concluded that not all variables have the same level of influence in all places (2, 15). This uncertainty in the influence of geometric variables on accidents has prompted researchers to develop mathematical models to better understand the relationship.

Mathematical models enable highway agencies to select design standards that are essential to highway safety and to allow comparison among alternative designs that will optimize the overall safety of the highway system under limited resources and other constraints. These models can also be used to test the sensitivity of accident rates to changes in a specific geometric variable.

1.2 Study Objectives and Methodology

The primary objective of this research was to establish statistically valid mathematical models that can explain the variations in accidents as a function of geometric variables. Engineers and planners will be able to use these models to examine the sensitivity of accident rates to any of the independent variables, and then decide on the best improvement strategy that satisfies their goals and objectives.

The secondary objective was to investigate the transferability of model parameters over time and space and develop a methodology to simplify the transfer process. This process would enable engineers to employ general models to predict accidents at any location, simply by updating it with a small set of sample data. In order to meet these objectives, the following tasks were undertaken.

Task 1. Study previously identified geometric variables influencing the occurrence of accidents (literature search). For this task, a thorough literature search was conducted and several articles on accident prediction models and geometry were collected and reviewed. The variables found by the earlier researchers to have an influence on the occurrence of accidents were identified.

Task 2. Collect and analyze data. Highway Safety Information System (HSIS) records for an 87-mile section of highway 89 and 91 in Utah were obtained from the Federal Highway Administration (FHWA) and the data files were decoded according to the study needs. Accidents observed in the test section were then plotted against certain geometric variables identified in Task 1 to test their significance.

Task 3. Identify statistically significant geometric variables within the study corridor. The "Proc GLM" procedure in SAS (version 6.04) was employed to test the statistical significance of the variables. GLM procedure calculates the Student t-values for each variable and tests the significance at the required level. The multi-colinearity of the variables was examined using the SAS package.

Task 4. Examine previous accident prediction models on the basis of underlying statistical modelling principles. During the literature search it became evident that the forms of the models and the statistical procedure for calibrating models differed from researcher to researcher. In this task the model form and the calibrating procedure best suited for this study corridor were determined.

Task 5. Calibrate and verify the selected model forms. The forward stepwise regression models were calibrated using 1985-89 data. Three separate models were calibrated for the curve sections, tangent sections, and the entire corridor respectively. These models were verified by using them to predict the accidents in 1990.

Task 6. Compare calibrated model parameters/coefficients with parameters of the models from other studies to determine the potential for model transfer. The models of other researchers were employed to predict accidents at the study site so that the transferability and validity over space and time could be assessed.

Task 7. Examine the procedures available for updating parameters and test the accuracy of updated models. From the literature search it was found that several parameter updating techniques can be employed to enhance the predictive accuracy of transferred models. The Bayesian parameter updating approach was chosen and employed to update the parameters of the models calibrated in northern Utah by combining with sample data from southern Utah. The updated models were then tested for the spatial transferability. The temporal transfer was tested by updating the northern Utah models by combining with cross-sectional data from the same test section.

CHAPTER 2

DATA COLLECTION AND ANALYSIS

2.1 Study Area

Among the five states that currently have HSIS for analysis, Utah is considered to be the state that has the most complete information on highway geometries (27). Therefore, for this study, an 87-mile segment of US Highway 89 and 91 from Brigham City to Bear Lake in northern Utah was selected.

The study section from Brigham City to the city of Logan is a part of US 91 and it passes through the Sardine Canyon. The Logan to Bear Lake section of US 89 passes through the Logan Canyon. The study corridor of US 91 starts at mile point zero at Brigham City and ends at mile point 45.20 in the city of Logan. US 89 starts at mile point 374.24 in Logan and ends at mile point 415.84 at Bear Lake.

To limit the study to two-lane rural sections, the sections with four lanes, and sections in the City of Logan and to the immediate south were deleted. Only the sections from mile point 3.78 to mile point 15.75 on US 91 and sections from mile point 374.24 to mile point 376.95, mile point 377.31 to mile point 411.77, and mile point 411 to mile point 415.84 on US 89 were considered.

There were a few sections in the study corridor with climbing (auxiliary/truck) lanes. Since climbing lanes are an integral part of the two-lane highways in mountainous terrain, they were considered to be a variable design option and were included in the data set for further analysis.

The final database had 296 sections in all, made up of 143 tangent sections and 153 curved sections.

2.2 Database

Data pertaining to 6 years (1985-1990) were made available by the FHWA from their Highway Safety Information System (HSIS) records for Utah. They were contained in four separate files: grade, road, curve, and accidents.

The grade file contained 1987 information on vertical grades in the form of section length in miles, beginning mile point, end mile point, percent grade, direction of grade, and route number.

The road data file consisted of information on AADT, access, beginning point, end point, county, weighted design speed, federal aid system, functional classification, lane width, median type, number of lanes, percent of trucks in off peak, one-way or two-way factor, section length pavement condition, percent passing distance, present service rating, percent of commercial vehicles in peak, accident year, route number, type of sign posted, section length in miles, shoulder type (right side), speed limit, state code, and terrain type.

The curve data file contained 1987 information on length of curves in miles, route number, beginning and ending mile points of curve, direction of curve, and degree of curvature.

The accident files consisted of information for the years 1985 through 1990. The information included the accident route code, accident route, accident mile point, accident year, accident month, accident day, time, number of vehicles involved in the accident, accident severity index, accident type in coded form, light condition in coded form, weather, collision type in coded form, object struck in coded form, road effect in coded form, number of injured in accident, accident severity, total people involved, traffic control, time of accident, accident recoded by, people treated by, and accident mile point. The coding was done as per accident report form standards.

2.3 Compilation of Data Files

For the purposes of this study, the curve file, road file, and grade files are combined manually into one file and named as the "roadway inventory" file. This file consisted of information on section length in feet, section length in miles, route number, degree of curvature, direction of the curve (to the left or right), beginning and ending mile points, AADT, number of lanes, pavement type, pavement condition, and percent grade.

The three accident files containing records from 1985-87, 1988-89, and 1990 were then combined with the roadway inventory file so that accidents in each section matched with the road characteristics in that section. This became a long and tedious process for several reasons.

First, because accidents were coded as separate entries in the HSIS, and a given section could have several accidents of different types in a given year. The number of accidents during each of the 6 years (1985-90) had to be aggregated and entered separately in a new column so that they correspond to the appropriate road section. Second, some of the variables were at times defined by characters and other times defined by numbers or blank spaces. To overcome these inconsistencies, the data base was recoded and rearranged.

The programs that combined roadway inventory file and 1985-87, 1988-89, and 1990 accident file are shown in Appendix A, B, and C, respectively. Appendices A, B, and C also show some of the variables that are recoded as zero when information was missing.

The program that combined all the accident files along with the inventory file is given in Appendix D.

2.4 Description of Variables in the Data Files

Section Length (L)

Section lengths were given in miles in the database. The distribution of the lengths of all the tangents and curves is shown in Table 1. The test corridor consisted of 296 sections, of which 143 were tangent sections and the rest were curve sections. It can be seen from the Table 1 that over 70% of the tangent sections in the corridor were 1,000 feet or less. As for curves, over 60% were 1,000 feet or less, and the curved segments made up approximately 32% of the corridor length.

One important feature that was missing from the data was information on spirals. The spirals were contained in the tangents without being identified. But because terrain within the corridor does not permit long spirals, errors due to this aggregation could be regarded as minimal.

The 5-year (1985-89) accidents in percentages for each of the above roadway categories are also shown in Table 1. On the tangents, the largest share of accidents (approximately 30%) occurred in sections less than 3,000 feet. The next clustering was observed in the 5,000 to 7,000 feet sections. On the curves, over 50% of the curve accidents occurred on sections less than 1,000 feet.

However, in relation to the proportion of sections in each of the above categories, the propensity is for accidents to increase as section length increases. This can be seen from the ratio of the total number of accidents per million vehicles to the total number of sections in the interval of 0.5 miles as shown in Figure 1.

Degree of Curvature (D)

Since tangent sections in the HSIS base were identified as having a degree curvature equal to zero, it permitted separate analyses to be performed of accidents on curves and tangents.

The degree of curvature in the selected corridor varied from 4^o to 30^o. The distribution of degrees of curvature and accidents on curves in the study corridor is depicted in Figure 2.

Contrary to previous findings, most curve accidents seem to have occurred on 4^o to 6^o curves. But, in terms of total accidents in the corridor, less than 15% of the accidents have occurred in these curves.

Table 2 shows the number of accidents that occurred from 1985 through 1989 and the number of curves with a given degree of curvature. Figure 3 shows the distribution of accidents in relation to a relative safety index (explained later), which indicates that the accidents increase with the degree of curvature.

Vertical Grade (G)

The study corridor traverses two major canyons (Sardine and Logan) in northern Utah. Consequently, there were many segments where a horizontal curve is connected directly or through a short tangent section to a vertical curve. These sections were not identified in the HSIS records and were difficult to find without extensive field surveys or referring to state DOT inventories. Therefore, no distinction was made between sections on the basis of this condition.

The distribution of vertical grades and accidents within the grades is shown in Figure 4. The number of accidents that occurred from 1985 through 1989 at sites with a given grade (in percentage) in tangents and curves, and the number of sections for each grade are given in Table 3. It can be seen that the grade varies from 0 to 7% and most of the sections are of 2% grade.

Number of Lanes (N)

There were a few sections in the study corridor with climbing (auxiliary/truck) lanes (third lane). Since climbing lanes are an integral part of two-lane highways in mountainous terrain, it was considered to be a potential variable that could explain the occurrence of accidents, and was included in the analysis. It is also evident that 97.3% of the accidents between 1985 and 1989 occurred in two-lane sections and only 2.7% in three lane sections.

Table 4 shows the number of accidents that occurred from 1985 through 1989 on two- and three-lane road sections in tangents and curves separately.

Right Shoulder Width (RSW)

The nature of the terrain had determined the width of the shoulder in many of the sections through the canyons. The distribution of the widths and accidents is shown in Figure 5.

Table 5 shows the number of accidents that occurred from 1985 through 1989 on curves and tangents and the number of sections with a given width of shoulder. It is also evident from Table 5 that in both tangents and curves, more accidents occurred on 6-foot shoulders, though not many sections had 6-foot shoulders.

Traffic Volume (AADT)

Section traffic volumes were recorded in terms of AADT in the data files. However, the AADT changed little within the corridor. The major difference was noted in the section through the City of Logan, which was eliminated to satisfy the rural condition.

In view of this invariability and also the fact that it was more of an operational variable, AADT was incorporated into the dependent variable as the exposure variable. Table 6 shows the number of sections and the number of accidents observed in each year from 1985 through 1989 for tangents and curves.

2.5 Data Analysis

Selection of Variables

Prior research and the present study objectives were used as a guide to select variables for significance testing. The task was not difficult because there were no other truly random geometric variables in the database except those mentioned above.

Some other variables such as lane width and left shoulder width were either irrelevant or invariable. Sight distance was not included in the analysis since it was given in terms of percents and it could not have been included in our database so that it corresponds to the curve and tangent sections.

Independent Variables

The relationships between the independent variables and accidents were not clearly apparent from the individual plots. For example, it is seen from Figure 2 that accidents in the present case are not responding to the degree of curvature as it had been hypothesized by prior researchers. Moreover, the relationships found in previous studies were not exactly in concurrence with one another. Therefore, it was important to test the contribution of all forms and combinations of variables to accident occurrence.

The general linear model procedure (ProcGLM) of SAS was used to check the significance of the variables. ProcGLM procedure uses the method of least squares to fit general linear models (33). For this study, the factorial modeling (with interaction) procedure was adopted.

Degree of curvature, percent grade, right and left shoulder widths, number of lanes, and section length were tested for significance by computing t-values and p-values for each of the parameters. The output is given in Appendix E.

It can be seen from Appendix E that the section length is the most significant variable. Degree of curvature, section length, right shoulder width, and number of lanes are significant at the ten percent level. No t-value could be estimated for left shoulder width, which means that it is highly insignificant. Percent grade could not meet the ten percent significance level.

To minimize the errors due to multi-collinearity, the variables were screened once more by examining the correlation coefficients between the variables. As seen from Table 7, the coefficients are all less than 4%. This is a strong indication of the absence of multi-collinearity between the chosen variables. Even though percent grade could not fall within the chosen level of significance in terms of t-value, it was included in the final analysis since the study corridor is in mountainous terrain.

In summary, the independent variables selected for further analysis are section length in miles, degree of curvature, shoulder width on right side of the road in feet, grade in percentage, and combinations of some of them. Table 8 shows the variables used in some of the previous studies and the variables used in this study.

Dependent Variable

The dependent variable was the expected accident rate (per million vehicles per year) estimated as follows:

$$\underline{\text{Total Accidents (1985-89)}*10^{6}}$$

(5x365xAADT)

CHAPTER 3

MODEL DEVELOPMENT

3.1 Functional Forms of Models

Many model forms have been tested in previous safety research. These can be broadly categorized as linear and nonlinear models expressed as:

$$Y = \sum a_i x_i$$
$$Y = \sum a_i b_i^{X_i}$$
$$Y = \sum a_i X_i^{b_i}$$
$$\text{LogY} = \sum a_i \text{LogX}_i$$
$$\text{LogY} = \sum a_i X_i^{bi}$$

where

Y = accident rate
X = independent variable
a & b = calibrated coefficients

In addition to the above, models with independent variables in the form of products of geometric and traffic characteristics have also been proposed. These variations in the forms suggest the variability in the significance and the interactions among variables from region to region, and over time.

3.2 Selection of Model Form and Calibration

The following model forms were chosen and calibrated separately with each of the three data sets corresponding to tangents, horizontal curves, and the entire corridor:

$$Y = \sum a_i X_i^{b_i} \text{.........(Model A)}$$

$$\text{LogY} = \sum a_i \text{LogX}_i \text{........(Model B)}$$

$$\text{LogY} = \sum a_i X_i^{bi\text{.........}} \text{(Model C)}$$

where

Y	=	accident rate
X	=	independent variable
a & b	=	calibrated parameters / coefficients

Since making a decision about which of the variables to include in a regression model is difficult, stepwise regression technique is sometimes used (35). This technique allows the computer to experiment with different combinations of the independent variable and give insight into the relations between the independent variables and the dependant variable. The problem with stepwise regression is that some of the variables which are not significant may enter the model, or variables which are significant may not enter the model. For the purpose of this study, it was decided to employ the "forward stepwise" GLM procedure to screen the variables and calibrate the traffic accident prediction models.

GLM also enables the reduction of confusion created by the inclusion of all the variables.

The forward stepwise procedure begins with no variables in the model. For each of the independent variables, the ProcGLM can be used to calculate the F-statistic. The F-statistic signifies the variable's contribution to the explanatory capability of the model if included.

The forward stepwise function compares the F-statistics to the level of significance (in this case the level specified is 0.1) that is specified in the model statements. If no F-statistics have values greater than the specified level, the process ends. Otherwise, the process adds the variable that has the largest F-statistic to the model. It then calculates F-statistics again for the variables still remaining outside the model, and the process continues. Thus, independent variables are added one by one to the model until none of the remaining variables are significant. The results of the forward stepwise multiple regression analysis for each of the data sets are given in Table 9.

3.3 Model Verification

The models developed were verified by using them to predict accidents in 1990 and comparing the outcome with the observed accidents. The predicted values of the present models were then regressed with the observed accidents in 1990 to test the predictive accuracy of the model.

Corridor Models

From Table 9, it can be seen that Model A has an R^2 value of 0.74. The predicted accidents by this model, when regressed with the observed accidents of 1990, could explain 56.31 (R^2) percent of the variation in accidents. The total number of accidents predicted by this model for 1990 is 85.86% (predicted number of accidents is 164 and the observed number of accidents was 191) of the accidents observed in 1990. This model predicts a mean accident rate of 0.5524 accidents per million vehicles per year, whereas the observed mean accident rate was 0.6426 accidents per million vehicles per year.

However, from Table 10 it can be seen that this model predicts 269 sections to have an accident rate of one accident per million vehicles per year in the year 1990. But it was observed in 1990 that 209 sections had a zero accident rate and only 23 sections had an accident rate of one.

Model B has an R^2 value of 0.38. The predicted accidents for 1990 by this model when regressed with observed accidents in 1990 gave an R^2 value of 0.2424. This model predicted about 6% more accidents than accidents observed in 1990. The model predicts a mean accident rate of 0.141 for 1990, whereas the observed mean rate in 1990 was 0.133 accidents per million vehicles per year.

However, from Table 11 it can be seen that this model predicts 108 sections to have an accident rate of 0.2 accidents per million vehicles per year but 209 sections were observed to have an accident rate of zero in 1990 and only seven sections had an accident rate of 0.2.

Model C has an R^2 value of 0.516. The number of predicted accidents by this model for 1990 when regressed with observed accidents in 1990 could explain 33.39% of the variation in accidents observed in 1990. The model predicts about 6% more accidents than observed in 1990 with an estimated mean accident rate of 0.14088 per million vehicles per year, whereas the observed accident rate in 1990 was 0.133 per million vehicles per year.

However, from Table 12 it can be seen that this model predicts only two sections to have an accident rate of zero when 209 sections were observed to have an accident rate of zero in 1990.

Tangent Models

From Table 9 it can be seen that Model A has an R^2 value of 0.843. The predicted accident by this model for 1990, when regressed with the observed accidents of 1990, could explain 66.80% (R^2) in the variation of accidents. This model predicts 17.89% (predicted number of accident = 145 and observed number of accidents = 123) more accidents than those that occurred in 1990. The model predicts a mean accident rate of 0.733, whereas the observed mean accident rate in 1990 was 0.855 accidents per million vehicles per year.

However, from Table 10 it can be seen that Model A predicts 115 sections to have an accident rate of one accident per million vehicles per year in the year 1990. But it was observed in 1990 that only nine sections had an accident rate of one, and 98 sections had an accident rate of 0.

From Table 9 it can be seen that Model B has an R^2 value of 0.43. The predicted accident values when regressed with the observed accidents in 1990 could explain about 34.74% of the variations in accidents observed in 1990. The model could predict about 6.15% more accidents than those observed in 1990. The predicted mean accident rate for 1990 was 0.1715 whereas the observed accident rate was 0.1612 per million vehicles per year.

However, from Table 11 it can be seen that this model predicts 24 sections to have an accident rate of zero, but it was observed in 1990 that 98 sections had an accident rate of zero.

Model C has an R^2 value of 0.84 as seen from Table 9. The predicted accident values by this model when regressed with observed accidents in 1990 could explain 33.39% of the variation in accidents. The model could predict about 6.15% more accidents than those observed in 1990. The mean accident rate observed was 0.1715 where as the observed rate was 0.1612 accidents per million vehicles per year.

However, from Table 12 it can be seen that this model predicts no sections to have an accident rate of zero but 98 sections were observed.

Curve Models

From Table 9 it can be seen that Model A has an R^2 value of 0.28. The predicted accident values by this model for 1990 when regressed with observed accident values in 1990 could explain 20.93% of the variation in accidents. This model predicts 13.24% (predicted number of accidents = 59, and observed number of accident = 68) fewer accidents than those that occurred in 1990. The predicted mean accident rate for 1990 was 0.3838, whereas the observed accident rate was 0.4439 accidents.

However, from Table 10 it can be seen that this model predicts 129 sections to have an accident rate of one accident per million vehicles per year for the year 1990 but it was observed that 111 sections had a zero accident rate in 1990, and only 14 sections had an accident rate of one.

Model B has an R^2 value of 0.18. The predicted accident values by this model for 1990 when regressed with observed accidents in 1990 could explain about 24.24% of the variations in accidents. This model could predict about 5.22% more accidents than those observed in 1990 with a mean accident rate of 0.1122, whereas the observed accident rate was 0.1066 accidents per million vehicles per year.

However, from Table 11 it can be seen that this model predicts 74 sections to have accident rate of 0.2, but 111 sections were observed to have an accident rate of 0 in 1990 and only 5 sections had an accident rate of 0.2.

Model C has an R^2 value of 0.28. The predicted accidents by this model for 1990 when regressed with observed accident in 1990 could explain only 16.5% of the variation in accidents. The model could predict about 5.22% more accidents than those were observed in 1990. The mean predicted accident rate was 0.1120, whereas the observed rate was 0.1066 accidents per million vehicle per year.

However, from Table 12 it can be seen that this model predicts 70 sections to have an accident rate of 0.1, but 111 sections were observed to have an accident rate of zero in 1990 and only one section had an accident rate of 0.1.

It can be seen in Table 9 that the model form of $Y = \sum a_i X_i^{b_i}$ (Model A) has more independent variables than the model form $\mathrm{Log}Y = \sum a_i \mathrm{Log}X_i$ (Model B). Even though

LogY= $\sum a_i X_i^{b_i}$ (Model C) has more variables, its R^2 value is less than the Model A form. Among all the model forms, Model A has the highest predictability in terms of R^2 value.

Model A is also easier to understand and interpret. Hence, it was taken to be the best model form for the study corridor and was further examined. All the models examined in detail are presented in Table 13.

Programs that calibrated these models and the output are given in Appendices F, G, and H.

3.4 Model Validity Over Time and Space

The best fit models for the study corridor and frequently used models from other studies are shown in Table 14. It is evident from this Table that, in spite of certain similarities in the explanatory variables, best-fit models in the present case are of different forms than those derived in previous studies.

Tables 15a, 15b, and 15c show the variations in the influence of the variables on accident occurrence when models were calibrated for individual years, for corridor, and for tangents and curves, respectively. These models for each of the years were calibrated using multiple regression procedure. The variations in the influence of variables from year to year illustrate the randomness of accidents from year to year, even at the same site.

For example, from Table 15a, it can be seen that the sign for the coefficient of shoulder width right (RSW) is positive for the year 1985, then negative until the year 1988, and again positive in 1989.

A positive sign for the coefficient indicates that the number of accidents increases as the magnitude of the corresponding independent variable increases (8). Any coefficient with a negative sign indicates that the accident rate decreases as the magnitude of the variable increases.

Caution must be exercised while interpreting the sign of the coefficient. This is the sign of the variable in the presence of other variables, but this sign may change with changes in other variables or when examined independently.

On the other hand, it can be observed that the sign of the variable "section length" (L) has remained positive all through the years, indicating that the contributing nature remained the same. It might be a coincidence that the coefficient of L has increased gradually from 1985 through 1988.

3.5 Applying Previous Models to Utah

Disregarding the goodness-of-fit and the significance of the variables, it was decided to test the predictive accuracy of a previous model using data from 1990.

The most reasonable models for this comparison are those of Zegeer (12) [also given in Table 14], which were developed as a part of a national program with data from several states. Although these models are not of the same form as those calibrated in this study, at present they are known to more state transportation officials than any other models. Hence, the cross section and curve models were chosen to test the transferability of the forms and/or the coefficients.

Attempts to calibrate the same forms were unsuccessful. The main reason is that some input data such as shoulder type and spiral information for the curve model were unavailable and the number of accidents in certain categories was small.

Even the predictions with the uncalibrated models were completely astray. For instance, both models were able to explain less than 5% of the variations in accidents, and none of the variables were significant.

Thus, there is reason to believe that the previous traffic accident prediction models are site specific and time specific. It appears that the models developed elsewhere cannot be employed to predict the accidents at other sites without revisions. This prompted the search for procedures available to update models and parameters that would enable transfers spatially and temporally.

CHAPTER 4

PARAMETER UPDATING AND MODEL TRANSFERABILITY

4.1 Rationale

The success of traffic accident prediction models depends on the quantity and quality of the data used to calibrate the models. The amount of data required for a good model to be calibrated is always a big question with no precise answers. Answers vary depending on experience, and often on the availability of funds and time. Thus, most models calibrated with limited data in the past have proved to be inconsistent in terms of the predictive accuracy.

One of the characteristics that determines the versatility of statistical models is the extent of the environment over which it would be valid. If a model can be easily calibrated with a small data set, then it can be fully exploited and used in many locations. But accident prediction models presented to-date vary so much that it is unlikely that any one of them will fit the conditions in many places.

This is partly attributable to the inconsistencies and variations in data collection and recording systems in the different regions. Of course, one should be able to achieve higher levels of predictive accuracy if the data could be disaggregated further. But, this means that a roadway category like two-lane highways may have to be represented by a multitude of models due to the varied physical and operational conditions within the category.

With the advancement of statistical techniques, updating model parameters has become simpler. These techniques help the use of a prior model to predict present conditions with minimum data, thus cutting the cost of data collection and time. In other words, these techniques could be used to update model parameters to better reflect the changes that took place since their first calibration or new environments.

Transferability is of two types--temporal and spatial. Updating parameters is useful for both temporal and spatial transfer of traffic accident prediction models. Temporal transfer is one in which a model calibrated at a particular site is employed at the same site at a later date to predict accidents. Spatial transfer is one in which a model calibrated at a particular site is employed at a different site with similar characteristics.

The failure of models to predict precisely when they are transferred temporally may be due to differences in the behavior of influencing variables, weather, and driving patterns since the time of model calibration. The failure of models to predict precisely when they are transferred spatially may be attributable to the above-mentioned differences as well as socioeconomic and geophysical differences between the prior context and the new context. The nature of accidents may also vary drastically between sites. Thus, no model will be able to explain or predict accidents precisely at all sites or at the same site at all times.

Given this scenario, a search was made for a simple technique for updating parameters that can cut the expenditure on data collection, time, and manpower.

4.2 Common Parameter Updating Techniques

Parameter updating can be done in several ways. Also, the updating could be limited to a few constant/exponent terms or all constants in the model. Either way, the literature search revealed that updated models in most cases improved the predictive accuracy and the Bayesian updating approach performed better than all other approaches.

The advantages of updating were demonstrated in Artherton, et al. (28), who examined the transferability of disaggregate transportation planning models. They calibrated mode-choice models with 1967 (sample model) data from Washington, D.C. and later updated the parameters using 1963 (prior model) data from New Bedford, Massachusetts so as to test the spatial transferability.

The above authors adopted parameter updating procedures such as updating only the constant term with small data and the Bayesian method. It is reported that the models with the updated parameters had better predictive capability than the model developed at the site, and the Bayesian model performed better than all the other updated models.

Kumarage, et al. (29) tried model transferability techniques such as updating constant terms using sample data and multiplying the coefficients with some scalar values, and updating using Bayesian statistics. It was found that in all cases the updated models were superior than the prior models in terms of prediction.

The Bayesian technique has been used extensively in many fields of civil engineering. Raiffa, et al. (30) have shown that the Bayesian techniques can be employed to combine sample information with prior information to arrive at more accurate posterior (updated) information. The prior information is the original distribution of the parameter to be updated, and sample information related to the data on the parameter collected specifically for the study. These two distributions can then be combined to arrive at the Bayesian (updated) distribution of the parameter in question.

Artherton, et al. (28) assumed that both prior and posterior distributions of the parameters to be normally distributed and hence the updated parameters are expressed as:

$$\theta_2 = [(\theta_1 / \sigma_1^2) + (\theta_s / \sigma_s^2)] / [(1 / \sigma_1^2) + (1 / \sigma_s^2)]$$

$$\sigma_2 = [(1 / \sigma_1^2) + (1 / \sigma_s^2)]$$

where

θ_1 = original model parameter

θ_s = sample model parameter
θ_2 = updated model parameter
σ_1 = standard deviation of the original parameter
σ_s = standard deviation of the sample parameter
σ_2 = standard deviation of the updated parameter

θ_2, the updated parameter, would be a weighted average of the original parameter, θ_1, and the parameter estimated from the sample, θ_s; the weights would be the inverse of the respective variances. This procedure also offers an opportunity to introduce subjective judgments into the model estimation process where sample information is lacking.

4.3 Model Development Procedure Selection

Updating parameters of the models obtained by the stepwise regression procedure is cumbersome at times due to the lack of information on the mean and variance values when the variables are not significant. These parameters, which are not significant at all times, could not be updated easily without certain assumptions. Thus, the multiple regression procedure was adopted for this part of the study. The advantage of the multiple regression procedure is that all the independent variables could be brought into the picture, at all times, thus easing the effort of computations and minimizing the errors due to the assumptions.

Provo Study Area

Provo Canyon, a two-lane rural road section in southern Utah, was selected to test the transferability of the models developed in northern Utah. This road section is US 189, beginning at mile point 8.15 and ending at mile point 14.37.

The Provo database was also obtained from FHWA HSIS data files. It was arranged similarly to the database in northern Utah. This test section consisted of 36 tangents and 39 curves. The degree of curvature of the curves varied between 4° and 21°.

4.4 Bayesian Updating for Temporal Transfer

Bayesian updating involves integrating the models calibrated in the Logan area (northern Utah) with data from 1985-87 to represent 1990 conditions. For this purpose, it is very important to have a complete prior model. A complete model is one which is calibrated with a large data set and has good predictive accuracy. Such complete models are referred to in this study as "prior models" and the coefficients in the model are termed "prior parameters." For this part of the study, the following three prior models were specified and tested:

Model I — A model calibrated only with 1985 data

Model II — A model calibrated using data from 1985-87

Model III — A model calibrated using data from 1985 through 1989.

The constants and regression parameters of the above three models related to the corridor, tangents, and curves were updated as discussed earlier and the values at each stage of the updated process are presented in Table 16a, 16b and 16c, respectively. The columns titled NC (NC = New Coefficients) in the tables are the updated coefficients.

Table 17 shows prior Model III, and updated models of Model I and Model II.

For example, in Table 16a the parameter of section length (L) in the first column, titled Coeff 85, is the prior parameter and the value. In the second column, titled Coeff 88, is the sample parameter. These values are plugged in the Bayesian updating formula to derive the posterior parameter, given in the third column titled NC (85-88).

The value in the third column now becomes the prior parameter and the value in the fourth column, titled Coeff 89, becomes the sample parameter. These values are plugged into the Bayesian updating formula to estimate a posterior parameter, which is given in the final column titled NC (85-89). Likewise all the parameters are updated. The values in the last column are the posterior parameters or updated parameters of the final updated Model II given in Table 17.

4.5 Model Verification

The models given in Table 17 were used to predict the accidents in the year 1990. The predicted values by the models were then regressed with the observed values to check the amount of variation explained by the models. The results of this exercise are given in Table 18.

Corridor Models

The accidents predicted by Model III when regressed with the accidents observed in 1990 could explain 54% of the variation. The accidents predicted by Model I and Model II when regressed with the observed accidents gave R^2 values of 0.54 and 0.56, respectively. Here, all the models could explain almost the same amount of variation.

However, from Table 18, it can be seen that the updated Model I predicted the same number of accidents as were observed (predicted accidents = 190, and observed accidents = 190). The updated Model II predicted 35 fewer accidents than the observed, and Model III predicted 26 fewer accidents than the observed. Figure 6 shows the plot of number of sections per each predicted accident rate by the updated Model I, Model II, Model III, and observed accidents.

Tangent Models

The accidents predicted by Model III when regressed with accidents observed in 1990 gave an R^2 value of only 0.17, whereas the updated Model I and updated Model II gave R^2 values of 0.63 and 0.66, respectively. From Table 18, it can be seen that the updated Model II predicts five accidents short of the observed (predicted accidents = 117, and observed = 122) and the updated Model I predicts 50 accidents more than the observed and the Model III predicts 18 fewer accidents than the observed. Figure 7 shows the plot of number of sections per each predicted accident rate by updated Model I, Model II, Model III, and observed accidents.

Curve Models

The accidents predicted by Model III, the updated Model I, and the updated Model II when regressed with the observed accidents could all explain only 18% of the variation in 1990 accidents. But from Table 18, it can be seen that all three models predicted only 10 (observed accidents = 68) fewer accidents than the observed. Figure 8 shows the plot of the number of sections per each accident rate predicted by updated Model I, Model II, Model III, and observed accidents.

4.6 Bayesian Updating for Spatial Transfer

Spatial transfer can be defined as the application of a model in a context other than the calibrated context. In this section the models calibrated in northern Utah (Logan area) were updated with data from southern Utah (Provo) using the Bayesian updating method to test the spatial transferability of models.

Model Verification

Table 19 shows the models calibrated using 1989 Provo data. The parameters of these models were treated as “sample parameters” in this study and northern Utah parameters were treated as “prior parameters.”

The updated parameters at each step for corridor, tangent, and curve are depicted in Tables 20a, 20b, and 20c. The updated models are given in Table 21.

Corridor Model

The 1989 model of Provo has an R^2 value of 0.3518. The predicted accidents when regressed with the accidents observed in 1990 could explain 12.32% of variation. Accidents predicted for 1990 by the updated model when regressed with 1990 observed accidents could explain only 7.65% of the variation. The updated parameters at each step are shown in Table 20a.

From Table 22 it can be seen that the mean observed accident rate for 1990 is 0.25 accidents per million vehicles per year with a standard deviation of 0.412. The 1989 model predicted a mean accident rate of 0.18 accidents with a standard deviation of 0.155, whereas the updated model predicted only 0.07 accidents per million vehicles per year with a standard deviation of 0.25. Figure 9 shows the expected and observed number of sections experiencing different accident rates.

Tangent Model

The 1989 model of Provo has an R^2 value of 0.4932. The predicted values, when regressed with the observed accidents of 1990, explained 2.37% of variation in accidents in 1990. Accidents predicted for 1990 by the updated model when regressed with 1990 observed accidents explained 6.11% of the variations in accidents. The updated parameters at each step are given in Table 20b.

From Table 22 it can be seen that the mean observed accident rate for 1990 is 0.28 accidents per million vehicles per year with a standard deviation of 0.462. The 1989 model predicted a mean accident rate of 0.158 accidents with a standard deviation of

0.238, whereas the updated model predicted 0.298 accidents per million vehicles per year with a standard deviation of 0.162. Figure 10 shows the expected and observed number of tangent sections experiencing different accident rates.

Curve Model

The 1989 model of Provo has an R^2 value of 0.41. The predicted values when regressed with the accidents observed in 1990 could explain 12.6% of variation in accidents in 1990. Accidents predicted for 1990 by the updated model when regressed with 1990 observed accidents explained 23.05% of the variability. The updated parameters at each step can be seen from Table 20c.

From Table 22 it can be seen that the mean observed accident rate for 1990 is 0.22 accidents per million vehicles per year with a standard deviation of 0.363. The 1989 model predicted a mean accident rate of 0.20 accidents with a standard deviation of 0.285, whereas the updated model predicted 0.13 accidents per million vehicles per year with a standard deviation of 0.29. Figure 11 shows the observed and expected number of curved sections experiencing different accident rates.

CHAPTER 5

CONCLUSIONS

5.1 Discussion

The models developed in this study concurred with the findings of prior studies in some respects. However, the level of influence of horizontal geometry on accidents that became highly apparent in the work of Zegeer, et al. (3) and Reinfurt, et al. (9) was not evident in the present case. At the aggregate level, when the entire corridor was considered as the analysis unit, the variation in accident rate (accidents per million vehicles) was explained mostly by the exposure variables, section length, and a combination of section length with other independent variables.

For instance, in the best-fit model for the aggregate data shown in Table 13, section length explained 70% of the accidents compared to degree of curvature, which explained less than 1%. Furthermore, over 50% of the accidents occurred on curves of less than 8 degrees, and the mean tangent accident rate (20.3 accidents/million vehicles/mile/year) in the study corridor was not significantly different from the corresponding curve rate (22.7).

On the other hand, if one considered the accidents in proportion to the curves, there was certainly evidence to hypothesize that crash probability increases with degree of curvature. To illustrate this, a relative safety index (RSI) was defined as follows:

$$RSI = \frac{\text{Total Accident Rate at Curvature = D degree}}{\text{Total Number of Sections at Curvature = D degree}}$$

As seen from Figure 4, RSI increases as degree of curvature increases. In other words, there is a disproportionately high share of accidents in those sections with high degrees of curvature. But the trend is not uniform. In fact, there is a disproportionately high share of accidents in 4 and 5 degree curves.

5.2 Conclusion

The study reported here confirmed that road geometry can explain a substantial portion of the variation in accidents on two-lane rural highways. It was also found that disaggregation of data does not always increase the predictive accuracy of a model. For instance, it was found that the curve (disaggregate) model was able to explain only 28% of the variation compared to 74% by the corridor (aggregate) model. The model for tangent sections (another disaggregate model), on the other hand, was more credible than all the aggregate models.

The other principal finding was that the aggregate and disaggregate models are transferable over short time spans, but their transferability over space is not clearly evident. For example, in spite of being widely cited as global models, Zegeer, et al.'s (3) models were able to explain less than 5% of the variation in accidents at the cross section or on the curves when employed to predict accidents in Utah. Of course, a few missing variables might have been a cause. But Zegeer, et al. (3) have demonstrated that curvature and section length are the most significant among the variables. Thus, the models should have been able to explain a larger part of the variation in accidents in Utah since information on the degree of curvature and section length was available.

The impact of horizontal curvature on accidents was found to be much less significant than what was noted in previous studies. Thus, if curve flattening results in a significant elongation of curve length and an increase in operating speed, the safety benefits noted in the previous literature (24, 27, 31, 33) may not be too realistic. This question calls for a further examination of the relationship between accident rate and degree of curvature.

In general, it could be said that there is a need for considerably more work in the accident modelling area. It may need to start with some standard definitions or criteria for disaggregating road sections so that models and results are easily transferable. If a reasonable level of uniformity can be achieved, parameter estimation and transferability would become practical. The use of modelling principles in transportation safety analysis will become commonplace. Ultimately, agencies responsible for transportation will be able to make informed decisions as opposed to educated guesses about the influence of geometric elements on traffic accidents.

REFERENCES

1. American Association of State Highway and Transportation Officials. A policy on geometric design of highways and streets. Washington, D.C., 1965.

2. Perkins, David D, and Harold T. Thomas. Candidate accident surrogative for highway safety analysis. Transportation Research Record 905. 1983. 147-152 p.

3. Hirsh, M., J.N, Prashakar, and M.B. Akiva. New approach to geometric design of highways. TRR 1100.

4. Zegeer, Charles V., and Forest M. Counsil. Safety effectiveness of highway design features.

5. Zegeer, Charles V., Robert C. Deen, and Jesse G. Mayes. Effect of lane and shoulder widths on accident reduction on rural two-lane roads. TRR 806 1981. 33-43 p.

6. Neuman, Timothy R. Design risk analysis. TRB 1992 Washington, D.C. 1992.

7. Zegeer, Charles V., J. Hummer, D. Reinfurt, L. Herf, and W.W. Hunter. Safety Effects of cross-section design for two-lane roads. FHWA/RD-87/008.

8. Cribbins P.D., J.M Arey, and J.K. Donaldson. Effects of selected roadway and operational characteristics on accidents on multilane highways. HRR 188. 1967.

9. Glauz, W.D., K.M. Bauer, and D.J. Migletz. Expected traffic conflict rates and their use in predicting accidents. TRR 1026. 1985.

10. Allen, Terrence M. A factor analysis of accident records. HRR 79. 1965.

11. Schoppert, D. W. Predicting traffic accidents from roadway elements of rural two-lane highways with shoulders. HRB 158. 1967.

12. Blensly, R.C., and J.A. Head. Statistical determination of effect of paved shoulder width on traffic accident frequency. HRB 240. 1968.

13. Goldstein, L.G., and J.N. Mosel. A factor study of drivers attitudes, with further study on driver aggression. HRB 172. 1967.

14. Versace, John. Factor analysis of roadway and accident data. HRB 240. 1968.

15. Dart, Olin K. Jr., and Lawrence J. Mann. Louisiana State University. Relationship of rural highway geometry to accident rates in Louisiana. HRR 312. 1970.

16. Mulinazzi, T.E., and H.L. Michael. Correlation of design characteristics and operational controls with accident rates on urban rates in urban arterials. Purdue University. Joint Highway Research Project Report 35. Dec 1967.

17. Levonian, E., H.W. Case, and R. Gregory. Prediction of recorded accidents and violations using nondriving predictors. HRR 4. 1963.

18. Lawson, Stephen D. Description and Predictions of accidents on urban radial routes. Traffic Engineering Control. Vol 27 No 6. June 1986.

19. Tamburri, Thomas N., and Richard N. Smith. The safety index: A method of evaluating and rating safety benefits. HRR 332. 1971.

20. Davis, D.I. Shoulder improvements on two-lane roads. TRR 737. 1979.59-60 p.

21. Fatal and injury accident rates on Federal-aid and other highway systems. FHWA, USDOT. Sept 1978.

22. Cleveland, Donald E., Lidia P. Kostyniuk, and Kuo-Liang Ting. Geometric design element groups and high-volume two-lane rural highway safety. TRR 960. 1984.

23. Raff, M.S. Interstate highway accident study. HRB 74. 1953.

24. Kihlberg, J.K., and K.J. Tharp. Accident rates as related to design elements of rural highways. NCHRP Report 47. HRB 1968.

25. Sonquist, J.A., E.L. Baker, and J.N. Morgan. Searching for structure. Institute for social research. University of Michigan. Ann Arbor. 1973.

26. Harman, H.H. Modern factor analysis. University of Chicago Press, Chicago. 1967.

27. Miaou, Shaw-Pin and Harry, Lum. A statistical evaluation of the effects of highway geometric design on truck accident involvements. TRB 1993. Washington, D.C. Paper No: 930789.

28. Artherton, Terry J., and Moshe.Akiva. Transferability and updating of disaggregate travel demand models. TRR 1976. 12-18 p.

29. Kumarasinghe, A.S., and, S.C.Wirasinghe. Updating of parameters in aggregate total demand models. Canadian Journal of Civil Engineering. Vol 19. No: 2. April 1992.

30. Raiffa, H., and R. Schlaifer. Applied statisticaldecision theory. Graduate School of Business Administration. Cambridge. 1961.

31. Zegeer, Charles V., Richard Stewart, Forrest M. Council, and Donald W. Reinfurt. Safety effects of geometric improvements on horizontal curves. TRB 92. Washington, D.C. 1992. Paper No: 920893.

32. Roy Jorgensen and Associates. Evaluation of criteria for safety improvements on the highway. Gaitherburg, Md. 1966.

33. Neuman, Timothy R., Glennon, John C.,and Saag, James B. Accident analyses for highway curves. TRR 923. 65-72 p.

APPENDIX A

Program to Combine Roadway Inventory File and 1985-87 Accident File

```
Libname save ‘C:\Reddy\’;
Data Road;
Infile ‘road.dat’ end = endoff;
Input nobs, sectlng sect_lng rtnum $ deg_curv dir_curv $ begin end aadt no_lane
pavetype pavecond pct_sight rtsight lanewid shldtype shldwidr shldwidl medtype medwid
pctsight spdlmt terrain;
Data Acc857;
Set save.UTAC857;
 if acc_type = ‘A’ then acc_type = ‘10’;
if acc_type = ‘R’ then acc_type = ‘11’;
if acc_type = ‘L’ then acc_type = ‘12’;
if acc_type = ‘D’ then acc_type = ‘13’;
if acc_type = ‘O’ then acc_type = ‘14’;
Data Loops;
Set Acc857;
Do i = 1 to 311;
Set Road Point = i;
If rtnum = a_rout then do;
If begin <= a_milept/100 <  end then output;end;
Data Combine;
Merge Road Loops;
File ‘C:\Reddy\GA857.DAT’;
Put nobs, begin, end, sect_lng rtnum 4 deg_curv dir_curv $ aadt sectlng no_lane pavetype
pavecond pct_grad rtsign lane wid shldtype shldwidr shldwidl medtype medwid pctsight
spdlmt terrain a_route $ a_milept accyr accmo accday day_wk hour time numvehs accsev
acc_type rdsurf noccsev tot_peop locatn local rdchar;
Proc Sort; By Begin;
Proc Print;
```

APPENDIX B

Program to Combine Roadway Inventory File and 1988-89 Accident File

```
Libname save 'C:\Reddy\';
Data Road;
Infile 'road.dat' end = endoff;
Input nobs, sectlng sect_lng rtnum $ deg_curv dir_curv $ begin end aadt no_lane
pavetype pavecond pct_sight rtsight lanewid shldtype shldwidr shldwidl medtype medwid
pctsight spdlmt terrain;
Data Acc889;
Set save.UTAC889;
if acc_type = 'A' then acc_type = '10';
if acc_type = 'R' then acc_type = '11';
if acc_type = 'L' then acc_type = '12';
if acc_type = 'D' then acc_type = '13';
if acc_type = 'O' then acc_type = '14';
Data Loops;
Set Acc889;
Do i = 1 to 311;
Set Road Point = i;
 If rtnum = a_rout then do;
If begin <= a_milept/100 <  end then output;end;
Data Combine;
Merge Road Loops;
File 'C:\Reddy\GA889.DAT';
Put nobs, begin, end, sect_lng rtnum 4 deg_curv dir_curv $ aadt sectlng no_lane pavetype
pavecond pct_grad rtsign lanewid shldtype shldwidr shldwidl medtype medwid pctsight
spdlmt terrain a_route $ a_milept accyr accmo accday day_wk hour time numvehs accsev
acc_type rdsurf noccsev tot_peop locatn local rdchar;
Proc Sort; By Begin;
Proc Print;
```

APPENDIX C

Program to Combine Roadway Inventory File and 1990 Accident File

```
Libname save ‘C:\Reddy\’;
Data Road;
Infile ‘road.dat’ end = endoff;
Input nobs, sectlng sect_lng rtnum $ deg_curv dir_curv $ begin end aadt no_lane
pavetype pavecond pct_sight rtsight lanewid shldtype shldwidr shldwidl medtype medwid
pctsight spdlmt terrain;
Data Acc90;  Set save.UTAC90;
if acc_type = ‘A’ then acc_type = ‘10’;
if acc_type = ‘R’ then acc_type = ‘11’;
if acc_type = ‘L’ then acc_type = ‘12’;
if acc_type = ‘D’ then acc_type = ‘13’;
if acc_type = ‘O’ then acc_type = ‘14’;
Data Loops;
Set Acc90;
Do i = 1 to 311;
Set Road Point = i;
If rtnum = a_rout then do;
If begin <= a_milept/100 <  end then output;end;
Data Combine;
Merge Road Loops
File ‘C:\Reddy\GA90.DAT’;
Put nobs, begin, end, sect_lng rtnum 4 deg_curv dir_curv $ aadt sectlng no_lane pavetype
pavecond pct_grad rtsign lanewid shldtype shldwidr shldwidl medtype medwid pctsight
spdlmt terrain a_route $ a_milept accyr accmo accday day_wk hour time numvehs accsev
acc_type rdsurf noccsev tot_peop locatn local rdchar;
Proc Sort; By Begin;
Proc Print;
```

APPENDIX D

Program to Combine All Accident Files Along with Roadway Inventory Files

```
Libname save 'C:\Reddy\';
Data ACC857;
InFile 'C:\Reddy\GA857.DAT';
InPut nobs, begin, end, sect_lng rtnum 4 deg_curv dir_curv $ aadt sectlng no_lane
pavetype pavecond pct_grad rtsign lanewid shldtype shldwidr shldwidl medtype medwid
pctsight spdlmt terrain a_route $ a_milept accyr accmo accday day_wk hour time
numvehs accsev acc_type rdsurf noccsev tot_peop locatn local rdchar;

Data ACC889;
InFile 'C:\Reddy\GA889.DAT';
InPut nobs, begin, end, sect_lng rtnum 4 deg_curv dir_curv $ aadt sectlng no_lane
pavetype pavecond pct_grad rtsign lanewid shldtype shldwidr shldwidl medtype medwid
pctsight spdlmt terrain a_route $ a_milept accyr accmo accday day_wk hour time
numvehs accsev acc_type rdsurf noccsev tot_peop locatn local rdchar;

Data ACC90;
InFile 'C:\Reddy\GA90.DAT';
InPut nobs, begin, end, sect_lng rtnum 4 deg_curv dir_curv $ aadt sectlng no_lane
pavetype pavecond pct_grad rtsign lanewid shldtype shldwidr shldwidl medtype medwid
pctsight spdlmt terrain a_route $ a_milept accyr accmo accday day_wk hour time
numvehs accsev acc_type rdsurf noccsev tot_peop locatn local rdchar;
Data Accidents;
Merge GA857 GA889 GA90; By Begin;
Proc Sort; By Begin;
Proc Print;
```

APPENDIX E

Program to Perform GLM

```
libname save 'C:\Reddy\';
data road;
infile 'labs.dat';
input sectlng sect_lng rtnum $ deg_curv dir_curv $ begin end aadt no_lane pavetype
pavecond pct_grad rtsign lanewid shldtype shldwidr shldwidl medtyp medwid pctsight
spdlmt terrain totalacc mrate accrate acc90 acrate90 accn859 accn90;
tar = totalacc*1000000/(5*365*aadt);

if no_lane lt 4 then output;

proc glm;
model tar = deg_curv sect_lng pct_grad shldwidr shldwidl no_lane;
proc sort;by begin;
proc print;
```

APPENDIX F

Program to Calibrate Corridor Model

```
libname save 'C:\Reddy\';
data road;
infile 'labs.dat';
input   sectlng sect_lng rtnum $ deg_curv dir_curv $ begin end aadt no_lane pavetype
        pavecond pct_grad rtsign lanewid shldtype shldwidr shldwidl medtyp medwid
        pctsight spdlmt terrain totalacc mrate accrate acc90 acrate90 accn859 accn90;
tar = totalacc*1000000/(5*365*aadt);
tar90 = acc90*1000000/(365*aadt);
x1 = deg_curv*sect_lng;
x2 = deg_curv*pct_grad;
x3 = sect_lng*pct_grad;
x4 = deg_curv*shldwidr;
x5 = sect_lng*shldwidr;
x6 = shldwidr*pct_grad;
x7 = shldwidr*pct_grad*deg_curv;
x8 = shldwidr*pct_grad*sect_lng;
x9 = shldwidr*deg_curv*sect_lng;
x10 = pct_grad*deg_curv*sect_lng;
x11 = pct_grad*shldwidr*sect_lng*deg_curv;
if no_lane lt 4 then output;
proc reg;
model tar=    sect_lng pct_grad shldwidr deg_curv x1 x2 x3 x4 x5 x6 x7 x8 x9 x10 x11/
              selection = stepwise
p r sls = .1;
output out = s p = pred r = resid;
proc sort;by begin;
proc print;
```

This is the output of the above program

Stepwise Procedure for Dependent Variable TAR

Step 4   Var. X11 Entered R-square = 0.737   C(p)= 16.350

	DF	S.S	Mean Square	F	Prob>F
Regression	4	258.32	64.58052	204.49	0.0001
Error	291	91.89	0.31580		
Total	295	350.221			

Variable	Parameter Estimate	Standard Error	Type II S.S	F	Prob>F
INTERCEP	0.0092	0.0471	0.01	0.03	0.8539
SECT_LNG	3.4681	0.1745	124.61	394.00	0.0001
DEG_CURV	0.0163	0.0055	2.709	8.58	0.0037
X5	-0.2018	0.0361	9.851	31.20	0.0001
X11	-0.0059	0.0034	0.942	2.98	0.0852

Bounds on condition number: 3.521733, 40.26664

All variables in the model are significant at the 0.1000 level. No other variable met the 0.1500 significance level for entry into the model.

Summary of Stepwise Procedure for Dependent Variable TAR

Variable Entered	Partial R**2	Model R**2	C(p)	F	Prob>F
SECT_LNG	0.6987	0.6987	55.1483	681	0.0001
X5	0.0311	0.7299	21.2693	33.7731	0.0001
DEG_CURV	0.0050	0.7349	17.4510	5.5621	0.0190
X11	0.0027	0.7376	16.3508	2.9838	0.0852

APPENDIX G

Program to Calibrate Tangent Model

```
libname save 'C:\Reddy\';
data road;
infile 'labs.dat';
input   sectlng sect_lng rtnum $ deg_curv dir_curv $ begin end aadt no_lane pavetype
        pavecond pct_grad rtsign lanewid shldtype shldwidr shldwidl medtyp medwid
        pctsight spdlmt terrain totalacc mrate accrate acc90 acrate90 accn859 accn90;
tar = totalacc*1000000/(5*365*aadt);
tar90 = acc90*1000000/(365*aadt);
x1 = deg_curv*sect_lng;
x2 = deg_curv*pct_grad;
x3 = sect_lng*pct_grad;
x4 = deg_curv*shldwidr;
x5 = sect_lng*shldwidr;
x6 = shldwidr*pct_grad;
x7 = shldwidr*pct_grad*deg_curv;
x8 = shldwidr*pct_grad*sect_lng;
x9 = shldwidr*deg_curv*sect_lng;
x10 = pct_grad*deg_curv*sect_lng;
x11 = pct_grad*shldwidr*sect_lng*deg_curv;
if no_lane lt 4 and deg_curv eq 0 then output;
proc reg;
model tar=     sect_lng pct_grad shldwidr deg_curv x1 x2 x3 x4 x5 x6 x7 x8 x9 x10 x11/
               selection = stepwise p r sls = .1;
output out = s p = pred r = resid;
proc sort;by begin;
proc print;
```

This is the output of the above program

Stepwise Procedure for Dependent Variable TAR

Step 2 Var X5 Entered R-square = 0.843 C(p) = -0.0751

	DF	S.S	Mean Square	F	Prob>F
Regression	2	238.89	119.4459	376.96	0.0001
Error	140	44.361	0.31686		
Total	142	283.25			

Variable	Parameter Estimate	Standard Error	Type II S.S	F	Prob>F
INTERCEP	0.0553	0.0531	0.343	1.08	0.2998
SECT_LNG	3.4198	0.1814	112.61	355	0.0001
X5	-0.200	0.0374	9.057	28.6	0.0001

Bounds on condition number: 3.535091, 14.14036

All variables in the model are significant at the 0.1000 level. No other variable met the 0.1500 significance level for entry into the model.

Summary of Stepwise Procedure for Dependent Variable TAR

Variable Entered	Partial R**2	Model R**2	C(p)	F	Prob>F
SECT_LNG	0.811	0.811	25.8	606	0.0001
X5	0.0320	0.843	-0.1	28.5	0.0001

APPENDIX H

Program to Calibrate Curve Model

```
libname save 'C:\Reddy\';
data road;
infile 'labs.dat';
input   sectlng sect_lng rtnum $ deg_curv dir_curv $ begin end aadt no_lane pavetype
        pavecond pct_grad rtsign lanewid shldtype shldwidr shldwidl medtyp medwid
        pctsight spdlmt terrain totalacc mrate accrate acc90 acrate90 accn859 accn90;
tar = totalacc*1000000/(5*365*aadt);
tar90 = acc90*1000000/(365*aadt);
x1 = deg_curv*sect_lng;
x2 = deg_curv*pct_grad;
x3 = sect_lng*pct_grad;
x4 = deg_curv*shldwidr;
x5 = sect_lng*shldwidr;
x6 = shldwidr*pct_grad;
x7 = shldwidr*pct_grad*deg_curv;
x8 = shldwidr*pct_grad*sect_lng;
x9 = shldwidr*deg_curv*sect_lng;
x10 = pct_grad*deg_curv*sect_lng;
x11 = pct_grad*shldwidr*sect_lng*deg_curv;
if no_lane lt 4 and deg_curv gt 0 then output;
proc reg;
model tar=    sect_lng pct_grad shldwidr deg_curv x1 x2 x3 x4 x5 x6 x7 x8 x9 x10 x11/
              selection = stepwise p r sls = .1;
output out = s p = pred r = resid;
proc sort;by begin;
proc print;
```

This is the output of the above program

Stepwise Procedure for Dependent Variable TAR

Step 7   Var.X7 Removed   R-square = 0.275   C(p) = 6.0966

	DF	**S.S**	**Mean Square**	**F**	**Prob>F**
Regression	5	15.942	3.18854	11.16	0.0001
Error	147	42.015	0.28581		
Total	152	57.957			

Variable	Parameter Estimate	Standard Error	Type II S.S	F	Prob>F
INTERCEP	-0.2579	0.129	1.1276	3.95	0.0489
SECT_LNG	3.8205	0.972	4.4145	15.4	0.0001
X1	0.3707	0.086	5.2105	18.2	0.0001
X2	0.0110	0.002	4.6081	16.1	0.0001
X4	-0.003	0.001	2.2546	7.89	0.0057
X10	-0.119	0.024	6.7189	23.5	0.0001

Bounds on condition number: 7.359586, 92.79664

All variables in the model are significant at the 0.1000 level. No other variable met the 0.1500 significance level for entry into the model.

Summary of Stepwise Procedure for Dependent Variable TAR

Variable Entered	Partial R**2	Model R**2	C(p)	F	Prob>F
X1	0.117	0.117	30.1381	20.0	0.0001
X4	0.027	0.144	26.6170	4.7	0.0305
X10	0.033	0.177	21.9063	5.99	0.0155
X7	0.050	0.227	13.6857	9.65	0.0023
SECT_LNG	0.031	0.258	9.3686	6.17	0.0141
X2	0.025	0.284	6.1979	5.19	0.0240
X7	0.009	0.275	6.0966	1.90	0.1692

APPENDIX I

TABLES

Table 1. Distribution of section lengths and accidents on tangents and curves in percentage

Tangent Section Length (feet)	% of Total Tangent Sections	% Total Accidents on Tangents	Curve Sections Length (feet)	% of Total Curve Sections	%Total Accidents on Curves
1,000.00	70.63	15.64	1,000.00	65.35	53.73
2,000.00	11.89	7.50	2,000.00	22.88	33.58
3,000.00	8.39	8.80	3,000.00	11.77	12.70
4,000.00	1.40	0.70			
5,000.00	0.70	0.84			
6,000.00	2.80	10.61			
7,000.00	1.40	17.04			
8,000.00	0.70	1.82			
9,000.00	0.00	0.00			
10,000.00	0.00	0.00			
11,000.00	0.70	10.75			
12,000.00	0.00	0.00			
13,000.00	0.00	0.00			
14,000.00	0.00	0.00			
15,000.00	0.00	0.00			
16,000.00	0.00	0.00			
17,000.00	0.00	0.00			
18,000.00	0.70	4.89			
19,000.00	0.70	21.37			

Table 2. Distribution of curves and accidents in percentage

Degree of Curvature	% of Curve Sections	%Accidents in 1985	in 1986	in 1987	in 1988	in 1989
4	12.4	17.5	9.2	21.6	20.3	17.8
5	9.8	15.0	4.6	17.6	23.0	15.6
6	7.2	2.5	0.7	2.0	4.1	0.0
7	3.9	0.0	0.0	0.0	4.1	0.0
8	5.2	5.0	0.7	2.0	6.8	8.9
9	7.2	0.0	0.7	3.9	1.4	4.4
10	6.5	7.5	0.0	0.0	4.1	4.4
11	7.8	7.5	2.0	13.7	0.0	6.6
12	5.2	5.0	2.0	2.0	2.7	8.8
13	3.3	0.0	0.7	3.9	0.0	0.0
14	5.2	2.5	3.3	5.9	1.4	4.4
15	4.6	2.5	2.6	7.8	0.0	2.2
16	4.6	5.0	2.0	7.8	12.2	6.6
18	0.7	2.5	0.0	0.0	0.0	0.0
19	0.7	0.0	0.0	0.0	0.0	0.0
20	5.9	12.5	3.9	7.8	8.1	4.4
21	1.3	2.5	0.0	0.0	1.4	0.0
22	2.0	0.0	1.3	0.0	0.0	0.0
23	2.0	5.0	1.3	0.0	5.4	4.4
24	1.3	0.0	0.0	0.0	1.4	0.0
28	1.3	0.0	0.0	0.0	1.4	0.0
29	0.7	5.0	0.7	3.9	1.4	2.2
30	1.3	2.5	0.7	0.0	1.4	8.8

Table 3. Distribution of grade and accidents in percentage

Tangent % Grade	% Tangent Sections	% Accidents in 1985	in 1986	in 1987	in 1988	in 1989
0	1.4	0.9	1.9	0.6	0.6	0.0
2	55.6	17.5	22.6	19.5	18.5	19.3
2.5	0.7	0.0	0.0	12.7	0.0	0.7
3	7.7	41.2	45.3	44.8	45.1	50.0
4	12.7	21.9	19.8	13.0	18.5	15.0
5	10.6	6.1	15.1	11.7	6.4	7.9
6	9.9	5.3	4.7	2.6	1.2	0.7
7	1.4	7.0	8.5	5.8	9.8	6.4

Curves % Grade	% Curves Sections	% Accidents in 1985	in 1986	in 1987	in 1988	in 1989
0	2.0	7.5	1.9	7.5	1.4	2.2
2	60.8	22.5	41.5	50.9	39.2	44.4
3	2.6	7.5	13.2	13.2	13.5	4.4
4	10.5	10.0	9.4	0.0	9.5	8.8
5	12.4	30.0	24.5	20.8	23.0	28.9
6	10.5	17.5	5.7	1.9	8.1	6.6
7	1.3	5.0	3.8	5.7	5.4	4.4

Table 4. Distribution of lanes and accidents in percentage

Tangents # of Lanes	**% Tangent Sections**	**% Accidents**				
		in 1985	**in 1986**	**in 1987**	**in 1988**	**in 1989**
2	96.5	98.2	96.0	97.5	96.5	100
3	3.5	1.8	4.0	3.5	3.5	0.0

Curves # of Lanes	**% Curves Sections**	**% Accidents**				
		in 1985	**in 1986**	**in 1987**	**in 1988**	**in 1989**
2	96.1	97.5	94.3	98.0	95.9	97.83
3.9	2.5	5.7	2.0	4.1	2.2	

Table 5. Distribution of right side shoulder width and accidents in percentage

Tangent R. Shoulder Width	% Tangent Sections	% Accidents in 1985	in 1986	in 1987	in 1988	in 1989
0	4.9	3.5	1.8	5.0	3.5	7.9
1	4.9	1.8	2.7	1.9	2.9	2.9
2	2.8	23.7	21.2	22.4	19.1	23.6
4	5.6	0.9	1.8	2.5	1.7	1.4
6	4.9	57.9	65.5	55.3	63.0	57.9
8	8.5	6.1	11.5	7.5	5.2	4.3
9	8.3	6.1	6.2	5.6	4.6	2.1

Curves R. Shoulder Width	% Curve Sections	% Accidents in 1985	in 1986	in 1987	in 1988	in 1989
0	5.2	5	7.5	11.8	8.1	8.9
1	4.6	0.0	1.9	0.0	2.7	0.0
2	54.9	47.5	45.3	39.2	41.9	51.1
4	4.6	2.5	0.0	0.0	0.0	0.0
6	4.6	30.0	30.2	31.4	35.1	24.4
8	7.8	2.5	5.7	3.9	4.1	2.2
9	18.3	12.5	9.4	13.7	8.1	13.3

Table 6. Distribution of AADT and accidents in percentage

Tangents AADT	% Tangent Sections	% Accidents in 1985	in 1986	in 1987	in 1988	in 1989
1230	13.4	2.6	3.2	7.5	1.7	0.0
1280	11.3	6.1	0.0	2.5	1.7	19.3
1415	5.6	3.5	2.4	3.7	2.8	0.7
1635	3.5	3.5	4.8	6.8	9.7	50.0
1835	26.8	11.4	12.8	8.1	7.4	15.0
3140	32.4	10.5	12.0	11.2	10.8	7.9
10450	4.9	57.9	59.2	55.3	61.9	0.7
14465	2.1	4.4	5.6	5.0	4.0	6.4

Curves AADT	% Curves Sections	% Accidents in 1985	in 1986	in 1987	in 1988	in 1989
1230	13.7	7.5	3.8	9.8	10.8	6.7
1280	10.5	17.5	5.7	5.9	6.8	13.3
1415	4.6	0	1.9	0	2.7	0
1635	2.6	0	3.8	3.9	5.4	6.7
1835	28.1	22.5	32.1	23.5	21.6	26.7
3140	33.3	22.5	20.8	23.5	17.6	20
10450	4.6	30	30.2	31.4	35.1	24.4
14465	2.6	0	1.9	2	0	2.2

Table 7. R^2 values to test multi-collinearity

VARIABLE	Section Length	Degree of Curve	R. Shoulder Width	% Grade
Section Length		0.026	0.001	0.0026
Degree of Curve	0.0260		0.006	0.0001
R. Shoulder Width	0.001	0.006		0.0328
Percent Grade	0.0026	0.0001	0.0328	

Table 8. Geometric variables used in previous models and variables selected for this study

Author's Name (ref.#)	Variables
Zegeer et al. (12)	AADT, lane-width, width of paved shoulder, width of unpaved shoulder, median, roadside hazard rating for highway segments, and terrains
Zegeer, et al. (13)	Length of curve, volume of vehicles, degree of curve, spiral, and width of roadway on the curve.
Zegeer, et al. (13)	Lane width, shoulder width, and width of stabilized component of shoulder.
Dart, et al. (14)	Percent trucks, traffic volume ratio, cross slope, (traffic volume ratio), (traffic volume ratio) (traffic conflicts), (lane width), (traffic conflict), and (shoulder width) (horizontal alignment).
Present Study	Section length (L), Degree of curve (D), Percent grade (G), Shoulder width on right side (SWR), (D)(L), (D)(G), (D)SWR),(L)(SWR), (L)(G), (G)(SWR), (D)(SWR)(L), (D)(SWR)(G), (SWR)(G)(L), (D)(G)(L), and (SWR)(G)(D)(L)

Table 9. Accident prediction model forms

Corridor

Model A
AR = 0.0092 + 0.016 (D) + 3.5 (L) - 0.02 (L)(SWR) - 0.006(L)(D)(G)(SWR)
$R^2 = 0.74$

Model B
Log (AR) = 1.02 + 0.77 Log (L) - 0.1 Log [(SWR)(D)(L)]
$R^2 = 0.38$

Model C
Log (AR) = 0.004 + 0.0071 (D) - 0.87 (L)(SWR) - 0.003 ((G)(SWR)(L)(D)
$R^2 = 0.74$

Tangents

Model A
AR = 0.1 + 3.4 (L) - 0.2(L)(SWR)
$R^2 = 0.843$

Model B
Log (AR) = 1 + 0.74 Log (L)
$R^2 = 0.43$

Model C
Log (AR) = 0.02 + 1.5 (L) - 0.1(L)(SWR)
$R^2 = 0.84$

Curves

Model A
AR = - 0.3 + 3.8 (D) + 0.37 (D)(L) + 0.011(D)(G) + 0.004(D)(SWR) - 0.12(G)(D)(L)
$R^2 = 0.28$

Model B
Log (AR) = 0.5 + 0.4 Log (D)(L) -2 Log (SWR)(G)(D)
$R^2 = 0.18$

Model C
Log (AR) = - 0.1 + 1.7 (L) + 0.16 (D)(L) + 0.005 (D)(G) - 0.1((G)(D)(L)
$R^2 = 0.28$

Table 10. Comparison of Model A predicted values and observed values

AR	Corridor Predicted	1990 Observed	Tangent Predicted	1990 Observed	Curve Predicted	1990 Observed
0	0	209	0	98	17	111
1	269	23	115	9	129	14
2	15	26	16	11	7	15
3	5	25	5	13	0	12
4	4	4	3	3	0	0
5	1	1	1	4	0	0
6	0	1	0	1	0	0
7	0	1	0	1	0	1
8	0	0	0	0		
9	1	0	1	1		
11	0	0	0	0		
12	0	0	0	0		
13	0	0	0	0		
14	0	1	0	1		

Table 11. Comparison of Model B predicted values and observed values

AR	Corridor Predicted	1990 Observed	Tangent Predicted	1990 Observed	Curve Predicted	1990 Observed
0	29	209	24	98	7	111
0.1	90	2	26	1	59	1
0.2	108	7	40	2	74	5
0.3	37	14	20	6	13	8
0.4	21	18	19	6	0	12
0.5	8	14	8	6	0	8
0.6	3	12	4	9	0	3
0.7	0	11	1	7	0	4
0.8	0	5	0	5	0	0
0.9	0	2	0	1	0	1
1.0	0	0	0	0		
1.1	0	1	0	1		
1.2	0	1	0	1		

Table 12. Comparison of Model C predicted values and observed values

AR	Corridor Predicted	1990 Observed	Tangent Predicted	1990 Observed	Curve Predicted	1990 Observed
0	2	209	0	98	7	111
0.1	132	2	54	1	70	1
0.2	118	7	60	2	59	5
0.3	31	14	14	6	15	8
0.4	2	18	4	6	2	12
0.5	4	14	4	6	0	8
0.6	2	12	2	9	0	3
0.7	3	11	3	7	0	4
0.8	0	5	0	5	0	0
0.9	0	2	0	1	0	1
1.0	0	0	0		0	
1.1	1	1	1		1	
1.2	0	1	0		1	
1.3	0	0	0		0	
1.4	1	0	1		0	

Table 13. Best accident prediction models for northern Utah

Corridor

$AR = 0.0092 + 0.016 (D) + 3.5 (L) - 0.02 (L)(SWR) - 0.006(L)(D)(G)(SWR)$
$R^2 = 0.74$

Tangents

$AR = 0.1 + 3.4 (L) - 0.2(L)(SWR)$
$R^2 = 0.843$

Curve

$AR = - 0.3 + 3.8 (D) + 0.37 (D)(L) + 0.011(D)(G) + 0.004(D)(SWR) - 0.12(G)(D)(L)$
$R^2 = 0.28$

Table 14. Comparison of models developed for Utah and previous models

Models developed for Utah

Corridor

$$AR = 0.0092 + 0.016\,(D) + 3.5\,(L.) - 2\ \ (L)(SWR) - 0.006\,(L)(D)(SWR)(G)$$
$R^2 = 0.74$

Curves

$$AR = -0.3 + 3.8\,(L) + 0.37\,(D)(L) + 0.011\,(D)(G) + 0.004\,(D)(SWR) - 0.012\,(L)(G)(D)$$
$R^2 = 0.28$

Tangents

$$AR = 0.1 + 3.4\,(L) - 2.0(L)(SWR)$$
$R^2 = 0.84$

AR=Accident rate per thousand vehicles per year
D=Degree of curve
L=Section length in miles
G=Percent grade
SWR=Right shoulder width in feet

Zegeer, et al. Model (11)

Corridor

$$AR = 40.299(0.7329)^{L}\,(0.8497)^{S}\,(1.0132)^{LS}\,(0.7727)^{P}\,(1.0213)^{LP}$$

AR=Number of ROR and OD accidents per million vehicle miles
L=Lane width in feet
S=Shoulder width in feet (including stabilized and unstabilized components)
P= Width in feet of stabilized component of shoulder ($0 <= P =< S$: P=0 for unstabilized shoulders and P = S for full-width stabilization)

Zegeer, et al. Model (12)

Curves

$$AR = [(1.552)(L)(V) + 0.014(D)(V) - (0.012)(S)(V)](0.978)^{w-30}$$

A=Number of total accidents on the curve in a 5 - year period
L=Length of the curve in miles(or fraction of a mile)
V= Volume of vehicles in million vehicles in a 5 - year period passing through the curve(in both directions)
D=Degree of curve
S=Presence of spiral, where S = 0 if no spiral exists, and S = 1 if there is a spiral
W=Width of the roadway on the curve in feet

Table 15a. Accident prediction models for individual years for corridor

	1985	1986	1987	1988	1989
Constant	-0.222	-0.002	0.0985	-0.170	-0.087
Section Length (L)	1.292	2.324	3.1920	5.644	5.564
% Grade (G)	0.148	0.017	0.0270	-0.093	-0.0317
R. Shoulder Width	0.0489	-0.0278	-0.1400	-0.0018	0.0356
Degree of Curve (D)	0.0166	0.0230	0.0388	0.0483	0.0082
(D)*(L)	0.0135	0.2236	0.1225	-0.1127	-0.2666
(D)*(G)	0.0000051	-0.0072	-0.015	0.0024	0.0088
(L)*(G)	-0.004650	-0.21700	0.02090	-0.05131	-0.0076
(D)*(RSW)	-0.002480	-0.00467	-0.00263	-0.00567	0.00324
(L)*(RSW)	0.06770	0.01800	-0.16974	-0.4995	-0.5148
(RSW)*(G)	-0.026400	0.01500	0.00297	0.029162	-0.0077

Table 15b. Accident prediction models for individual years for tangents

	Coeff 85	Coeff 86	Coeff 87	Coeff 88	Coeff 89
Constant	-0.05867	-0.01116	0.194156	-0.37513	-0.12736
(L)	0.9711	2.19799	2.9014	5.443	5.429
(G)	0.157	0.039	0.0307	-0.0395	-0.00084
(SWR)	0.0469	-0.02531	-0.02084	0.037	0.044
(L)(G)	0.061	-0.1956	0.095877	0.03297	0.0245
(L)(SWR)	0.0862	0.027	-0.1651	-0.5148	-0.51546
(SWR)(G)	-0.035	0.0127	0.001399	0.0177	-0.01213

Table 15c. Accident prediction models for individual years for curves

	Coeff 85	**Coeff 86**	**Coeff 87**	**Coeff 88**	**Coeff 89**
Constant	-0.67388	-0.2755	-0.48	-0.5214	-0.6129
(L)	1.8531	6.2717	7.684	14.7	12.712
(G)	0.1586	-0.0419	0.08	-0.08399	-0.0687
(RSW)	0.0523	-0.01934	-0.02121	-0.06894	0.0307
(D)	0.0367	0.02689	0.0477	0.0377	0.0234
(D)(L)	0.099448	0.139578	0.14	-0.3269	-0.518
(D)(G)	0.000528	-0.00271	-0.01232	0.0109	0.0156
(L)(G)	-0.613	-0.608	-1.1818	-1.7156	-0.8909
(D)(RSW)	-0.00681	-0.00422	-0.0048	-0.005	-0.00024
(L)(RSW)	0.29962	-0.31288	-0.04033	-0.33848	-0.5787
(RSW)(G)	-0.00899	0.026336	0.016056	0.046499	0.002549

Table 16a. Bayesian parameter updating of corridor models

Model I	Coeff 85	Coeff 86	NC(85-86)	Coeff 87	NC(85-87)	Coeff 88	NC(85-88)	Coeff 89	NC(85-89)
Constant	-0.2222	-0.0017	-0.10624	0.0985	-0.05217	-0.1709	-0.07529	-0.0871	-0.07765
(L)	1.2929	2.3248	1.836643	3.192	2.194018	5.644	2.86517	5.564	3.403252
(G)	0.14814	0.017	0.079178	0.027	0.065407	-0.09277	0.034613	-0.0317	0.021389
(SWR)	0.04892	-0.0278	0.008896	-0.014	0.002945	-0.0018	0.002033	0.0355	0.007505
(D)	0.01659	0.023	0.02	0.0388	0.02489	0.04832	0.029397	0.0083	0.024963
(D)(L)	-0.0135	0.2236	0.11185	0.1225	0.114657	-0.1127	0.070539	-0.2666	0.003528
(D)(G)	0.00005	0.0072	-0.00372	-0.0153	-0.0067	0.00239	-0.00496	0.0089	-0.00219
(L)(G)	-0.0047	-0.217	-0.11646	0.0209	-0.08022	-0.05131	-0.07459	0.0076	-0.06125
(D)(SWR)	0.00248	-0.0047	-0.00362	-0.00263	-0.00335	-0.00567	-0.00379	0.0003	-0.00298
(L)(SWR)	0.0677	0.018	0.041554	-0.1698	-0.01423	-0.49951	-0.10875	0.5148	-0.18965
(SWR)(G)	-0.0264	0.015	-0.00398	0.0029	-0.00219	0.0292	0.003782	-0.008	0.001529

Model II	**Coeff 85-87**	**Coeff 88**	**NC (85-88)**	**Coeff 89**	**NC (85-89)**
Constant	-0.04187	-0.1709	0.0747	-0.08708	-0.07775
(L)	2.269893	5.644	3.128606	5.056402	3.602319
(G)	0.064327	-0.09277	0.024412	-0.0317	0.010646
(SWR)	0.002245	-0.0018	0.001238	0.035557	0.008159
(D)	0.026278	0.048319	0.031847	0.008242	0.025751
(D)(L)	0.110904	-0.1127	0.054362	-0.2666	-0.02405
(D)(G)	-0.00745	0.002389	0.00496	0.008865	-0.00152
(L)(G)	0.06698	-0.05131	-0.06299	-0.0076	-0.0494
(D)(SWR)	-0.00326	-0.00567	-0.00388	0.000324	-0.00283
(L)(SWR)	-0.02788	-0.49951	-0.1479	-0.51477	-0.23791
(SWR)(G)	-0.00276	0.196327	0.00536	-0.00773	0.002148

Table 16 b. Bayesian parameter updating of tangent models

Model I	Coeff 85	Coeff 86	NC(85-86)	Coeff 87	NC(85-87)	Coeff 88	NC(85-88)	Coeff 89	NC(85-89)
Constant	-0.0587	-0.0112	-0.03264	0.19415	0.022238	-0.375	-0.06072	-0.1274	1.831067
(L)	0.9711	2.198	1.641523	2.9014	1.946508	5.443	2.678405	5.429	3.761319
(G)	0.157	0.039	0.092863	0.0307	0.07775	-0.0395	0.053133	-0.00084	0.532388
(SWR)	0.0469	-0.02531	0.00745	-0.0208	0.000601	1.037	0.00822	0.044	0.390041
(L)(G)	0.061	-0.1956	-0.07889	0.09588	-0.03668	0.03297	-0.02212	0.0245	1.112098
(L)(SWR)	0.0862	0.027	0.053886	-0.1651	0.000903	-0.5148	-0.10718	-0.5155	0.433751
(SWR)(G)	-0.035	0.0127	-0.00893	0.0014	-0.00643	0.0177	-0.00138	-0.01213	0.144993

Model II	**Coeff 85-87**	**Coeff 88**	**NC (85-88)**	**Coeff 89**	**NC (85-89)**
Constant	0.04144	-0.37513	-0.06798	-0.12736	-0.08336
(L)	2.0235	5.443	2.921665	5.429	3.571105
(G)	0.0759	-0.0395	0.04562	0.00084	0.033449
(SWR)	0.00028	0.037	0.009843	0.044	0.018638
(L)(G)	-0.0127	0.032976	-0.0008	0.0245	0.005718
(L)(SWR)	-0.0173	-0.5148	-0.14813	-0.51546	-0.24323
(SWR)(G)	-0.00699	0.0177	-0.0005	-0.01213	-0.00351

Table 16 c. Bayesian parameter updating of curve models

Model I	Coeff 85	Coeff 86	NC(85-86)	Coeff 87	NC(85-87)	Coeff 88	NC(85-88)	Coeff 89	NC(85-89)
Constant	-0.6739	-0.2755	-0.47872	-0.48	-0.47909	-0.5214	-0.48665	-0.6129	-0.509
(L)	1.8531	6.2717	4.017484	7.684	5.09934	14.7	6.81314	12.712	7.8579
(G)	0.1586	-0.042	0.06044	0.08	0.0662	-0.084	0.0394	-0.0687	0.020272
(SWR)	0.0523	-0.01934	0.017534	-0.02121	0.006177	-0.06894	-0.00716	0.0307	-0.00048
(D)	0.0367	0.0269	0.031899	0.0477	0.036558	0.0377	0.036762	0.0234	0.034392
(D)(L)	0.0995	0.1396	0.199109	0.14	0.125292	-0.3269	0.044677	-0.518	-0.05488
(D)(G)	0.00053	-0.0027	-0.00096	-0.01232	-0.00413	0.0109	-0.00154	0.0156	0.00138
(L)(G)	-0.613	-0.608	-0.61055	-1.1818	-0.77909	-1.7156	-0.94625	-0.8909	-0.93645
(D)(SWR)	-0.0068	-0.004	-0.00554	-0.0048	-0.00533	-0.005	-0.00527	-0.00024	-0.00439
(L)(SWR)	0.2996	-0.3129	-0.0004	-0.04033	-0.01219	-0.33848	-0.07043	-0.5787	-0.16047
(SWR)(G)	-0.009	0.0264	0.008316	0.01606	0.010	60.0465	0.01701	0.00255	0.01445

Model II	**Coeff 85-87**	**Coeff 88**	**NC (85-88)**	**Coeff 89**	**NC (85-89)**
Constant	-0.4765	-0.5214	-0.48737	-0.6129	-0.51572
(L)	5.2696	14.7	7.551639	12.7102	8.716785
(G)	0.0656	-0.08399	0.029621	-0.0687	0.007512
(SWR)	0.0039	-0.06894	-0.0137	0.0307	-0.00367
(D)	0.037	0.0377	0.037165	0.0234	0.034117
(D)(L)	0.1264	-0.3269	0.016752	-0.518	-0.104
(D)(G)	-0.0048	0.0109	-0.00097	0.0156	0.00276
(L)(G)	-0.801	- 1.7156	-1.02212	-0.8909	-0.9925
(D)(SWR)	-0.0053	0.005	-0.00523	-0.00024	-0.0041
(L)(SWR)	-0.01786	-0.33848	-0.09546	-0.5787	-0.20463
(SWR)(G)	0.011134	0.046499	0.0197	0.002549	0.015824

Table 17. Bayesian accident prediction models

Corridor

Model I

AR = -0.7765 + 3.403 (L) + 0.0213 (G) + 0.0075 (SWR) + 0.025 (D) + 0.0035 D)(L) - 0.0022 (D)(G) - 0.061 (L)(G) - 0.003 (D)(SWR) -0.2 (L)(SWR) .+0.00153 (SWR)(G)

Model II

AR = -0.07775 + 3.602(L) +0.010646 (G) + 0.0082 (SWR) + 0.025 (D) - 0.0241 (D)(L) - 0015 (D)(G) - 0.049 (L)(G) - 0.00283 -(D)(SWR) - 0.23791 (L)(SWR) + 0.0022(SWR)(G)

Model III

AR = -0.09525 + 3.7343 (L) + 0.01726 (G) + 0.003 (SWR) + 0.02969 (D) - 0.0196 (D)(L) - 0.0023 (D)(G) - 0.0786 (L)(G) -0.002989 (D)(SWR) - 0.21049 (L)(SWR) + 0.00395 (SWR)(G)

Tangents

Model I

AR = 1.831 + 3.76 (L) + 0.532 (G) + 0.39 (SWR) + 1.112 (L)(G) + 0.434 (L)(SWR) + 0.145 (SWR)(G)

Model II

AR = -0.083 + 3.57(L) + 0.033(G) + 0.0186(SWR) + 0.0057 (L)(G) - 0.243(L)(SWR) - 0.0004(SWR)(G)

Model III

AR = - 0.106 + 3.51(L) + 0.044(G) + 0.0144(SWR) - 0.0177(L)(G) - 0.021(L)(SWR) - 0.003(SWR)(G)

Curves

Model I

AR = - 0.51 + 7.858(L) + 0.02(G) - 0.005(SWR) + 0.034(D) - 0.055 (D)(L) + 0.0014(D)(G) -0.94(L)(G) - 0.005(D)(SWR) - 0.16(L)(SWR) + 0.014 (SWR)(G)

Model II

AR = 0.5157 + 8.717 (L) + .0075(G) - 0.0037(SWR) + .034 (D) - 0.104 (D)(L) + .003(D)(G) -1(L)(G) - .004(D)(SWR) - 0.21(L)(SWR) +0.016(SWR)(G)

Model III

AR = -0.494 + 8.64 (L) + 0.0117 (G) - 0.017 (SWR) + 0.034 (D) - 0.087 (D)(L) + 0.0024 (D)(G) - .07 (L)(G) - 0.004(D)(SWR) -0.157 (L)(SWR) + 0.019 (SWR)(G)

Table 18. Comparison of Bayesian updated model predictions & observed accidents

Type	Min.	Max.	Mean	Sum of Accidents
Corridor				
Model I	0	14.27	0.645	190
Model II	0.0011	9.63	0.523	155
Model III	0	9.957	0.555	164
Observed 1990	0	13.41	0.645	190
Tangents				
Model I	0	18.771	1.211	172
Model II	0.019	10.666	0.822	117
Model III	0	9.886	0.737	104
Observed 1990	0	13.405	0.861	122
Curves				
Model I	0	1.199	0.357	55
Model II	0	1.312	0.372	57
Model III	0	1.302	0.385	58
Observed 1990	0	5.97	0.444	68

Table 19. Accident prediction models developed for Provo, Utah using 1989 data

Corridor

AR = 0.1497 + 10.966(L) - 0.171(G) - 0.0239(SWR) - 0.0263(D) + 0.0436 (D)(L) 0.007(D)(G) - 0.852(L)(G) + 0.002(D)(SWR) - 1.127(L)(SWR) +0.0275(SWR)(G)

R^2 = 0.3518

Tangent

AR = -0.203 + 19.815(L) - 0.0492(G) - 0.0033(SWR) - 3.02(L)(G) -1.796(L)(SWR) + 0.02325 (SWR)(G)

R^2 = 0.4932

Curve

AR = 0.862 + 23.918(L) - 0.126(G) - 0.095(SWR) - 0.218(D) - 0.206 (D)(L) + 0.022(D)(G) - 1.307(L)(G) + 0.0272(D)(SWR) - 2.456(L)(SWR) -0.0036(SWR)(G)

R^2 = 0.4095

Table 20a. Bayesian method of parameter updating of southern Utah corridor model

	Prior Parameter N. Utah	Prior STD N. Utah	Sample Parameter S. Utah	Sample STD S. Utah	New Parameter
Constant	-0.0953	2.7609	0.14971	6.56032	-0.0584
(L)	3.73426	6.3477	10.966	50.1185	3.8484
(G)	0.01726	0.78107	-0.1707	1.10724	-0.0452
(SWR)	0.003	0.51523	-0.024	0.97973	-0.0028
(D)	-0.0297	0.2429	-0.263	0.37595	-0.0287
(D)(L)	-0.0196	1.14337	0.04362	0.56938	0.03105
(D)(G)	-0.0024	0.0495	0.00715	0.05383	0.002
(L)(G)	-0.786	1.8795	-0.8519	7.4003	-0.1255
(D)(SWR)	-0.003	0.03079	0.00179	0.04681	-0.0015
(L)(SWR)	-0.2149	0.75766	-1.1273	0.67406	-0.0015
(SWR)(G)	0.00395	0.17624	0.02753	0.16032	0.11859

Table 20b. Bayesian method of parameter updating of southern Utah tangent model

	Prior Parameter N. Utah	Prior STD N. Utah	Sample Parameter S. Utah	Sample STD S. Utah	New Parameter
Constant	-0.1056	2.290	-0.7004	16.2813	2.268
(L)	3.5082	4.6998	41.112	118.288	3.567
(G)	0.0439	0.6656	-0.0882	3.52994	0.0394
(SWR)	0.0144	0.4874	0.0387	2.4649	0.01534
(L)(G)	-0.0177	1.3908	-5.9719	23.224	-0.039
(L)(SWR)	-0.2078	0.5422	-3.666	11.7171	-0.2152
(SWR)(G)	-0.0028	-0.1812	0.0461	0.51539	0.00259

Table 20c. Bayesian method of parameter updating of southern Utah curve model

	Prior Parameter N. Utah	**Prior STD N. Utah**	**Sample Parameter S. Utah**	**Sample STD S. Utah**	**New Parameter**
Constant	-0.4935	4.797	0.8611	11.759	-0.3002
(L)	8.635	30.097	23.918	99.741	9.9106
(G)	0.01177	1.1765	-0.1264	1.5963	-0.0369
(SWR)	-0.0174	0.73053	-0.0949	1.7727	-0.0287
(D)	0.0342	0.2744	-0.2175	0.8921	0.01247
(D)(L)	-0.0848	1.5925	-0.2058	2.5756	-0.1182
(D)(G)	0.0024	0.0547	0.02215	0.08419	0.00826
(L)(G)	0-1.0737	5.32203	-1.307	10.0934	-1.1244
(D)(SWR)	-0.0041	0.0396	0.02716	0.12408	-0.0012
(L)(SWR)	-0.1573	3.77044	-2.4562	14.7612	-0.2981
(SWR)(G)	0.01925	0.18034	-0.0036	0.18829	0.0083

Table 21. Bayesian updated models for southern Utah

Corridor

AR = -0.584 + 3.84843(L) - 0.0452(G) - 0.0028(SWR) - 0.0287(D) + 0.03105(D)(L) + 0.002(D)(G) - 0.1255(L)(G) - 0.0015(D)(SWR) - 0.23(L)(SWR) + 0.01685(SWR)(G)

Tangents

AR = -0.1172 + 3.56749(L) + 0.03939(G) + 0.01534(SWR) - 0.039(L)(G) -0.2152(L)(SWR) + 0.00259(SWR)(G)

Curves

AR = -0.3002 + 9.91059(L) - 0.0369(G) - 0.0287(SWR) + 0.01247(D) - 0.1182(D)(L) + 0.00826(D)(G) - 1.1244(L)(G) - 0.0012(D)(SWR) - 0.2981(L)(SWR) +0.0083(SWR)(G)

Table 22. Comparison of Bayesian updated model predictions & observed accidents

Type	Minimum Accident Rate	Maximum Accident Rate	Mean Accident Rate	Total Accident Rate
Corridor				
1989 Model	0	0.7888	0.23213	17.4098
Updated Model	0	0.08028	0.0734	5.504
Observed 1990	0	2.5265	0.25019	18.7645
Tangent				
1989 Model	0	0.8245	0.15839	5.7021
Updated Model	0.1286	0.83896	0.2979	10.727
Observed 1990	0	2.5265	0.28121	8.09522
Curve				
1989 Model	0	0.94746	0.20042	7.81633
Updated Model	0	1.38398	0.12898	5.03034
Observed 1990	0	1.2632	0.22156	9.31463

APPENDIX J

FIGURES

Figure 1. Distribution of accidents with section length

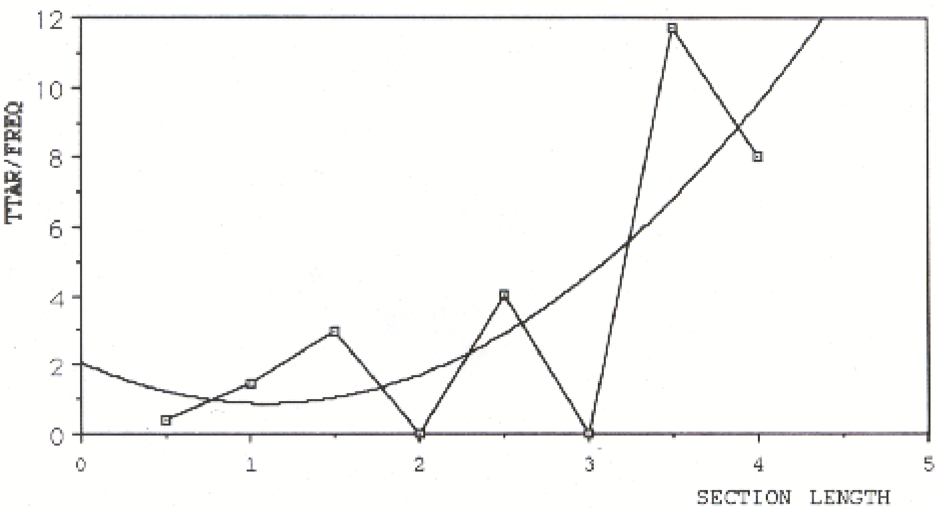


Figure 2. Distribution of degree of curvature and accidents

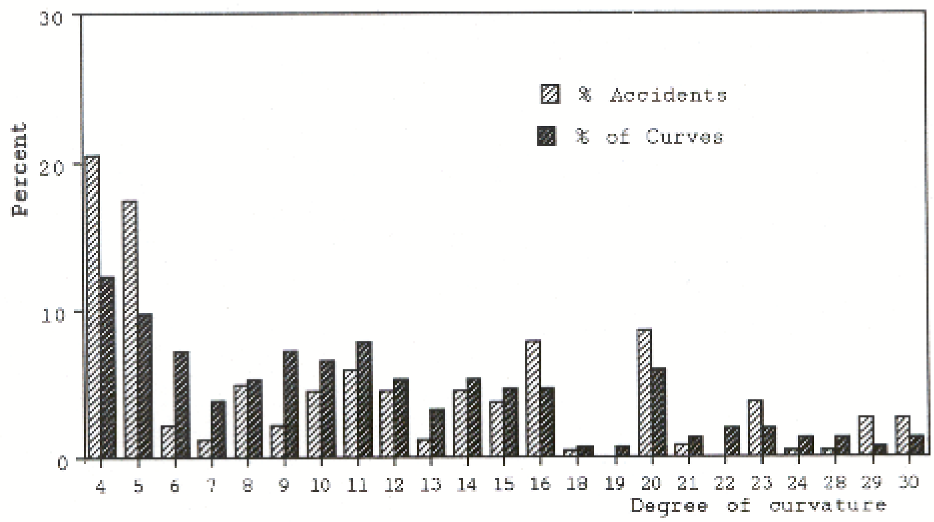


Figure 3. RSI for degree of curvature

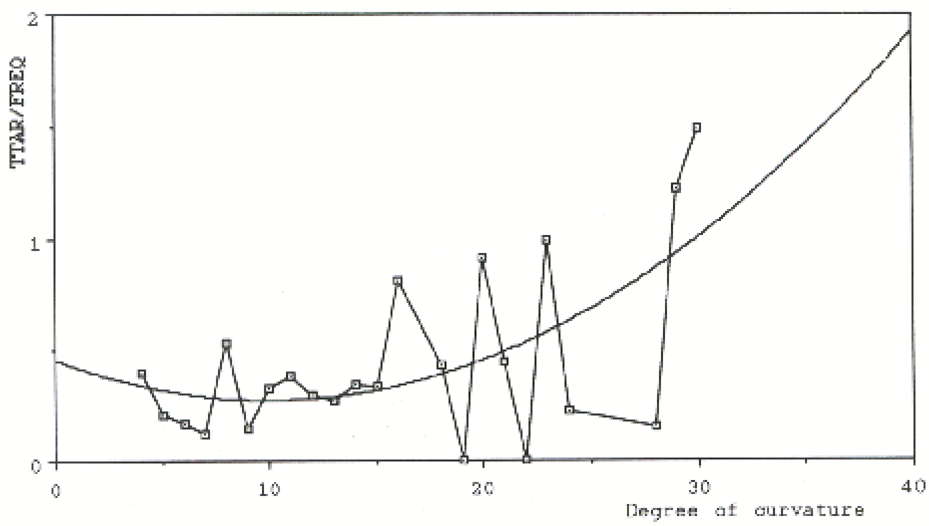


Figure 4. Distribution of vertical grade, frequency, and accident rate

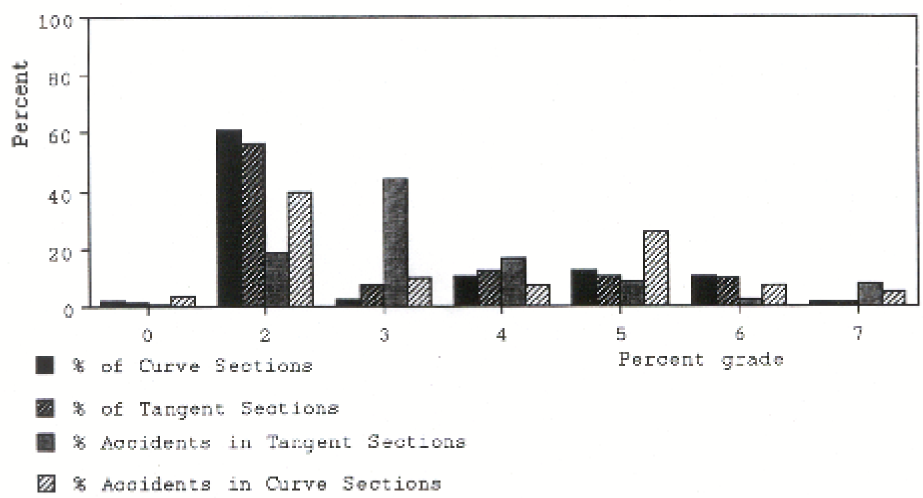


Figure 5. Distribution of shoulder width and accidents

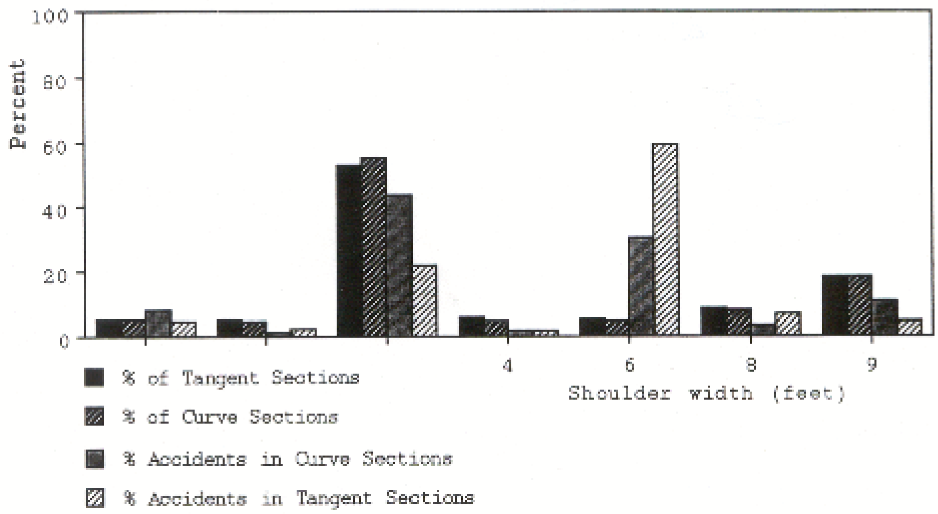


Figure 6. Comparison of accident predictions for corridor

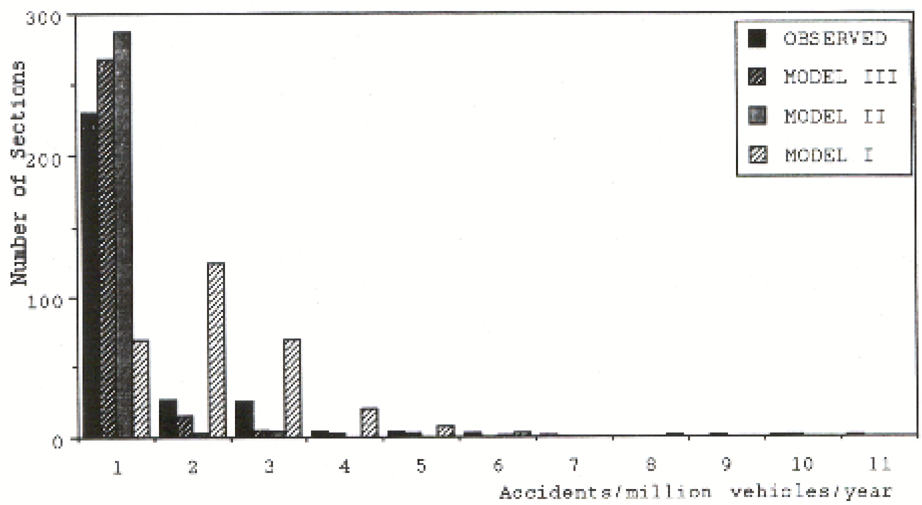


Figure 7. Comparison of accident predictions for tangents

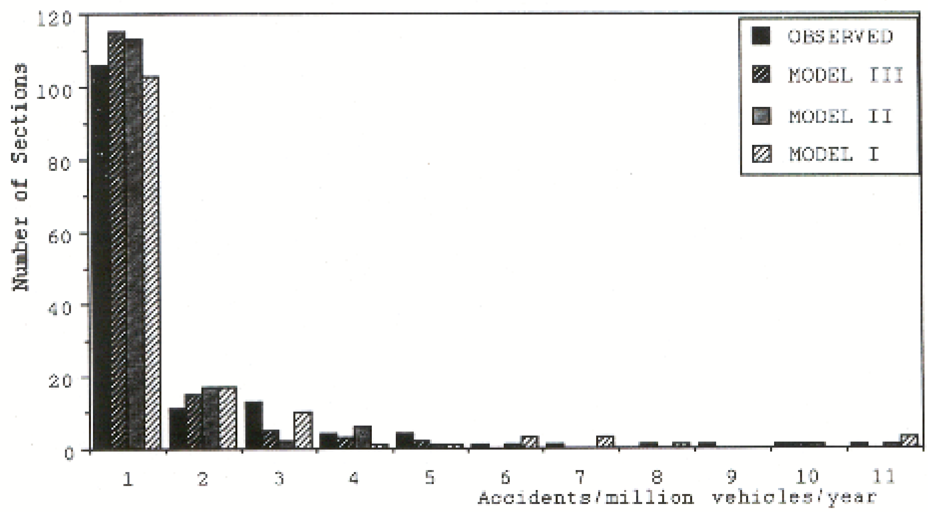


Figure 8. Comparision of accident predictions for curves

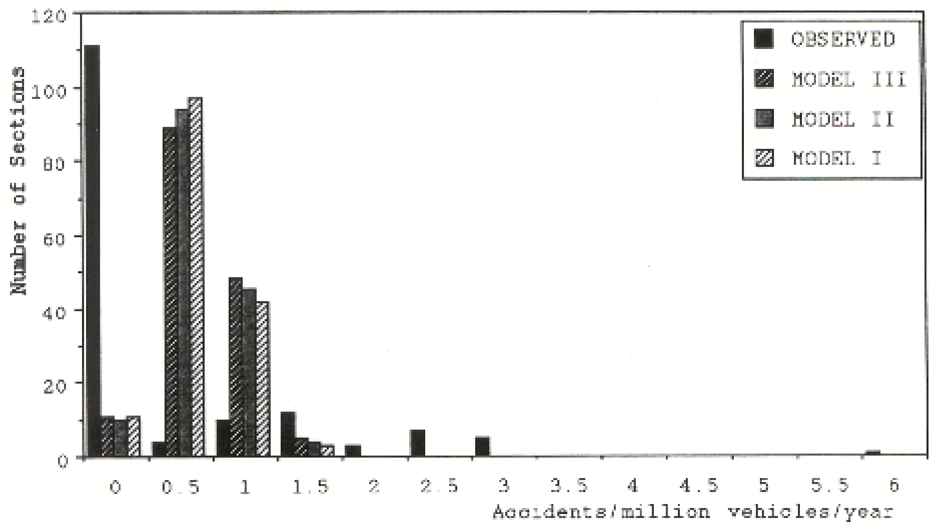


Figure 9. Comparison of predicted accidents for corridors

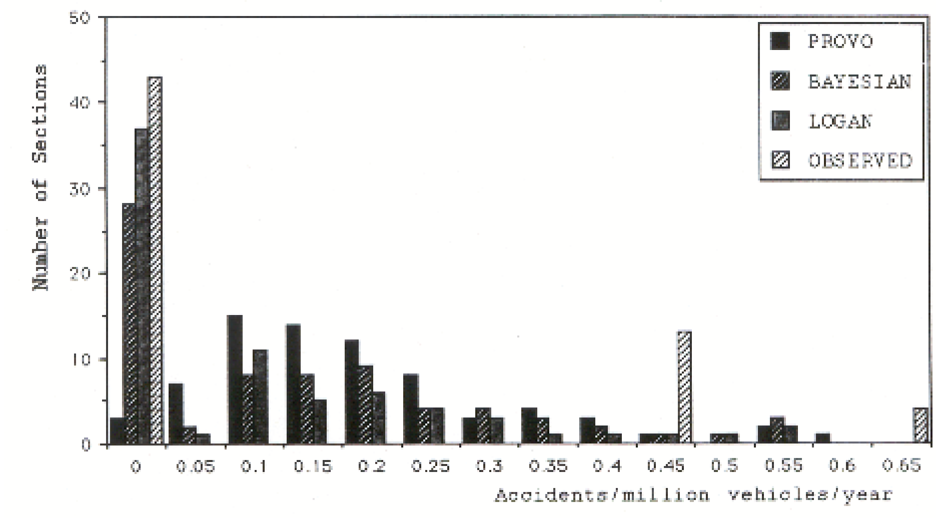


Figure 10. Comparison of accident predictions for tangents.

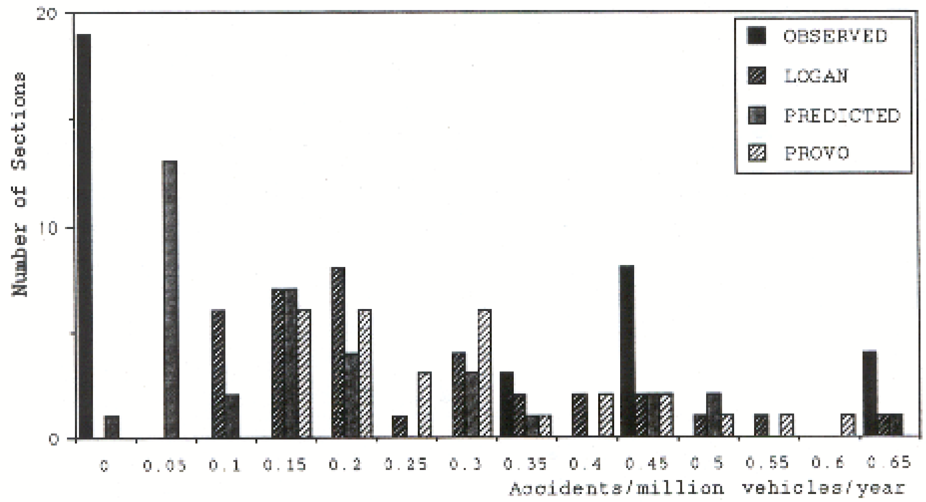
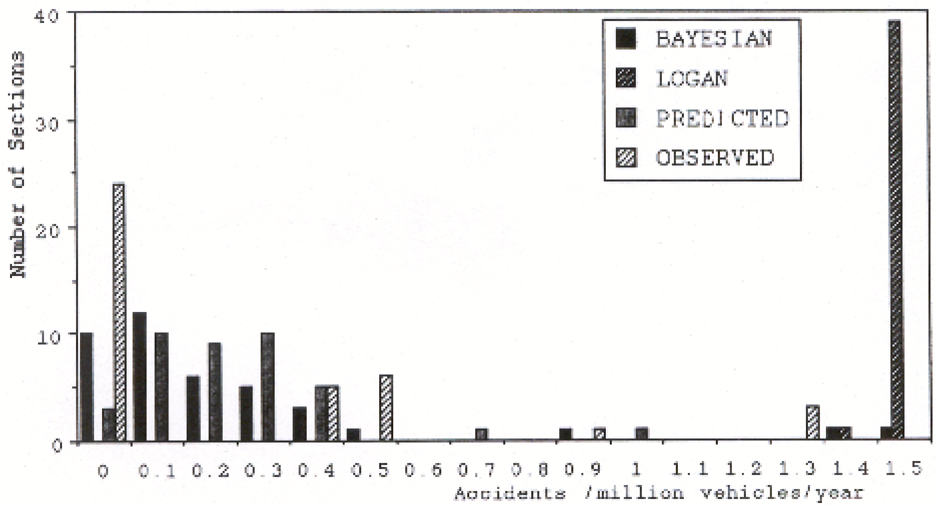


Figure 11. Comparison of accident predictions for curves.



APPENDIX K

RELATED RESEARCH

Geometric Design Variables

After examining several previous studies, Perkins, et al. (2) found that accidents are related to several highway system characteristics that can be classified into two categories.

Operational. Traffic volume, major and minor road volumes, opposing traffic volume, percentage of diverging traffic, traffic mix, volume ratio, posted speed, operating speed, speed difference, speed variation, lateral placement, traffic conflicts, erratic maneuvers, cycle length, signal phasing, number of phases, total stopped vehicle delay, and red and yellow light violations.

Non-Operational. Degree of curvature, frequency of curves, grade, grade continuity, surface cross slope, sight distance, visibility of signal and sign, pavement width, lane width, approach width, pavement shoulder presence, shoulder width, percentage shoulder reduction (at bridges), median width, ratio of bridge width to pavement width, taper length, number of lanes dropped, length of deceleration lane, bridge safety index, structural adequacy of guard rail and bridge rail, access control, number of commercial drive ways per mile, and number of intersections per mile.

After examining the effect of each of the above characteristics, they concluded that:

(1) Average daily traffic (ADT) is the most significant characteristic in the case of injury accidents and fatalities at signalized intersection.

(2) Nonsignalized intersections with higher posted speed limits (50 to 55 mph) are prone to more accidents.

(3) The wider the pavements, the lesser the accidents.

(4) Shoulder width is not significant at curves.

Hirsh, et al. (3) stated that the influence of geometric variables on accidents depends on traffic volume, traffic composition, driver performance, vehicle characteristics, perception-reaction time, and speed distribution.

AASHTO (1) categorized the variables influencing accident occurrence under human factors, geometry, and vehicle. Zegeer, et al. (4) found that about 50 roadway related features could be affect accidents. They included cross sectional elements such as, lane width, shoulder width, shoulder type, road side features (side slope, clear zone, placement and type of road side obstacles), bridge width, median width, number of lanes, passing lanes, and left turning lanes.

Miaou, et al. (27) suggested that the potential factors that make vehicle accident rates differ from one roadway class to another are physical nature of the roadway, such as geometric design, roadway markings, traffic signs, the type of incurred travel, traffic control, and traffic conditions.

Zegeer, et al. (5) found that accidents per mile decrease with the increase in average annual daily traffic (AADT) because higher volumes are associated with higher classes of roads, which normally have wider lanes and shoulders, and less and more gradual curvature than lower-volume facilities. They mentioned that, through lane widening, run-off-the-road and opposite direction accidents can be decreased. They also mentioned that the number of access points per kilometer is associated with accident rates.

Neuman, et al. (33) analyzed 3,304 curve sites from four states (Illinois, Florida, Ohio, and Texas) and focused on the incremental accident effects of five basic variables (average daily traffic (ADT), degree of curvature, length of curve, roadway width, and shoulder width). Analysis of covariance (AOCV) was used to study these incremental effects. This procedure provided a framework that considered both the direct effects of each variable and all of the potential interaction effects between variables. It was found that all variables, except ADT, have a significant influence on accidents.

Cribbins, et al. (8) collected and evaluated field data for 92 highway sites, and records of 6,000 accidents that occurred at these sites. They examined eight variables in detail: (1) median width, (2) speed limit, (3) volume, (4) level of service, (5) access point index, (6) intersection openings per mile, (7) signalized openings per mile, and (8) median openings per mile, for analysis. It was found that the number of access-points, the number of intersections, medians excluding intersections have a heavt influence on accident occurrence.

Dart (15) studied 10 variables to identify their relationship to accidents. They were (1) percent of trucks, (2) traffic volume ratio, (3) lane width, (4) shoulder width, (5) pavement cross slope, (6) horizontal alignment, (7) vertical alignment, (8) percentage of continuous obstruction, (9) marginal obstruction per mile, and (10) traffic access points per mile. It was found from this study that the cross slope and traffic volume ratio were significant.

Previous Accident Prediction Models

Neuman, et al. (6) presented two accident prediction models developed for a FHWA research study (7). The first model is for predicting safety effects using cross sectional elements as independent variables. This model, given below, is recommended as the basis for studying the sensitivity of accident rates to lane and shoulder width combinations.

$$A = 0.0019(ADT)^{0.8824}(0.8786)^{W}(0.9192)^{PA}(0.9316)^{UP}(1.2365)^{H}(1.3221)^{TER2}$$

where

A	=	number of run-off road, head-on, opposite-direction sideswipe,and same-direction sideswipe accidents per mile per year (termed “related accidents”)
ADT	=	two-directional average daily traffic volume (vpd)
W	=	lane width in feet
PA	=	width of paved shoulder in feet
UP	=	width of unpaved (gravel, turf, earth) shoulder in feet
H	=	median roadside hazard rating for the highway segment, measured subjectively on a scale of 1 (least hazardous) to 7 (most hazardous)
TER2	=	1 for mountainous terrain, zero otherwise

The second model, given below, is recommended for assessing the safety effects of curvature:

$$A = [(1.552)(L)(V) + 0.014(D)(V) - (0.012)(S)(V)](0.978)^{W-30}$$

where

A	=	number of total accidents on the curve in a 5-year period
L	=	length of the curve in miles (or fraction of mile)
V	=	volume of vehicles in million vehicles in a 5-year period passing through the curve (in both directions)
D	=	degree of curvature
S	=	presence of spiral where S = 0 if no spiral exists, and S = 1 if spiral exists
W	=	width of the roadway on the curve in feet

With these models, a series of tables expressing accident reduction factors for various combinations of geometric features is provided. According to the authors, those factors helped them estimate the safety effectiveness of any proposed design. Moreover, they found no significant relationship between accidents and vertical alignment, stopping sight distance, and horizontal alignment consistency.

Zegeer, et al. (31) conducted a study of accidents that occurred over 5 years and calibrated the following curve accident prediction models:

$$\text{Total accident rate} = 1.94 + 0.24\,D - 0.26\,W - 0.25\,S$$

$$\text{Total accidents/curve} = (ADT)(L)\,(1.94 + 0.24\,D - 0.026\,W - 0.25\,S)$$

where

D	=	degree of curvature
W	=	lane width
S	=	spiral

L = section length
ADT = average daily traffic

Cribbins, et al. (8) performed a multiple-regression analysis to evaluate the simultaneous effects of all the site characteristics on accident frequency. A Student's t-value was calculated for each regression coefficient. This t-value indicated whether the variable corresponding to the coefficient had a significant effect on the accident rate or not.

The method of least squares was used to determine the existence of any linear relationships of the form Y = A+BX. The values of the sum of squares, R^2, and standard errors of Y were calculated. If these values were high enough to give some indication of linear dependence, then they were included in the multivariate model. After a stepwise regression analysis, only five variables were found to be significant. The final equation is given as:

$$Y = -28.34191 + 0.00011X_1 + 3.28169X_3 + 0.34218X_6 + 0.0005X_7 + 7.34777X_8$$

$$R^2 = 0.69$$

where

X_1 = access-point index
X_3 = signalized openings per mile
X_6 = speed limit
X_7 = volume
X_8 = level of service

Another equation relating total accidents to injury accidents given in the article is as follows:

$$Y = -0.02216 + 0.30227X$$

where

Y = total accidents
X = injury accidents

Cribbins, et al. (8) suggested that the number of access points, the number of intersections, and the number of medians, excluding intersections, have a substantial influence on accident occurrence. It was concluded that whenever storage lanes are installed at a median opening, the accident rate is not significantly affected by the number of openings, excluding intersections, median width, speed-limit, or ADT.

Glauz, et al. (9) proposed that expected accidents could be predicted for an intersection by using conflict data from that intersection with the help of accident and conflict data of other intersections of the same class (signalization and volume level). The two equations given are:

$$A_o = C_o R$$
$$Var(A_o) = Var(C)Var(R) + C_o^2 Var(R) + Var(C)$$

where

A_o = expected number of accidents
Co = expected conflict rate obtained from the field study at the intersection
Ro = estimate of the accident/conflict ratio for that class of intersections

The following minimum variance prediction model was also given by the authors:

$$A_m = [A_o / Var(A_o) + A_a / Var(A_a)]Var(A_m)$$
$$Var(A_m) = 1/[1/Var(A_o) + 1/Var(A_a)]$$

It was mentioned that the usable conflict types are signalized intersections (same direction, opposing left turn), unsignalized intersections (through cross traffic from the left and right), and unsignalized intersections, medium volume only (opposing left turns and left turn same direction).

Allen (10) dealt with eight major factors that contribute to the occurrence of accidents. He suggested that the multivariate analysis technique can be used to analyze the data. The principal multivariate method used in this study was a factor analysis of the correlation coefficients.

The factors were extracted by the principal components method and rotated by the matrix method and the varimax method. The biquartamin method was also used since those methods required orthogonal (independent) factors. He suggested to refer to the correlation matrix and to compare the correlations between the variables with the loadings on the factor to interpret the data.

According to Allen (10), Schoppert, et al. (11) and Blensly, et al. (12) also used multivariate analysis to study the effects of roadway characteristics on the accident rates of a section of highway in Oregon. Goldstein, et al. (13) analyzed attitude items and related these factors to self-reported accidents and violations. Versace (14) also used the factor-analysis technique to analyze a portion of Oregon data.

Dart (15) calculated the squares, first order interactions of 10 variables, and used regression analysis to determine the contribution of the variables to total accidents on wet roads, dry roads, accidents during the day, and night. The regression model given is as follows:

$$Y = 41.32-1.23X_1-0.54X_2 - 0.67X_6+ 0.03X_1X_2 + 0.03X_2X_6 - 0.0009X_2X_9 + 0.026X_2X_{11} - 0.12X_4X_{11} + 0.009X_5X_9$$

where

Y = total accidents per 100 million vehicle miles
X_1 = percentage of trucks
X_2 = traffic volume ratio or peak hour volume/service volume at level "B".
X_6 = cross slope
X_1X_2 = (percentage of trucks)(traffic volume ratio)
X_2X_6 = (traffic volume ratio)(cross slope)
X_2X_9 = (traffic volume ratio)(horizontal alignment)
X_2X_{11} = (traffic volume ratio)(traffic conflicts)
X_4X_{11} = (lane width)(traffic conflicts)
X_5X_9 = (shoulder width)(horizontal alignment)

In the above study, Dart (15) found that cross slope effect is very significant in the Louisiana environment because of high rainfall. In all cases, the traffic volume ratio played an important role. An illustration of the study done in this field is also given in this article. Correlation procedures were used by Blensly (12) to evaluate the relationship between the paved shoulder width and accident occurrence, and variance measures were employed to analyze the difference between the average accident frequency on sections with narrower paved shoulders (4 ft or less) and the average accident frequency on sections with wider paved shoulders (8 ft or more).

The analysis of covariance established that when the effect of average daily traffic (ADT) was controlled, the mean numbers of property damage and total accidents are significantly higher on sections with wide paved shoulders than on sections with narrow paved shoulders in the 1,000 to 5,600 ADT range.

Schopperts (11) used a sample of 1,400 miles of two-lane highways and analyzed the data using the multiple correlation technique. Roy Jorgensen and Associates and Westat Research, Inc. (32) used regression analysis to develop accident prediction models and also appealed for accident record data to be coordinated with existing highway geometric data.

Mulinazzi, et al. (16) used multiple regression analysis techniques to relate accident rates to design characteristics and concluded that the accident rate increases with the number of frictional points per mile (the sum of the number of approaches to the arterial, intersection, and driveways).

In his article, Levonian (17) emphasized that multivariate analysis not only provides for statistical control, but also retains the reliability associated with the entire sample. He said that the prediction of accidents could be enhanced when several predictors were used and the results could be interpreted very easily by this method.

Lawson (18) did a study on the accident records of Birmingham radial routes according to the land use adjacent to the road, the type of carriageway, and the daily traffic flows on the roads. He adopted the generalized analysis of variance technique to find the significance of the terms in the data. This enabled him to find the significance of each factor on accidents while controlling the other factors.

Lawson (18) determined the significance of each variable by removing or adding it and examining the change in the scaled deviance. This difference in scaled deviance was then compared with 95 % of the chi-squared distribution. He also noticed that Poisson and negative binomial distribution techniques were superior to normal distributions in describing accident data. A software package called "GLIM" was used for the above analysis.

According to Lawson (18), the widespread availability of packages such as GLIM meant that it is convenient to investigate the nature of the data before analysis. He maintains that one can then select the appropriate error structure rather than make the assumptions of normality associated with many analysis techniques.

Tamburri, et al. (19) in their study of accident predictions and cost analysis assumed future accident rate to have the same ratio as the present accidents to traffic volume. This methodology was applied to both freeways and rural expressways (not urban expressways). For other existing roads, the accident rate was assumed to be the same as the current rate regardless of the traffic increases. When lanes were added without improvements to geometrics, it was assumed that the accident rate of the widened freeway will be a given percentage of the rate it would have been without widening. Some of the accident prediction models given in the article are as follows:

Type of Facility	Prediction Model
Existing conventional highway and urban expressway	Current rate of existing highway
Existing rural expressway	$(A/B).C^a$
Existing Freeway	$(A/B).C^a$

where

A = current accident rate
B = statewide average accident rate for current ADT
C = statewide average accident rate for future ADT
a = exponent

The above author also suggested two alternatives to compensate for small samples. First is to extend the time period, provided there was no substantial highway or environmental changes during extended periods. Second is to use the average accident rate of a longer section with similar characteristics to predict accidents.

Cleveland, et al. (22) adopted advanced multivariate analysis techniques to study the interactions between the variables and the accidents. Michigan FHWA skid data and Michigan State Route data were used in the analysis.

The first stage of the analysis consisted of a statistical examination of data to identify the variables that contributed most to the variation of accident occurrence by using the automatic interaction detection (AID) approach suggested by Sonquit, et al. (25). According to them, AID is a scheme that searches for dichotomous split level for an independent variable that gives the maximum improvement in the ability to predict values of the dependent variable through unexplained variance reduction. They concluded that the ADT was more helpful in the prediction of models than the conventional vehicle mile exposure rate.

The second stage of the Cleveland, et al. (22) study identified reasonable groups of correlated variables using the factor analysis method proposed by Harman (26). The third stage involved grouping interrelated roadside and geometric elements into reasonable bundles. The fourth step determined the ability of the bundles to explain the variations in accidents, and compared it with the variations explained by individual geometric design elements.

The fifth stage explored the characteristics of the segments with the best and the worst accident experience. Finally, in the sixth stage, illustrative categorical models were developed. The models were given in the form of tables rather than as equations. It was stated that the models developed explained 63% of the variation in the total accidents in the Michigan-FHWA SID data and 54% of the off-road accident variance in the Michigan State Route data set.