PB99-142713

Accident Models for Two-Lane Rural Roads: Segments and Intersections

PUBLICATION NO. FHWA-RD-98-133

OCTOBER 1998

U.S. Department of Transportation
Federal Highway Administration

Research and Development
Turner-Fairbank Highway Research Center
6300 Georgetown Pike
McLean, VA 22101-2296

REPRODUCED BY: NTIS
U.S. Department of Commerce
National Technical Information Service
Springfield, Virginia 22161

FOREWORD

This report is a direct step for the implementation of the Accident Analysis Module in the Interactive Highway Safety Design Model (IHSDM). The Accident Analysis Module is expected to estimate the safety impact of two-lane rural highway characteristics for existing and new projects. Several accident models are developed to estimate accident frequencies. The three main models are for road segments (with non-intersection accidents), one-way stop-controlled intersections with three legs, and two-way stop-controlled intersections with four legs. This report describes the collection, analysis, and modeling of accidents on rural roads in Minnesota (1985-1989) and Washington State (1993-1995).

Models of the Poisson type, negative binomial type, and extended negative binomial type are developed, and advanced statistical techniques are applied to assess the explanatory value of the models in the presence of Poisson randomness and overdispersion. The models derived from these data indicate that exposure and traffic counts are the chief highway variables contributing to accidents. Other variables that affect accidents on road segments are: lane width, shoulder width, horizontal and vertical alignments, roadside conditions, and driveway density. Other variables that affect accidents at intersections are: vertical and horizontal alignments, roadside conditions, number of driveways, posted speed, approach angles, and turning lanes.

A. George Ostensen, Director
Office of Safety and Traffic Operations
Research and Development

PROTECTED UNDER INTERNATIONAL COPYRIGHT
ALL RIGHTS RESERVED.
NATIONAL TECHNICAL INFORMATION SERVICE
U.S. DEPARTMENT OF COMMERCE

Reproduced from best available copy.

NOTICE

This document is disseminated under the sponsorship of the Department of Transportation in the interest of information exchange. The United States Government assumes no liability for its contents or use thereof. This report does not constitute a standard, specification, or regulation.

The United States Government does not endorse products or manufacturers. Trade and manufacturer's names appear in this report only because they are considered essential to the object of the document.

Technical Report Documentation Page

1. Report No.	2.	3. Recipient's Catalog No.
FHWA-RD-98-133	PB99-142713	

4. Title and Subtitle	5. Report Date
ACCIDENT MODELS FOR TWO-LANE RURAL ROADS: SEGMENTS AND INTERSECTIONS	October 1998
	6. Performing Organization Code
7. Author(s)	8. Performing Organization Report No.
Andrew Vogt and Joe G. Bared	
9. Performing Organization Name and Address	10. Work Unit No. (TRAIS)
Pragmatics, Inc. 8301 Greensboro Drive, Suite 225 McLean, Virginia 22102	3A5A
	11. Contract or Grant No. DTFH61-94-Z-00124
12. Sponsoring Agency Name and Address	13. Type of Report and Period Covered
Office of Safety and Traffic Operations R&D Federal Highway Administration 6300 Georgetown Pike McLean, Virginia 22101-2296	Final Report August 1994 - September 1997
	14. Sponsoring Agency Code

15. Supplementary Notes

Contracting Officer's Technical Representative (COTR): Joe G. Bared, HSR-20

16. Abstract

This report describes the collection, analysis, and modeling of accident and roadway data pertaining to segments and intersections on rural roads in the States of Minnesota (1985-1989) and Washington (1993-1995). The segments are on two-lane roads, and the intersections are three-legged and four-legged intersections of such roads, stop-controlled on the minor legs. Data were acquired from the Highway Safety Information System, photologs, construction plans, and State data bases. More than 1,300 segments and more than 700 intersections are included in the final samples on which the modeling is based. Variables collected include accident counts, traffic exposure, surface and shoulder width, Roadside Hazard Rating, number of driveways, channelization, horizontal and vertical alignments, intersection angles, speed limits, and commercial traffic percentage.

Models of Poisson type, negative binomial type, and extended negative binomial type (the latter due to Shaw-Pin Miaou) are developed, and advanced statistical techniques are applied to assess the explanatory value of the models in the presence of Poisson randomness and overdispersion.

The models derived from these data indicate that exposure and traffic counts are the chief highway variables contributing to accidents, but that surface and shoulder width, roadside conditions, and alignments are also significant, especially in the segment models. Unexpected behavior of intersection angle, Roadside Hazard Rating, number of driveways, and channelization in the intersection models is worthy of note. Despite the incompleteness of the data and uncertainties in the values of some variables, the quantity, quality, and variety of the data give the models both descriptive and predictive value.

17. Key Words	18. Distribution Statement
Highway safety, Accident prediction models, Negative binomial regression, Extended negative binomial models, Highway geometric design	No restrictions. This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	179	

Form DOT F 1700.7 **Reproduction of completed page authorized**

ACKNOWLEDGMENTS

The authors express appreciation to staff members of the Federal Highway Administration's Turner-Fairbank Highway Research Center, the Highway Safety Information System, the Minnesota and Washington Departments of Transportation, the Midwest Climate Center, and Pragmatics, Inc. (especially Highway Engineers Rakesh Gupta and Jian Song). We also thank Ezra Hauer and Shaw-Pin Miaou for assistance.

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
		AREA		
in^2	square inches	645.2	square millimeters	mm^2
ft^2	square feet	0.093	square meters	m^2
yd^2	square yards	0.836	square meters	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	square kilometers	km^2
		VOLUME		
fl oz	fluid ounces	29.57	milliliters	mL
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.028	cubic meters	m^3
yd^3	cubic yards	0.765	cubic meters	m^3
NOTE: Volumes greater than 1000 l shall be shown in m^3.				
		MASS		
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams (or "metric ton")	Mg (or "t")
		TEMPERATURE (exact)		
°F	Fahrenheit temperature	5(F-32)/9 or (F-32)/1.8	Celcius temperature	°C
		ILLUMINATION		
fc	foot-candles	10.76	lux	lx
fl	foot-Lamberts	3.426	candela/m^2	cd/m^2
		FORCE and PRESSURE or STRESS		
lbf	poundforce	4.45	newtons	N
lbf/in^2	poundforce per square inch	6.89	kilopascals	kPa

APPROXIMATE CONVERSIONS FROM SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
m	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
		AREA		
mm^2	square millimeters	0.0016	square inches	in^2
m^2	square meters	10.764	square feet	ft^2
m^2	square meters	1.195	square yards	yd^2
ha	hectares	2.47	acres	ac
km^2	square kilometers	0.386	square miles	mi^2
		VOLUME		
mL	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m^3	cubic meters	35.71	cubic feet	ft^3
m^3	cubic meters	1.307	cubic yards	yd^3
		MASS		
g	grams	0.035	ounces	oz
kg	kilograms	2.202	pounds	lb
Mg (or "t")	megagrams (or "metric ton")	1.103	short tons (2000 lb)	T
		TEMPERATURE (exact)		
°C	Celcius temperature	1.8C + 32	Fahrenheit temperature	°F
		ILLUMINATION		
lx	lux	0.0929	foot-candles	fc
cd/m^2	candela/m^2	0.2919	foot-Lamberts	fl
		FORCE and PRESSURE or STRESS		
N	newtons	0.225	poundforce	lbf
kPa	kilopascals	0.145	poundforce per square inch	lbf/in^2

* SI is the symbol for the International System of Units. Appropriate rounding should be made to comply with Section 4 of ASTM E380.

(Revised September 1993)

TABLE OF CONTENTS

TABLE OF CONTENTS
(continued)

Page

TABLE OF CONTENTS
(continued)

Page

LIST OF TABLES

LIST OF TABLES
(continued)

LIST OF FIGURES

INDEX OF VARIABLES

INDEX OF VARIABLES
(continued)

1. INTRODUCTION

Estimating the number of accidents that may result for a given highway design is a matter of great interest to the highway engineering community. Numerous studies have been performed in this area (see McGee et al.[1] and references cited therein) with the aim of determining the effects of different design elements and their relative importance. Since safety is a primary consideration in highway design, the safety consequences of highway design features have been and will remain a matter of continuing interest.

The present study was undertaken in connection with the development of the Interactive Highway Safety Design Model (IHSDM). The IHSDM is envisioned as a set of tools to assist the highway designer. In particular it is expected to include an Accident Analysis Module that will relate accidents to highway variables along segments and at intersections. Rural roadways tend to have high accident rates,[2] and adequate models for these roadways are especially desirable. This study focuses on segments of rural two-lane roads and on three- and four-legged intersections on such roads, stop-controlled on the minor leg or legs.

The study makes use of Highway Safety Information System (HSIS) data for two States, Minnesota and Washington. Accident data (including both severity and type), traffic data, lane and shoulder width data, and some alignment data are available in HSIS files. Data were also obtained from photologs and, in the case of Minnesota, construction plans. These data include horizontal and vertical alignments, channelization, driveways, and Roadside Hazard Rating. The latter is a measure of sideslope and clear zone proposed by Zegeer et al. (1987).[3]

The analysis and modeling on the data sets have been performed with SAS® software. SAS includes a variety of procedures for summarizing univariate and multivariate statistics and for modeling the relationship between a variable such as number of accidents and covariates such as traffic volumes and highway design variables.

Accident models are typically of Poisson and generalized linear form. The number of accidents

---

[1] McGee, H.W., Hughes, W.E., and Daily, H., "Effect of Highway Standards on Safety," National Cooperative Highway Research Program, Report 374, Transportation Research Board, National Research Council, Washington, D.C., 1995.

[2] Tessmer, J.M., "Rural and Urban Crashes: A Comparative Analysis," Technical Report DOT-HS-808-450, National Highway Traffic Safety Administration, United States Department of Transportation, Washington, D.C., 1996.

[3] Zegeer, C.V., Hummer, J., Reinfurt, D., Herf, L., and Hunter, W., "Safety Cost-Effectiveness of Incremental Changes in Cross-Section Design – Informational Guide," FHWA-RD-87-094, Federal Highway Administration, Washington, D.C., 1987.

in a given space-time region is regarded as a random variable that takes values 0, 1, 2, ... with probabilities obeying the Poisson distribution. A characteristic feature of this distribution is that the variance, or mean squared deviation of this variable, is equal to its mean. The mean number of accidents is assumed to be an exponential applied to a suitable linear combination of highway variables. Thus the model falls under the heading of a generalized linear model. The exponential function guarantees that the mean is positive.

More recently negative binomial models, a variant of the Poisson, have been used in accident modeling. Such models generalize the Poisson form by permitting the variance to be overdispersed, equal to the mean plus a quadratic term in the mean whose coefficient is called the overdispersion parameter. When this parameter is zero, a Poisson model results. When it is larger than zero, it represents variation above and beyond that due to the highway variables present in the model. Such variation is due to accident-related factors pertaining to drivers, vehicles, and location not encompassed by the highway variables. The LIMDEP® software package, or SAS-based programs, can be used to develop negative binomial models.

In addition, Shaw-Pin Miaou has developed an "extended" negative binomial model that permits variables with multiple values along a roadway to be treated in disaggregate form, value by value, rather than in aggregate form, by averages over the whole roadway. Highway segments are not truly homogeneous even if shoulder widths, lane widths, speed limits, and the like stay constant along them. Other variables, such as horizontal and vertical alignments, are subject to variation within the typical segment. The extended negative binomial model aims to capture the effect of such inhomogeneities.

In the following chapters the literature is reviewed; the data collection methodology is described in detail; the data analysis is presented; accident models of Poisson, negative binomial, and extended negative binomial type are exhibited; and validation and additional analyses are performed. The modeling chapter includes logistic modeling of accident severities on the Minnesota data. The last chapter presents the final models (obtained earlier in Tables 27 and 35) in the form of equations and exhibits associated Accident Reduction Factors. Two appendices offer additional information about the Minnesota population and the final model equations in metric form, respectively.

Some of the results in this report are to be found in the article by Vogt and Bared (1998).[4]

---

[4] Vogt, A., and Bared, J.G., "Accident Models for Two-Lane Rural Segments and Intersections," Transportation Research Board, TRR 1635, Washington, D.C., 1998, to appear.

2. LITERATURE REVIEW

This chapter surveys the modeling literature pertaining to highway segments and intersections and reviews variables used in past studies. It also includes a discussion of artificial neural networks.

SEGMENT MODELS

Miaou et al. (1993)[5] used a model of Poisson type to estimate accidents along highway segments. Although the model was applied to truck accidents, it is applicable to other vehicles on a highway. Poisson regression provides one of the most suitable models because vehicle accidents are discrete rare events and accident counts are nonnegative integers. Accidents are usually positively skewed because of the high proportion of highway segments without accidents. Poisson regression models provide an easy linkage to probability, as opposed to other commonly used models such as multiple linear regression. The form of the model is:

$$P(y_i) = \frac{(\lambda_i v_i)^{y_i} e^{-\lambda_i v_i}}{y_i!}$$

where

y_i is the number of trucks involved in accidents on the i-th two-lane undivided highway segment for a given exposure;

$P(y_i)$ is the probability that y_i trucks will be involved in accidents;

λ_i is the mean accident rate (in number of trucks per million truck-miles) on the i-th segment; and

v_i is the truck exposure (in millions of truck-miles) on the i-th segment.

In this formulation λ_i is estimated by:

$$\lambda_i = \exp(0.0818 + 0.1022x_{1i} + 0.0949x_{2i} + 0.0426x_{3i} + 0.0341x_{4i} - 0.0263x_{5i}) .$$

---

[5] Miaou, S.-P., Hu, P.S., Wright, T., Davis, S.C., and Rathi, A.K., "Development of Relationship between Truck Accidents and Geometric Design: Phase I," FHWA-RD-91-124, Federal Highway Administration, Washington, D.C., 1993.

For the i-th segment

x_{1i} = Average daily traffic per lane (in thousands of vehicles)
x_{2i} = Horizontal curvature (in degrees per hundred feet)
$x_{3i} = x_{2i} \times$ horizontal curve length (in miles)
x_{4i} = Deviation of stabilized outside shoulder width per direction from 12 ft (in feet)
x_{5i} = Percent trucks in traffic stream.

The estimated value of λ_i is always non-negative and is represented by a loglinear function of explanatory variables x_{ji} related to geometry, traffic, and other highway characteristics. With respect to the underlying Poisson assumption that the mean equals the variance, the model for two-lane rural segments is not very satisfactory since the estimated ratio of variance to mean, 1.36, is not close to one. A negative binomial regression model was proposed to allow for overdispersion, with variance equal to mean μ_i plus an extra term of the form $K(\mu_i)^2$. The quantity K is the overdispersion parameter. The regression coefficients in the negative binomial model are similar to those of the Poisson model. However, the negative binomial allows for additional variance representing the effect of omitted variables.

Poisson and negative binomial modeling techniques are believed to be robust and quite suitable for accident modeling. One weakness of the above model, though, is the minuscule frequency of truck accidents, since they constitute a very small proportion of total accidents, even though the highway sample of 14,731 lane-miles extending over a 5-year period is large. Another weakness may be ascribed to a highly significant variable, truck ADT(Average Daily Traffic). This variable was acquired from the Highway Performance and Monitoring System (HPMS), a separate data source that was integrated with the original data. Whether the values of truck ADT were sufficiently local to represent the truck traffic on a given segment adequately is not known.

The report of Luyanda et al.[6] utilized a variety of multivariate statistical techniques to investigate relationships between the major factors of rural highway conditions and accident occurrences. Cluster analysis, discriminant analysis, factor analysis, and linear regression were applied in stepwise fashion. Highway segments were divided into three groups: multi-lane segments, two-lane segments in flat and rolling terrain, and two-lane segments in hilly terrain. Comparisons were made between groups and within groups. Within the multi-lane segments, the significant variables identified by discriminant analysis were different from those identified by stepwise regression. For the other two groups, the R^2 values[7] were disappointingly low, 0.23 and 0.07, respectively. The report should be

---

[6] Luyanda, F., Smith R.W., Padron, M., Resto, P., Gutierrez, J., and Fernandez, L., "Multivariate Statistical Analysis of Highway Accident and Highway Conditions," University of Puerto Rico, Mayaguez School of Engineering Research Center, Report DOT-RSPA-DMA-50/84/9, Puerto Rico, 1983.

[7] R^2 is the coefficient of multiple determination, defined in Chapter 5 below.

regarded as exploratory because of uncertainties in accident location and the small sample size. Although the results of the discriminant analysis seem to be reliable, they do not give a safety evaluation, but rather a classification by grouping. The assumption of linearity in the regression analysis is simplistic and should be refined. Moreover, highway segments and intersections were not differentiated to permit classification of accidents into segment accidents or intersection accidents.

The reports of Zegeer et al. (1986), Mak (1987), and Zegeer et al. (1991)[8] applied regression techniques to develop accident models for two-lane roads. The model for cross-section safety on two-lane highways proposed by Zegeer et al. (1986) is:

$$A = 0.0019(ADT)^{0.8824}(0.8786)^{W}(0.9192)^{PA}(0.9316)^{UP}(1.2365)^{H}(0.8822)^{TER1}(1.3221)^{TER2}$$

where

A = number of accidents per mile per year
ADT = two-directional average daily traffic
W = lane width in feet
PA = width of paved shoulder in feet
UP = width of unpaved shoulder in feet
H = median roadside hazard rating
$TER1$ = 1 for flat terrain, 0 otherwise
$TER2$ = 1 for mountainous terrain, 0 otherwise.

---

[8] Zegeer, C.V., Hummer, J., Reinfurt, D., Herf, L., and Hunter, W., "Safety Effects of Cross-Section Design for Two-Lane Roads," FHWA-RD-87-008, Federal Highway Administration, Washington, D.C., 1986.

Mak, K.K., "Effect of Bridge Width on Safety," State of the Art Report 6, Relationship Between Safety and Key Highway Features – A Synthesis of Prior Research, Transportation Research Board, National Research Council, Washington, D.C., 1987.

Zegeer, C.V., Stewart, R., Reinfurt, D., Council, F., Neuman, T., Hamilton, E., Miller, T., and Hunter, W., "Cost Effective Geometric Improvements for Safety Upgrading of Horizontal Curves," FHWA-RD-90-021, Federal Highway Administration, Washington, D.C., 1991.

The accidents considered in this model are single vehicle accidents, head-on accidents, and same and opposite direction sideswipe accidents.

A quadratic model for accidents on bridges was developed by Mak (1987):

$$Y = 0.50 - 0.061RW + 0.0022(RW)^2$$

where

Y = number of accidents per million vehicles
RW = relative bridge width (bridge width minus width of traveled way) in feet.

Zegeer et al. (1991) developed a model for accidents on horizontal curves:

$$A_h = [1.552L \times V + 0.14D \times V - 0.12S \times V](0.978)^{(W - 30)}$$

where

A_h = total number of accidents on a horizontal curve in a 5-year period
L = length of the curve (in miles)
V = volume of vehicles in a 5-year period (in millions of vehicles)
D = degree of curve (in degrees per hundred feet)
S = 1 for a spiral curve, 0 for no spiral
W = roadway width including shoulder widths (in feet).

The last-mentioned study, Zegeer et al. (1991), reviewed data base characteristics, determined the important variables through a preliminary analysis, and then proceeded to model building. The preliminary analysis made use of several multiple linear regression models to identify significant or "important" variables. The authors reported that a linear accident rate model was much better than a log-linear model. For a nonlinear model they adopted and reparametrized an existing model.[9] This model was a hybrid, with both linear and nonlinear components. Although the required statistical assumptions were not fully stated, use of the least-squares method was based on the assumption that

[9] Designing Safer Roads: Practices for Resurfacing, Restoration, and Rehabilitation, Transportation Research Board, SR 214, Washington, D.C., 1986.

the residuals would follow a normal or log-normal distribution. Because accident distributions are skewed to the right, normality is not a tenable assumption.

Arguing that previous efforts were not sufficiently successful in attributing accidents to individual geometric elements and traffic characteristics, Kuo-Liang and Chin-Lung (1988)[10] explored a technique that purported to remove the assumptions of normality and linearity. Their model was developed for two-lane rural roads. A technique called Automatic Interaction Detection (AID) was used to group roadway segments by selected or created categories of explanatory variables. These categories of variables maximize the difference between group sums of squares. Then a model was developed by the Multiple Analysis Classification (MAC) technique of the following form:

$$Y_{ij...n} = Y + A_i + B_j + \ldots E_{ij...n}$$

where

$Y_{ij...n}$ = the score of unit n that falls in category i of predictor A, category j of predictor B, etc.
Y = grand mean of the dependent variable
A_i = the effect of membership in the i-th category of predictor A
B_j = the effect of membership in the j-th category of predictor B
...
$E_{ij...n}$ = error term for this unit.

This method, though in part innovative, is still a variation on simple linear regression and accounts for only 33% of the total variance. The low predictive power may also be due to the lack of a horizontal alignment variable and small sample size.

Durth (1989)[11] used risk analysis to perform highway safety evaluation. This is quite different from conventional approaches to accident analysis and modeling. The method is well-known in the fields of nuclear power plants and chemical factories. Based on research in Germany from 1986, the claim is made that risk analysis can be successfully applied to traffic safety. A risk model relies on diverse information in modular and hierarchical form from different branches of sciences (medicine, mechanical engineering, civil engineering, psychology, etc.). It reconstructs known dependencies and identifies relationships that need to be verified. Although the method may be promising, the

---

[10] Kuo-Liang, T., and Chin-Lung, Y., "A Predictive Accident Model for Two-Lane Rural Highways in Taiwan," Republic of China, Traffic Safety Theory and Research Methods, Session 4, Statistical Analysis and Models, Amsterdam, 1988.

[11] Durth, W., "Risk Analysis in Highway Engineering," Proceedings of Strategic Highway Research Program and Traffic Safety on Two Continents, VTI Report 351A, Gothenburg, Sweden, 1989.

report of Durth does not clearly describe the substance of the research. Nor does it indicate how to develop the stated dependencies and how to verify them practically.

Kulmala and Roine (1988)[12] developed models for Finnish roads. They assumed a Poisson error distribution and intended their models to be used for prediction. Their typical model form was:

$$A = K \times S^a \times \exp(\Sigma_i\, b_i x_i)$$

where

A = total number of fatal and injury accidents on a segment
S = exposure in vehicle-kilometers
x_i = explanatory variables such as surface width in meters, percentage of the segment length for which passing sight distance exceeds 300 meters, percentage of heavy vehicles, average curvature, and an interaction variable (pavement and speed limit).

This multiplicative Poisson regression model is comparable to that of Miaou et al. (1993).

SEGMENT VARIABLES

Average Daily Traffic (ADT)

ADT is one of the most significant variables in predicting accidents, yet it is not controllable. Many models have used traffic exposure as a dependent variable although its relationship with accident counts is not fully linear. In general, it is recommended to use ADT as an independent variable for greater accuracy because it interacts with other controllable variables, and it measures the effect of traffic flow intensity (Hauer, 1994).[13]

Lane Width, Shoulder Width, and Shoulder Type

Modeling approaches vary from study to study, and techniques of data collection and analysis likewise vary. Thus the effect of lane width and shoulder width on accident frequency has some variation in different studies. Generally it has been found that accident rates decrease when lane and

---

[12] Kulmala, R., and Roine, M., "Accident Prediction Models for Two-Lane Roads in Finland," Technical Research Centre of Finland, Traffic Safety Theory and Research Methods, Session 4, Statistical Analysis and Models, Amsterdam, 1988.

[13] Hauer, E., "On Two Uses of Exposure," Paper Presented at the Transportation Research Board Annual Meeting, Washington, D.C., 1994.

shoulder widths increase. The report by Zegeer et al. (1986) on the effect of cross-section for two-lane rural roads indicated that a paved shoulder widening of 2 feet per side reduces accidents by 16%, while reports of Miaou et al. (1993) and Zegeer et al. (1986) found reductions of 8% and 6.6%, respectively. The latter two reports take into account horizontal curvature and curve length as explanatory variables, while the former does not explicitly include horizontal alignment. Luyanda et al. (1983) showed that shoulder type, an amalgam that includes width and surface type, is a significant variable but did not define this variable in detail. The synthesis of Jorgensen (1978)[14] reported a negative relationship between accidents and shoulder width for two-lane rural highways on the basis of studies done primarily in the 1950's and 1960's. Variation of shoulder width for Interstate Highways and other freeways exists mostly along the inside shoulder, and older reports indicate that accidents increase as the inside shoulder width increases, contrary to the findings of Miaou et al. (1993). The increase of accidents with inside shoulder width may be due to emergency parking on wider shoulders or to insufficient accident history in the older studies.

Horizontal and Vertical Alignment

Horizontal and vertical alignment can be expressed in alternative ways to capture the effect of individual curves (disaggregate) or a sequence of curves (aggregate). Examples of measures of horizontal curvature are as follows:

- Horizontal curvature change rate - Miaou et al. (1993)

$$CCR = \Sigma_{i < k} |\theta\{i+1\} - \theta\{i\}|$$

where $\theta\{i\}$ is the degree of curve (degrees per hundred feet) of the i-th horizontal curve on a segment, recorded as positive if to the right and negative if to the left in the increasing roadway direction, and k is the number of horizontal curves on the segment. If k = 1, *CCR* is set to zero. *CCR* is an aggregate measure, while $\theta\{i\}$ is a disaggregate measure.

- Average curvature - Polus (1980)[15]

$$AC = \frac{\Sigma_i \alpha\{i\}}{L}$$

where L is the segment length in miles and $\alpha\{i\}$ is the absolute horizontal angle between the i-th and

---

[14] Roy Jorgensen Associates, "Cost and Safety Effectiveness of Highway Design Elements," National Cooperative Highway Research Program, Report 197, Transportation Research Board, National Research Council, Washington, D.C., 1978.

[15] Polus, A., "The Relationship of Overall Geometric Characteristics to the Safety Level of Rural Highways," Traffic Quarterly, 34(4), 1980.

(i+1)-th tangents, in degrees. Here AC is aggregate and $\alpha\{i\}$ is disaggregate. Vertical grade variables can be expressed similarly. Researchers have used both aggregate explanatory variables (Polus, 1980; Kulmala and Roine, 1988) and disaggregate ones (Miaou et al. 1993; Zegeer et al., 1991) in the modeling process, although aggregate variables are not directly helpful to designers who are improving individual curves. Nevertheless, aggregate variables are useful as surrogates in evaluating alignment safety. In most of the referenced reports, the results confirm the common sense opinion that sharper and longer curves result in more accidents, regardless of whether the statistical techniques applied are multiple linear regression or generalized linear models.

Roadside and Terrain Condition

When roadside features such as slopes, guardrails, trees, poles, etc. are considered separately, the portion of accident rates explained by roadside features is weak. The reports by Graham and Harwood (1982)[16] and Zegeer et al. (1986) indicate this drawback. Zegeer et al. (1991) reported that mountainous terrain type has a negative effect on safety. Zegeer et al. (1987), as noted in Chapter 1, packaged the roadside variables in a subjective measure called Roadside Hazard Rating based on visual evaluation of clear zone and sideslope. Roadside Hazard Rating takes numerical values from one to seven. This measure "indicates the accident damage likely to be sustained by errant vehicles on a scale from one (low likelihood of an off-road collision or overturn) to seven (high likelihood of an accident resulting in a fatality or severe injury)." On a segment length with variable hazards, an average or middle value is assigned.

Speed

Various attempts have failed to find relationships between accidents and speed, whether the latter is design speed, posted speed, or operating speed. One of the few models where speed is considered comes from Finland (Kulmala and Roine, 1988). A report of Fridstrøm et al. (1995)[17] indicates that a change in posted speed lowered fatal accidents in Denmark.

Driveways

The influence of driveway accidents was highlighted by two studies (Fee et al., 1970; McGuirk and

---

[16] Graham, J.L., and Harwood, D.W, Effectiveness of Clear Recovery Zones, National Cooperative Highway Research Program, Report 247, Transportation Research Board, National Research Council, Washington, D.C., 1982.

[17] Fridstrøm, L., Ifver, J., Ingebrigsten, S., Kulmala, R., and Thomsen L.K., "Measuring the Contribution of Randomness, Exposure, Weather, and Daylight to the Variation in the Road Accident Counts," Accident Analysis and Prevention, 27(1): 1-20, 1995.

Staterly, 1976).[18] Driveway density and driveway spacing were found to be significant safety factors. McGuirk and Staterly (1976) developed a linear model for accident rates Y:

$$Y = 7.728 - 0.055X,$$

where X is driveway spacing. Figure 1, illustrating the relationship of accidents to driveway density, appears in Cirillo (1992),[19] and was taken from the report of Fee et al. (1970).

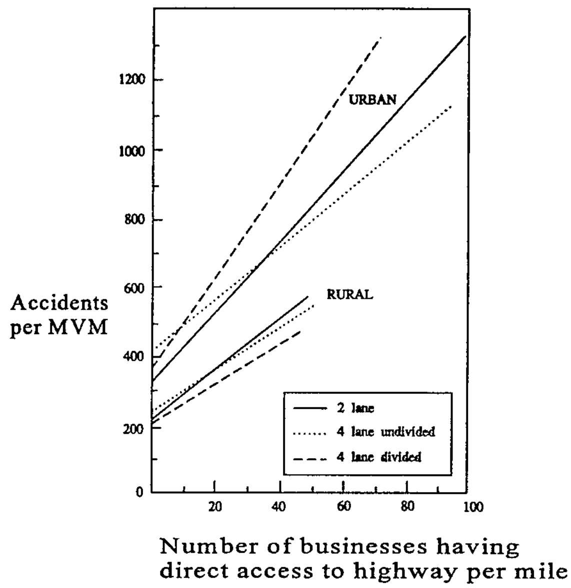


FIGURE 1. ACCIDENT RATES ON NON-INTERSTATE HIGHWAYS FOR SELECTED HIGHWAY TYPES BY NUMBER OF BUSINESSES PER MILE (Cirillo, 1992)

---

[18] Fee, J.A., Beatty, R.L., Dietz, S.K., Kaufman, S.F., and Yates, S.F., "Interstate System Accident Research Study 1," U.S. Government Printing Office, Washington, D.C., 1970.

McGuirk, W.W., and Staterly, G.T., Jr., "Evaluation of Factors Influencing Driveway Accidents," Transportation Research Board, TRR 601, Washington, D.C., 1976.

[19] Cirillo, J.A., "Safety Effectiveness of Highway Design Features, Volume I: Access Control," FHWA-RD-91-044, Federal Highway Administration, Washington, D.C., 1992.

INTERSECTION MODELS

The methodology and statistical techniques used in a series of three reports (Lau and May, 1989; Lau and May, 1988; Naclerio et al., 1989)[20] on signalized and unsignalized intersections are of interest to intersection modelers. Accident prediction models were developed to identify locations where accident experience was more frequent or more severe than normal, and to evaluate the safety consequences of alternative improvements. Factors and highway characteristics reported in the California data base were included in the model: accident data, traffic volumes, intersection features, and control types. However, variables such as degree of horizontal curvature and rate of vertical curvature, believed to be important, were not included. Unlike other partial studies, these models encompass all types of intersections, and the methodology addresses the successive stages of planning, design, and site improvement.

Three types of accident severity were modeled separately: fatal, injury, and property damage only. Collision types such as angle, rear-end, etc., that may further explain the cause of accidents were missing from the model. A nonparametric statistical modeling technique known as the Classification Regression Tree (CART) was used to group intersections by significance of prediction. The response variable was number of accidents per year, with traffic volume used only as an explanatory variable. The CART technique has particular applicability to categorical and discontinuous variables. However, the classification obtained was not sufficiently detailed to reveal the effect of individual highway factors. For injury accidents, nine groups of signalized intersections were identified, and eight groups were identified for property damage only accidents. The model for fatal accidents was not reliable, with a correlation coefficient of only 0.009. As a starting point for the analysis of relationships, intersections are categorized by highway functional classification into groups that are assumed to perform differently. The potential for application to optimization, i.e., to help the designer choose highway characteristics that will minimize the expected number of accidents, was noted but no application was made. Another caveat of this methodology is implied in its tendency to produce a grouping not much different from the existing conventional State grouping.

---

[20] Lau, M.-K., and May, A.D., "Accident Prediction Model Development for Unsignalized Intersections: Final Report," University of California at Berkeley, Institute of Transportation Studies, Report UCB-ITS-RR-89-12, Berkeley, California, 1989.

Lau, M.-K., and May, A.D., "Accident Prediction Model Development: Signalized Intersections, Final Report," University of California at Berkeley, Institute of Transportation Studies, Report UCB-ITS-RR-88-7, Berkeley, California, 1988.

Naclerio, M.T., Kruger, P, and May, A.D., "Accident Prediction Models for Signalized and Unsignalized Intersections. Addendum," University of California at Berkeley, Institute of Transportation Studies, Report UCB-ITS-RR-89-17, Berkeley, California, 1989.

Hauer et al. (1988)[21] developed accident prediction models for signalized intersections by maneuver patterns (15 defined conflict patterns) before the occurrence of accidents. Each pattern involved at most two conflicting flows. A typical model form is as follows:

$$E(m) = b_0(F_1)^{b_1}(F_2)^{b_2}$$

where

$E(m)$ = expected number of accidents for maneuver pattern m
F_1 = traffic flow of turning movement 1
b_1 = power of F_1
F_2 = traffic flow of turning movement 2
b_2 = power of F_2.

Equations were derived for each of the 15 pre-accident patterns to compute the expected number of accidents. These equations can also be used to estimate the kinds of accident caused by traffic flow patterns. Their design consequences are limited because they are based exclusively on traffic flow variables, and these are uncontrollable. Unlike traffic flow patterns, physical elements such as channelization and alignment are manageable safety improvements. On the other hand, the models are negative binomial in form. This form, as the authors indicate, has the attractive feature that it can be modified by empirical Bayesian techniques to incorporate actual experience at an individual intersection.

Garber and Srinivasan (1991)[22] used traffic flow (left-turn volumes) movements as explanatory variables for predicting accidents during peak-hours and otherwise. Besides safety evaluations, these models are favorable for improvements such as installing left turn lanes and adding protected phasing. Despite high R^2 values, the simple linear regression models used in this study are inadequate for discrete events such as accidents that have a very low mean and are not normally distributed. Moreover, these models predict accidents for elderly drivers, a small segment of the driver population.

---

[21] Hauer E., Ng, J. C.N., and Lovell, J., "Estimation of Safety at Signalized Intersections," Transportation Research Board, TRR 1185, Washington, D.C., 1988.

[22] Garber, N.J., and Srinivasan, R., "Risk Assessment of Elderly Drivers at Intersections: Statistical Modeling," Transportation Research Board, TRR 1325, Washington, D.C., 1991.

INTERSECTION VARIABLES

Traffic Flow

Traffic flows (ADT) have often been used as measures of exposure or as explanatory variables in modeling accidents at intersections. Many accident studies have used intersection accident rates in the form of accidents per million entering vehicles (Kuciemba and Cirillo, 1992).[23] This type of rate has been used for safety performance evaluations and safety comparisons even though it does not take into account the magnitude of conflicting movements. Another common way to measure intersection accident rates is in accidents per unit time. McDonald (1966)[24] exhibited a model relating accident frequency (accidents per year) to a product of powers of the cross-road and major road entering ADT.

$$N = 0.000783(V_m)^{0.455}(V_c)^{0.633}$$

where

N = number of accidents per year
V_m = major road ADT in vehicles per day
V_c = cross-road ADT in vehicles per day.

Leong (1973)[25] proposed comparable but simpler models of the form:

$$N = k(V_m V_c)^a.$$

---

[23] Kuciemba, S.R., and Cirillo, J.A., "Safety Effectiveness of Highway Design Features, Volume V: Intersections," FHWA-RD-91-048, Federal Highway Administration, Washington, D.C., 1992.

[24] McDonald, J.W., "Relationship Between Number of Accidents and Traffic Volumes at Divided Highway Intersections," National Research Council, Highway Research Board Report 74: 7-17, Washington, D.C., 1966.

[25] Leong, W.H.J., "Relationship Between Accidents and Traffic Volumes at Urban Intersections," Journal of Australian Road Research Board, 5(3): 72-90, 1973.

A method for handling exposure measures developed by Surti (1965)[26] was applied by Hakkert and Mahalel (1978)[27]. The latter authors proposed that accident frequency is linearly related to an exposure measure X, called index flow, calculated as the sum of the products of the flows at each of 24 conflict points defined by Surti. The model for urban intersections is as follows:

$$N = 2.27 + 0.000112X.$$

Hauer et al. (1988), as already noted, used traffic flows for each conflict pattern to predict accidents, found different functional forms and coefficients for different patterns, and addressed the short-comings of simple models of intersection accidents in terms of flows. The need for detailed models by pattern is presumably greater for signalized intersections than it is for stop-controlled minor roads with low traffic.

Control Type

The safety effect of converting to all-way stop was contradictory in two papers (Lovell and Hauer, 1986; Persaud, 1986).[28] Lovell and Hauer affirmed the benefit of converting to four-way stop, while Persaud rejected its effectiveness. King and Goldblatt (1975)[29] concluded that signalization reduces right-angle accidents but increases rear-end accidents, with no significant change in total accident-related disutility.

---

[26] Surti, V.H., "Accident Exposure for At-Grade Intersections," Traffic Engineering, 36 (3): 26-27 and 53, 1965.

[27] Hakkert, A.-S., and Mahalel, D., "Estimating the Number of Accidents at Intersections from a Knowledge of the Traffic Flows on the Approaches," Accident Analysis and Prevention, 10: 69-79, 1978.

[28] Lovell, J., and Hauer, E., "The Safety Effect of Conversion to All-Way Stop Control," Transportation Research Board, TRR 1068, Washington, D.C., 1986.

Persaud, B.N., "Safety Migration, the Influence of Traffic Volumes, and Other Issues in Evaluating Safety Effectiveness - Some Findings on Conversion of Intersections to Multiway Stop Control," Transportation Research Board, TRR 1068, Washington, D.C., 1986.

[29] King, G.F., and Goldblatt, R.B., "Relationship of Accident Patterns to Type of Intersection Control," Transportation Research Board, TRR 540, Washington, D.C., 1975.

Sight Distance and Alignment

Three reports relate intersection sight distance (ISD) to accidents (David and Norman 1975; Wu, 1973; Moore and Humphreys, 1975).[30] David and Norman reported that an increase in sight radius reduces the number of accidents. Sight radius was defined to be an average of all intersection sight distances at 50 feet from the intersection. Thus sight radius is not equivalent to the ISD defined in the AASHTO Design Manual, the so-called "Green Book."[31] Wu cited the safety effect of clear vision and poor vision at both rural and urban signalized intersections. Clear and poor vision are qualitative descriptors as opposed to precise quantitative measures of ISD. Bared and Lum (1992),[32] in a presentation on the safety effectiveness of intersection design elements, concluded that sight distance and other alignment variables are important at intersections. Among others, Urbanik et al. (1989)[33] affirmed the significance of sight distance on crest vertical curves at intersections. Intersection sight distance will be indirectly considered in this study by surrogate variables: horizontal curvature, vertical curvature, and Roadside Hazard Rating.

ARTIFICIAL NEURAL NETWORKS

Artificial neural network applications have recently received considerable attention. The methodology of modeling, or estimation, is somewhat comparable to statistical modeling (Smith,

---

[30] David, N., and Norman, J.R., "Motor Vehicle Accidents in Relation to Geometric and Traffic Features of Highway Intersection," Volume II, FHWA-RD-76-129, Federal Highway Administration, Washington, D.C., 1975.

Wu, Y.S., "Effect of Clear Vision Right-of Way on Traffic Accidents at Urban and Rural Signalized Intersections," Report TSD-228-73, Department of State Highways, Michigan, 1973.

Moore, W L. Jr., and Humphreys, J.B., "Sight Distance Obstructions on Private Property at Urban Intersections," Transportation Research Board, TRR 541, Washington, D.C., 1975.

[31] A Policy on Geometric Design of Highways and Streets, American Association of State Highway and Transportation Officials (AASHTO), Washington, D.C., 1994.

[32] Bared, J.G., and Lum, H., Safety Evaluation of Intersections Design Elements (Pilot Study), Transportation Research Board Conference Presentation, Washington, D.C., 1992.

[33] Urbanik, T., II, Hinshaw, W., and Fambro, D.B., "Safety Effects of Limited Sight Distance on Crest Vertical Curves," Transportation Research Board, TRR 1208, Washington, D.C., 1989.

1993).[34] Neural networks should not, however, be heralded as a substitute for statistical modeling, but rather as a complementary effort (without the restrictive assumption of a particular statistical model) or an alternative approach to fitting non-linear data.

A typical neural network (shown in Figure 2) is composed of input units X_1, X_2, ... corresponding to independent variables (in our case, highway or intersection variables), a hidden layer known as the first layer, and an output layer (second layer) whose output units Y_1, ... correspond to dependent variables (expected number of accidents per time period).

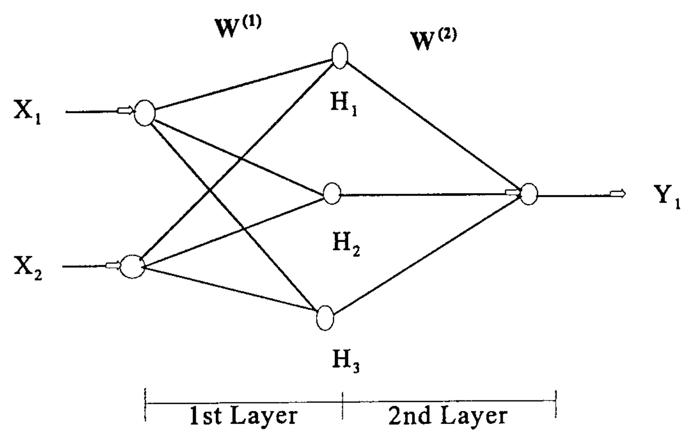


FIGURE 2. A TYPICAL NEURAL NETWORK

In between are hidden units H_1, H_2, ... corresponding to intermediate variables. These interact by means of weight matrices $W^{(1)}$ and $W^{(2)}$ with adjustable weights. The values of the hidden units are obtained from the formulas:

$$H_j = f\left(\sum_k W_{jk}^{(1)} X_k \right)$$
$$Y_i = f\left(\sum_j W_{ij}^{(2)} H_j \right).$$

One multiplies the first weight matrix by the input vector $X = (X_1, X_2, ...)$, and then applies an activation function f to each component of the result. Likewise the values of the output units are obtained by applying the second weight matrix to the vector $H = (H_1, H_2, ...)$ of hidden unit values,

[34] Smith, M., "Neural Networks for Statistical Modeling," Van Nostrand, New York, 1993.

and then applying the activation function f to each component of the result. In this way one obtains an output vector $Y = (Y_1, Y_2, \ldots)$.

The activation function *f* is typically of sigmoid form and may be a logistic function, hyperbolic tangent, etc.:

$$f(u) = \frac{1}{1 + e^{-u}}, \qquad f(u) = \frac{e^{u} - e^{-u}}{e^{u} + e^{-u}}.$$

Usually the activation function is taken to be the same for all components but it need not be.

Values of $W^{(1)}$ and $W^{(2)}$ are assumed at the initial iteration. The accuracy of the estimated output is improved by an iterative learning process in which the outputs for various input vectors are compared with targets (observed frequency of accidents) and an average error term E is computed:

$$E = \frac{\sum_{n=1}^{N} (Y^{(n)} - T^{(n)})^2}{N}.$$

Here

N = Number of highway sites or observations
$Y^{(n)}$ = Estimated number of accidents at site n for $n = 1, 2, \ldots, N$
$T^{(n)}$ = Observed number of accidents at site n for $n = 1, 2, \ldots, N$.

After one pass through all observations (the training set), a gradient descent method may be used to calculate improved values of the weights $W^{(1)}$ and $W^{(2)}$, values that make E smaller. After reevaluation of the weights with the gradient descent method, successive passes can be made and the weights further adjusted until the error is reduced to a satisfactory level. The computation thus has two modes, the mapping mode, in which outputs are computed, and the learning mode, in which weights are adjusted to minimize E. Although the method may not necessarily converge to a global minimum, it generally gets quite close to one if an adequate number of hidden units are employed.

The most delicate part of neural network modeling is generalization, the development of a model that is reliable in predicting future accidents. Overfitting (i.e., getting weights for which E is so small on the training set that even random variation is accounted for) can be minimized by having two validation samples in addition to the training sample. According to Smith (1993), the data set should be divided into three subsets: 40% for training, 30% to prevent overfitting, and 30% for testing. Training on the training set should stop at the epoch when the error E computed on the second set begins to rise (the second set is not used for training but merely to decide when to stop training). Then the third set is used to see how well the model performs. The cross-validation helps to optimize the fit in three ways: by limiting/optimizing the number of hidden units, by limiting/optimizing the number of iterations, and by inhibiting network use of large weights.

The major advantages and disadvantages of neural networks in modeling applications are as follows:

Advantages

- There is no need to assume an underlying data distribution such as usually is done in statistical modeling.
- Neural networks are applicable to multivariate non-linear problems.
- The transformations of the variables are automated in the computational process.

Disadvantages

- Minimizing overfitting requires a great deal of computational effort.
- The individual relations between the input variables and the output variables are not developed by engineering judgment so that the model tends to be a black box or input/output table without analytical basis.
- The sample size has to be large.

The disadvantages appear to outweigh the advantages, particularly in view of the black box effect.

3. DATA COLLECTION

This chapter discusses the populations on which the study is based and how samples were selected from these populations, how sample data were collected, and the limitations on the quality of the sample data. Table 1 gives a list of the chief variables collected.

THE POPULATIONS AND SAMPLE SELECTION

The States for which data were obtained are Minnesota and Washington. Both of these States are included in the Highway Safety Information System (HSIS), and both States have relatively well-maintained data bases. In addition, data for recent years (1985 through 1994 for Minnesota and 1993 through 1995 for Washington) were available, or became available in the course of the study. For Washington a shortcoming was the unavailability of a separate intersection file.

The populations from which the samples were drawn were rural segments of two-lane roads and rural three- and four-legged intersections of two-lane roads stop-controlled on the minor road. The roads had to be present in State and HSIS databases, and thus the segment road or major road was always a State highway. Roads with unusually low traffic were not included, and other reasonable constraints were imposed. Samples were picked from the population in part randomly and in part systematically. Since the purpose of this study was not to summarize the population of each State, but rather to obtain insight into the effects of different variables, observations were selected with some view to achieving variety in traffic volumes, roadway width, and terrain.

Minnesota Segments

The sample of Minnesota segments was prepared as follows:

i) HSIS files of homogeneous segments of State roads for two time periods, 1985-1987 and 1988-1989, were obtained and the constraints below were imposed.

rural two-lane, two-way, paved road
17 feet < surface width ≤ 24 feet
left and right shoulder width differing by 2 feet or less
average of left and right shoulder width ≤ 12 feet
segment length > 0.1 mile
segment present in both time periods with characteristics unchanged
5-year average daily traffic (ADT) > 5 vehicles
5-year average daily commercial traffic > 5 vehicles

ii) The resulting population consisted of 3,308 segments. Some statistics, derived from HSIS data, on this population are presented in Appendix 1. Median values of ADT, segment length, surface width, and shoulder width were obtained for the population and used to classify segments by high

Preceding page blank

versus low ADT, high versus low segment length, high versus low surface width, and high versus low shoulder width. The population was then divided into 16 bins on the basis of whether each of the four variables was high or low. The resulting bins varied in size from 13 segments to 679 segments. Thirteen segments were randomly selected from each bin, along with a hundred other segments randomly selected from the remaining population as a whole, and these formed a pilot study sample of 308 segments.

iii) The pilot study sample was eventually enlarged by the addition of 416 more segments so that all members of the six smallest bins were included in the sample. The sizes of these six bins ranged from 13 segments to 45 segments. The selection method for the final sample was equivalent to exhaustion of the first six bins, a random choice of 45 segments from each of the remaining bins, and a random choice of a hundred additional segments from the remaining bins without regard to bin membership. The resulting sample consisted of 724 segments.

iv) For each of these segments an attempt was made to obtain photolog data (signage, Roadside Hazard Rating, driveways, intersections, speed limits) at FHWA and in Minnesota and to extract vertical and horizontal alignments along the segments as they were in the years 1985-1989 from construction plans in Minnesota. After much investigation and double-checking, relatively complete data could be acquired for 619 segments. These constituted the final sample. The remaining segments were removed because photologs or construction plans were unavailable or were seriously incomplete, because significant regrading or realignment had been done in the time period 1985-1989, or in a few cases because photologs revealed that the segments were not two-lane roads. One segment was removed because the ADT was 22,710 vehicles per day, substantially higher than that of all others roads in the study.

Minnesota Intersections

The samples of Minnesota intersections were prepared as follows:

i) HSIS files of intersections with main line a State road for two time periods, 1985-1987 and 1988-1989, were obtained and the constraints below were imposed.

rural environment
main line a U.S. trunk highway or Minnesota trunk highway
main line and cross-street two-lane, two-way road
stop sign on minor road, thru on main line
17 feet $<$ surface width $\leq$ 24 feet
intersection present in both time periods with characteristics unchanged
number of legs three or four
main line has two legs
main line does not change direction at intersection by more than 45°
traffic data on major and minor roads obtained in 1982 or later

three-legged intersections of types tee or wye[35]
four-legged intersections of types right angle or skewed crossing

ii) The resulting populations consisted of 949 three-legged intersections and 1,156 four-legged intersections. See Appendix 1 for statistics concerning these two populations. Median values of main line ADT and minor road ADT were obtained for each population and used to classify intersections by high versus low major road ADT, and high versus low minor road ADT. Each population was then divided into four bins numbered 00 to 11, based on whether each of the two variables was high or low. 1 means high, 0 low, and the first number refers to major road ADT, the second to minor road ADT. The resulting bins had the sizes shown below.

Minnesota Intersections

	Three-legged				Four-legged			
Bin	Final Sample	%	Population	%	Final Sample	%	Population	%
00	103	26.5	274	28.9	84	25.7	359	31.1
01	99	25.5	200	21.1	79	24.2	229	19.8
10	90	23.1	201	21.2	87	26.6	215	18.6
11	97	24.9	274	28.9	77	23.5	353	30.5
Total	389	100.0	949	100.0	327	100.0	1,156	100.0

iii) Initially pilot study samples of 25 intersections were chosen randomly from within each of the eight bins. Examination of photologs showed that intersections in three of the bins failed to satisfy the constraints in disproportionately large numbers. So 10, 5, and 7 extra intersections were chosen randomly from the bins 3-legged 10, 3-legged 11, and 4-legged 10, respectively. Thereafter in the course of ensuing months an additional 100, then 160, and then 200 intersections were chosen randomly from the 3-legged bins in equal numbers; while an additional 100, and then 160 were chosen likewise from the 4-legged bins. The total sample of 3-legged intersections consisted of 100 + 10 + 5 + 100 + 160 + 200 = 575 intersections. The total sample of 4-legged intersections consisted of 100 + 7 + 100 + 160 = 367 intersections.

iv) For each of these intersections an attempt was made to obtain photolog data (signage, Roadside Hazard Rating, driveways, turning lane/bypass lane data, speed limits) at FHWA and in Minnesota, and to extract vertical and horizontal alignments for curves any portion of which were within 764 feet of an intersection along the main line from construction plans in Minnesota. The information was

[35]According to the "Green Book," A Policy on Geometric Design of Highways and Streets, American Association of Highway and Transportation Officials, 1994, p. 836, an intersection is of tee type when two legs form a through road and the third leg enters at a nonacute angle, of wye type if all three legs have a through character or the angle with the third leg is small.

for the intersections as they were in the years 1985-1989. Relatively complete data could be acquired for 389 three-legged intersections and 327 four-legged intersections. The remainder were eliminated because photologs showed that they did not satisfy the constraints, or plans were unavailable for them, or the intersections had significant construction during 1985-1989.

Washington Segments

The sample of Washington State segments was prepared as follows:

i) HSIS files of homogeneous segments of State roads for the years 1993 and 1994 were obtained and the constraints below were imposed:

rural two-lane, two-way, paved road
17 feet < surface width ≤ 24 feet
left and right shoulder width differing by 2 feet or less
average of left and right shoulder width ≤ 12 feet
segment length > 0.1 mile
segment present in both time periods with characteristics unchanged
2-year average daily traffic (ADT) > 5 vehicles
no vertical curves of zero length with change of grade of 1% or more
no horizontal curves of zero length with angular change of 1° or more

Unlike Minnesota, horizontal and vertical alignment data were available for Washington State in separate HSIS Horizontal and Vertical Curve files.

ii) The resulting population consisted of 6,144 segments. Median values of ADT, segment length, surface width, and shoulder width were obtained for this population. The median segment length was 0.36 miles (considerably lower than Minnesota's median of 0.5695 miles). The segments were classified by high versus low ADT, high versus low segment length, high versus low surface width, and high versus low shoulder width, with the medians as the division points except for segment length for which 0.600 miles was used. The population was then divided into 16 bins on the basis of whether each of the four variables was high or low. The resulting bins varied in size from 87 segments to 913 segments.

iii) 61 segments were picked randomly from each of the 16 bins, for a total of 976 segments. An additional 25 segments were picked for which the TERRAIN variable had the value "mountainous."

iv) On the basis of videotape reviews, further examination of alignment variables, and an enlargement of the time frame to include the year 1995, the sample was reduced to a total of 712 segments. Some segments were eliminated because the videotapes showed that they did not meet the constraints (e.g., the environment was urban or the number of lanes had changed) or the alignment data contained anomalies such as a significant difference between the outgoing grade of one vertical curve and the incoming grade of the next. Others were omitted because in Washington State, unlike Minnesota, most segments begin and end with an intersection. After 250 feet were removed from one or both ends of segments in such cases, it was found that a significant number of segments no longer met the requirement that their length was greater than 0.1 miles. In addition, 1995 HSIS Washington State files became available at a relatively late stage of the study and the sample was further trimmed when the requirement was imposed that the segment also be present in the 1995 files with chief characteristics unchanged.

Washington Intersections

There are no HSIS intersection files for Washington State nor does the State maintain separate intersection files. Washington State videotapes were, however, accompanied by logs indicating the locations and names of all cross-streets along each State route. Since ADT data for county and local roads were not readily available, it was decided to note intersections of State roads found in the videotapes and satisfying the same constraints as the Minnesota data. This was not done for all Washington State videotapes, but only for ones being reviewed to extract data for the segment sample. A total of 431 intersections were reviewed by this method.

The Washington State Department of Transportation provided a log of intersections for which it had ADT data on the cross-streets. The intersections in this log were intersections on State roads together with intersections in the Highway Performance Monitoring System. In addition, by inspecting HSIS road files, the Project Team was able to match major and minor State roads in some other cases to get ADT data. However, for some of the intersections no reliable estimate of cross-street ADT could be obtained. In addition, inspection of videotapes showed that some of the intersections failed to satisfy the intersection constraints imposed in Minnesota (e.g., they were not rural). When traffic, alignment, and roadway data were assembled, and incomplete observations removed, the resulting data sets, "opportunity" samples rather than a random samples, consisted of 181 three-legged intersections and 90 four-legged intersections.

HOW DATA WERE COLLECTED

Data were extracted from HSIS data files for Minnesota and Washington, from photologs for Minnesota and videotapes for Washington, and from construction plans at the Minnesota Department of Transportation. In addition, weather data for the state of Minnesota were acquired from the Midwest Climate Center. A number of small-scale investigations were also done that made use of other data provided by personnel at the respective Departments of Transportation.

HSIS data are stored in SAS data bases. The needed data elements were extracted and assembled into SAS data sets representing the study populations with identifiers for each population bin. Random numbers were used to prepare SAS data sets representing the study samples (with the exception of the Washington intersections). Other sample data were recorded manually on specially prepared data sheets from photologs, videotapes, and plans. These were entered into SAS data sets that were merged with the HSIS data to obtain the full sample data sets.

Numerous data checks were done at each stage . Second and sometimes third viewings of photologs, videotapes, and plans occurred, as well as consistency checks on SAS data base entries and some checks on the HSIS files themselves. Variables such as Roadside Hazard Rating were determined by two and sometimes three different individuals to minimize subjectivity.

HSIS Data

Accident data, traffic data, vertical and horizontal alignment data for Washington State, and other geometric data were extracted from HSIS files. These data were used in part to constrain the populations so that segments were on two-lane paved rural roads where segment lengths, surface widths, shoulder widths, ADT, and commercial ADT fell within prescribed ranges, while intersection geometries were three-legged or four-legged with all legs two-lane and two-way rural roads.

The data elements for the samples are those shown in Table 1. In the case of Washington State vertical and horizontal alignment data were obtained from HSIS files, but for Minnesota they were obtained from construction plans.

Minnesota Photologs

Photologs for the State of Minnesota were examined at FHWA's Turner-Fairbank Highway Research Center. In some cases photologs were not available at FHWA, but were found and examined at the Minnesota Department of Transportation (MNDOT) in Saint Paul, Minnesota. The photologs were used to verify HSIS data (e.g., rural environment, two lanes, stop sign on minor road), to determine Roadside Hazard Rating, to count driveways and intersections within a segment, to determine channelization at intersections, and to note posted regulatory and advisory speeds when seen.

Minnesota Construction Plans

Construction plans obtained in the Plan Office of MNDOT provided horizontal and vertical alignment data as well as the angle between legs at intersections. Location of plans was an arduous task, requiring that true beginning and ending mileposts of a segment or reference point of an intersection be matched up to the correct stations, that a control section be determined from a separate book, that a card file of projects by segment be consulted to discover any projects and project numbers, and then that the corresponding project plan sheets be recovered and verified. Plans were then copied and were examined in detail at a later time.

Washington Videotapes

Videotapes for the State of Washington's roadways were purchased from the Washington Department of Transportation and were reviewed at PRAGMATICS. Like the Minnesota photologs, the video tapes were used to verify the correctness of the HSIS data and to obtain Roadside Hazard Rating, speeds, numbers of driveways, and channelization. In addition, they were used to estimate the angle between legs at an intersection.

Weather Data

Weather data were acquired for Minnesota intersections. The Midwestern Climate Center (MCC) in Illinois provided a listing of the nine Climate Districts in Minnesota, each of which is relatively homogeneous in its weather conditions. Weather data for each District are available based on averages of reports from local weather stations, many of which are run by volunteers. In Northern Minnesota the stations are sparser than elsewhere in the State. The percentages of dry, wet, snow/slush, and ice/pack snow days, respectively, for each year from 1985 to 1989 by Climate District were provided at a nominal charge. PRAGMATICS, Inc. staff attached these to segments and intersections falling within the corresponding Climate District.

Modeling of the Minnesota data did not show the weather to be significant, possibly because the weather variable could not be localized to a level below the Climate District. Consequently, weather data were not acquired for Washington State.

Miscellaneous Investigations

Aerial photographs were consulted in both Minnesota and Washington for possible use in estimating horizontal alignment, intersection angles, and intersection channelization. The Photogrammetric Unit of MNDOT provided contact prints for 12 out of 20 requested intersections at a scale of 1" = 100'; the other eight were not available. Washington State provided a few sample prints of aerial photographs at a scale of 1" = 2,000'. Curvatures and angles could be readily made out from the Washington photos, but channelization at intersections was not readily ascertainable. Since the information could be obtained in other ways, not all intersections and segments were available in aerial photographs, and the cost was high in Washington State, it was decided not to acquire such photos for the full samples.

Minnesota has nine Highway Districts. Each Highway District Office was queried for information about a sample of intersections (channelization installation dates, age of stop signs on minor roads). Age of stop signs is thought to be related to reflectivity and visibility. All nine Districts responded and provided some information, including sketches of the intersections. In all cases the channelization (turning and/or acceleration lanes) was installed prior to 1985, but exact installation dates were not available. Likewise the dates of stop sign installations were not generally available, but the District Offices indicated that stop signs were replaced on a 10-year schedule.

Queries were also made in Minnesota about traffic data and commercial traffic data, as well as the availability of traffic data on county roads, and in both Minnesota and Washington about underreporting of accidents. Results are reported below.

TABLE 1. Variables collected in the study

	MINNESOTA SEGMENTS			
	Variable	Meaning	Units	Source
Identifiers	m_sysnbr	Route number		HSIS
	true_beg	true beg. milepost	miles	HSIS
	true_end	true end milepost	miles	HSIS
	beg_sta	beg. station	hundreds of feet	Plans
	end_sta	end station	hundreds of feet	Plans
Traffic	ADT	average daily traffic	vehicles per day	HSIS
	com_avg	average daily heavy vehicle traffic	vehicles per day	HSIS
Miscel-laneous	LW	lane width	feet	HSIS
	SHW	shoulder width	feet	HSIS
	RHR	Roadside Hazard Rating	1, 2, 3, 4, 5, 6, 7	Photologs
	nodrwy, noint	number of driveways, number of intersections		Photologs
	shl_typ	shoulder type		HSIS
	light	yes or no if lighting/no lighting		Photologs
	terrain	flat, rolling, or mountainous		Photologs
Weather	dd, wd, ss, ips	number of dry, wet, snow/slush, ice\packsnow days	days per year	MCC
Horizontal alignment	pc{i}	beg. station of curve no. i	hundreds of feet	Plans
	pt{i}	end station of curve no. i	hundreds of feet	Plans
	DEG{i}	degree of curve, curve no. i	degrees per 100 ft	Plans
	dir{i}	direction, left or right, curve no. i		Plans
Vertical alignment	b{i}	beg. station of curve no. i	hundreds of feet	Plans
	e{i}	end station of curve no. i	hundreds of feet	Plans
	g{i}	grade no. i (prior to curve no. i)	percent	Plans

Variables explicitly used in models are in capital letters; 1 mi = 1.61 km, 1 ft = 0.3048 m

TABLE 1. Variables collected in the study (continued)

	MINNESOTA SEGMENTS, continued			
	Variable	Meaning	Units	Source
Speed	advspd	advisory speed	miles per hour	Photologs
	regspd	regulatory speed	miles per hour	Photologs
	speed	posted speed (accident sites only)	miles per hour	HSIS
Accident data	TOTACC	total number of non-intersection accidents in 1985-9, 1990-3		HSIS
	fatal, injury, nonincap, possinj, injunk, propdam	no. of fatal, injury, non-incapacitating, possible injury, injury unknown, and property damage only non-intersection accidents		HSIS
	rearend, sswipe, leftturn, rorleft, rtangle, riteturn, rorright, headon, sswipopp, other, unknown	no. of rearend, sideswipe, left turn, run-off-road left, right angle, right turn, run-off-road right, headon, sideswipe opposite, other, and type unknown accidents		HSIS

1 mi = 1.61 km

TABLE 1. Variables collected in the study (continued)

	MINNESOTA THREE-LEGGED AND FOUR-LEGGED INTERSECTIONS			
	Variable	Meaning	Units	Source
Identifiers	int_synb	Route number		HSIS
	refpnt	nominal milepost of intersection center	miles	HSIS
	true_sta	station of intersection center	hundreds of feet	Plans
Traffic	int1	average daily traffic on major road	vehicles per day	HSIS
	int2	average daily traffic on minor road	vehicles per day	HSIS
Miscel-laneous	RHRI	Roadside Hazard Rating within ±250 ft on major road	1, 2, 3, 4, 5, 6, 7	Photologs
	ND	number of driveways within ±250 ft on major road		Photologs
	light	yes or no if lighting or no lighting		Photologs
	terrain	flat, rolling, or mountainous		Photologs
Weather	dd, wd, ss, ips	number of dry, wet, snow/slush, ice\packsnow days	days per year	MCC
Horizontal alignment on major road	pc{i}	beg. station of curve no. i (if any portion of curve is within ±764 ft of intersection center along major road)	hundreds of feet	Plans
	pt{i}	end station, curve no. i	hundreds of feet	Plans
	DEG{i}	degree of curve, curve no. i	degrees per hundred feet	Plans
	dir{i}	direction, left or right, curve no. i		Plans
Vertical alignment on major road	b{i}	beg. station of curve no. i (if any portion of curve is within ±764 ft of intersection center along major road)	hundreds of feet	Plans
	e{i}	end station of curve no. i	hundreds of feet	Plans
	g{i}	grade no. i (prior to curve no. i)	percent	Plans

1 mi = 1.61 km, 1 ft = 0.3048 m

TABLE 1. Variables collected in the study (continued)

	MINNESOTA THREE-LEGGED AND FOUR-LEGGED INTERSECTIONS, continued			
	Variable	Meaning	Units	Source
Speed on major road	advspd	advisory speed	miles per hour	Photologs
	regspd	regulatory speed	miles per hour	Photologs
	ap_spd	posted approach speed, both legs	miles per hour	HSIS
Accident data	TOTACC	number of intersection accidents or intersection-related accidents occurring within ±250 feet of intersection on major road during 1985-1989, 1990-93		HSIS
	fatal, injury, nonincap, possinj, injunk, propdam	no. of fatal, injury, non-incapacitating, possible injury, injury unknown, and property damage only accidents		HSIS
	rearend, sswipe, leftturn, rorleft, rtangle, riteturn, rorright, headon, sswipopp, other, unknown	no. of rearend, sideswipe, left turn, run-off-road left, right angle, right turn, run-off-road right, headon, sideswipe opposite, other, and type unknown accidents		HSIS

1 mi = 1.61 km

TABLE 1. Variables collected in the study (continued)

	Variable	Meaning	Units	Source
	MINNESOTA THREE-LEGGED INTERSECTIONS ONLY			
Angle	angle	angle between increasing direction of major road and third leg	degrees	Plans
	dir_ang	direction of third leg (left or right) from increasing dir. of major road		Plans
Channel-ization	tlml	yes or no whether a right turn lane does or does not exist on major road		Photologs
	tlcs	yes or no whether a right turn/acceleration lane does or does not exist on the minor road		Photologs
	bypass	yes or no whether a bypass lane does or does not exist on the major road (opposite the minor road)		Photologs
	MINNESOTA FOUR-LEGGED INTERSECTIONS ONLY			
Angle	l_angle	angle between increasing direction of major road and left leg of minor	degrees	Plans
	r_angle	angle between increasing direction of major road and right leg of minor	degrees	Plans
Channel-ization	tlml1	yes or no whether a right turn lane does or does not exist along increasing direction of major road		Photologs
	tlml2	yes or no whether a right turn lane does or does not exist along decreasing direction of major road		Photologs
	l_tlcs	yes or no whether a right turn/acceleration lane does or does not exist on the left leg of the minor road		Photologs
	r_tlcs	yes or no whether a right turn/acceleration lane does or does not exist on the right leg of the minor road		Photologs

TABLE 1. Variables collected in the study (continued)

	WASHINGTON SEGMENTS			
	Variable	Meaning	Units	Source
Identifiers	rte_nbr	Route number		HSIS
	begmp	beg. milepost	miles	HSIS
	endmp	end milepost	miles	HSIS
Traffic	ADT	average daily traffic	vehicles per day	HSIS
	com_avg	average daily heavy vehicle traffic	vehicles per day	HSIS
Miscel-laneous	LW	lane width	feet	HSIS
	SHW	shoulder width	feet	HSIS
	RHR	Roadside Hazard Rating	1, 2, 3, 4, 5, 6, 7	Photologs
	nodrwy	number of driveways		Photologs
	noint	number of intersections		Photologs
	light	yes or no if lighting or no lighting		Photologs
	terrain	flat, rolling, or mountainous		Photologs
Horizontal alignment	pc{i}	beg. milepost of curve no. i	miles	HSIS
	pt{i}	end milepost of curve no. i	miles	HSIS
	rad{i}	radius of curve, curve no. i	feet	HSIS
	dir{i}	direction, left or right, curve no. i		HSIS
Vertical alignment	b{i}	beg. milepost of curve no. i	miles	HSIS
	e{i}	end milepost of curve no. i	miles	HSIS
	g{i}	incoming grade no. i	percent	HSIS
	h{i}	outgoing grade no. i	percent	HSIS

1 mi = 1.61 km, 1 ft = 0.3048 m

TABLE 1. Variables collected in the study (continued)

	WASHINGTON SEGMENTS, continued			
	Variable	Meaning	Units	Source
Speed	advspd	advisory speed	miles per hour	Photologs
	regspd	regulatory speed	miles per hour	Photologs
	spd_limt	posted speed	miles per hour	HSIS
	hspd{i}	speed on horizontal curve no. i	miles per hour	HSIS
	vspd{i}	speed on vertical curve no. i	miles per hour	HSIS
Accident data	TOTACC	total number of non-intersection accidents in 1993-5		HSIS
	fatal, injury, nonincap, possinj, injunk, propdam	no. of fatal, injury, non-incapacitating, possible injury, injury unknown, and property damage only non-intersection accidents		HSIS
	RORACC	number of run-off-road accidents		HSIS

1 mi = 1.61 km

TABLE 1. Variables collected in the study (continued)

	WASHINGTON THREE-LEGGED AND FOUR-LEGGED INTERSECTIONS			
	Variable	Meaning	Units	Source
Identifiers	rte_nbr	Route number		HSIS
	arm	accumulated milepost of intersection center	miles	HSIS
Traffic	ADT1	average daily traffic on major road	vehicles per day	HSIS
	ADT2	average daily traffic on minor road	vehicles per day	HSIS
Miscel-laneous	RHRI	Roadside Hazard Rating within ±250 ft on major road	1, 2, 3, 4, 5, 6, 7	Photologs
	ND	number of driveways within ±250 ft on major road		Photologs
	light	yes or no if lighting or no lighting		Photologs
	terrain	flat, rolling, or mountainous		Photologs
Horizontal alignment on major road	pc{i}	beg. milepost of horizontal curve no. i (if any portion of curve is within ±764 ft of intersection center along major road)	miles	HSIS
	pt{i}	end milepost, curve no. i	miles	HSIS
	rad{i}	radius of curve, curve no. i	feet	HSIS
	dir{i}	direction, left or right, curve no. i		HSIS
Vertical alignment on major road	b{i}	beg. milepost of vertical curve no. i (if any portion of curve is within ±764 ft of intersection center along major road)	miles	HSIS
	e{i}	end milepost of vertical curve no. i	miles	HSIS
	g{i}	grade no. i	percent	HSIS

1 mi = 1.61 km, 1 ft = 0.3048 m

TABLE 1. Variables collected in the study (continued)

	WASHINGTON THREE-LEGGED AND FOUR-LEGGED INTERSECTION, continued			
	Variable	Meaning	Units	Source
Speed on major road	advspd	advisory speed	miles per hour	Photologs
	regspd	regulatory speed	miles per hour	Photologs
	ap_spd	posted approach speed	miles per hour	HSIS
Accident data	TOTACC	number of intersection accidents or intersection-related accidents occurring within ±250 feet of intersection on major road during 1985-9, 1990-3		HSIS
	fatal, injury, nonincap, possinj, injunk, propdam	no. of fatal, injury, non-incapacitating, possible injury, injury unknown, and property damage only accidents		HSIS
	rearend, sswipe, leftturn, rorleft, rtangle, riteturn, rorright, headon, sswipopp, other, unknown	no. of rearend, sideswipe, left turn, run-off-road left, right angle, right turn, run-off-road right, headon, sideswipe opposite, other, and type unknown accidents		HSIS
	RORACC	number of run-off-road accidents		HSIS

1 mi = 1.61 km

TABLE 1. Variables collected in the study (continued)

	Variable	Meaning	Units	Source
	WASHINGTON THREE-LEGGED INTERSECTIONS ONLY			
Angle	angle	angle between increasing direction of major road and third leg	degrees	Photologs
	dir_ang	direction of third leg (left or right) from increasing dir. of major road		Photologs
Channel-ization	tlml	yes or no whether a right turn lane does or does not exist on major road		Photologs
	tlcs	yes or no whether a right turn/acceleration lane does or does not exist on the minor road		Photologs
	bypass	yes or no whether a bypass lane does or does not exist on the major road (opposite the minor road)		Photologs
	WASHINGTON FOUR-LEGGED INTERSECTIONS ONLY			
Angle	l_angle	angle between increasing direction of major road and left leg of minor	degrees	Photologs
	r_angle	angle between increasing direction of major road and right leg of minor	degrees	Photologs
Channel-ization	tlml1	yes or no whether a right turn lane does or does not exist along increasing direction of major road		Photologs
	tlml2	yes or no whether a right turn lane does or does not exist along decreasing direction of major road		Photologs
	l_tlcs	yes or no whether a right turn/acceleration lane does or does not exist on the left leg of the minor road		Photologs
	r_tlcs	yes or no whether a right turn/acceleration lane does or does not exist on the right leg of the minor road		Photologs

LIMITATIONS ON DATA QUALITY

As noted, numerous checks were performed on the data. Examples of such checks were repeated reviews of plans and photologs, comparisons of values of multiple variables for consistency (for example, radius of curvature versus degree of curve), use of computer programs to flag unusually large values of variables, and to confirm that ordering was preserved (beginning milepost comes earlier than end milepost for each curve). However, the accuracy of the data was limited by a number of inherent factors discussed below.

Accident Data

Accident data were obtained from HSIS files.

Segment accidents were required to be "non-intersection" accidents, i.e., accidents that did not occur at intersections and were not intersection related. Intersection accidents were accidents at intersections in the database and all intersection-related accidents occurring within ± 250 feet of an intersection in the database. In the Minnesota data, a variable called "INTERSE" was used in the segment database to exclude accidents with the values "intersection" or "intersection-related" and in the intersection databases to include accidents with precisely these values. In Washington a variable called "LOC_TYPE" was used in the segment database to eliminate all accidents coded as: at intersection and related, intersection related but not at intersection, at intersection but not related, driveway within intersection. Likewise, "LOC_TYPE" was used to retain precisely these accidents when they were within 250 feet of the intersection under study. Accidents occurring on the minor road at an intersection approach were typically coded to the major road at the intersection.

Severities were also recorded for each accident, while accident types (run-off-road, etc.) were recorded for Minnesota. In the case of Washington, accident types were not recorded since the accident file has elaborate subcategories that differ significantly from those of Minnesota. An exception was made in the case of run-off-road accidents. A Washington State variable called "V1EVENT2" in the HSIS file was used to estimate whether an accident was of run-off-road type: If the accident was a single vehicle accident in which the vehicle struck an appurtenance or other object, overturned, ran into a ditch or river or over an embankment (these are categories in the file), it was taken to be a run-off-road accident.

Underreporting of accidents was a matter of some concern. In both States during the time periods under consideration, accidents involving either injuries or property damage of $500 or more had to be reported. In Minnesota the reporting threshold rose to $1,000 as of August 1, 1994. The amount of any underreporting is a matter of speculation (one source in Minnesota thought there might be one minor unreported accident for each reported one because accident-prone drivers wish to avoid both penalties for intoxication and insurance premium increases).

The reliability of the reported accident characteristics depends on the acumen of the reporting officer or official and witnesses as well as on the comparability of variables between the two States.

Traffic Data

The HSIS traffic variables in Table 1, ADT and com_avg, derive from Minnesota and Washington traffic count data.

ADT data for the Minnesota segments appear to have been reliably estimated on a timely basis. Two multi-year data sets, 1985-1987 and 1988-1989, and four annual data sets, 1990, 1991, 1992, and 1993, were available for this study. The traffic data in these sets seem to have been based on measurements and calculations, e.g., interpolation and/or extrapolation both along roads and in time. The HSIS Guidebook dated October 1993 notes that traffic data on major roads are collected on a two-year cycle, and on minor rural roads on a four-year cycle, and that growth factors are applied for the years in which measurements are not made.

According to MNDOT manual counts, including detailed classification of vehicle types, are done at about a thousand sites around the State. In a manual count a person stands at the roadside and counts and classifies every vehicle that passes over a 16-hour period (from 6 AM to 10 PM on a weekday). One hundred of the sites, the major ones, are counted every 2 years; and another 900 every 6 years. Every 2 years estimates are produced of ADT and commercial ADT throughout the State. Count locations do not exist on every segment but are averaged from those of adjacent segments along relatively homogeneous roads. A count might be done once in, say, 6 miles in some places.

The vehicle types that are summarized under the variable com_avg in Table 1 are heavy vehicles, defined as those with two or more axles and six or more tires. On roads with low traffic, about 25% of the heavy vehicle traffic consists of five-axle semis, usually with 18 wheels; on roads with high traffic about 75% is five-axle semis. A twin trailer (cab + tractor + trailer + another trailer) with perhaps five or six axles, along with most three-axled trucks without tractors, would be counted as a heavy vehicle but not a semi. The variable com_avg is thought not to be as accurate as ADT.

Minnesota intersection traffic data are somewhat less reliable than segment traffic data. The intersection files from Minnesota give traffic counts for both the major and minor roads, along with the year in which these data were acquired. Not only are the years quite variable from intersection to intersection, varying from 1976 to 1992, but very few of them appear to have been updated between the 1985-1989 time period files and the 1990-1993 time period files. Traffic counts had been made only once in the years from 1987 to 1993 and annual files just repeated the value of an earlier year. In other cases no traffic counts had been made since 1986 or earlier

In view of this unreliability, efforts were made to determine a growth rate factor that could be used to update traffic counts to the time periods of interest. MNDOT personnel reported that population growth rates did not relate in a simple fashion to traffic flow (so traffic counts on an intersection could not be updated from one year to the next by a population growth multiplier). Sometimes traffic counts will be higher when new development and construction is going on and then will ease off when the buildings and houses are occupied. A program was written to extract a growth rate by least squares from traffic data for segments near the intersection and thereafter use the year of intersection traffic count to extrapolate to an ADT for the years 1987 and mid-1991. The Minnesota intersection traffic variables used in the modeling and validation below, ADT1 and ADT2, were derived from int1 and int2 by means of this program.

Washington State traffic data became available at a relatively late stage of this study but only for segments and for some intersections along segments. The traffic data were based on upstream traffic counts, but in some cases the count stations were rather far upstream, 10 or more miles. The Project Team considered averaging a downstream count and an upstream count when the upstream count was at a significant distance, but decided against it in order to maintain conformity with HSIS files. The chief concern with these data, apart from the distance of count stations, is that routes, alternate routes, and each half of certain divided highways have similar labels and considerable programming is required to ensure that a count lies on a route of interest rather than a related one. According to the HSIS Washington Guidebook, a small number of the count stations are permanent and a large number of others are used for 72-hour counts every second or third year. The counts for com_avg are considered to be less reliable than the overall counts, in part because they are based on fewer stations. Washington State Department of Transportation personnel observed that the truck counts are done on weekdays, that com_avg is based on this figure, and that it might be better to take the weekday figure and add 10% to 20% to get the overall weekly value. It was also noted that the percentage of truck traffic on a road can vary from 4% to 17% at different times of year, chiefly because of seasonal variation in the nontruck traffic.

Alignment Data

Horizontal and vertical alignment data came from construction plans in the case of Minnesota and from HSIS horizontal and vertical curve files in the case of Washington.

The Minnesota plans varied in age from a few years prior to 1985 to approximately 1920. Special effort was made to determine that these plans showed the latest alignment or realignment and that no realignment was done during the time periods under study. Nonetheless it is possible that some roads were realigned and that plans were never conveyed to the Minnesota Plan Office. The Plan Office plans are primarily Federal aid projects, and State and County aid projects sometimes do not get recorded at the State Plan Office. In addition to location problems (discussed below), problems sometimes arose because of illegibility of markings on the plan and inconsistencies between alternative measures (e.g., radius versus degree of curve, or beginning and end of curve versus length of curve) written on the plan. These were typically resolved by a judgment as to which number was most plausible. A few horizontal curves had spiral transitions at beginnings and/or ends. These were not recorded but a judgment was made as to a beginning and endpoint for a single idealized horizontal curve. A very small fraction, 2% or less, of vertical curves were represented in the plans as angle points, where the grade changes without a transition, typically a small change. Our initial understanding was that no such transitions occurred on Minnesota major roads and these points were edited so that a transition curve of 50 feet was introduced. Later, visiting Minnesota engineers reported that angle points do occasionally occur on main roads.

The Washington State alignment data were represented by a Horizontal Curve file and a Vertical Curve file. Many segments and intersections were eliminated from the sample because of anomalies in the values in these files, but the ones that remained also had minor anomalies. Because of rounding errors in the original Washington data (not enough significant digits kept) some curves appeared to overlap, and editing had to be done to restore plausible beginning and ending points for curves. In addition in some cases there were small differences between the ending grade of one vertical curve and the beginning grade of the next. When the intervening stretch was treated as a straightaway during the modeling, its grade was taken to be the average of the two neighboring grades. A few angle points occurred for both horizontal and vertical curves with small grade changes or small angle change. Curve lengths were adjusted to 50 feet for these exceptional cases.

Location Uncertainties

Minnesota data compilation was hampered by the fact that HSIS files, Minnesota photologs, and Minnesota construction plans use three different ways of measuring distance: true mileposts, nominal mileposts, and control stations. HSIS variables begmp and endmp and true_beg and true_end refer respectively to nominal beginning and ending mileposts and true beginning and ending distances of segments. Both the Minnesota photologs and the Minnesota accident data are keyed to nominal mileposts rather than true distances, and the primary usage of true_beg and true_end is to calculate segment length. The milepost of an accident in the accident files is nominal rather than true distance, and the tenths of a mile shown on Minnesota photologs are nominal mileposts not true distance. This was confirmed by MNDOT personnel and by comparison of photologs with the Minnesota List-Trumile-File for Trunk Highways. This latter book, a print-out of a file (our copy was dated September 1, 1988) obtained in Minnesota, had a listing of all State highways along with reference posts (i.e., nominal mileposts), true distances, and control stations, most of the entries effective as of 1977 (but with some updates as recent as 1983).

Control stations, used in the construction plans, are local numbers, in hundreds of feet, and may be equated to nominal mileposts by use of the just mentioned file. Many plans contain station adjustments (places where a gap in the stations occurs) and converting back and forth between the various units is an art. This conversion is especially difficult for intersections. The intersection reference point, the nominal milepost of the intersection center, is sometimes not adequately tied to construction plans or to features on the photologs: station numbers of nearby landmarks are occasionally either wrong or absent, and interpolation adds a further source of error. Plans, sometimes of ancient vintage, do not show an intersection or expected landmark, or else are ambiguous (two or more intersections or landmarks shown in the plan are plausible candidates for the sought after one). This is particularly true of three-legged intersections since these are the least well-marked, least documented, and least significant data class.

Linking a particular intersection to its photolog and to a particular site on a plan involves a comparison among four different numbers: the reference point for the intersection, the distance recorded on the photolog, the true distance recorded by the State, and the station number in the construction plans. Sometimes discrepancies occur among these numbers: the intersection may be at a slightly different point than expected in the photolog, or it may be several hundred feet away from its expected location in the plan. When the plan does not show an intersection in the near vicinity of the expected spot, an identifiable landmark must be found to verify locations and in some cases this is quite difficult.

For Washington State data, distances are measured in ARM's (accumulated route miles). The ARM is a true milepost, used in all of the HSIS files: roadway, traffic, accident, and alignment. Only the videotapes are in nominal mileposts, but a logbook permits unambiguous translation back and forth. Discrepancies were rare, perhaps because Washington Department of Transportation personnel had already resolved them. The only issue of concern was rounding errors, noted above.

A final caveat with respect to location concerns the accident data. MNDOT indicated that the accident data reviewers attempt to locate a nearby physical feature mentioned in the police report. They then determine the reference point for that feature and add an adjustment, typically a few hundred feet, to get to the accident site. The reviewers aim to get within 50 feet of the true accident site. They also assign a reliability code to their estimate.

Time Uncertainties

HSIS traffic and roadway data, the Minnesota construction plan data, and the photolog data are all supposed to apply to the time intervals under consideration. Rural areas might be expected to change more gradually than urban and suburban areas. However, some variables such as traffic data are based on averages of discrete observations that may not be representative. Others, including Minnesota intersection traffic data discussed above, may be out of date. Photolog years in Minnesota vary from 1987 to 1990 and in Washington from 1993 to 1995; changes in the number of driveways, speed limits, channelization, etc., may have occurred before or after the photolog was obtained.

For validation of the Minnesota model, 1990-1993 data were used. Since construction plans and photologs for the new time period were unavailable, some variables could not be re-measured. So it was assumed that these were generally unchanged.

Miscellaneous Limitations

Data acquired from the photologs were subject to various limitations. Minnesota photologs in reels and CD-ROM's offered a larger visual field than the videotapes acquired from Washington State. On the other hand, the latter were accompanied by audio that indicated signage and roadside features and gave the numbers on sometimes otherwise unreadable speed limit signs. The Washington voice-over also provided intersecting street and route names and was accompanied by a written log. In both cases some effort was required to verify that minor roads had stop signs, to determine channelization, and to assess whether a driveway had been seen along the road. Driveways, for example, can sometimes be mistaken for footpaths. In addition, for Washington State the photologs were used to estimate angle of intersection between major and minor roads, and limited visibility along minor roads made this difficult.

Roadside Hazard Rating was determined from the photologs. Different observers would not always agree on the value of this subjective variable (values of two, and sometimes three, independent observers were averaged, and photologs were re-inspected in some cases). The hazard rating sometimes varied substantially along a segment. With regard to intersections, it was more difficult to arrive at values in the vicinity of Washington State intersections since the roadsides at these intersections tended to be less rural than their Minnesota counterparts (small town streets rather than country roads), and the proper rating to assign to a roadside business or residence was not always evident.

Weather data collected by the Midwest Climate Center, as already noted, were limited by the fact that they were not sufficiently local.

The treatment of intersections along a segment was not quite consistent between Minnesota and Washington. In Minnesota very few segments began or ended at an intersection, and for the few that did (thought to be less than 5%) no attempt was made to remove, say, 250 feet from the segment and shorten it by omitting the intersection vicinity. In Washington most of the segments began and/or ended with an intersection, and all such segments were shortened by removal of 250 feet at each end where an intersection was encountered. On the other hand, no internal intersections were removed from the segments in either State. In Washington 95% of the segments contained no internal intersections, but in Minnesota more than half of the segments contained at least one intersection. This means that in Minnesota accidents along segments are more likely to include accidents that happened near intersections (although they would not be intersection-related or at an intersection).

It should also be noted that some desirable variables were omitted from the study altogether, e.g., superelevations, alignments on minor roads, actual speeds, and sight distances. To some extent the latter are represented in, or can be reconstructed from, horizontal and vertical alignment as well as Roadside Hazard Rating, but a direct unambiguous measurement is lacking. Also excluded, of course, are detailed information about drivers and vehicles on the road; accident circumstances such as time of day, week, and year; and weather at the time and place of an accident. To some extent demographic conditions such as ages of drivers and law enforcement practices are incorporated in the STATE variable (see below).

SUMMARY

Minnesota and Washington State data were constrained to lie on rural two-lane roads with segment length 0.1 miles or longer with both segments and intersections having reasonable bounds on ADT. Other reasonable constraints were also imposed, including relatively complete and consistent data for the time periods of interest. Many observations from the original populations were lost when these constraints were imposed, but good-sized samples remained. The Washington intersection samples, "opportunity" samples, were smaller than the other samples and it is not known how representative they are of the population of Washington State intersections.

Data collected include: accident counts, exposure and ADT, lane and shoulder widths, Roadside Hazard Rating, number of driveways, horizontal and vertical alignments, commercial traffic percentage, weather (in Minnesota), intersection angles and channelization, and speed limits. These data are often estimates based on averages and are subject to some uncertainties in location and time. ADT's are based on observations at selected sites, interpolation, and/or extrapolation, and are particularly crude estimates in the case of intersections. In view of the importance of ADT in the modeling, the crudity of these estimates should serve as a caution.

Driver and vehicle characteristics were not collected, nor were such design variables as sight distances and minor road alignments.

Despite shortcomings in quality and completeness, the data obtained provide a relatively diverse and comprehensive basis for analysis and modeling.

4. ANALYSIS

To analyze the data acquired for the segments and intersections, a variety of new variables were developed based on the originally collected variables. It has already been noted that the traffic variables used for modeling the Minnesota intersections were obtained from the original variables by applying growth factors from nearby segments. There was significant variation in the number and size of vertical and horizontal curves from segment to segment and from one intersection to the next. Thus aggregate variables were developed for vertical and horizontal alignment to summarize alignment data and permit direct comparison of one observation with another. Other variables were developed for such items as exposure, driveway density, and intersection density. A speed variable was developed from the multiple speed variables collected.

For both the new variables and the old, univariate statistics were compiled showing their distributions in each data set. In preparation for the modeling effort, bivariate comparisons were also done to reveal correlations between variables and to clarify relationships among variables.

In this chapter we discuss the new variables and exhibit and review the univariate and bivariate statistics for both old and new variables. See the Index of Variables, at the beginning of this report, for a comprehensive listing of variables used in the modeling.

NEW VARIABLES

Accident Variables

Accident data for all data sets includes information on severities. So, in addition to the variable TOTACC for all non-intersection accidents along a segment and all intersection accidents within 250 feet of an intersection, a variable, INJACC, excluding property damage only accidents was introduced. INJACC counts fatal accidents and the various types of injury accidents (fatal + injury + non-incapacitating + possible injury). In the case of Minnesota some logistic modeling of severities was also done to determine the probability that an accident is severe. This made use of a severity variable Y defined on an accident database developed at the same time as the Minnesota segment and intersection data sets. This variable had value 1 if an accident was in one of the first two classes (fatal or injury) and value 0 otherwise (non-incapacitating, possible injury, or property damage only).

Run-off-road accidents are described by the variable RORACC. In Minnesota this is the sum of run-off-road left accidents and run-off-road right accidents. In Washington it was obtained indirectly from the HSIS variable V1EVENT2, as explained earlier.

Traffic Variables

A variable seg_lng, representing segment length in miles, is used to develop an exposure variable

EXPO for segments. Seg_lng is obtained from true_beg and true_end in Minnesota and from begmpr and endmpr in Washington data (begmpr and endmpr are begmp and endmp with 250 feet removed if the segment begins or ends at an intersection). The variable EXPO is then given by:

$$EXPO = \frac{ADT \times 365 \times (number\ of\ years\ in\ time\ period) \times seg\_lng}{10^6}$$

The units of EXPO are millions of vehicle-miles (MVM).

The Minnesota and Washington intersection traffic variables are ADT1 and ADT2. These represent estimated average daily traffic on the major and minor road, respectively. As noted already, for Minnesota these variables are derived by applying growth factors to the Minnesota traffic variables, which tend to be somewhat out of date. In addition, a variable CINDEX, conflict index, is used for Minnesota intersection accident severity modeling. CINDEX is defined to be the ratio of average daily traffic entering the intersection from the minor road to average daily traffic entering the intersection from both minor and major road. CINDEX is given by:

$$CINDEX = \begin{matrix} \dfrac{ADT2}{ADT1 + ADT2} \ for\ four\text{-}legged\ intersections, \\ \dfrac{(1/2)ADT2}{ADT1 + (1/2)ADT2} \ for\ three\text{-}legged\ intersections. \end{matrix}$$

Commercial traffic is represented in both segment and intersection databases by the variable T:

$$T = \frac{100 \times (com\_avg)}{ADT}.$$

Horizontal Alignment Variables

For horizontal curves DEG{i}, the degree of curve in degrees per hundred feet, is an important variable. It was present in the Minnesota data, while in the Washington data it had to be computed from the familiar formula:

$$DEG\{i\} = \frac{18{,}000}{\pi \times rad\{i\}},$$

where the radius is in feet.

Various criteria were considered to determine how horizontal curves that were not entirely within a segment would be treated. One possible approach was to restrict attention to horizontal curves whose midpoints lie in the segment. This possibility was explored. However, the approach ultimately adopted was to regard a horizontal curve as eligible if any portion of it overlapped the segment. Variables associated with individual eligible horizontal curves are:

$$WH\{i\} = \frac{length\ of\ portion\ of\ horizontal\ curve\ no.\ i\ within\ segment}{seg\_lng}$$

and

$$whm\{i\} = \frac{length\ of\ horizontal\ curve\ no.\ i}{seg\_lngh},$$

where seg_lngh is the segment length increased by adding on any portions of horizontal curves that fall outside the segment. These dimensionless weights are two different ways of taking into account the fact that horizontal curves may lie partly inside a segment and partly outside (or can even properly contain the segment). If two-thirds of the curve is inside, WH{i} has a numerator equal to two-thirds the numerator of whm{i} while the latter has a denominator equal to the denominator of WH{i} plus one-third the curve length plus lengths of portions of any other horizontal curve that lie outside. These weights are intrinsically non-negative, summing to a number less than or equal to 1.

Although in the final model for segments the variable WH{i} appears explicitly and each horizontal curve makes a separate contribution, in general the curves have to be aggregated in some fashion. The following aggregate variables are used in some segment models:

$$\begin{aligned}
H &= \textstyle\sum_i WH\{i\} \times DEG\{i\} \\
HM1 &= \textstyle\sum_i whm\{i\} \times DEG\{i\} \\
HM1.5 &= \textstyle\sum_i whm\{i\} \times (DEG\{i\})^{1.5} \\
HM2 &= \textstyle\sum_i whm\{i\} \times (DEG\{i\})^{2}.
\end{aligned}$$

For the study of horizontal curves at intersections, each intersection was treated as a segment extending ± 250 feet along the major road from the intersection center or sometimes ± 764 feet. Two hundred fifty feet (or approximately 75 meters) is a typical length of an acceleration lane onto the major road, while 764 feet (approximately 233 meters) is a typical distance required for a vehicle turning onto a major road from a minor leg to achieve reasonable speed. Horizontal curves were considered eligible if they met this artificial segment. Aggregate variables of the following form were considered:

$$HI = \frac{\sum_i DEG\{i\}}{Number\ of\ horizontal\ curves\ overlapping\ intersection\ center\ \pm 250\ feet}$$

$$HEI = \frac{\sum_j DEG\{j\}}{Number\ of\ horizontal\ curves\ overlapping\ intersection\ center\ \pm 764\ feet},$$

where the sum is over the corresponding curves. HI and HEI (E for extended) are the unweighted averages of the degrees of curvature of the corresponding curves.

Vertical Alignment Variables

Vertical alignment variables are subject to some of the same considerations as horizontal alignment variables.

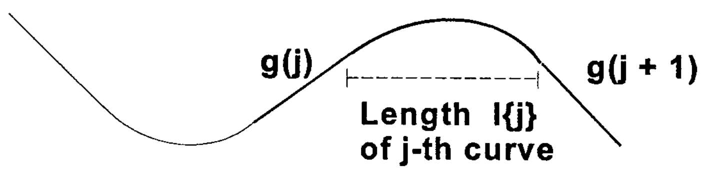


FIGURE 3. A VERTICAL CURVE

A basic variable associated with each vertical curve is V{j}:

$$V\{j\} = \frac{absolute\ value\ of\ change\ of\ grade\ at\ j\text{-}th\ vertical\ curve}{length\ l\{j\}\ of\ j\text{-}th\ vertical\ curve\ in\ hundreds\ of\ feet},$$

with units of percent per hundred feet. Change of grade $\Delta g\{j\}$ equals g{j} - g{j+1} for the Minnesota data and g{j} - h{j} for the Washington data and l{j} is the length of the curve in hundreds of feet. Likewise a weight is associated with each individual curve that meets a segment, namely WV{j}:

$$WV\{j\} = \frac{length\ of\ portion\ of\ vertical\ curve\ no.\ j\ within\ segment}{seg\_lng}.$$

The aggregate variables VC, VM, VMC, and VMCC were used for segment models:

$$VC = \sum_j WV\{j\} \times V\{j\} \quad (crests\ only)$$

$$VM = \frac{\sum_j V\{j\}}{seg\_lngv} \quad (all\ vertical\ curves)$$

$$VMC = \frac{\sum_j V\{j\}}{seg\_lngvc} \quad (crests\ only)$$

$$VMCC = \frac{\sum_j V\{j\}}{seg\_lngvcc} \quad (crests\ of\ type\ I\ only)\ .$$

Crest curves are vertical curves for which the grade decreases (positive to negative, positive to less positive, negative to more negative), and crests of type I are crests for which the grade changes sign.[36] The last three variables are unweighted averages of the V{j} variable, and their denominators equal seg_lng plus the length of portions of the corresponding curves that lie outside the segment. The units of the denominators are miles. Variables for sag curves, for vertical curves with grade increases, and for sags of type III (with sign change) were also considered separately in Minnesota, but were not as significant as the crest variables.

For intersections three vertical variables were considered:

$$VCI = \frac{\sum_j V\{j\}}{Number\ of\ vertical\ crest\ curves\ overlapping\ intersection\ center\ \pm 250\ feet}$$

$$VI = \frac{\sum_j V\{j\}}{Number\ of\ vertical\ curves\ overlapping\ intersection\ center\ \pm 250\ feet}$$

$$VEI = \frac{\sum_j V\{j\}}{Number\ of\ vertical\ curves\ overlapping\ intersection\ center\ \pm 764\ feet}\ .$$

These sums are over the stipulated vertical curves, and hence VCI, VI, and VEI are unweighted averages of V{j} for each type of curve.

Complementary to vertical curves are sections of uniform grade and these also were used in the modeling for Minnesota and Washington segments. On such sections there is a constant absolute grade GR{k}. In Minnesota this was readily obtainable, but in Washington there were cases where h{k-1} and g{k} did not agree. Although other options were considered, for simplicity the segment section from e{k-1} to b{k} was treated as if it were of uniform grade with absolute grade GR{k}

[36] See the "Green Book," A Policy on Geometric Design of Highways and Streets, American Association of State Highway and Transportation Officials (AASHTO), Washington, D. C., 1994, p. 281.

= |(h{k-1}+g{k})/2|. In addition to GR{k}, each such section had a variable WG{k}:

$$WG\{k\} = \frac{\text{length of portion of uniform grade section no. } k \text{ within segment}}{seg\_lng}$$

A composite variable GR was defined:

$$GR = \sum_k WG\{k\} \times GR\{k\} \quad ,$$

where the sum is over all uniform grade sections overlapping with the segment.

Angle Variables

An angle variable DEV, representing the average deviation from 90°, was defined by:

$$DEV = \begin{cases} |angle - 90| & \text{if intersection is three-legged} \\ \dfrac{|l\_angle - 90| + |r\_angle - 90|}{2} & \text{if intersection is four-legged.} \end{cases}$$

Two more angle variables are also used. DEV15 is a variable discovered empirically that seems to be negatively correlated with accidents on four-legged intersections. Another intersection angle variable considered in this study, suggested by E. Hauer, is HAU:

$$DEV15 = \frac{(DEV - 15)^2}{100}$$

$$HAU = \begin{cases} angle - 90 & \text{if } dir\_ang \text{ is right at a three-legged intersection} \\ 90 - angle & \text{if } dir\_ang \text{ is left at a three-legged intersection} \\ \dfrac{r\_angle - l\_angle}{2} & \text{at a four-legged intersection.} \end{cases}$$

The variable HAU is a signed variable. See Figures 4 and 5 below. For a three-legged intersection with the angle to the right of the increasing direction, HAU is positive when the angle is larger than 90°, as in 4(a), and HAU is negative when the angle is smaller than 90°, as in 4(b). If the angle is to the left of the increasing direction (see Figure 5), 180° minus the angle becomes the new angle and HAU is defined as ((180 - angle) - 90) = (90 - angle), as above. For four-legged intersections, as in 4(c), it is the average of the two three-legged values (and thus 90° cancels out). Figure 5 illustrates the calculation of HAU in a variety of cases. It is thought[37] that turns from the far lane of the major road may be less accident prone in situation 4a) than in situation 4b), so that positive values of HAU correspond to fewer accidents.

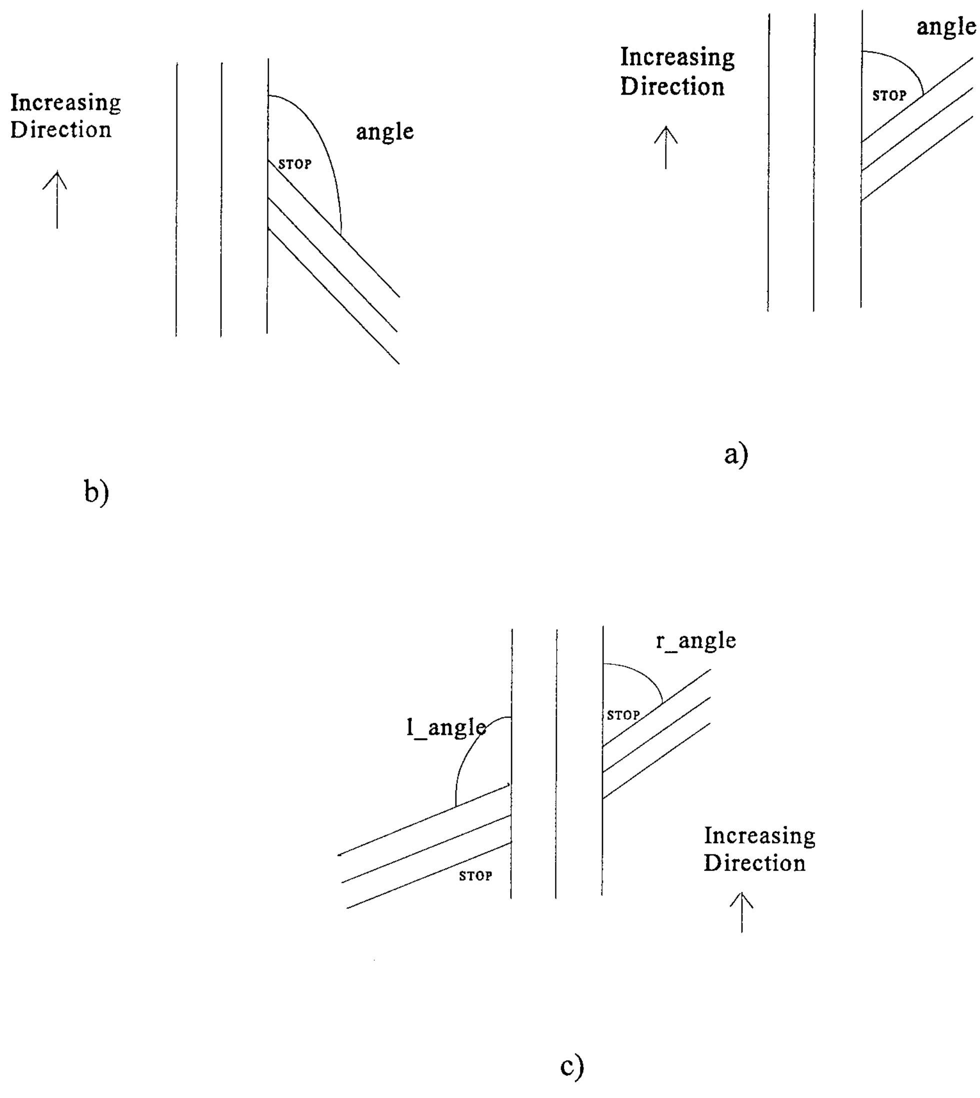


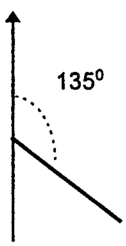
FIGURE 4. INTERSECTION ANGLE GEOMETRIES

---

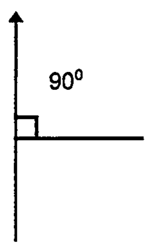
[37]Kulmala, R., "Safety at Three- and Four-Arm Junctions: Development and Application of Accident Prediction Models," VTT Publication 233, Technical Research Centre of Finland, Espoo, 1995.

For 3-legged Intersections:

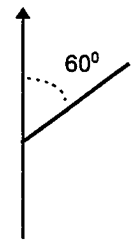
A. Minor road to right of major road in direction of increasing mileposts



HAU = angle - 90
= 135 - 90
= 45

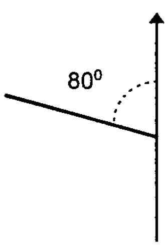


HAU = angle - 90
= 90 - 90
= 0

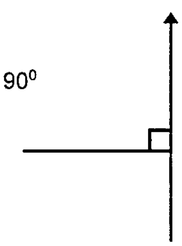


HAU = angle - 90
= 60 - 90
= -30

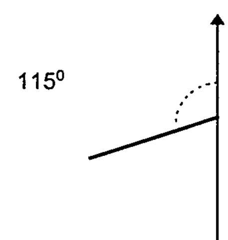
B. Minor road to left of major road in direction of increasing mileposts



HAU = 90 - angle
= 90 - 80
= 10

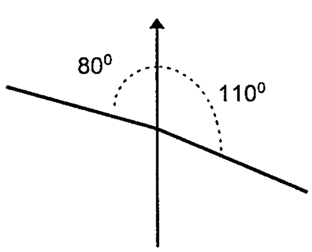


HAU = 90 - angle
= 90 - 90
= 0



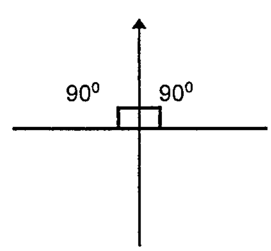
HAU = 90 - angle
= 90 - 115
= - 25

For 4-legged Intersections:

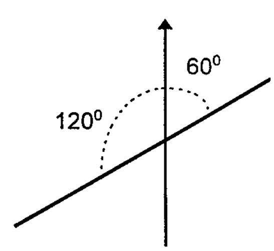


(minor road not straight)

HAU = (r_angle - l_angle) / 2
= (110 - 80) / 2
= 15



HAU = (r_angle - l_angle) / 2
= (90 - 90) / 2
= 0



HAU = (r_angle - l_angle) / 2
= (60 - 120) /2
= -30

FIGURE 5. EXAMPLES OF CALCULATION OF THE ANGLE VARIABLE HAU

Miscellaneous Variables

Some other segment variables included in the study are TOTWIDTH, DD, INTD, STATE, and SPD:

$$
\begin{aligned}
TOTWIDTH &= 2 \times (LW + SHW) \\
DD &= \frac{nodrwy}{seg\_lng} \\
INTD &= \frac{noint}{seg\_lng} \\
STATE &= \begin{cases} 0 & \text{if Minnesota} \\ 1 & \text{if Washington} \end{cases} \\
SPD &= \text{average speed along segment.}
\end{aligned}
$$

SPD is an amalgam of advisory and posted speeds seen on some roads together with HSIS speeds. Advisory and regulatory speeds, if seen on photologs, were given preference. However, photolog speeds were not collected for some Minnesota segments, were missing for others even when the photolog was searched a few miles outside the segment, and had multiple values in some cases when seen (i.e., changes in speed along a direction, different speeds in opposing directions, a difference between regulatory and advisory speed). Minnesota HSIS speeds were for accident sites only (at the same segment or a nearby one). For Washington data, a posted speed variable was obtained from the HSIS roadway file, together with speeds for each horizontal and vertical curve from the HSIS alignment files. Averaging these to achieve a single number could not be done without some subjectivity.

Other intersection variables are RT and SPDI:

$$
RT = \begin{cases} 1 & \text{if one or more right turn lanes exist on the major road} \\ 0 & \text{if no right turn lane exists on the major road} \end{cases}
$$

$$
SPDI = \text{average incoming speed at intersection along major road.}
$$

SPDI is an amalgam of mainline speeds observed at intersections, averaged by approach where possible.

Finally, two weather variables NONDRYP and SNP were devised for use with the Minnesota data:

$$
\begin{aligned}
NONDRYP &= \text{fraction of nondry days in } 1985\text{–}89 \\
SNP &= \text{fraction of snow days in } 1985\text{–}89.
\end{aligned}
$$

UNIVARIATE STATISTICS

Tables 2 through 7 indicate the behavior of the chief variables on the six data sets: segments, three-legged intersections, and four-legged intersections in both Minnesota and Washington. It is instructive to make comparisons among these tables and in the case of Minnesota to compare the sample data with the population data in Appendix 1.

Minnesota versus Washington

Accidents tend to be more serious in Washington State than in Minnesota for segments and intersections, and the accident rate (accidents per MVM) on segments is much higher in Washington than in Minnesota. The accident rates appear to be comparable in the two States on intersections, but this may be somewhat misleading since the conflict index is lower for Washington than Minnesota. There also appears to be a higher percentage of run-off-road accidents in Washington. (This may be due to the indirect method employed to count Washington run-off-road accidents.)

There is more traffic in Washington on segments and major intersection approaches, and a higher density of driveways. Both of these suggest that the Washington data sets are less rural than those of Minnesota. Annual exposure (MVM per year) is about the same on average in both States, and this is accounted for by the fact that segment lengths are shorter on average in Washington.

Roadside Hazard Rating tends to be higher in Washington, with steeper grades. Washington averages for horizontal and vertical alignment are the same as or higher than Minnesota's, but Washington tends to have fewer curves than Minnesota both on segments and in the vicinity of intersections. This may reflect historical differences in highway design practice and/or in the principles used to label roadway segments as segments. Likewise, Minnesota appears to have more angular variation at intersections than Washington (perhaps due in part to data shortcomings), and more turning lanes on the major road. Minnesota has wider shoulders than Washington, but Washington has more that are paved. These differences may also reflect design considerations and history.

Segments versus Intersections

Accidents at intersections tend to be more serious than those on segments, and accidents at intersections are more frequent (if an intersection is regarded as a segment 500 feet long), other things being equal. ADT rises as one goes from segments to major roads of three-legged intersections to major roads of four-legged intersections. The tables also show that three-leggeds tend to have more horizontal curvature than four-leggeds, but that vertical alignment tends to be about the same in three-leggeds and four-leggeds.

Minnesota Sample versus Population

The Minnesota samples are quite comparable in the distribution of severities and the percentage of run-off-road accidents to their counterparts in the Minnesota populations represented in Appendix 1. With respect to segments, we can also compare ADT, commercial vehicle percentage, and lane width and find that they are quite similar between the sample and the population. Shoulder width and shoulder type between sample and population are also similar although there seems to be a slight tendency for the population of segments to have less shoulder width (albeit more of it paved) than the sample does.

TABLE 2. Summary Statistics: 619 Segments, Minnesota State
Two-Lane Rural Roads, 1985 -1989

Variable and Abbreviation	Min.	Max.	Median	Mean	Freq	%Zero
Total Number of Accidents (TOTACC)	0	58	1	2.74	1694	36.3
Total Number of Injury Accidents (INJACC)	0	23	0	0.99	614 (36.25%)	58.3
Total Number of Run-Off-Road Accidents (RORACC)	0	15	0	0.88	547 (32.29%)	61.6
Severity: Fatal = K Injury = A Non-incap = B Poss-inj = C Prop-dam = P					32 (1.9%) 89 (5.2%) 256 (15.1%) 237 (14.0%) 1080 (63.7%)	
Accident Rate (0.6656 TOTACC/MVM)	0	9.32	0.44	0.70		36.3
Injury Accident Rate (0.2413 INJACC/MVM)	0	4.66	0	0.25		58.3
Average Daily Traffic = ADT	208	15,162	1,866	2,402		
Segment Length = seg_lng (miles)	0.1	8.237	0.659	1.14		
Exposure over five years = EXPO (MVM)	0.13	68.32	2.25	4.11		
Commercial Vehicle Percentage = T (%)	1.90	26.86	9.87	10.45		
Lane Width = LW (feet)	10	12	12	11.54		
Shoulder Width = SHW (feet)	0	12	8	7.08		0.3
Shoulder Type None Gravel or Stone Composite Paved					2 (0.3%) 341 (55.1%) 35 (5.7%) 241 (38.9%)	

1 mi = 1.61 km, 1 ft = 0.3048 m

TABLE 2. Summary Statistics: 619 Segments, Minnesota State (continued) Two-Lane Rural Roads, 1985 -1989

Variable and Abbreviation	Min.	Max.	Median	Mean	Freq	%Zero
Roadside Hazard Rating = RHR = 1 2 3 4 5 6, 7	1	6	2	2.14	174 (28.1%) 248 (40.1%) 141 (22.8%) 48 (7.8%) 6 (1%) 2 (0.3%), 0	
Driveway Density = DD	0	100	3.73	6.58		22.5
Intersection Density = INTD	0	22.7	1.14	2.60		31.7
Light Yes No					5 (0.8%) 614 (99.2%)	
Terrain Flat Rolling Mountainous Missing (not noted)					249 (46.2%) 28 (4.5%) 1 (0.2%) 341 (55.5%)	
Degree of Curve = H = $\sum WH\{i\}DEG\{i\}$	0	7.50	0.078	0.51		33.4
Crest Curve Grade rate = VC = $\sum WV\{i\}(\lvert\Delta g\{i\}\rvert/l\{i\})$ (crests only)	0	0.89	0.037	0.067		16.5
Absolute Grade = GR = $\sum WG\{i\}GR\{i\}$ (%)	0	4.46	0.24	0.38		1.9
Speed = SPD (mph)	20	55	55	48.7		
Snow Percentage = SNP	22.9	36.9	32.5	29.4		
Non-dry Percentage = NONDRYP	41.0	56.6	45.5	47.4		

1 mi = 1.61 km, 1 ft = 0.3048 m

TABLE 3. Summary Statistics: 712 Segments, Washington State
Two-Lane Rural Roads, 1993 -1995

Variable and Abbreviation	Min.	Max.	Median	Mean	Freq	%Zero
Total Number of Accidents (TOTACC)	0	29	1	2.40	1706	37.6
Total Number of Injury Accidents (INJACC)	0	13	0	1.11	790 (46.31%)	53.8
Total Number of Run-Off-Road Accidents (RORACC)	0	19	1	1.39	993 (58.21%)	48.6
Severity: Fatal = K Injury = A Non-incap = B Poss-inj = C Prop-dam = P					39 (2.3%) 130 (7.6%) 381 (22.3%) 240 (14.1%) 916 (53.7%)	
Accident Rate (1.0228 TOTACC/MVM)	0	33.60	0.649	1.096		37.6
Injury Accident Rate (0.4736 INJACC/MVM)	0	9.65	0	0.495		53.8
Average Daily Traffic = ADT	159	17,766	2,239	3,352		
Segment Length = seg_lng (miles)	0.1	13.233	0.554	0.75		
Exposure over three years = EXPO (MVM)	0.04	22.4	1.31	2.34		
Commercial Vehicle Percentage = T (%)	1.55	52.22	11.73	13.04		
Lane Width = LW (feet)	9	12	11	11.37		
Shoulder Width = SHW (feet)	0	10	5	5.01		0.8
Shoulder Type Missing or other Gravel or Stone Composite Paved					8 (1.1%) 72 (10.1%) 230 (32.3%) 402 (56.5%)	

1 mi = 1.61 km, 1 ft = 0.3048 m

TABLE 3. Summary Statistics: 712 Segments, Washington State (continued)
Two-Lane Rural Roads, 1993 -1995

Variable and Abbreviation	Min.	Max.	Median	Mean	Freq	%Zero
Roadside Hazard Rating = RHR = 1 2 3 4 5 6 7	1	7	3	3.67	38 (5.3%) 152 (21.4%) 181 (25.4%) 109 (15.3%) 128 (18.0%) 73 (10.3%) 31 (4.4%)	
Driveway Density = DD	0	85.07	6.12	10.12		18.1
Intersection Density = INTD	0	17.3	0	0.12		97.5
Light Yes No					21 (2.9%) 691 (97.1%)	
Terrain Flat Rolling Mountainous					157 (22.1%) 485 (68.1%) 70 (9.8%)	
Degree of Curve = H = $\sum$WH{i}DEG{i}	0	30.55	0.319	1.028		36.7
Crest Curve Grade rate = VC = $\sum$WV{i}(|Δg{i}|/l{i}) (crests only)	0	1.997	0.026	0.068		36.8
Absolute Grade = GR = $\sum$WG{i}GR{i} (%)	0	6.92	0.494	0.92		13.1
Speed = SPD (mph)	21.9	55	55	50.5		

1 mi = 1.61 km, 1 ft = 0.3048 m

TABLE 4. Summary Statistics: 389 Three-Legged Intersections, Minnesota State Two-Lane Rural Roads, 1985 - 1989

Variable and Abbreviation	Min.	Max.	Median	Mean	Freq	%Zero
Total Number of Accidents (TOTACC)	0	39	0	1.35	524	51.9
Total Number of Injury Accidents (INJACC)	0	17	0	0.59	229 (43.70%)	69.9
Total Number of Run-Off-Road Accidents (RORACC)	0	4	0	0.12	45 (8.59%)	90.7
Severity: Fatal = K Injury = A Non-incap = B Poss-inj = C Prop-dam = P					8 (1.5%) 26 (5.0%) 84 (16.0%) 111 (21.2%) 295 (56.3%)	
Accident Rate (TOTACC per million entering vehicles)	0	3.08	0	0.16		51.9
Accident Rate (0.269 TOTACC /YEAR)	0	7.8	0	0.269		51.9
Injury Accident Rate (0.118 INJACC/YEAR)	0	0.8	0	0.118		69.9
Average Daily Traffic on Major Road = ADT1	201	19,413	2,313	3,687		
Average Daily Traffic on Minor Road = ADT2	4.5	4,206	240	413		
Conflict Index = CINDEX	0.002	0.442	0.049	0.077		
Angular Deviation from 90° = DEV (degrees)	0	90	0	13.4		50.6
Roadside Hazard Rating = RHRI = 1 2 3 4 5 6,7	1	5	2	2.11	98 (25.2%) 184 (47.3%) 74 (19.0%) 32 (8.2 %) 1 (0.3%) 0 (0.0%)	
Number of Driveways = ND	0	9	1	1.26		37.5

1 mi = 1.61 km, 1 ft = 0.3048 m

TABLE 4. Summary Statistics: 389 Three-Legged Intersections, Minnesota State (continued)

Two-Lane Rural Roads, 1985 - 1989

Variable and Abbreviation	Min.	Max.	Median	Mean	Freq	%Zero
Light — Yes No					1 (0.3%) 388 (99.7%)	
Terrain — Flat Rolling Missing (not noted)					115 (29.6%) 28 (7.2%) 246 (63.2%)	
Degree of Curve = HI = $(1/n)\sum DEG\{i\}$	0	29	0	1.21		54.0
Crest Curve Grade Rate = VCI = $(1/m)\sum(\lvert\Delta g\{i\}\rvert/l\{i\})$ (crests only)	0	4.39	0	0.14		52.7
Speed = SPDI (mph)	22.5	55	55	52.7		
Turning Lanes on Main Road — None Right Turn Bypass Lane Both					216 (55.5%) 119 (30.6%) 8 (2.1%) 46 (11.8%)	
Right Turn/Acceleration Lane on Minor Leg — Yes No					8 (2.1%) 381 (97.9%)	
Snow Percentage = SNP	22.9	36.9	28.4	29.1		
Non-dry Percentage = NONDRYP	41.0	55.6	46.7	47.5		

1 mi = 1.61 km, 1 ft = 0.3048 m

TABLE 5. Summary Statistics: 181 Three-Legged Intersections, Washington State Two-Lane Rural Roads, 1993 - 1995

Variable and Abbreviation	Min.	Max.	Median	Mean	Freq	%Zero
Total Number of Accidents (TOTACC)	0	9	0	1.02	184	58.6
Total Number of Injury Accidents (INJACC)	0	7	0	0.470	85 (46.20%)	72.4
Total Number of Run-Off-Road Accidents (RORACC)	0	5	0	0.204	37 (20.11%)	85.1
Severity: Fatal = K Injury = A Non-incap = B Poss-inj = C Prop-dam = P					2 (1.1%) 12 (6.5%) 37 (20.1%) 34 (18.5%) 99 (53.8%)	
Accident Rate (TOTACC per million entering vehicles)	0	1.00	0	0.135		58.6
Accident Rate (0.339 TOTACC/YEAR)	0	3	0	0.339		58.6
Injury Accident Rate (0.157 INJACC/YEAR)	0	2.333	0	0.157		72.4
Average Daily Traffic on Major Road = ADT1	897	15,995	4,838	5,780		
Average Daily Traffic on Minor Road = ADT2	4	7,529	196	573		
Conflict Index = CINDEX	0.0003	0.366	0.020	0.052		
Angular Deviation from 90° = DEV (degrees)	0	55	0	8.93		95.6
Roadside Hazard Rating = RHRI = 1 2 3 4 5 6, 7	1	6	3	3.3	7 (3.9%) 39 (21.5%) 51 (28.2%) 61 (33.7%) 21 (11.6%) 2 (1.1%), 0 (0.0%)	
Number of Driveways = ND	0	12	1	1.486		37.0

1 mi = 1.61 km, 1 ft = 0.3048 m

TABLE 5. Summary Statistics: 181 Three-Legged Intersections, Washington State (continued)

Two-Lane Rural Roads, 1993 - 1995

Variable and Abbreviation	Min.	Max.	Median	Mean	Freq	%Zero
Light Yes No					35 (19.3%) 146 (80.7%)	
Terrain Flat Rolling Mountainous					42 (23.2%) 124 (68.5%) 15 (8.3%)	
Degree of Curve = HI = (1/n)∑DEG{i}	0	22.2	0	1.22		68.0
Crest Curve Grade Rate = VCI = (1/m)∑(|Δg{i}|/l{i}) (crests only)	0	4.32	0	0.16		68.0
Speed = SPDI (mph)	23.75	55	55	52.1		
Turning Lanes on Main Road None Right Turn Bypass Lane Both					143 (79.0%) 12 (6.6%) 12 (6.6%) 14 (7.7%)	
Right Turn/Acceleration Lane on Minor Leg Yes No					4 (2.2%) 177 (97.8%)	

1 mi = 1.61 km, 1 ft = 0.3048 m

TABLE 6. Summary Statistics: 327 Four-Legged Intersections, Minnesota State Two-Lane Rural Roads, 1985 - 1989

Variable and Abbreviation	Min.	Max.	Median	Mean	Freq	%Zero
Total Number of Accidents (TOTACC)	0	16	1	1.51	494	39.8
Total Number of Injury Accidents (INJACC)	0	9	0	0.77	253 (51.2%)	59.9
Total Number of Run-Off-Road Accidents (RORACC)	0	2	0	0.092	30 (6.1%)	92.4
Severity: Fatal = K Injury = A Non-incap = B Poss-inj = C Prop-dam = P					18 (3.6%) 40 (8.1%) 96 (19.4%) 99 (20.0%) 241 (48.8%)	
Accident Rate (TOTACC per million entering vehicles)	0	2.87	0.201	0.323		39.8
Accident Rate (0.302 TOTACC/YEAR)	0	3.2	0.2	0.302		39.8
Injury Accident Rate (0.155 INJACC/YEAR)	0	1.8	0	0.155		59.9
Average Daily Traffic on Major Road = ADT1	174	14,611	2,620	2,238		
Average Daily Traffic on Minor Road = ADT2	6.9	3,414	192	308		
Conflict Index = CINDEX	0.003	0.637	0.103	0.142		
Angular Deviation from 90° = DEV (degrees)	0	75	0.6	9.9		37.6
Roadside Hazard Rating = RHRI = 1 2 3 4 5 6, 7	1	6	2	2.02	97 (29.7%) 157 (48.0%) 51 (15.6%) 16 (4.9%) 5 (1.5%) 1 (0.3%), 0 (0.0%)	
Number of Driveways = ND	0	6	0	0.62		67.6

1 mi = 1.61 km, 1 ft = 0.3048 m

TABLE 6. Summary Statistics: 327 Four-Legged Intersections, Minnesota State (continued)
Two-Lane Rural Roads, 1985 - 1989

Variable and Abbreviation		Min.	Max.	Median	Mean	Freq	%Zero
Light	Yes No					1 (0.3%) 326 (99.7%)	
Terrain (not noted)	Flat Rolling Missing					166 (50.8%) 20 (6.1%) 141 (43.1%)	
Degree of Curve = HI = (1/n)∑DEG{i}		0	9	0	0.49		59.9
Crest Curve Grade Rate = VCI = (1/m)∑(|Δg{i}|/l{i}) (crests only)		0	2.94	0.025	0.152		48.0
Speed = SPDI (mph)		30	55	55	54.0		
Right Turn Lanes on Main Road Turn Turns	None One Right Two Right					155 (47.4%) 30 (9.2%) 142 (43.4%)	
Right Turn/Acceleration Lanes on Minor Legs None Both						326 (99.7%) 1 (0.3%)	

1 mi = 1.61 km, 1 ft = 0.3048 m

TABLE 7. Summary Statistics: 90 Four-Legged Intersections, Washington State Two-Lane Rural Roads, 1993 - 1995

Variable and Abbreviation	Min.	Max.	Median	Mean	Freq	%Zero
Total Number of Accidents (TOTACC)	0	18	1	2.83	255	42.2
Total Number of Injury Accidents (INJACC)	0	13	0	1.77	159 (62.35%)	53.3
Total Number of Run-Off-Road Accidents (RORACC)	0	2	0	0.24	22 (8.63%)	78.9
Severity: Fatal = K Injury = A Non-incap = B Poss-inj = C Prop-dam = P					5 (2.0%) 20 (7.8%) 72 (28.2%) 62 (24.3%) 96 (37.6%)	
Accident Rate (TOTACC per million entering vehicles)	0	1.73	0.131	0.328		42.2
Accident Rate (0.944 TOTACC/YEAR)	0	6	0.333	0.944		42.2
Injury Accident Rate (0.589 INJACC/YEAR)	0	4.33	0	0.589		78.9
Average Daily Traffic on Major Road = ADT1	1,143	17,205	6,540	7,381		
Average Daily Traffic on Minor Road = ADT2	6	3,165	416	718		
Conflict Index = CINDEX	0.001	0.480	0.0646	0.0934		
Angular Deviation from 90° = DEV (degrees)	0	45	0	2.47		88.9
Roadside Hazard Rating = RHRI = 1 2 3 4 5 6, 7	1	5	3	2.82	9 (10.0%) 23 (25.6%) 38 (42.2%) 15 (16.7%) 5 (5.6%) 0 (0.0%), 0 (0.0%)	
Number of Driveways = ND	0	7	0	1.11		53.3

1 mi = 1.61 km, 1 ft = 0.3048 m

TABLE 7. Summary Statistics: 90 Four-Legged Intersections, Washington State (continued)

Two-Lane Rural Roads, 1993 - 1995

Variable and Abbreviation	Min.	Max.	Median	Mean	Freq	%Zero
Light Yes No					33 (36.7%) 57 (63.3%)	
Terrain Flat Rolling Mountainous					28 (31.1%) 60 (66.7%) 2 (2.2%)	
Degree of Curve = HI = (1/n)∑DEG{i}	0	6.50	0	0.497		78.9
Crest Curve Grade Rate = VCI = (1/m)∑(|Δg{i}|/l{i}) (crests only)	0	3.585	0	0.185		66.7
Speed = SPDI (mph)	22.5	55	55	51.0		
Right Turn Lanes on Major Road None One Right Turn Two Right Turns					53 (58.9%) 8 (8.9%) 29 (32.2%)	
Right Turn/Acceleration Lanes on Minor Legs None Both					89 (98.9%) 1 (1.1%)	

1 mi = 1.61 km, 1 ft = 0.3048 m

BIVARIATE STATISTICS

In this section tables are exhibited that indicate the correlation coefficient between accident count and one other highway variable. A positive coefficient indicates that as the highway variable increases accident counts do also; a negative coefficient indicates that as one variable increases the other tends to decrease. When a relationship is pronounced significant in this discussion, it means that the P-value is small (say, under 15%, and usually under 5%). The P-value is the probability that the sample correlation would have the given magnitude or greater when the true correlation in the population is zero. Thus significant relationships are ones that provide strong evidence that the two variables are correlated on the population from which the sample comes.

A major limitation of bivariate statistics is that the relationship between one variable and another may be masked or appear in a misleading light when a few especially influential variables such as ADT are present and their effect is ignored. The effect of a geometric variable, for example, on accidents when ADT is held constant is best revealed by the modeling to be discussed later since the modeling attempts to assess the combined contributions of all variables. With this caveat, bivariate statistics for accidents versus other variables are presented in Tables 8, 9, and 10. In Tables 11, 12, and 13 some of the significant correlations of highway variables with one another are also shown (in qualitative form rather than quantitative).

Segment Accidents

The most pronounced correlations with accidents, applicable in both Minnesota and Washington, are as follows:

positive correlation	negative correlation
EXPO	T
ADT	
SEG_LGN	
RHR	
GR	

Horizontal and vertical alignment also correlate positively with accidents but are not consistently significant. Some variables yield opposite signs from one State to the other, notably, lane and shoulder width, each of which is negatively correlated with accidents in Minnesota and positively in Washington. The consistent negative correlation of truck percentage suggests that trucks avoid the most dangerous roads. The weather variables in Minnesota are not significant.

If the accidents are restricted to serious accidents or run-off-road accidents, the same relationships persist with slight changes. The negative correlation of truck percentage is less significant. On the other hand, for run-off-road accidents both horizontal alignment H and grade GR are more significant.

Three-legged Intersection Accidents

Accidents at three-legged intersections show the following relationships:

positive correlation

ADT1
ADT2
RT

Horizontal and vertical alignment or driveways nearby generally contribute positively to accident counts but not in a consistently significant manner. Turning lanes are often installed at intersections with high turning volumes and high accident counts, but it is not clear why a right turn lane on the mainline would correlate positively with accidents while the conflict index would show much less significance (in Minnesota). Bad weather is marginally significant at Minnesota three-leggeds.

Serious accidents and run-off-road accidents show the same pattern although major road ADT is not significant for run-off-road accidents.

Four-legged Intersection Accidents

The significant correlations in this case are:

Positive correlation

ADT1
ADT2
CINDEX

The Minnesota data, but not the Washington data, show expected dependencies on channelization, alignment, Roadside Hazard Rating, number of driveways, as well as (weak) positive dependence on bad weather.

Serious and run-off-road accidents behave likewise, but major road ADT is not significant for run-off-road accidents.

TABLE 8. Bivariate Statistics: Segment Accidents versus Other Variables

	Estimated Correlation Coefficients and P-values					
	619 Minnesota Segments, 1985-1989			712 Washington Segments, 1993-1995		
	TOTACC	INJACC	RORACC	TOTACC	INJACC	RORACC
EXPO	0.76745 0.0001	0.72472 0.0001	0.60551 0.0001	0.70743 0.0001	0.64720 0.0001	0.54719 0.0001
ADT	0.27530 0.0001	0.22595 0.0001	0.11532 0.0041	0.37440 0.0001	0.36273 0.0001	0.21071 0.0001
T Truck %	-0.09217 0.0218	-0.06136 0.1273	-0.05460 0.1749	-0.05653 0.1319	-0.05478 0.1442	-0.00644 0.8638
SEG_LGN	0.44942 0.0001	0.45724 0.0001	0.48321 0.0001	0.30915 0.0001	0.27959 0.0001	0.32665 0.0001
LW Lane width	-0.08012 0.0463	-0.09321 0.0204	-0.12262 0.0022	0.02502 0.5051	0.00689 0.8545	0.02252 0.5485
SHW Shoulder width	-0.10800 0.0072	-0.10056 0.0123	-0.18391 0.0001	0.04740 0.2065	0.02646 0.4808	0.00250 0.9469
TOTWIDTH	-0.12749 0.015	-0.12374 0.0020	-0.21336 0.0001	0.05064 0.1771	0.02668 0.4772	0.00744 0.8429
RHR Hazrat	0.20682 0.0001	0.16669 0.0001	0.21610 0.0001	0.14740 0.0001	0.11561 0.0020	0.17778 0.0001
DD Drwyrate	-0.04493 0.2643	-0.04191 0.2978	-0.08898 0.0268	0.04818 0.1991	0.04837 0.1974	-0.01197 0.7499
INTD Intrate	-0.10648 0.0080	-0.11133 0.0056	-0.12677 0.0016	-0.02564 0.4945	0.00355 0.9247	-0.00229 0.9514
H Hor	0.04330 0.2821	0.04837 0.2294	0.10057 0.0123	0.09732 0.0094	0.06497 0.0832	0.14451 0.0001
HM1 Adj. Hor	0.02686 0.5048	0.03130 0.4369	0.08140 0.0429	0.07953 0.0339	0.05023 0.1807	0.12619 0.0007
HM1.5 Adj. Hor to 1.5 power	0.04812 0.2319	0.05584 0.1653	0.11283 0.0049	0.06542 0.0811	0.03098 0.4092	0.09895 0.0082
HM2 Adj. Hor to 2nd power	0.06678 0.0969	0.08138 0.0430	0.13349 0.0009	0.04389 0.2422	0.01040 0.7817	0.06730 0.0727

TABLE 8. **Bivariate Statistics: Segment Accidents versus Other Variables (continued)**

	Estimated Correlation Coefficients and P-values					
	619 Minnesota Segments, 1985-1989			712 Washington Segments, 1993-1995		
	TOTACC	INJACC	RORACC	TOTACC	INJACC	RORACC
VC Crests	0.15054 0.0002	0.11163 0.0054	0.19049 0.0001	0.00865 0.8178	-0.00585 0.8762	0.03563 0.3425
VM Adj. Vert	0.17305 0.0001	0.14311 0.0004	0.25085 0.0001	0.04373 0.2438	0.03569 0.3417	0.04593 0.2209
VMC Adj. Crests	0.16106 0.0001	0.13484 0.0008	0.23772 0.0001	0.04929 0.1890	0.03467 0.3556	0.06381 0.0889
VMCC Adj. Crests of Type I	0.12476 0.0019	0.10869 0.0068	0.19460 0.0001	0.01097 0.7702	-0.00574 0.8784	0.04101 0.2745
GR Abs. Grade	0.09618 0.0167	0.04945 0.2193	0.12483 0.0019	0.07741 0.0389	0.04929 0.1889	0.11074 0.0031
SPD Speed	0.07167 0.0748	0.06674 0.0971	0.04099 0.3086	-0.03082 0.4116	-0.02020 0.5906	-0.03805 0.3106
SNP Snow %	-0.01842 0.6474	0.02900 0.4714	0.01945 0.6291		not collected	
NONDRYP Nondry %	0.00181 0.9642	0.04549 0.2584	0.04124 0.3057		not collected	

TABLE 9. Bivariate Statistics: 3-Legged Intersection Accidents versus Other Variables

	Estimated Correlation Coefficients and P-values					
	389 MN Intersections, 1985-1989			181 WA Intersections, 1993-1995		
	TOTACC	INJACC	RORACC	TOTACC	INJACC	RORACC
ADT1	0.52037 0.0001	0.48556 0.0001	0.22233 0.0001	0.27304 0.0002	0.28700 0.0001	0.03473 0.6425
ADT2	0.40714 0.0001	0.36869 0.0001	0.19482 0.0001	0.41803 0.0001	0.28528 0.0001	0.22967 0.0019
CINDEX	0.00491 0.9232	-0.02749 0.5888	0.10347 0.0414	0.27266 0.0002	0.19051 0.0102	0.19287 0.0093
DEV from 90°	0.07855 0.1219	0.02930 0.5645	0.04963 0.3289	-0.05725 0.4440	-0.06922 0.3545	-0.07399 0.3222
DEV15 Adj dev from 90°±15°	0.06977 0.1696	0.02944 0.5627	0.01872 0.7128	-0.04358 0.5602	-0.05515 0.4609	-0.06044 0.4190
HAU Signed dev	0.11514 0.0231	0.10063 0.0473	0.15847 0.0017	-0.02998 0.6887	-0.00162 0.9828	0.01919 0.7976
RHRI Hazrat	0.13436 0.0080	0.13260 0.0088	0.01454 0.7750	0.05133 0.4926	0.00348 0.9629	0.07482 0.3168
ND No. Drwy ±250 ft	0.02207 0.6643	0.03361 0.5087	0.01447 0.7760	0.10166 0.1733	0.09845 0.1873	-0.07505 0.3153
HI Hor to ±250 ft	0.07944 0.1178	0.05850 0.2497	0.15707 0.0019	0.05174 0.4891	0.01442 0.8473	0.14196 0.0566
HEI Hor to ±764 ft	0.09646 0.0573	0.09223 0.0692	0.19928 0.0001	0.05586 0.4551	0.04412 0.5554	0.10672 0.1527
VCI Crests to ±250 ft	0.03295 0.5171	0.02387 0.6388	-0.02423 0.6338	-0.00235 0.9749	0.01825 0.8073	-0.00689 0.9266
VI Vert to ±250 ft	0.09520 0.0607	0.04164 0.4128	0.07861 0.1217	0.08123 0.2770	0.06945 0.3529	0.06602 0.3772
VEI Vert to ±764 ft	0.02400 0.6369	-0.01233 0.8084	0.03026 0.5519	0.12292 0.0992	0.07211 0.3347	0.05967 0.4249
SPDI Speed	-0.07340 0.1485	-0.07720 0.1285	-0.039 0.4320	-0.09219 0.2171	-0.07893 0.2909	0.02244 0.7643
RT Right turn lane on major road	0.23441 0.0001	0.22678 0.0001	0.10163 0.0451	0.21566 0.0035	0.15009 0.0437	0.23483 0.0015
SNP Snow %	0.08129 0.1094	0.08143 0.1088	-0.02859 0.5740		not collected	
NONDRYP Nondry %	0.08868 0.0806	0.08400 0.0981	-0.01188 0.8154		not collected	

1 mile = 1.61 km, 1 ft = .3048 m

TABLE 10. Bivariate Statistics: 4-Legged Intersection Accidents versus Other Variables

	Estimated Correlation Coefficients and P-values					
	327 Minnesota Intersections, 1985-1989			90 Washington Intersections, 1993-1995		
	TOTACC	INJACC	RORACC	TOTACC	INJACC	RORACC
ADT1	0.54437	0.49973	0.40133	0.25595	0.24971	-0.01675
	0.0001	0.0001	0.0001	0.0149	0.0176	0.8755
ADT2	0.61418	0.57735	0.27390	0.39618	0.33811	0.24331
	0.0001	0.0001	0.0001	0.0001	0.0011	0.0208
CINDEX	0.13832	0.13583	-0.03482	0.24574	0.19834	0.30749
	0.0123	0.0140	0.5303	0.0196	0.0609	0.0032
DEV	-0.04303	-0.04538	-0.06918	0.06761	0.03049	-0.09349
from 90°	0.4380	0.4135	0.2122	0.5266	0.7754	0.3808
DEV15	-0.10460	-0.10775	-0.07555	-0.00113	-0.00299	-0.03156
Adj dev from 90°±15°	0.0588	0.0516	0.1729	0.9916	0.9777	0.7678
HAU	-0.06632	-0.04573	-0.03414	0.09522	0.03804	0.03457
Signed dev	0.2317	0.4099	0.5384	0.3720	0.7219	0.7464
RHRI	0.10842	0.05967	0.13430	-0.16309	-0.16003	0.02006
Hazrat	0.0501	0.2820	0.0151	0.1246	0.1319	0.8511
ND	0.18270	0.14527	0.11849	0.03186	0.07011	-0.06316
No. Drwy ±250 ft	0.0009	0.0085	0.0322	0.7656	0.5114	0.5543
HI	0.16615	0.19496	0.12018	-0.20082	-0.20821	-0.10994
Hor to ±250 ft	0.0026	0.0004	0.0298	0.0577	0.0489	0.3023
HEI	0.17134	0.19274	0.11271	-0.15453	-0.14205	-0.06438
Hor to ±764 ft	0.0019	0.0005	0.0417	0.1459	0.1817	0.5466
VCI	0.12097	0.09643	0.02668	0.02163	0.05819	-0.11022
Crests to ±250 ft	0.0287	0.0816	0.6307	0.8397	0.5859	0.3011
VI	0.07644	0.03342	0.06234	-0.07992	-0.05111	-0.08611
Vert to ±250 ft	0.1679	0.5470	0.2610	0.4540	0.6324	0.4197
VEI	0.04494	0.00352	0.07422	-0.08297	-0.06621	-0.10088
Vert to ±764 ft	0.4180	0.9495	0.1806	0.4369	0.5353	0.3441
SPDI	-0.09505	-0.11989	-0.02332	0.09481	0.09333	0.22014
Speed	0.0861	0.0302	0.6744	0.3741	0.3816	0.0371
RT	0.21059	0.21229	0.07658	0.11450	0.09124	0.13312
Right turn lanes on major road	0.0001	0.0001	0.1671	0.2826	0.3924	0.2110
SNP	0.06533	0.08155	0.02198			
Snow %	0.2387	0.1412	0.6922		not collected	
NONDRYP	0.06827	0.08149	0.03825			
Nondry %	0.2182	0.1415	0.4906		not collected	

1 mile = 1.61 km, 1 ft = .3048 m

TABLE 11. Correlations between Segment Variables in MN and WA Samples

VARIABLE	POSITIVE CORRELATES	NEGATIVE CORRELATES
ADT	SHW, TOTWIDTH	T, SEG_LGN
T Truck %	SEG_LGN, SHW, TOTWIDTH, SPD	ADT, SNP, NONDRYP
SEG_LGN	T, RHR, SPD, SNP, NONDRYP	ADT, DD, INTD, SHW
LW Lane width	SPD	
SHW Shoulder width	ADT, T, SPD	RHR, H, VC, GR
TOTWIDTH	ADT, T, SPD	RHR, H, VC, GR
RHR Roadside Hazard Rating	SEG_LGN, H, VC, GR, SNP, NONDRYP	SHW, TOTWIDTH, SPD
DD Drwyrate	INTD	T, SEG_LGN, SPD
INTD Intrate	DD	SEG_LGN
H Hor	RHR, VC, GR	SEG_LGN, SHW, TOTWIDTH, SPD
VC Crests	RHR, H, GR	T, TOTWIDTH, SPD
GR Absolute grade	RHR, H, VC	SHW, TOTWIDTH, SPD
SPD Speed	T, SEG_LGN, LW, SHW, TOTWIDTH	RHR, DD, H, VC, GR
SNP, NONDRYP (MN only)	SEG_LGN, RHR, H	T

NOTE: Segment length (SEG_LGN), Roadside Hazard Rating (RHR), Speed (SPD), and Truck Percentage (T) show strong correlation with a large number of variables. Segment lengths tend to be longer in rural areas and this accounts for the negative correlation with ADT, driveway density, and intersection density. The Roadside Hazard Rating and Speed variables also show expected correlates. The behavior of the Truck Percentage variable suggests that teamsters favor routes with certain characteristics and/or that such routes are more likely to have commercial development nearby.

TABLE 12. Correlations between 3-Legged Intersection Variables in MN and WA Samples

VARIABLE	POSITIVE CORRELATES	NEGATIVE CORRELATES
ADT1	ADT2, ND, SNP, NONDRYP	CINDEX, SPDI
ADT2	ADT1, CINDEX, ND, HI, RT	SPDI
CINDEX	ADT2, HI	ADT1, SPDI, SNP, NONDRYP
DEV from 90°	RHRI	
RHRI Roadside Hazard Rating	DEV, HI, VI	
ND No. of Drwys ± 250 ft	ADT1, ADT2, HEI, SNP, NONDRYP	SPDI
HI Hor. to ± 250 ft	ADT2, CINDEX, RHRI, VCI, VI, VEI	SPDI
HEI Hor. to ± 764 ft	ADT2, CINDEX, RHRI, ND, VI, VEI	SPDI
VCI Crests to ± 250 ft	HI, VI, VEI	SPDI
VI Vert. to ± 250 ft	RHRI, HI, HEI, VCI, VEI	SPDI
VEI Vert. to ± 764 ft	RHRI, HI, HEI, VCI, VI	SPDI
SPDI Speed		ADT1, ADT2, CINDEX, ND, HI, HEI, VCI, VI, VEI
RT Right Turn Lane on Major Road	ADT2	
SNP, NONDRYP (MN only)	ADT1, ND	CINDEX

1 mile = 1.61 km, 1 ft = .3048 m

NOTE: Perhaps the fact of chief interest in Table 12 (the 3-legged intersections) is the negative correlation between posted speed and the other variables of interest. In Table 13 (the 4-legged intersections) speed plays a similar role but not quite so marked.

TABLE 13. Correlations between 4-Legged Intersection Variables in MN and WA Samples

VARIABLE	POSITIVE CORRELATES	NEGATIVE CORRELATES
ADT1	ADT2, DEV, HI, SNP, NONDRYP	CINDEX
ADT2	ADT1, CINDEX, RT	
CINDEX	ADT2, RT	ADT1
DEV from 90°	ADT1	SPDI
RHRI Roadside Hazard Rating	VI, VEI	
ND No. of Drwys ± 250 ft	SNP, NONDRYP	SPDI
HI Hor. to ± 250 ft	HEI, SNP, NONDRYP	
HEI Hor. to ± 764 ft	HI, RT, SNP, NONDRYP	SPDI
VCI Crests to ± 250 ft	VI, VEI	SPDI
VI Vert. to ± 250 ft	RHRI, VCI, VEI	
VEI Vert. to ± 764 ft	RHRI, VCI, VI	SNP, NONDRYP
SPDI Speed		DEV, ND
RT Right Turn Lanes on Major Road	ADT2, CINDEX, HEI	
SNP, NONDRYP (MN only)	ADT1, DEV, ND, HI. HEI	VEI, SPDI

1 mile = 1.61 km, 1 ft = .3048 m

Other Bivariate Relationships

Bivariate relationships between highway variables are also in evidence as might be expected. In Tables 11, 12, and 13 above we indicate relationships in which the correlation coefficient has the same sign in both Minnesota and Washington and the correlation is strongly significant in both States (P-value typically less than 5%) or strongly significant in one State and moderately significant in the other (P-value typically less than 15%). We omit obvious correlations (e.g., between different vertical measures).

In the case of weather variables (SNP and NONDRYP) the correlation is for Minnesota data, the only State where weather data were collected. The weather variables show some surprising correlations in the intersection samples. See Table 14 below. These correlations have no counter-

TABLE 14. Correlations between Weather and Minnesota Highway Variables

Correlation coefficient and P-value	Minnesota 3-legged intersection sample		Minnesota 4-legged intersection sample	
	ADT1	ND	ADT1	ND
NONDRYP	.21201, .0001	.12608, .0128	.12202, .0274	.21916, .0001
SNP	.19164, .0001	.13523, .0076	.09611, .0827	.21555, .0001

parts in the segment data. The direct implication, however frivolous it may be, is that rural intersections with high major road ADT or with nearby driveways tend to have more rain and snow than other rural intersections. The correlation of weather with minor road ADT is not significant.

SUMMARY

A wide variety of variables have been introduced in this chapter to facilitate the modeling in the next.

The summary univariate statistics for these variables (Tables 2 through 7) indicate that most of them show a good range of values that will provide variation for the modeling. Exceptions are: lighting along the segments (the vast majority have none), right turn/acceleration lanes on the minor legs of intersections (most have none), and intersection angle deviation from 90° on Washington State intersections. Most Washington intersection angles are 90°, perhaps in part because photolog estimates had to be used in Washington State and are much cruder than those obtained from Minnesota plans.

Bivariate statistics indicate that commercial traffic on two-lane segments correlates negatively with accidents while surface width and lane width have unexpected effects in Washington State. Traffic is the dominant variable for intersections, but the existence of a right turn lane on the major road correlates positively with accidents on three-legged intersections.

Bivariate relationships between accident variables and highway variables should be interpreted with caution: they may indicate that the highway variable correlates with a another influential highway variable. Modeling with several variables simultaneously may permit greater insight into the relative effects of different highway variables.

5. MODELING

In this chapter the modeling effort is described. The chapter begins with a discussion of Poisson and negative binomial modeling and goodness-of-fit measures. Then models are developed for the Minnesota and Washington segments and the behavior of the variables is examined. We pass then to an extended negative binomial model developed by Shaw-Pin Miaou that attempts to capture the effect of variation along a roadway. In our case this can be applied to horizontal curvatures, vertical curves, and straightaway grades along the segments. The extended negative binomial methodology is applied to the Minnesota segments, to the Washington segments, and then jointly to the Combined segments with a variable for the State. Thereafter Poisson and negative binomial models are developed for the four intersection data sets and for the combined intersection data sets. Most of the models attempt to represent the mean total number of accidents (TOTACC), but we also include a few models of serious accidents (INJACC) as well. Finally logistic regression models for accident severity are developed and evaluated.

POISSON AND NEGATIVE BINOMIAL MODELING TECHNIQUES

The Poisson and Negative Binomial Models

Poisson and negative binomial models, with parameters a generalized linear function of covariates, are by now a well-accepted method of modeling discrete rare events such as roadway accidents. See Miaou and Lum (1993).[38] It is assumed that accidents occurring on a particular roadway or at a particular intersection are independent of one another and that a certain mean number of accidents per unit time is characteristic of the given site and of other sites with the same properties. The mean itself is assumed to depend on highway variables. Since the mean must be greater than zero, it is taken to have a generalized linear form given by:

$$\mu_i = \exp(\beta_0 + \sum_{j=1}^{n} x_{ij}\beta_j) \tag{5.1}$$

where μ_i is the mean number of accidents to be expected at site number i in a given time period, x_{i1}, x_{i2}, ..., x_{in}, are the values of the highway variables at site number i during that time period, and β_0, β_1, ..., β_n are coefficients to be estimated by the modeling.

[38] Miaou, S-P., and Lum, H., "Modeling Vehicle Accidents and Highway Geometric Design Relationships," Accident Analysis and Prevention, 25(6): 689-709, 1993.

Preceding page blank

In the Poisson distribution the variance in the number of accidents at a site is equal to the mean μ_i The Poisson model takes the form:

$$P(y_i) = \frac{\exp(-\mu_i)(\mu_i)^{y_i}}{y_i!}$$

where $P(y_i)$ is the probability of y_i accidents at the given site. The negative binomial distribution adds a quadratic term to the variance representing overdispersion. The negative binomial model takes the form:

$$P(y_i) = \frac{\Gamma(y_i+\frac{1}{K})}{y_i!\ \Gamma(\frac{1}{K})} \left(\frac{K\mu_i}{1+K\mu_i}\right)^{y_i} \left(\frac{1}{1+K\mu_i}\right)^{\frac{1}{K}}$$

where K is the overdispersion parameter and the variance is:

$$\mu_i + K(\mu_i)^2.$$

As pointed out by Dean and Lawless (1989)[39] the negative binomial allows for extra-Poisson variation due to other variables not included in the model. Hauer et al. (1988) propose that μ_i is to be regarded as the grand mean of a family of sites with the same highway variables x_{ij}, each site having Poisson-distributed accidents. If K equals 0, the negative binomial reduces to the Poisson model. The larger the value of K the more variability there is in the data over and above that associated with the mean μ_i .

The coefficients β_j are estimated by maximizing the log-likelihood function $L(\beta)$ for the Poisson distribution:

$$L(\beta) = \sum_i (y_i \log \mu_i - \mu_i - \log y_i!). \tag{5.2}$$

Here $\beta = (\beta_0, \beta_1, ..., \beta_n)$ is the vector of coefficients, y_i is the observed accident count for segment

---

[39] Dean, C., and Lawless, J.F., "Tests for Detecting Overdispersion in Poisson Regression Models," Journal of the American Statistical Association, 84 (406): 467-472, 1989.

or intersection no. i, and μ_i is given by (5.1). The value of β that maximizes (5.2) is the estimated coefficient vector $\hat{\beta}$. The value of μ_i that it yields, denoted by $\hat{y}_i$, is the estimated mean accident count.

For the negative binomial distribution the estimated coefficient vector and $\hat{y}_i$, along with an estimate $\hat{K}$ for K, are obtained by maximizing $L(\beta, K)$:

$$L(\beta,K) =$$
$$\sum_i [(\sum_{j=0}^{y_i} \log(1+Kj)) - \log(1+Ky_i) + y_i \log\mu_i - (y_i+\frac{1}{K})\log(1+K\mu_i) - \log(y_i!)]. \tag{5.3}$$

For convenience the same letters will often be used for both the parameters and their estimated values, i.e., hats ^ will be omitted.

Model Evaluation - Overdispersion

A decision about whether the Poisson form is appropriate can be based on one of several statistics. As noted in SAS Technical Report P-243[40] the deviance of a model m is:

$$D^m = 2\ (L^f - L^m)$$

where L^f is the log-likelihood (5.2) that would be achieved if the model gave a perfect fit ($\mu_i = y_i$ for each i, and K = 0) and L^m is the log-likelihood (5.2 or 5.3) of the model under consideration ($\mu_i = \hat{y}_i$). If the latter model is correct, D^m is approximately a chi-squared random variable with degrees of freedom equal to the number n of observations minus the number p of parameters.

A value of the deviance greatly in excess of n - p suggests that the model is overdispersed due to missing variables and/or non-Poisson form. Thus when deviance divided by degrees of freedom

$$\frac{D^m}{n-p}$$

[40] SAS Technical Report P243, SAS/STAT Software: The GENMOD Procedure, Release 6.09, SAS Institute Inc., Cary, North Carolina, 1993.

is significantly larger than 1, overdispersion is indicated.

Likewise, the Pearson chi-square statistic, defined by

$$\chi^2 = \sum_{i=1}^{n} \frac{(y_i - \hat{y}_i)^2}{\hat{y}_i} ,$$

is an approximately chi-squared random variable with mean n - p for a valid Poisson model. If

$$\frac{\chi^2}{n - p}$$

is significantly larger than 1, overdispersion is also indicated.

On the assumption that the basic form of the model is correct, Dean and Lawless (1989) recommend yet another statistic T_1 to test the hypothesis that the model is a Poisson model against the alternative that it is overdispersed. When the null hypothesis K = 0 is true and the number of observations is large, the statistic

$$T_1 = \frac{\sum_i \left((y_i - \hat{y}_i)^2 - y_i\right)}{\sqrt{2 \cdot \sum_i (\hat{y}_i)^2}}$$

is approximately a standard normal random variable. If T_1 is large positive, the hypothesis K = 0 is rejected, the data are considered to be overdispersed, and a negative binomial model with K positive is an alternative candidate model.

Model Evaluation - Goodness of Fit

In addition to a plausible basis for the underlying distributional assumptions, three important tests for an acceptable model are the following:

- The estimated regression coefficient for each covariate should be statistically significant, i.e., one should be able to reject the null hypothesis that the coefficient is zero;

- Engineering and intuitive judgments should be able to confirm the validity and practicality of the sign and rough magnitude of each estimated coefficient; and

- Goodness-of-fit measures and statistics, such as R-squared (the coefficient of determination), the deviance, and the Pearson chi-square, should indicate that the variables do have explanatory and predictive power.

The modeling of the data in this study was done using SAS and LIMDEP software. Along with approximate maximum likelihood estimates for the regression coefficients, these software packages yield estimates of the standard error for each coefficient. From these, P-values can be computed for the null hypothesis that the true value of some regression coefficient is zero. The z-score of the estimated coefficient is the estimated coefficient minus zero, divided by the estimated standard error. The P-value is the probability that a normal random variable has an absolute value larger than the z-score obtained. If the P-value is small, we have good evidence that the corresponding variable is significant, that the difference between the coefficient estimate and zero arises not from chance but from a systematic effect.

Goodness-of-fit measures associated with Poisson-type models have been introduced and reviewed by Fridstrøm et al. (1995)[41] and Miaou (1996).[42]

The R-squared goodness-of-fit measures, used to estimate the percentage of variation explained by a regression model, are somewhat controversial. Different R-squared measures may yield substantially different answers, or even answers larger than 1, particularly for models that are not linear. See the article of Kvalseth (1985).[43] Until recently, R-squared measures appropriate for Poisson or negative binomial models had not been established. Fridstrøm et al. (1995) developed several alternative goodness-of-fit methodologies for generalized Poisson regression models. Four of these approaches are used here to evaluate goodness-of-fit.

The first approach is based on the ordinary R-squared, or coefficient of determination, used in linear regression models:

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \tag{5.4}$$

---

[41] Fridstrøm, L., Ifver, J., Ingebrigsten, S., Kulmala R., and Thomsen L.K., "Measuring the Contribution of Randomness, Exposure, Weather, and Daylight to the Variation in the Road Accident Counts," Accident Analysis and Prevention, 27(1): 1-20, 1995.

[42] Miaou, S-P., "Measuring the Goodness-of-Fit of Accident Prediction Models," Federal Highway Administration, Report No. FHWA-RD-96-040, Washington, D.C., 1996.

[43] Kvalseth, T.O., "Cautionary Note About R^2," The American Statistician, Amer. Stat. Assoc., 39(4): 279-285, 1985.

where

y_i = observed accident count for highway segment or intersection no. i

$\bar{y}$ = average accident count for the sample

$\hat{y}_i$ = estimated mean accident count for observation no. i

The numerator in the second term (of 5.4) is the variation not explained by the model. In a perfectly specified and estimated Poisson model (variance equal to mean), the most that can be explained of the given data is expected to be P^2, where

$$P^2 = 1 - \frac{\sum_i \hat{y}_i}{\sum_i (y_i - \bar{y})^2} . \tag{5.5}$$

The numerator in the second term (of 5.5) is unexplainable Poisson variation, random variation to be expected when independent events of mean frequency $\hat{y}_i$ occur. Thus, the scaled R-squared R_P^2 is the proportion of potentially explainable systematic variation that can be explained from the causal factors considered.

$$R_P^2 = \frac{R^2}{P^2} \tag{5.6}$$

Two additional approaches of Fridstrøm et al., the weighted R-squared and the Freeman-Tukey R-squared, are similar. The weighted R-squared is the same as the ordinary R-squared except that the sum-of-squares in both numerator and denominator is divided by the predicted mean $\hat{y}_i$. For the weighted R-squared the counterparts of (5.4), (5.5), and (5.6) are:

$$R_w^2 = 1 - \frac{\sum_i \frac{(y_i - \hat{y}_i)^2}{\hat{y}_i}}{\sum_i \frac{(y_i - \bar{y})^2}{\hat{y}_i}} \tag{5.7}$$

$$P_w^2 = 1 - \frac{n}{\sum_i \frac{(y_i - \bar{y})^2}{\hat{y}_i}} \tag{5.8}$$

$$R_{PW}^2 = \frac{R_w^2}{P_w^2} \tag{5.9}$$

where n is the total number of observations in the sample. If $\hat{y}_i$ is the true mean for observation number i and y_i is a Poisson variable, then $(y_i - \hat{y}_i)/\surd(\hat{y}_i)$ is a variable with mean zero and standard deviation 1. Note that the numerator in (5.7) is the Pearson chi-square statistic for a Poisson model.

The Freeman-Tukey R-squared transforms the variable y_i (assumed to be a Poisson variable with mean $\hat{y}_i$) to one that is approximately normal. The counterparts of (5.4), (5.5), and (5.6) are:

$$R_{FT}^2 = 1 - \frac{\sum_i \hat{e}_i^2}{\sum_i (f_i - \bar{f})^2} \tag{5.10}$$

$$P_{FT}^2 = 1 - \frac{n}{\sum_i (f_i - \bar{f})^2} \tag{5.11}$$

$$R_{PFT}^2 = \frac{R_{FT}^2}{P_{FT}^2} \tag{5.12}$$

where

$f_i = \sqrt{y_i} + \sqrt{y_i + 1}$ = Freeman-Tukey transform of y_i

$\bar{f}$ = sample mean of f_i

and $\hat{e}_i = \sqrt{y_i} + \sqrt{y_i + 1} - \sqrt{4\hat{y}_i + 1}$.

The variable $\hat{e}_i$ is approximately a standard normal random variable (at least for $\hat{y}_i$ larger than 1).

The three measures introduced so far are strongly oriented toward Poisson models. Indeed because they do not explicitly include an overdispersion parameter they seem inappropriate for negative binomial models. But a fourth approach is tailored to the negative binomial.

The fourth approach, the Log-Likelihood R-squared, is based on the deviance D^m of the model. Fridstrøm et al. propose the following measures:

$$R_D^2 = 1 - \left(\frac{\frac{D^m}{n-k-1}}{\frac{D^0}{n-2}}\right) \tag{5.13}$$

$$P_D^2 = 1 - \left(\frac{\frac{D_E^m}{n-k}}{\frac{D^0}{n-2}}\right) \tag{5.14}$$

$$R_{PD}^2 = \frac{R_D^2}{P_D^2} \tag{5.15}$$

Here D^0 is the deviance of a model with only two parameters, the constant term (intercept) and the overdispersion parameter; k is the number of parameters of the model m under consideration (not including the overdispersion parameter in the model); and D_E^m is the expected value of the deviance in the case when a Poisson model with the same means $\hat{y}_i$ as the model m is the correct one. Roughly speaking, R_D^2 indicates how much explanatory power results from adding the highway characteristics and R_{PD}^2 represents this as a fraction of the highest possible expected explanatory power of any model with the same means as m.

For negative binomial and Poisson models Fridstrøm et al. regard R_{PD}^2 and R_{PFT}^2 with favor. They express reservations about R_P^2 and R_{PW}^2 : the first of these, being unnormalized, will make observations with large predicted means more influential, while the second tends to exaggerate the estimation errors associated with small predicted means.

Yet another measure of goodness-of-fit, this one advocated by Miaou (1996), is based explicitly on the overdispersion parameter.

$$R_K^2 = 1 - \frac{K}{K_{max}} \tag{5.16}$$

Here K is the overdispersion parameter estimated in the model, and K_{max} is the overdispersion parameter estimated in the negative binomial model discussed above, namely, the model with only a constant term and an overdispersion parameter. Based on simulations Miaou concluded that this measure shows promise. It is simple to calculate, it yields a value between 0 and 1, it has the proportionate increase property (Miaou proposes as a criterion that independent variables of equal importance, when added to a model, increase the value of the measure by the same absolute amount regardless of the order in which they are added), and it is independent of the choice of intercept term in the model.

SEGMENT MODELS

In this section we develop models for segments. The models are of Poisson type, negative binomial type, and extended negative binomial type. We discuss the choice of variables and explain the steps that lead to the final models presented. The choice of variables to retain, and the form in which to use them, are to some extent arbitrary since not all possibilities can be examined and some are more or less equivalent. The decisions are guided by criteria of simplicity (use of variables that are easily understood), comprehensiveness (inclusion of as many types of variables as possible), and significance (coefficients that are significantly different from zero according to statistical tests in one or more models). Many models can be generated, and we present here only a selection of models that illustrate the main phenomena and/or show the significant interactions.

In general, we will exhibit a formula for the mean number of accidents on a segment as a generalized linear function of highway variables. This formula will show the estimated coefficient of each variable in the model. In addition, we show the estimated standard error of the coefficient estimate and its P-value. The P-value is the probability that the estimated coefficient would have the value shown or any value farther from zero when the true coefficient is zero. A P-value of less than 5% is usually considered ample confirmation that the true coefficient is non-zero and that the estimated coefficient is significant. Later on, for the intersection models, we will liberalize this criterion considerably.

The State Variable

The STATE variable (value 0 for Minnesota, 1 for Washington) is used on all models that combine the two States. In effect it allows the constant or intercept term in each State to be different while constraining other coefficients to be the same. Including such a variable is equivalent to acknowledging that the accident experience of two different States is likely to be different on segments with the same traffic volumes and same highway characteristics. The STATE variable represents the demographics and habits of a different population of drivers in a different region and perhaps at a different era. Law enforcement practices, driver ages, and life styles may be quite different. Although the extra degree of freedom makes it easier to develop a combined model, it is of some interest when the coefficient of the State variable is insignificant (as it is in a few of the models below).

The Exposure Variable

For the segment modeling it is natural to include both segment length (seg_lng) and ADT as explanatory variables, and to expect that the number of accidents will be roughly proportional to the product of these factors times the time in days (365 days per year times 5 years in Minnesota or 3 years in Washington). Poisson models in Minnesota (Table 15) support this rough proportionality. If total number of accidents is modeled as a function of segment length and ADT, we obtain the following:

TABLE 15. Minnesota Segments, Poisson Models with Exposure Variables

Mean No. of Accidents = 5×(365/10^3)×exp{-.3916 + 1.0150 LSEG + .9765 LADT}			
Estimated standard error of coefficient estimates	.0448	.0278	.0344
P-value	.0001	.0001	.0001
Mean No. of Accidents = EXPO×exp{-.3934 - .0040 AVGM}			
Estimated standard error of coefficients estimates	.0382	.0278	
P-value	.0001	.6474	

1 mile = 1.61 km

where LSEG is the log of the segment length and LADT is the log of AVGM (ADT in 1000's of vehicles per day). The Minnesota standard errors are consistent with the conclusion that the true coefficients of LSEG and LADT are 1. The second model shows the effect of using EXPO as an offset (i.e., as a multiplier) but retaining AVGM. The Minnesota data do not support the retention of AVGM.

Similar tables for Washington State and the Combined data sets (Tables 16 and 17) indicate that LSEG and LADT have coefficients near 1 but still significantly different from 1 since the estimated standard errors are small. Also, if EXPO is taken as an offset and AVGM is retained, the latter is found to be significant. Although other choices could be made, the decision was made to use EXPO as an offset and exclude segment length as a separate variable, with the expectation that additional effects apparently due to segment length can be represented by other highway variables. AVGM was retained in some runs, although, as will be seen, it was not significant in the final model.

TABLE 16. Washington Segments, Poisson Models with Exposure Variables

Mean No. of Accidents = 3×(365/10^3)×exp{.1606 + .9121 LSEG + .8918 LADT}			
Estimated standard error of coefficient estimates	.0462	.0310	.0299
P-value	.0001	.0001	.0001
Mean No. of Accidents = EXPO×exp{.1674 - .0269 AVGM}			
Estimated standard error of coefficient estimates	.0390	.0059	
P-value	.0001	.0001	

1 mile = 1.61 km

TABLE 17. Combined Segments, Poisson Models with Exposure Variables

Mean No. of Accidents = (5 or 3)×(365/10^3)×exp{-.3282 + .9685 LSEG + .9296 LADT + .4450 STATE}				
Estimated standard error of coefficient estimates	.0346	.0206	.0226	.0366
P-value	.0001	.0001	.0001	.0001
Mean No. of Accidents = EXPO×exp{ -.3405 - .0200 AVGM + .4719 STATE}				
Estimated standard error of coefficient estimates	.0291	.0049	.0357	
P-value	.0001	.0001	.0001	

1 mile = 1.61 km

Lane Width and Shoulder Width

Wider lanes and wider shoulders should lower accidents. If we add these two variables to the Poisson models (Table 18), some notable differences are found between Minnesota and Washington. The lane width variable is seen to be of unexpected sign and insignificant in the Washington data.

TABLE 18. Poisson Models of Segments with Lane and Surface Width

MINNESOTA					
Mean No. of Accidents = EXPO×exp{3.2115 + .0202AVGM - .2501LW - .1183SHW}					
Estimated standard error of coefficient estimates	.4172	.0089	.0354	.0104	
P-value	.0001	.0222	.0001	.0001	
WASHINGTON					
Mean No of Accidents. = EXPO×exp{-.0093 - .0157AVGM + .0461LW - .0759SHW}					
Estimated standard error of coefficient estimates	.5270	.0063	.0464	.0110	
P-value	.9860	.0123	.3201	.0001	
COMBINED					
Mean No.of Accidents = EXPO×exp{1.5393 - .0079AVGM - .1117LW - .0915SHW + .2850STATE}					
Estimated standard error of coefficient estimates	.3236	.0050	.0277	.0075	.0606
P-value	.0001	.1108	.0001	.0001	.0001

1 mile = 1.61 km, 1 ft = .3048 m

In the last chapter we had already noted anomalies in the correlation between accidents and lane or shoulder width in Washington. Several factors contribute to this situation. One of them is the direct correlation between lane width and shoulder width that occurs in the Washington State data but not the Minnesota data. The correlation coefficients are given by:

Lane Width LW versus Shoulder Width SHW	MINNESOTA SEGMENTS	WASHINGTON SEGMENTS	COMBINED SEGMENTS
Correlation coefficient	-.06313	.11127	.07047
P-value	.1166	.0029	.0101

The P-values are estimated probabilities that the correlation coefficient estimates would have the values shown or values farther from zero if there were zero correlation between the variables on the populations from which the data sets are samples. Minnesota lane widths and shoulder widths have a slight but not especially significant negative correlation, while Washington lane widths and shoulder widths have a significant positive correlation. This is also reflected when we consider univariate statistics for LW, SHW, and TOTWIDTH:

State	Variable	Min	Max	Median	Mean
MN	Lane Width LW	10	12	12	11.54
	Shoulder Width SHW	0	12	8	7.08
	TOTWIDTH	20	48	38	37.22
WA	Lane Width LW	9	12	11	11.37
	Shoulder Width SHW	0	10	5	5.01
	TOTWIDTH	18	44	32	32.77

1 ft = .3048 m

Another relevant fact is the shoulder composition in each State:

MINNESOTA SHOULDERS			WASHINGTON SHOULDERS		
mixed bituminous	243	39.3%	asphalt	402	56.5%
gravel or stone	335	54.1%	bituminous	230	32.3%
composite	34	5.5%	gravel	72	10.1%
sod	5	.8%	curb	1	.1%
missing	2	.3%	missing	7	1.0%
	619	100.0%		712	100.0%

Washington shoulders tend to resemble the road surface more than Minnesota shoulders.
This suggests the possibility that a more appropriate variable than either lane width or shoulder width might be the variable TOTWIDTH, total width of road and shoulders. When the shoulder is paved, drivers may not make as much of a distinction between it and the road, and the combined width may be the only important variable. When variables are dependent, it is sometimes useful to replace them with one significant combination. Against this it can be argued that lane width and

shoulder width have different types of effects on accidents and that it is inappropriate to treat them as one additive variable. Indeed, in the final models we do not.

Table 19 exhibits some models with only TOTWIDTH.

TABLE 19. Poisson Models of Segments with TOTWIDTH

MINNESOTA				
Mean No. of Accidents = EXPO×exp{1.7994 + .0152AVGM - .0614TOTWIDTH}				
Estimated standard error of coefficient estimates	.1828	.0087	.0051	
P-value	.0001	.0816	.0001	
WASHINGTON				
Mean No. of Accidents = EXPO×exp{1.2141 - .0192AVGM - .0324TOTWIDTH}				
Estimated standard error of coefficient estimates	.1649	.0061	.0050	
P-value	.0001	.0015	.0001	
COMBINED				
Mean No of Accidents = EXPO×exp{1.3310 - .0078AVGM - .0464TOTWIDTH + .2853STATE}				
Estimated standard error of coefficient estimates	.1313	.0050	.0036	.0386
P-value	.0001	.1191	.0001	.0001
COMBINED (WITHOUT AVGM)				
Mean No. of Accidents = EXPO×exp{1.3480 - .0476TOTWIDTH + .2650STATE}				
Estimated standard error of coefficient estimates	.1309	.0035	.0365	
P-value	.0001	.0001	.0001	

1 mile = 1.61 km, 1 ft = .3048 m

Comparison of these models with those using LW and SHW suggests that replacing LW and SHW by TOTWIDTH plus an adjusted intercept yields similar explanatory value. However, because of the importance of these two geometric variables and the fact that in principle their values are independent, we retain both variables to the extent possible. In a few runs below TOTWIDTH is used instead to facilitate comparisons between the two States.

NOTE: Variables ACCRES = (Number of accidents minus predicted number from a Poisson model not using lane width LW) and LWRES = (LW minus predicted LW from a regression model using other highway variables) can be developed. Their correlation coefficients and associated P-values, not reproduced here, confirm that in Minnesota lane width has a significant independent negative effect on accident counts while in Washington lane width has an insignificant independent positive effect on accident counts.

Horizontal and Vertical Curve Variables

With the exception of the extended negative binomial models, in which individual horizontal and vertical curves were modeled, the horizontal variables used in this study have been the composites H, HM1, HM1.5, and HM2 and the vertical variables have been the composites VC, VM, VMC, and VMCC. All of these variables were found to be highly significant.

The only oddity is shown in Table 20 below and concerns the joint effect of H (average horizontal degree of curve) and VC (sum of crest % grade changes per hundred feet weighted by relative crest curve lengths).

In Table 20 the coefficients of the vertical and horizontal variables differ substantially between the two States and VC is insignificant in Washington with P-value .1854. If one replaces VC by VMC, an alternative measure of crest curves that sums the crest % grade changes per hundred feet over all crests and divides by segment length, the vertical variable becomes significant and its model coefficient stabilizes somewhat (but the horizontal variable H still shows dramatic change in its coefficient). See Table 21. There is of course strong correlation between the horizontal and vertical variables in both States.

Segment Variables		MINNESOTA	WASHINGTON	COMBINED
Horizontal Measure H versus Crest Measure VC	Correlation coefficient	.21320	.38635	.33840
	P-value	.0001	.0001	.0001
Horizontal Measure H versus Crest Measure VMC	Correlation coefficient	.26423	.36362	.32581
	P-value	.0001	.0001	.0001

It is possible that unimportant reweighting is occurring among variables that measure essentially

TABLE 20. Poisson Models of Segments with TOTWIDTH, H, and VC

MINNESOTA					
Mean No.of Accidents = EXPO×exp{.9330 - .0422TOTWIDTH + .1849H + 1.6051VC}					
Estimated standard error of coefficient estimates	.1983	.0052	.0248	.2376	
P-value	.0001	.0001	.0001	.0001	
WASHINGTON					
Mean No. of Accidents = EXPO×exp{.7692 - .0257TOTWIDTH + .0985H + .2596VC}					
Estimated standard error of coefficient estimates	.1731	.0051	.0082	.1960	
P-value	.0001	.00001	.0001	.1854	
COMBINED					
Mean No.of Accidents = EXPO×exp{.9169 - .0385TOTWIDTH + .0954H + .7770VC + .2387STATE}					
Estimated standard error of coefficient estimates	.1344	.0036	.0077	.1345	.0370
P-value	.0001	.0001	.0001	.0001	.0001

1 mile = 1.61 km, 1 ft = .3048 m

the same thing. In Washington 63.2% of the segments contain crest curves versus 83.5% of Minnesota's. However, the mean values of VC and VMC are higher in Washington and their standard deviations are much higher. It is perhaps not surprising that there would be differences between Washington and Minnesota in the coefficient estimates, but it is surprising that VC and VMC behave differently in Washington. VMC roughly measures the number of crests per mile (if one assumes that they all have about the same grade change per hundred feet), while VC measures the average grade change per hundred feet and assigns zero grade change to portions where no crest exists. VMC will be large if there are crests with large grade change per hundred feet, but VC will damp these down if they occur over short lengths (because they will be weighted by length).

Because vertical and horizontal alignment are in principle independent and both are very important, we will retain both. We do this despite the fact that the correlation coefficients are considerably larger and more significant than those between lane width and shoulder width in Washington (which

TABLE 21. Poisson Models of Segments with TOTWIDTH, H, and VMC

MINNESOTA					
Mean No. of Accidents = EXPO×exp{.9039 - .0397TOTWIDTH + .1840H + .0544VMC}					
Estimated standard error of coefficient estimates	.2027	.0054	.0248	.0081	
P-value	.0001	.0001	.0001	.0001	
WASHINGTON					
Mean No. of Accidents = EXPO×exp{.6895 - .0240TOTWIDTH + .0926H + .0395VMC}					
Estimated standard error of coefficient estimates	.1743	.0051	.0085	.0094	
P-value	.0001	.00001	.0001	.0001	
COMBINED					
Mean No. of Accidents = EXPO×exp{.7478 - .0340TOTWIDTH + .0928H + .0538VMC + .2503STATE}					
Estimated standard error of coefficient estimates	.1373	.0036	.0075	.0059	.0369
P-value	.0001	.0001	.0001	.0001	.0001

1 mile = 1.61 km, 1 ft = .3048 m

led us to introduce the combined variable TOTWIDTH). But in some runs we replace VC with VMC. The relationship between the vertical and horizontal measure will be reconsidered below when we use the extended negative binomial model, which takes into account individual curves on a segment.

Grade, Roadside Hazard Rating, Driveway Density, and Other Variables

Other variables systematically investigated in connection with model development include GR (average absolute straight-away grade), RHR (Roadside Hazard Rating), DD (driveway density), SPD (speed), T (commercial traffic %), and INTD (intersection density). Weather variables (NONDRYP and SNP) were also investigated in Minnesota.

The weather variables can be dismissed at once. Both NONDRYP and SNP had negative regression coefficients in models and were not significant. A higher percentage of bad weather tends to accompany a decreased number of accidents, but the P-values are large. In a few runs SNP is marginally significant. Because the weather variable was not local but pertained to a large Weather District in the State of Minnesota and because of its relative insignificance, it was dropped from the modeling and was not collected in Washington State. See Shankar et al.[44] for a study of weather variables in Washington State that indicates sufficiently local weather can be significant.

Among the remaining variables, SPD is not significant in either State nor in the combined data set. This may in part reflect lack of variation in the speed data, as well as the quality of the speed data (speeds were not collected on some segments, but were later reconstructed from HSIS files).

GR is very significant in both States. The other variables are significant in one State or the other (but not both) and significant in the modeling of the combined data sets. One curiosity is that T has a negative coefficient in Minnesota and is not significant, but has a significant positive coefficient in Washington.

The P-values for these variables in Poisson runs on the combined data sets (with other variables LW, SHW, H, VC, and STATE; and with EXPO as an offset variable) are:

VARIABLE	P-value
GR	.0001
RHR	.0001
DD	.0107
INTD	.0563
T	.0697
SPD	.4118

Next we attempt to include combinations of these variables in a combined Poisson model for both States. When this is done, GR and RHR do well, as do GR and DD, and GR and T. GR, RHR, and DD do well together (although STATE gets a P-value of .1417 in this case); and GR, RHR, and

[44] Shankar, V., Mannering, F., and Barfield, W., "Effect of Roadway Geometrics and Environmental Factors on Rural Freeway Accident Frequencies," Accident Analysis and Prevention, 27(3): 371-389, 1995.

INTD do well together.

Thus it is certainly appropriate to include GR and RHR in the model and at least one other variable. INTD measures intersection density. However, intersection accidents and intersection-related accidents are excluded from the accident variable in the segment models. For this reason, any effect of INTD will be indirect and INTD is not strictly comparable to DD (driveway density). This rules out a sum of DD and INTD as a measure. If GR, RHR, DD, and INTD are all included in the model, they have the respective P-values .0001, .0001, .0001, and .1863. We conclude that INTD does have an independent effect distinct from that of DD, but not sufficiently significant to include in the model.

The situation is similar with the commercial traffic variable T. It appears to be significant for the combined data set, but not sufficiently – when other variables are present – for inclusion in the model.

Table 22 shows resultant Poisson models for Minnesota and Washington. The anomalous behavior of lane width and VC in Washington exhibited in Table 15 has already been discussed. However, we should note the insignificance of Roadside Hazard Rating RHR in Minnesota. An interesting set of correlations exists with a bearing on the insignificance of RHR in Minnesota and the peculiar behavior of lane width LW in Washington.

Correlation coefficient and P-value	MINNESOTA SEGMENTS	WASHINGTON SEGMENTS	COMBINED SEGMENTS
Lane Width LW versus Roadside Hazrat RHR	-.01141, .7769	.11555, .0020	-.1202, .6613
Shoulder Width SHW versus Roadside Hazrat RHR	-.23729, .0001	-.14910, .0001	-.33705, .0001
TOTWIDTH versus Roadside Hazrat RHR	-.23563, .0001	-.11560, .0001	-.32559, .0001

RHR in Minnesota has a mean of 2.14 and a standard deviation of .97, while in Washington its mean is 3.67 and standard deviation 1.57. Roadside Hazard Rating is higher and more variable in Washington State. The insignificance of RHR in Minnesota in part relates to the absence of variation. The unexpected sign of the lane width coefficient in Washington likewise may be in part due to its correlation with the quite variable magnitudes of RHR in Washington. When the data from the two States are combined, this correlation becomes insignificant and the coefficients of LW and RHR both attain more plausible values.

In Table 22 most coefficients for the combined model are intermediate between those of the two States. The most prominent anomalies are the negative sign of lane width in Washington, the

TABLE 22. Poisson Models for Segment Accidents

Regression Coefficients (Estimated Standard Error and P-value in parentheses)

Variables (offset = exposure EXPO)	Minnesota 1985-89	Washington 1993-95	Combined
Intercept	2.0693 (.4371, .0001)	-.9719 (.5444, .0742)	.7064 (.3290, .0318)
AVGM (ADT/1,000)	.0128 (.0090, .1559)	-.0210 (.0067, .0017)	-.0112 (.0052, .0322)
Lane Width LW	-.1994 (.0359, .0001)	.0678 (.0480, .1577)	-.0869 (.0280, .0001)
Shoulder Width SHW	-.0792 (.0111, .0001)	-.0390 (.0117, .0008)	-.0599 (.0078, .0001)
Roadside Hazard Rating RHR	.0044 (.0273, .8706)	.0650 (.0171, .0001)	.0703 (.0141, .0001)
Driveway Rate DD	.0089 (.0033, .0075)	.0119 (.0023, .0001)	.0095 (.0019, .0001)
Degree of Curve H	.1363 (.0283, .0001)	.0783 (.0099, .0001)	.0711 (.0089, .0001)
Crest VC	1.1905 (.2634, .0001)	.2090 (.2073, .3135)	.6843 (.1455, .0001)
Absolute Grade GR	.2459 (.0598, .0001)	.0779 (.0234, .0009)	.1009 (.0213, .0001)
State (MN = 0, WA = 1)	--	--	.0909 (.0453, .0447)
n, p $D^m/(n - p)$, $\chi^2/(n - p)$	619, 9 1.6827, 1.6596	712, 9 1.6525, 1.7179	1331, 10 1.7135, 1.7422
T_1	13.55	12.04	22.71
R^2, P^2, R_P^2	.7379, .8890,.8300	.6287, .8138,.7726	.6611, .8610, .7778
R_W^2, P_W^2, R_{PW}^2	.8300, .8960, .9263	.7641, .8609, .8875	.7886, .8777, .8984
R_{FT}^2, P_{FT}^2, R_{PFT}^2	.6426, .7609, .8446	.5846, .7049, .8293	.5999, .7341, .8172

TABLE 23. Additional Poisson Models for Segment Accidents

Regression Coefficients (Estimated Standard Error and P-value in parentheses)

Variables (offset = exposure EXPO)	Minnesota 1985-89	Washington 1993-95	Combined
Intercept	2.1930 (.4438, .0001)	.0378 (.2034, .8526)	.7048 (.3293, .0323)
AVGM (ADT/1,000)	--	-.0252 (.0066, .0001)	--
Lane Width LW	-.1856 (.0350, .0001)	TOTWIDTH	-.0918 (.0281, .0011)
Shoulder Width SHW	-.0757 (.0106, .0001)	-.0135 (.0054, .0116)	-.0664 (.0077, .0001)
Roadside Hazard Rating RHR	--	.0726 (.0169, .0001)	.0662 (.0143, .0001)
Driveway Rate DD	.0092 (.0033, .0050)	.0102 (.0024, .0001)	.0097 (.0019, .0001)
Degree of Curve H	.1445 (.0278, .0001)	.0701 (.0101, .0001)	.0720 (.0089, .0001)
Crest VC in MN, Combined; VMC in WA	1.2257 (.2567, .0001)	.0378 (.0101, .0002)	.6999 (.1450, .0001)
Absolute Grade GR	.2438 (.0582, .0001)	.0740 (.0235, .0016)	.1077 (.0214, .0001)
SNP in MN; T in Combined	-.8851 (.5938, .1361)	--	.0070 (.0029, .0153)
STATE	--	--	.0418 (.0448, .3500)
n, p $D^m/(n - p)$, $\chi^2/(n - p)$	619, 8 1.6796, 1.6361	712, 8 1.6396, 1.6774	1331, 10 1.7126, 1.7592
T_1	14.54	12.04	22.55
R^2, P^2, R_P^2	.7297, .8890,.8208	.6279, .8138,.7716	.6607, .8610, .7673
R_W^2, P_W^2, R_{PW}^2	.8290, .8941, .9272	.7685, .8604, .8932	.7909, .8803, .8985
R_{FT}^2, P_{FT}^2, R_{PFT}^2	.6421, .7609, .8439	.5859, .7049, .8311	.6006, .7341, .8182

insignificance of Roadside Hazard Rating RHR in Minnesota, and the insignificance of the crest variable VC in Washington.

Table 23 shows a few variant Poisson models with characteristics of special interest. In Table 23 the insignificant variables from Table 22 are removed and other variables are introduced. In Minnesota AVGM and RHR have been removed, and SNP has been added (P-value = .1361). In Washington TOTWIDTH has replaced LW and SHW, and VMC has replaced VC. Also in Table 23 the combined data set is presented without AVGM but with the addition of T. The variable T is quite significant but STATE loses its significance (P-value = .3500).

Poisson versus Negative Binomial

For the models in Tables 22 and 23 the values of $D^m/(n - p)$, $\chi^2/(n - p)$, and T_1 are computed, along with several measures of goodness-of-fit. The goodness-of-fit measures indicate that the models have a good deal of explanatory power. However, the other statistics in all cases strongly support the conclusion that the data are overdispersed. In particular, the large values of T_1 establish this decisively. The sources of the overdispersion are presumably segment characteristics not included in the model. Some of these characteristics might be items not collected (e.g., sight distances, superelevations, local weather) that are possible to collect, but others are items well outside the scope of this study (e.g., driver characteristics).

Negative binomial models are a natural generalization of the Poisson that permit treatment of overdispersion. Such models can be developed with the software package LIMDEP or by trial and error with SAS and different choices of an overdispersion parameter. The negative binomial also has the advantage of lending itself nicely to application of empirical Bayesian techniques when past accident data are available at a site. An adjusted model can be developed with parameters partly derived from the past data and partly from the given negative binomial model. The new model makes use of the old but also allows the predictions of the old model to be tempered by actual experience on the roadway. See Hauer et al. (1988).

The phenomena noted in the earlier Poisson models occur in the negative binomial setting: differences between the behavior of AVGM, lane width LW, VC and VMC, and RHR from one State to the other; and marginal significance of INTD and T. So the analysis is not repeated. In general the estimated coefficients of variables are similar to what they were under the Poisson models. However, we have an estimate for one additional parameter, the overdispersion parameter K.

Table 24 shows four representative negative binomial models. The overdispersion parameters vary from 0.26 to 0.30. Variables that are omitted are not significant, and some that are retained are not as well – notably, intercept in three of the models, AVGM, and VC in the combined data set (and in Washington, not shown). AVGM is not at all significant in Minnesota, not very significant in Washington, and intermediate in the combined data set. Lane width has the wrong sign in Washington (not shown), and is less significant in the combined data set than it was in the Poisson

TABLE 24. Negative Binomial Models for Segment Accidents
Regression Coefficients (Estimated Standard Error and P-value in parentheses)

Variables (offset = exposure EXPO)	Minnesota 1985-89	Washington 1993-95	Combined	Combined Variant
Intercept	1.9456 (.6992, .0054)	.0358 (.2719, .8953)	.6883 (.4779, .1492)	.4733 (.4796, .3356)
AVGM (ADT/1,000)	--	-.0242 (.0137, .0787)	-.0109 (.0107, .3067)	--
Lane Width LW	-.1821 (.0573, .0015)	TOTWIDTH -.0127 (.0071, .0720)	-.0857 (.0405, .0343)	-.0700 (.0404, .0833)
Shoulder Width SHW	-.0800 (.0158, .0001)		-.0577 (.0106, .0001)	-.0569 (.0105, .0001)
Roadside Hazard Rating RHR	--	.0642 (.0254, .0116)	.0622 (.0219, .0046)	.0609 (.0219, .0055)
Driveway Rate DD	.0079 (.0042, .0630)	.0100 (.0035, .0045)	.0091 (.0027, .0007)	.0072 (.0026, .0067)
Degree of Curve H	.1421 (.0545, .0092)	.0735 (.0154, .0001)	.0856 (.0126, .0001)	.0772 (.0140, .0001)
VC (MN/COM) VMC (WA/COMV)	1.0495 (.4964, .0345)	.0333 (.0168, .0468)	.3748 (.2605, .1502)	.0394 (.0141, .0052)
Absolute Grade GR	.1990 (.0928, .0320)	.0800 (.0295, .0066)	.0976 (.0280, .0005)	.0941 (.0280, .0008)
State	--	--	.1420 (.0679, .0366)	.1427 (.0678, .0353)
n, p $D^m/(n - p - 1)$	619, 7 1.4938	712, 8 1.4767	1331, 10 1.4993	1331, 9 1.4922
K	.2657 (.0385, .0001)	.2821 (.0385, .0001)	.3022 (.0285,.0001)	.2943 (.0281,.0001)
R_K^2	.8609	.8302	.8310	.8354
R^2	.7251	.6268	.6489	.6669
R_D^2, P_D^2 R_{PD}^2	.3720, .5607 .6634	.3455, .5300 .6518	.3518, .5464 .6438	.3548,.5477 .6478

TABLE 25. Negative Binomial Models for Segment Injury Accidents
Regression Coefficients (Estimated Standard Error and P-value in parentheses)

Variables (offset = exposure EXPO)	Minnesota 1985-89	Washington 1993-95	Combined
Intercept	1.9998 (.8205, .0148)	-.2375 (.3511, .4988)	.1675 (.6108, .7839)
Lane Width LW	-.2458 (.0694, .0004)	TOTWIDTH -.0279 (.0089, .0017)	-.1155 (.0531, .0296)
Shoulder Width SHW	-.1053 (.0212, .0001)		-.0740 (.0143, .0001)
Roadside Hazard Rating RHR	--	.0506 (.0314, .1077)	.0410 (.0272, .1315)
Driveway Rate DD	--	.0065 (.0041, .1193)	.0054 (.0035, .1192)
Degree of Curve H	.2158 (.0667, .0012)	.0598 (.0194, .0020)	.0730 (.0161, .0001)
Crest VMC	--	.0405 (.0219, .0648)	.0399 (.0177, .0239)
Absolute Grade GR	--	.0725 (.0377, .0543)	.0574 (.0360, .1109)
State	--	--	.4149 (.0879, .0001)
n, p $D^m/(n - p - 1)$	619, 4 1.0702	712, 7 1.1593	1331, 9 1.1212
K	.2398 (.0786,.0023)	.2751 (.0682, .0001)	.2710 (.0518,.0001)
R^2_K	.8934	.8444	.8628
R^2	.5859	.4824	.5386
R^2_D, P^2_D R^2_{PD}	.3483, .4468 .7795	.3185, .4334 .7348	.3303, .4399 .7509

runs. The goodness-of-fit measures, including the ordinary R^2, yield no dramatic conclusions. R_K^2 is systematically larger than the others. All the measures suggest that the Minnesota coefficients account for Minnesota accidents a bit better than the other models.

Table 25 shows negative binomial models for serious accidents, based on the variable INJACC. Variables with little significance have been omitted and only those that are significant or marginally significant have been retained. The Minnesota model, with the fewest variables, once again has the highest goodness-of-fit. The coefficients are roughly comparable to those for the models for total number of accidents (TOTACC). Differences between the deviances D^m and R^2 as one passes from Table 24 (TOTACC) to Table 25 (INJACC) are not of importance. Both measures tend to give smaller values when observed data are near zero, and larger values when the observations are away from zero: INJACC has small or zero values more often than TOTACC.

The Extended Negative Binomial

Extended negative binomial models are a variant of negative binomial models in which the mean number of accidents μ_i at segment i is taken to have the form

$$\mu_i = \exp(\beta_0) \times \prod_{j=1}^{n} \left(\sum_{c=1}^{C_{ij}} w_{ijc} \exp(x_{ijc}\beta_j)\right) \tag{5.17}$$

instead of (5.1). With respect to the j-th highway variable, segment number i is decomposed into C_{ij} subsegments of relative lengths $\{w_{ijc} : c = 1, ..., C_{ij}\}$ where the variables x_{ij} take the respective putatively constant values $\{x_{ijc} : c = 1, ..., C_{ij}\}$. In effect this model slices up the segments into subsegments[45] where each variable is constant. The weights w_{ijc} are the relative lengths of the subsegments and add to 1. The value C_{ij} can be taken to be independent of i (and j) if the maximum number of subsegments in the data set is specified: for segments with fewer subsegments the extra weights can be set equal to zero. For some variables, all weights except one are set to zero, and the model behaves like an ordinary negative binomial model with respect to them.

An advantage of the extended negative binomial model is that it permits local variation along a roadway to be taken into account. Rather than summing local effects or averaging them, one in effect sums the accidents occurring on subsegments where conditions are constant. This gives the model form a scale independence: one may decompose segments into subsegments or aggregate adjacent segments without changing model form.

[45] The model treats different variables as if the subsegments with respect to each variable are independent, e.g., if one-third of the segment has a steep grade and one-half of the segment is on a horizontal curve, then one-sixth of the segment has both steep grade and a horizontal curve. Shaw-Pin Miaou, the author of the model, is developing a refinement that does not assume such independence.

TABLE 26. Extended Negative Binomial Models for Segment Accidents

Regression Coefficients (Estimated Standard Error and P-value in parentheses)

Variables (offset = exposure EXPO)	Minnesota 1985-89	Washington 1993-95	Combined
Intercept	2.0168 (.6593, .0022)	.0846 (.2883, .7692)	.6287 (.4993, .2080)
AVGM (ADT/1,000)	--	-.0239 (.0107, .0252)	-.0111 (.0897, .2099)
Lane Width LW	-.1843 (.0548, .0008)	TOTWIDTH -.0142 (.0077, .0669)	-.0829 (.0424, .0504)
Shoulder Width SHW	-.0812 (.0161, .0001)		-.0560 (.0116, .0001)
Roadside Hazard Rating RHR	--	.0689 (.0245, .0049)	.0665 (.0210, .0016)
Driveway Rate DD	.0089 (.0044, .0423)	.0119 (.0033, .0003)	.0091 (.0026, .0005)
Degrees of Curve DEG{i}	.0474 (.0133, .0003)	.0521 (.0085, .0001)	.0445 (.0078, .0001)
Crest Curve Rates V{j}	.4834 (.1416, .0006)	--	.4653 (.1255, .0002)
Absolute Grades GR{k}	.2404 (.0592, .0001)	.0894 (.0314, .0045)	.1047 (.0286, .0003)
State	--	--	.1585 (.0674, .0188)
n, p $D^m/(n - p - 1)$	619, 6 1.4980	712, 7 1.4877	1331, 10 1.5012
K	.2722 (.0457, .0001)	.3055 (.0460, .0001)	.3034 (.0331,.0001)
R_K^2	.8575	.8161	.8303
R^2	.7246	.5720	.6555

TABLE 27. Final Extended Negative Binomial Model for Segment Accidents

Regression Coefficients (Estimated Standard Error and P-value in parentheses)

Variables (offset = exposure EXPO)	Combined
Intercept	.6409 (.5008, .2006)
Lane Width LW	-.0846 (.0425, .0465)
Shoulder Width SHW	-.0591 (.0114, .0001)
Roadside Hazard Rating RHR	.0668 (.0211, .0015)
Driveway Rate DD	.0084 (.0026, .0011)
Degrees of Curve DEG{i}	.0450 (.0078, .0001)
Crest Curve Rates V{j}	.4652 (.1260, .0002)
Absolute Grades GR{k}	.1048 (.0287, .0003)
State	.1388 (.0659, .0351)
n, p $D^m/(n - p - 1)$	1331, 9 1.5012
K	.3056 (.0331, .0001)
R_K^2	.8291
R^2	.6547

TABLE 28. Extended Negative Binomial Models for Segment Injury Accidents

Regression Coefficients (Estimated Standard Error and P-value in parentheses)

<table>
<tr><th>Variables
(offset = exposure EXPO)</th><th>Minnesota
1985-89</th><th>Washington
1993-95</th><th>Combined</th></tr>
<tr><td>Intercept</td><td>1.7147
(.8860, .0530)</td><td>-.1571
(.3657, .6675)</td><td>.3534
(.6546, .5893)</td></tr>
<tr><td>Lane Width LW</td><td>-.2233
(.0735, .0024)</td><td rowspan="2">TOTWIDTH

-.0302
(.0095, .0015)</td><td>-.1306
(.0558, .0193)</td></tr>
<tr><td>Shoulder Width SHW</td><td>-.0996
(.0219, .0001)</td><td>-.0784
(.0150, .0001)</td></tr>
<tr><td>Roadside Hazard Rating RHR</td><td>--</td><td>.0568
(.0309, .0659)</td><td>.0598
(.0261, .0217)</td></tr>
<tr><td>Driveway Rate DD</td><td>--</td><td>.0085
(.0040, .0349)</td><td>.0062
(.0034, .0679)</td></tr>
<tr><td>Degrees of Curve DEG{i}</td><td>.0580
(.0116, .0001)</td><td>.0406
(.0107, .0001)</td><td>.0457
(.0091, .0001)</td></tr>
<tr><td>Crest Curve Rates V{j}</td><td>.5528
(.1364, .0001)</td><td>--</td><td>.4694
(.1687, .0054)</td></tr>
<tr><td>Absolute Grades GR{k}</td><td>--</td><td>.0823
(.0400, .0395)</td><td>--</td></tr>
<tr><td>State</td><td>--</td><td>--</td><td>.4309
(.0852, .0001)</td></tr>
<tr><td>n, p
$D^m/(n - p - 1)$</td><td>619, 6
1.0763</td><td>712, 6
1.3009</td><td>1331, 9
1.1308</td></tr>
<tr><td>K</td><td>.2482
(.0751, .0010)</td><td>.2951
(.0699, .0001)</td><td>.2880
(.0523,.0001)</td></tr>
<tr><td>R_K^2</td><td>.8899</td><td>.8320</td><td>.8542</td></tr>
<tr><td>R^2</td><td>.5926</td><td>.4750</td><td>.5277</td></tr>
</table>

As with the negative binomial the goal is to estimate the coefficient vector β and the overdispersion parameter K. Shaw-Pin Miaou made available a program that uses maximum likelihood to estimate these quantities. In Table 26 we show the results of the modeling.

In Table 26 AVGM and Roadside Hazard Rating RHR are strongly insignificant in Minnesota and so were removed. In Washington the crest variable V{j}, although having the correct sign, is strongly insignificant in the presence of the other variables and so was removed. In the combined data set AVGM (and the Intercept variable) are insignificant. When AVGM was removed and the commercial percentage variable T added, the estimated coefficient for T was positive but had a significance level of about 20%. When the speed variable SPD is added instead, it has a negative coefficient and a P-value of 50%.

Table 27 represents our final model for segments. It contains a large number of variables, all of them significant, and it represents the combined characteristics of rural segments in two States with a reasonable amount of variation in all variables.

Table 28 shows three extended negative binomial models for Injury Accidents. AVGM was insignificant in all three data sets. RHR and DD were insignificant in Minnesota. The straightaway grade variable GR{k} was not significant in Minnesota, and the crest vertical V{j} was not significant in Washington. Extended negative binomial runs with all variables present did not converge in the combined data set, but did when GR{k} was removed. A total of 36% of all reported segment accidents were Injury Accidents in Minnesota versus 46% in Washington, and this is reflected by the increase in the coefficient for State from Table 27 to Table 28.

INTERSECTION MODELS

Models for the three-legged and four-legged intersections in Minnesota and Washington are of Poisson and negative binomial type. Extended negative binomial models, appropriate for nonhomogeneous and variable stretches of road, are not attempted. The variables used to model accidents describe traffic volumes, horizontal and vertical alignment, channelization, roadside (driveways and hazard rating), intersection angle, and posted speed. Although sight distance is a desirable variable, data were not available. The alignment variables and hazard rating can be viewed as partial surrogates for sight distance.

Because the intersection models are based on fewer observations than the segment models, and the relationships revealed between accidents and intersection variables are less clear-cut, some adjustments are made in the criteria for retaining variables in the models. In order to identify design variables that influence accidents and are subject to control of designers, in many of the models P-values are allowed much greater range than in the segment models. Values as high as 30% occur in some models.

To some extent this represents a shift in methodology. For a P-value of 5%, under the null hypothesis that a particular variable has no influence and thus has zero as its true coefficient, there is one chance in 20 that the estimate for the coefficient will be as far away from zero as, or farther away than, it is found to be. With a P-value of 30%, under the null hypothesis there are three chances in 10 that the estimate will be as far from zero as, or farther than, the actual estimate. The estimated coefficient is viewed as a fluctuation from zero due to random errors in the data. However, there is no compelling reason why the null hypothesis should govern the analysis, especially when engineering judgment suggests that the variable under study has an influence on accident counts. A defensible alternative is to view the estimated coefficient arrived at by maximum likelihood methods as a "best guess" whose confidence interval is measured by the standard error of the estimate. Larger P-values correspond to larger confidence intervals, perhaps intervals that include zero, but the estimate itself summarizes the data better than assignment of a zero coefficient and removal of the variable from the model. Adopting the "best guess" viewpoint is a more aggressive, less conservative stance toward the investigation of the underlying reality. Permitting larger P-values may be thought of as a partial transition toward the latter stance: we still show some deference toward the null hypothesis, but we attend closely to the estimate offered by the model, more closely the smaller its standard error.

Tables 29 through 35 below exhibit the chief models of both Poisson and negative binomial type for both the three-legged and four-legged intersections. For comparability, number of years is used as an offset so that what is modeled is mean number of accidents per year. Estimated coefficients for each variable are shown, along with their standard errors and P-values. Some variables were considered in the preliminary analysis that may not appear in the Tables – variants of the variables used here, as well as weather variables SNP and NONDRYP in Minnesota (these had negative sign and were not very significant). Tables 36 and 37 exhibit models for Injury Accidents.

Traffic

The chief variables are major and minor road traffic – ADT1 and ADT2. In addition the variable CINDEX, conflict index, measuring the relationship between these two was considered. In preliminary runs it was not significant when used in addition to them, and it was less significant than either of them when used as a substitute for one of them. ADT1 and ADT2 have different relative effects in the three-legged versus the four-legged cases (cf. Table 35):

Variable	Estimated Coefficient	
	3-legged	4-legged
LADT1	.8052	.6026
LADT2	.5037	.6090

For four-legged intersections, major and minor road ADT have approximately equal influence, while for three-leggeds the major road ADT dominates. If one views a four-legged intersection as two three-legged intersections, admittedly an oversimplification, and accordingly halves the coefficient of LADT2 in the last column above, the effects are seen to be roughly compatible.

TABLE 29. Poisson Models, 3-Legged Intersections Accidents
Regression Coefficients (Standard error and P-value in parentheses)

Variables (Offset = number of years)	Minnesota 1985-89	Washington 1993-95	Combined
Intercept	-12.5714 (.8238, .0001)	-10.4414 (1.5325, .0001)	-12.1055 (.8241, .0001)
Log of ADT1	.8524 (.0560, .0001)	.6569 (.1386, .0001)	.8291 (.0511, .0001)
Log of ADT2	.4466 (.0461, .0001)	.5219 (.0628, .0001)	.4578 (.0367, .0001)
VCI (crests)	.3313 (.1301, .0109)	-.2430 (.1554, .1180)	-.0010 (.0957, .9915)
HI	.0473 (.0141, .0008)	-.0018 (.0260, .9458)	.0333 (.0124, .0073)
SPDI	.0190 (.0101, .0597)	.0062 (.0146, .6731)	.0151 (.0083, .0687)
Roadside Hazard Rating RHRI	.1788 (.0554, .0012)	.0995 (.0749, .1842)	.1712 (.0431, .0001)
No. Drwys ND	-.0441 (.0306, .1498)	-.0342 (.0426, .4215)	-.0436 (.0241, .0710)
Right Turn Lane RT	.2684 (.1068, .0119)	.1472 (.1814 .4172)	.2554 (.0909, .0050)
Angle HAU	.0060 (.0016, .0002)	-.0073 (.0100, .4620)	.0052 (.0016, .0008)
State (MN = 0, WA = 1)	--	--	-.2497 (.1071, .0198)
n, p $D^m/(n - p)$, $\chi^2/(n - p)$	389, 10 1.5388, 1.8818	181, 10 1.5867, 1.5900	570, 11 1.5554, 1.8344
T_1	18.25	7.38	21.22
R^2, P^2, R^2_P	.4653, .8375, .5556	.3298, .6844, .4819	.4203, .8147, .5159
R^2_W, P^2_W, R^2_{PW}	.6413, .8044, .7973	.5094, .6734, .7564	.5898, .7720, .7640
${R_{FT}}^2$, ${P_{FT}}^2$, R^2_{PFT}	.4722, .5568, .8481	.2702, .4090, .6606	.4130, .5206, .7933

TABLE 30. Poisson Models, 4-Legged Intersection Accidents

Regression Coefficients (Standard error and P-value in parentheses)

Variables (Offset = number of years)	Minnesota 1985-89	Washington 1993-95	Combined
Intercept	-10.5546 (.8711, .0001)	-10.7648 (1.4384, .0001)	-11.6312 (.8283, .0001)
Log of ADT1	.6517 (.0626, .0001)	.3710 (.1384, .0073)	.6064 (.0556, .0001)
Log of ADT2	.6089 (.0520, .0001)	.7934 (.0835, .0001)	.6739 (.0427, .0001)
VCI (crests)	.3805 (.1090, .0005)	-.0064 (.1171, .9565)	.2280 (.0777, .0033)
HI	.0334 (.0363, .3578)	-.4329 (.1188, .0003)	.0114 (.0308, .7106)
SPDI	.0166 (.0134, .2156)	.0630 (.0132, .0001)	.0415 (.0090, .0001)
Roadside Hazard Rating RHRI	-.0425 (.0508, .4026)	-.2050 (.0740, .0056)	-.0994 (.0411, .0156)
No. Drwys ND	.1165 (.0316, .0002)	.0546 (.0472, .2472)	.0919 (.0258, .0004)
Right Turn Lanes RT	-.0803 (.1119, .4736)	-.7261 (.1599 .0001)	-.2323 (.0886, .0087)
Angle HAU	-.0044 (.0024, .0701)	.0309 (.0079, .0001)	-.0016 (.0023, .4966)
State (MN = 0, WA = 1)	--	--	-.0629 (.1038, .5447)
n, p $D^m/(n - p)$, $\chi^2/(n - p)$	327, 10 1.3371, 1.3665	90, 10 3.1285, 2.8507	417, 11 1.8524, 1.8106
T_1	3.71	11.05	14.97
R^2, P^2, R_P^2	.6057, .7288, .8311	.4513, .8374, .5389	.4556, .7868, .5791
R_W^2, P_W^2, R_{PW}^2	.5635, .6705, .8404	.7564, .9039, .8369	.5695, .7558, .7535
R_{FT}^2, P_{FT}^2, R_{PFT}^2	.4807, .5081, .9460	.3813, .7792, .4894	.3700, .6183, .5985

TABLE 31. Negative Binomial Models, 3-Legged Intersection Accidents
Regression Coefficients (Standard error and P-value in parentheses)

Variables (Offset = number of years)	Minnesota 1985-89	Washington 1993-95	Combined
Intercept	-12.8114 (1.2566, .0001)	-11.3859 (2.8742, .0003)	-12.3250 (1.1872, .0001)
Log of ADT1	.8090 (.0658, .0001)	.7490 (.2492, .0027)	.8073 (.0632, .0001)
Log of ADT2	.5055 (.0715, .0001)	.5211 (.1022 .0001)	.5027 (.0561, .0001)
VCI (crests)	.2915 (.3025, .3353)	-.2115 (.2409, .3798)	.0758 (.1327, .5682)
HI	.0351 (.0334, .2935)	.0175 (.0527, .7399)	.0270 (.0250, .2800)
SPDI	.0253 (.0188, .1780)	.0100 (.0281, .7218)	.0188 (.0141, .1837)
Roadside Hazard Rating RHRI	.1653 (.0683, .0156)	.0681 (.1230, .5798)	.1372 (.0584, .0188)
No. Drwys ND	-.0293 (.0479, .5405)	-.0208 (.0756, .7835)	-.0270 (.0399, .4977)
Right Turn Lane RT	.2578 (.1402, .0660)	.1765 (.3598, .6238)	.2442 (.1265, .0537)
Angle HAU	.0047 (.0032, .1444)	-.0069 (.0242, .7736)	.0040 (.0033, .2355)
State (MN = 0, WA = 1)	--	--	-.1994 (.1578, .2064)
n, p $D^m/(n-p-1)$	389, 10 1.2959	181, 10 1.3731	570, 11 1.3277
K R_K^2	.4759(.1001,.0001) .7828	.7927(.3180,.0127) .6450	.5794(.0955,.0001) .7390
R^2	.4452	.3022	.4057
R_D^2, P_D^2 R_{PD}^2	.2878, .4585 .6278	.1751, .3919 .4468	.2609, .4463 .5847

TABLE 32. Negative Binomial Models, 4-Legged Intersection Accidents
Regression Coefficients (Standard error and P-value in parentheses)

Variables (Offset = number of years)	Minnesota 1985-89	Washington 1993-95	Combined
Intercept	-10.6729 (1.3603, .0001)	-10.9301 (3.7629, .0038)	-11.4840 (1.5737, .0001)
Log of ADT1	.6179 (.0847, .0001)	.3681 (.3828, .3364)	.5773 (.0985, .0001)
Log of ADT2	.6262 (.0730, .0001)	.9218 (.2280, .0001)	.6944 (.0795, .0001)
VCI (crests)	.3121 (.2490, .2101)	.0484 (.6446, .9401)	.2681 (.2147, .2118)
HI	.0441 (.0482, .3605)	-.3381 (.2142, .1144)	.0359 (.0477, .4519)
SPDI	.0222 (.0189, .2407)	.0507 (.0274, .0644)	.0399 (.0150, .0080)
Roadside Hazard Rating RHRI	-.0628 (.0579, .2786)	-.1997 (.1702, .2406)	-.1175 (.0587, .0454)
No. Drwys ND	.1295 (.0513, .0116)	-.0023 (.1316, .9858)	.1056 (.0501, .0351)
Right Turn Lanes RT	-.0557 (.1266, .6601)	-.7191 (.4662 .1230)	-.1627 (.1407, .2474)
Angle HAU	-.0052 (.0033, .1169)	.0384 (.0154, .0127)	-.0023 (.0039, .5534)
State (MN = 0, WA = 1)	--	--	.0094 (.1814, .9588)
n, p $D^m/(n - p - 1)$	327, 10 1.2920	90, 10 2.1620	417, 11 1.5457
K R^2_K	.2044(.0670,.0023) .8344	.9466(.2828, .0008) .6051	.5219 (.0849,.0001) .7187
R^2	.5882	.3366	.4313
R^2_D, P^2_D R^2_{PD}	.2981, .4052 .7357	.1197, .5290 .2262	.2653, .4799 .5529

TABLE 33. Additional Negative Binomial Models, Combined (MN/WA) Intersection Accidents

Regression Coefficients (Standard error and P-value in parentheses)

Variables (Offset = no. of years)	Combined 3-legged	Combined 4-legged
Intercept	-12.4698 (1.1151, .0001)	-11.0804 (1.5718, .0001)
Log of ADT1	.8046 (.0615, .0001)	.5834 (.0985, .0001)
Log of ADT2	.5002 (.0552, .0001)	.6839 (.0769 .0001)
VCI (crests)	--	.2714 (.2017, .1785)
HI	.0280 (.0248, .2587)	--
SPDI	.0216 (.0132, .1034)	.0298 (.0149, .0448)
Roadside Hazard Rating RHRI	.1412 (.0578, .0145)	--
No. Drwys ND	--	.0888 (.0524, .0899)
Right Turn Lane RT	.2461 (.1266, .0519)	-.1586 (.1390, .2538)
Angle (HAU for 3-leggeds, DEV for 4-leggeds)	.0037 (.0033, .2681)	-.0059 (.0047, .2190)
State (MN = 0, WA = 1)	-.2068 (.1574, .1890)	-.1335 (.1729, .4399)
n, p $D^m/(n - p - 1)$	570, 9 1.3243	417, 9 1.5448
K R_K^2	.5826(.0938, .0001) .7376	.5281(.0832,.0001) .7154
R^2	.4016	.4511
R_D^2, P_D^2 R_{PD}^2	.2628, .4484 .5862	.2658, .4811 .5524

TABLE 34. Additional Negative Binomial Models, Minnesota Intersection Accidents

Regression Coefficients (Standard error and P-value in parentheses)

Variables (Offset = number of years)	Minnesota, three-legged, 1985-89	Minnesota, four-legged, 1985-89	Minnesota, four-legged, 1985-89
Intercept	-11.2798 (.6343, .0001)	-9.5860 (.7397, .0001)	-9.4267 (.7632, .0001)
Log of ADT1	.7923 (.0619, .0001)	.6568 (.0829, .0001)	.6334 (.0881, .0001)
Log of ADT2	.4920 (.0683, .0001)	.5882 (.0691 .0001)	.6116 (.0695, .0001)
VCI (crests)	--	.3499 (.1931, .0699)	--
HI	--	--	.0719 (.0308, .0195)
Roadside Hazard Rating RHRI	.1944 (.0666, .0035)	--	--
No. Drwys ND	--	.1088 (.0459, .0177)	--
Right Turn Lane RT	.2822 (.1375, .0402)	--	--
Angle DEV	--	-.0105 (.0042, .0120)	-.0111 (.0042, .0083)
n, p $D^m/(n - p - 1)$	389, 5 1.3316	327, 6 1.2690	327, 5 1.2995
K R^2_K	.5377(.1024,.0001) .7546	.1854(.0611,.0024) .8498	.2293 (.0700,.0011) .8143
R^2	.3955	.6208	.5869
R^2_D, P^2_D R^2_{PD}	.2828, .4630 .6109	.3107, .4116 .7548	.2941, .4115 .7146

TABLE 35. Final Negative Binomial Models, Minnesota Intersection Accidents

Regression Coefficients (Standard error and P-value in parentheses)

Variables (Offset = number of years)	MN 3-leggeds, 1985-89	MN 4-leggeds, 1985-89
Intercept	-12.9922 (1.1511, .0001)	-10.4260 (1.3167, .0001)
Log of ADT1	.8052 (.0639, .0001)	.6026 (.0836, .0001)
Log of ADT2	.5037 (.0708, .0001)	.6091 (.0694 .0001)
VCI (crests)	.2901 (.2935, .3229)	.2885 (.2576, .2628)
HI	.0339 (.0327, .3004)	.0449 (.0473, .3431)
SPDI	.0285 (.0177, .1072)	.0187 (.0176, .2875)
Roadside Hazard Rating RHRI	.1726 (.0677, .0108)	--
No. Drwys ND	--	.1235 (.0519, .0173)
Right Turn Lane RT	.2671 (.1398, .0561)	--
Angle HAU	.0045 (.0032, .1578)	-.0049 (.0033, .1341)
n, p $D^m/(n - p - 1)$	389, 9 1.3200	327, 8 1.2874
K R^2_K	.4811(.0998,.0001) .7805	.2055(.0652,.0016) .8336
R^2	.4409	.5944
R^2_D, P^2_D R^2_{PD}	.2891, .4604 .6279	.3005, .4081 .7364

TABLE 36. Negative Binomial Models, 3-Legged Intersection Injury Accidents
Regression Coefficients (Standard error and P-value in parentheses)

Variables (Offset = no. of years)	Minnesota 1985-9	Washington 1993-5	Combined
Intercept	-13.0374 (1.7908, .0001)	-13.8430 (3.2641, .0001)	-12.9939 (1.6299, .0001)
Log of ADT1	.8122 (.0973, .0001)	.9037 (.2915, .0019)	.8357 (.0878, .0001)
Log of ADT2	.4551 (.0977, .0001)	.5445 (.1314 .0001)	.4840 (.0721, .0001)
VCI (crests)	.1869 (.3657, .6092)	-.1000 (.2787, .7196)	.0247 (.1773, .8893)
HI	.0335 (.0327, .3047)	-.0063 (.0739, .9316)	.0179 (.0294, .5424)
SPDI	.0156 (.0269, .5618)	.0165 (.0331, .6173)	.0125 (.0197, .5248)
Roadside Hazard Rating RHRI	.2065 (.0930, .0263)	-.0002 (.1505, .9990)	.1300 (.0757, .0858)
No. Drwys ND	-.0120 (.0714, .8671)	.0293 (.0840, .7276)	-.0044 (.0525, .9331)
Right Turn Lane RT	.3620 (.1814, .0460)	.1647 (.4034, .6830)	.2957 (.1590, .0629)
Angle HAU	.0051 (.0045, .2594)	.0016 (.0412, .9692)	.0046 (.0048, .3384)
State (MN = 0, WA = 1)	--	--	-.1299 (.1924, .4996)
n, p $D^m/(n - p - 1)$	389, 10 .9799	181, 10 .9546	570, 11 .9625
K R_K^2	.4935(.1818,.0066) .8208	.8166(.4144,.0488) .6500	.6219 (.1693,.0002) .7674
R^2	.4149	.1251	.3481
R_D^2, P_D^2 R_{PD}^2	.2520, .3687 .6835	.1917, .3126 .6134	.2441, .3601 .6778

TABLE 37. Negative Binomial Models, 4-Legged Intersection Injury Accidents
Regression Coefficients (Standard error and P-value in parentheses)

Variables (Offset = no. of years)	Minnesota 1985-9	Washington 1993-5	Combined
Intercept	-10.7829 (1.7656, .0001)	-12.5872 (4.5643, .0059)	-12.0196 (1.9399, .0001)
Log of ADT1	.6339 (.1055, .0001)	.4738 (.4945, .3380)	.5963 (.1187, .0001)
Log of ADT2	.6229 (.0870, .0001)	.9085 (.2459, .0002)	.6945 (.0947, .0001)
VCI (crests)	.2789 (.4623, .5464)	.1074 (.6848, .8754)	.2824 (.3469, .4156)
HI	.0729 (.0635, .2513)	-.6484 (.3838, .0911)	.0506 (.0637, .4264)
SPDI	.0112 (.0251, .6567)	.0651 (.0316, .0395)	.0377 (.0195, .0532)
Roadside Hazard Rating RHRI	-.1225 (.0720, .0889)	-.3189 (.2123, .1332)	-.2116 (.0762, .0055)
No. Drwys ND	.0857 (.0639, .1799)	.0303 (.1525, .8425)	.0900 (.0657, .1707)
Right Turn Lanes RT	.0451 (.1665, .7865)	-.9153 (.5273 .0826)	-.1273 (.1798, .4790)
Angle HAU	-.0043 (.0044, .3258)	.0360 (.0157, .0220)	-.0018 (.0052, .7339)
State (MN = 0, WA = 1)	--	--	.2487 (.2321, .2839)
n, p $D^m/(n - p - 1)$	327, 10 1.1051	90, 10 1.8042	417, 11 1.2989
K R^2_K	.1811(.1173,.1224) .8870	.9692(.3751,.0098) .6431	.6589 (.1499,.0001) .7470
R^2	.4929	.2139	.3420
R^2_D, P^2_D R^2_{PD}	.2414, .3316 .7279	.1472, .4844 .3040	.2404, .4186 .5744

Alignment, Channelization, and Speed

Two horizontal curve variables were used – HI and HEI – measuring degree of curvature out to 250 respectively 764 feet. These variables had unexpected sign and/or were insignificant in Washington State (for HI, see Tables 29, 30, 31, 32, 36, 37) but behaved somewhat better in Minnesota for both three-legged and four-legged intersections. HI was more stable than HEI, and so for comparability we elected to use HI as our horizontal variable in the runs shown.

Three vertical curve variables were considered – VCI, VI, and VEI. Each measures average grade change per hundred feet for vertical curves near the intersection. The first is for crests out to 250 feet, the second is for both crests and sags out to 250 feet, and the third is for both crests and sags out to 764 feet. In the Minnesota data – the larger of the two State data sets – VCI, the crest only variable and the vertical alignment variable most closely related to sight distance, was substantially more significant than VI and VEI, and hence was selected for inclusion in the runs presented here. On the Washington data the vertical curve variables tended to have unexpected sign and/or be very insignificant.

Several measures of channelization were used in the modeling, but the measure that proved most significant was RT, which takes the values 1 or 0 whether there is or is not at least one right turn lane on the major road. Other channelization variables – for bypass lanes on three-leggeds, zero, one, or two right turn lanes on four-leggeds, or acceleration lanes for the minor roads – were not significant and/or did not show much variation. Thus RT represents channelization in all runs. On three-legged intersections its coefficient was consistently positive and significant. It is not known whether this variable is a surrogate for high accident intersections (i.e., because many accidents tend to occur at the intersection, a right turn lane has been added) or a surrogate for high right turn major road traffic (and high left turn minor road traffic). On the four-legged intersections, the coefficient of RT tended to be negative but was not particularly significant.

The speed variable SPDI, an average of approach speeds – although negatively correlated with ADT, the alignment variables, and number of driveways – seemed to make an independent contribution to the accident frequency in all models.

Roadside Variables - Number of Driveways and Hazard Rating

Perhaps the most remarkable feature of the intersection models is the unexpected but systematic behavior of the variables ND, number of driveways, and RHRI, Roadside Hazard Rating. The coefficient of RHRI is positive at three-legged intersections while that of ND is negative. The reverse occurs on four-leggeds: the coefficient of RHRI is negative and that of ND is positive. Because of the unexpected negative signs, ND has been omitted from some three-legged runs and RHRI has been omitted from some four-legged runs.

With respect to driveways, perhaps drivers take more care when driveways are to be found in the neighborhood of a three-legged intersection, but insufficient additional care in the neighborhood of a four-legged intersection. Each driveway or intersection leg represents potential traffic and requires a share of driver attention. In the intersection data sets driveways actually occur at a larger percentage of three-legged intersections (62.5% in MN and 63% in WA according to Tables 4 and 5) than four-leggeds (32.4% in MN and 46.7% in WA according to Tables 6 and 7). At four-legged intersections, it might be argued that driveways are a third unexpected complication in addition to the two minor road legs, less easily integrated than two complications at a three-legged: a driveway and one minor leg.

With respect to hazard rating, an opposite and possibly inconsistent explanation might be offered: It may be that drivers underestimate roadside hazards at three-leggeds and relatively speaking overestimate them at four-leggeds. Roadside hazards such as obstacles and steep sideslope do not require the same kind of attention as potential traffic entry points. Perhaps such hazards are more likely to be properly attended to when both sides of the roadway have entry points and available accident avoidance tactics are more limited.

The Angle Variable

The variable HAU used in Tables 29 through 33 and 35 through 37 is a signed variable proposed by Ezra Hauer (see Figures 4 and 5). For a three-legged intersection HAU is positive when the angle is larger than 90° as in 4(a) and HAU is negative when the angle is smaller than 90° as in 4(b). On the basis of work of Kulmala (1995) it is thought that turns from the far lane of the major road may be less accident prone in situation 4(a) than in situation 4(b). Accordingly the coefficient of HAU in the three-legged intersection model would be negative (when HAU is positive accidents are less frequent; and when HAU is negative they are more frequent, it is proposed). Of course, there are other turns to be made: a turn from the near lane of the major road, and turns left and right from the minor road. The four-legged version of HAU is the average of the HAU variable for two three-legged intersections (one to the right, one to the left), and would likewise have a negative coefficient if accidents owing to far lane turns through large angles are predominant.

Tables 29 through 32 do not support any strong conclusion. Minnesota and Washington have opposite experience with the variable HAU. Minnesota angle data must be considered much more reliable, though, than Washington angle data. While Minnesota angles were determined from construction plans, those for Washington were very rough estimates made from photologs. Visibility of the direction of minor roads was extremely limited in the photologs. As Tables 4 through 7 indicate, for Minnesota three-leggeds 50.6% were reported as right angles versus 95.6% in Washington; for four-leggeds 37.6% were reported as right angles in Minnesota versus 88.9% in Washington. In the Minnesota Poisson models HAU is significant but the sign of its coefficient has unexpected value (positive) for the three-leggeds, although it behaves as expected for four-leggeds. Under the negative binomial models HAU is marginally significant for the Minnesota data with the same coefficient signs as for the Poisson.

The two other angle variables considered in this study are DEV (the absolute deviation from 90° of the angle, or the average of the two absolute deviations for the four-leggeds) and DEV15 (the squared difference between DEV and 15°, divided by 100). The behavior of these three variables on the Minnesota data is summarized below.

Accident Models for Minnesota **three**-legged intersections with ADT1 and ADT2 and one of the variables at right. P-values and signs of the estimated coefficients of the variable.	VARIABLE	Poisson Model	Negative Binomial Model
	DEV	.4906, +	.9955 , +
	DEV15	.4395, +	.9248, -
	HAU	.0006, +	.2391, +

Accident Models for Minnesota **four**-legged intersections with ADT1 and ADT2 and one of the variables at right. P-values and signs of the estimated coefficients of the variable.	VARIABLE	Poisson Model	Negative Binomial Model
	DEV	.0014, -	.0139, -
	DEV15	.0071, -	.0648, -
	HAU	.0748, -	.1419, -

Thus angle, however measured, is a significant variable at four-legged intersections, and HAU is significant (but the others are not) at three-leggeds.

DEV15 is an empirical variable developed in connection with study of the four-legged intersections. On some runs of Minnesota four-legged data it was more significant than DEV, suggesting that accident rates are highest at angles of 75° and 105°. It was also more significant than DEV on the combined Minnesota and Washington four-legged data. For reasons of simplicity we omit DEV15 from our tables, although we did use DEV on some four-legged runs (Tables 33 and 34).

Negative Binomial Models - Minnesota versus Washington

The statistics compiled in the lower rows of Tables 29 and 30 indicate that the Poisson models have definite explanatory power, especially the Minnesota models, but that they are nonetheless overdispersed. The values of T_1 should be approximately normally distributed about zero if the overdispersion parameter is zero, but the values instead tend to be large positive numbers. The scaled deviance and the scaled Pearson chi-square likewise have values indicative of overdispersion. Accordingly we pass to negative binomial models in Tables 31 through 37.

Tables 31 and 32 are negative binomial counterparts of Tables 29 and 30, with the same variables.

In general the Poisson and negative binomial models are consistent with one another: coefficients have the same sign and similar magnitudes. In most cases the P-value of coefficients increases, the individual variables are thus less significant, and the overdispersion parameter K, a stand-in for omitted variables, makes a significant contribution to all of the negative binomial models. In Washington the overdispersion parameters are larger than in Minnesota, and fewer variables are significant.

In particular, for the Washington three-legged models the marginally significant variables VCI and RHRI become insignificant as one passes from the Poisson model in Table 29 to the negative binomial model in Table 31. For the Washington four-legged models the variables ADT1, HI, SPDI, RHRI, and RT become less significant from Table 30 to Table 32, with ADT1 and RHRI becoming insignificant. Because it is well-accepted that ADT1 is an important variable, the quality of the data is called into question. The standard error for ADT1 is consistent with both a zero value and a much larger value (comparable to that of Minnesota).

For all intersections in this study, the traffic data are imperfect. In rural sites they typically are based on spot measurements (part of a day at a site along the road near the intersection). Although efforts are made to average the data, with daily, weekly, seasonal, and annual variation taken into account, and with attempts to localize the count to the vicinity of the intersection, the results are not very reliable. Examination of files for both Minnesota and Washington shows that reported ADT for rural intersections is often the same from year to year (with no evidence that new measurements have been made or that paper estimates have been revised). When traffic data are available for all legs, sometimes they do not make sense: the difference in ADT between the two legs of the major road has no obvious relation to the minor road ADT. Efforts were made in this study to correct imperfections in the Minnesota intersection ADT, but because the Washington data were not part of an established data base, no similar efforts could be made with them.

The Minnesota models are thus more trustworthy. Nonetheless, models for both sets of data, and for combined data, are included for comparison purposes. Where there is disagreement between Minnesota and Washington, the relevant variable should receive extra scrutiny and the evidence of Minnesota should be considered less conclusive than otherwise.

Additional Negative Binomial Runs

In Tables 33, 34, and 35 we exhibit additional negative binomial models for Minnesota and combined data.

Table 33 shows combined data for both States with variables that are significant or reasonably close to significant in the "best guess" spirit. For the three-leggeds, compare Table 33 with the last column in Table 31: VCI and ND have been omitted. Both are very insignificant and ND has unexpected sign (more driveways lead to fewer accidents). For the four-leggeds, compare Table 33 with the last column of Table 32: HI is very insignificant and has been omitted; RHRI, although significant, has unexpected sign (the more hazardous the roadside the fewer the accidents) and has also been omitted. The State variable is not significant in any of these runs, but has been retained nonetheless.

Table 34 shows Minnesota negative binomial runs where all but the most significant variables have been omitted. The results for the Minnesota three-leggeds are quite consistent with the Minnesota column of Table 31. For the four-leggeds either horizontal or vertical alignment can serve as significant explanatory variables but not both. Angular deviation DEV from 90° is also strongly significant; the fewer predicted accidents the greater the deviation. The runs in Table 34 keep only the most significant variables. Note that SPDI is not one of them; nor is HAU (but angle is represented by DEV).

Negative Binomial Models for Injury Accidents

We also exhibit negative binomial models for injury accidents (INJACC) in Tables 36 and 37. These tables are comparable to Tables 31 and 32 and show that the same coefficient magnitudes generally are to be found, although with reduced significance.

With respect to the three-legged INJACC runs, the most significant variables besides ADT are Roadside Hazard Rating RHRI and channelization RT (in Minnesota and the Combined data). This is similar to Table 31 where all accidents (TOTACC) are modeled.

With respect to the four-legged INJACC runs, RHRI is again significant but with unexpected sign, and this mirrors the behavior in Table 32 and elsewhere.

Final Intersection Models

The chief idiosyncrasies found in the various models are already present in the Poisson runs (Tables 29 and 30). We list some of these:

- driveways seem to decrease accidents at three-legged intersections;
- roadside hazards seem to decrease accidents at four-legged intersections;
- a major road right turn lane seems to increase accidents at three-legged intersections;
- the angle effect is variable from State to State and from three-legged to four-legged intersections;
- Washington coefficients are somewhat erratic in sign and the coefficient of ADT1 in the four-legged model is rather small relative to that of ADT2; and
- Washington models have lower R^2 values than the Minnesota models, and the combined models are intermediate.

In view of the small size of the Washington State sample (the combined models are generally dominated by the Minnesota data), the non-random and ad hoc character of the Washington intersections (an "opportunity" sample), the lesser quality of some of the collected Washington data (e.g., traffic and angle), and the insignificance of variables of interest (including the State variable), we take the Minnesota models as fundamental.

In particular, we offer the models in Table 35 as our final models for three-legged and four-legged intersections. These models are based exclusively on Minnesota data, and significant variables and marginally significant ones are included where we have allowed greater latitude for the alignment variables in the spirit of a "best guess" approach. In these runs the variables with unexpected signs (ND for the three-leggeds and RHRI for the four-leggeds) have been omitted. These models are the best we have to offer. Their shortcomings become apparent by comparing them with Tables 31 and 32, where more variables are included and both States are represented.

LOGISTIC MODELING

Logistic modeling was done in this study on the Minnesota data to determine whether the probability of a serious accident given that an accident has occurred can be related to highway and intersection variables. The variable INJACC counts the number of injury accidents (i.e., other than property damage only accidents) and includes accidents with non-incapacitating injuries and possible injuries, whereas the focus of the logistic modeling is serious accidents (fatal or injury accidents). All sites with zero accidents were excluded.

Although the results are inconclusive, we present them here since the methodology may be of interest.

Theory

Logistic regression is used to estimate probabilities for binary data or discrete ordinal data. In our case two severity classes are used: serious accidents and other accidents. The probability of an accident being severe is represented as a function of highway and intersection variables of generalized linear type,[46] typically a logistic function of a linear combination of these variables.

A variable Y for each accident is defined as follows:

---

[46] See Roadside Design Guide, American Association of State Highway and Transportation Officials, Washington, D.C., 1988; and Lau, M.Y.-K., and May, A.D., "Accident Prediction Model Development for Unsignalized Intersections: Final Report," University of California at Berkeley, Institute of Transportation Studies, Report No. UCB-ITS-RR-89-12, Berkeley, 1989.

$$Y = \begin{cases} 1 & \textit{if the accident type is fatal or injury} \\ 0 & \textit{if the accident type is non-incapacitating,} \\ & \textit{possible injury, or property damage only} \end{cases}$$

Then P_1 is the probability that Y has the value 1 given the value $x = (x_1,...,x_k)$ of the highway characteristics at the accident site . With the logistic function, the model takes the form

$$P_1 = \frac{\exp(\beta_0 + \Sigma_j \beta_j x_j)}{1 + \exp(\beta_0 + \Sigma_j \beta_j x_j)}$$

This functional form guarantees that P_1 will always be a number between 0 and 1. Since P_1 is the probability that an accident is severe (Y = 1) given the values of x, then $1 - P_1$ is the probability that an accident is not severe (Y = 0). The likelihood function for all the observed severities, derived from the binomial distribution under the assumption that the accidents are independent events, is:

$$l(\beta) = \Pi_i \pi(x_i)^{Y_i}[1 - \pi(x_i)]^{1-Y_i}$$

Here $x_i = (x_{i1}, ..., x_{ik})$ denotes the vector of highway variables at accident no. i and Y_i is 1 or 0 whether accident no. i is serious or not. Under the assumption that the model form is correct, the estimated coefficient vector β is the value of $\beta = (\beta_1, ..., \beta_k)$ that maximizes $l(\beta)$.

A measure of goodness of fit used on this model is the rank correlation (available in the SAS procedure LOGISTIC). All possible accident pairs with distinct severities are formed from the data, and then one calculates:

total = t = the total number of pairs

concordance = nc = the number of pairs for which the model predicts higher probability of a severe accident for the member of the pair that had the more severe accident

discordance = nd = the number of pairs for which the model predicts higher probability of a severe accident for the member of the pair that had the less severe accident

ties = t - nc - nd = the number of pairs with same predicted probability of a severe accident .

Probabilities are grouped into intervals of length .02 and are considered equal if they lie in the same interval. Finally one calculates
c = (nc + 0.5(t - nc - nd))/t.

The statistic c takes values between 0 and 1, and achieves the value .5 on average if a member of each pair is chosen with equal probability. Thus the farther above .5 c is the better the model.

Results

On the 619 Minnesota segments of this study in the time period 1985-89 there were a total of 1,694 accidents, 121 of them serious. The models that result from maximum likelihood techniques showed no significant variables other than commercial ADT percentage T. Horizontal alignment or vertical alignment, but not both, had positive coefficients but the P-values were insignificant (one form of horizontal, not shown here, had a P-value of .306). One typical run yielded equation (5.18):

$$P_1 = \frac{\exp(-3.006 + 0.041T + 0.031VMCC)}{1 + \exp(-3.006 + 0.041T + 0.031VMCC)} . \tag{5.18}$$

The P-values and statistic c are shown below.

TABLE 38. Logistic Model for Serious Accident Probability, Minnesota Segments

PARAMETER	ESTIMATE	P-value
Intercept	-3.0060	0.0001
Percent of commercial vehicles = T	0.0413	0.0310
Crests of Type I rate = VMCC	0.0314	0.5634
Concordance = 53.1%, Discordance = 41.6%, c = 55.8%		

The statistic c differs from 50% by an appreciable but modest amount.

For the three-legged Minnesota intersections, from 1985 to 1989, there were 524 accidents, 34 of them serious. Accident severity does not seem to be significantly affected by the value of the Conflict Index CINDEX. However, as equation (5.19) shows, horizontal alignment (out to 764 feet in each direction) tends to increase the severity, while severity is negatively influenced by vertical alignment (The variable VCEI is a variant of VCI, going out to 764 feet rather than 250 feet). Since there are very few serious accidents, this result contrary to expectation may reflect peculiarities in the sample.

$$P_1 = \frac{\exp(-2.39 - 2.51VCEI + .075HEI)}{1 + \exp(-2.39 - 2.51VCEI + .075HEI)} . \tag{5.19}$$

TABLE 39. Logistic Model for Serious Accident Probability, MN 3-Legged Intersections

PARAMETER	ESTIMATE	P-value
Intercept	-2.39	0.0001
Crest curve rate VCEI (out to ±764')	-2.5099	0.03
Horizontal curvature rate HEI (out to ± 764')	0.0753	0.09
Concordance = 60.4%, Discordance = 33.5%, c = 63.4%		

For the four-legged Minnesota intersections, from 1985 to 1989, there were 494 accidents, 58 of them serious. The model below was developed.

$$P_1 = \frac{\exp(-2.38 + 1.75CINDEX - 0.016DEV + 0.079RHRI)}{1 + \exp(-2.38 + 1.75CINDEX - 0.016DEV + 0.079RHRI)} .$$

Alignments were not at all significant. Instead the conflict index and the angular deviation from 90° were marginally so. Roadside Hazard Rating, although not significant, was also retained.

TABLE 40. Logistic Model for Serious Accident Probability, MN 4-Legged Intersections

PARAMETER	ESTIMATE	P-value
Intercept	-2.38	0.0001
Conflict index CINDEX	1.75	0.10
Angle DEV	-0.016	0.20
Roadside Hazard Rating RHRI	0.079	0.55
Concordance = 57.1%, Discordance = 40.3%, c = 58.4%		

SUMMARY

A variety of modeling techniques – Poisson, negative binomial, extended negative binomial, and logistic – have been applied in this chapter, along with measures of overdispersion, goodness-of-fit, and concordance. In general the Poisson models, negative binomial, and extended negative binomial models give mutually consistent values for regression coefficients. The T_1 statistic indicates that overdispersion is present and thus that negative binomial models are to be preferred. The logistic models are not particularly satisfactory, perhaps because of the relative infrequency of serious accidents and the relatively greater importance of missing variables.

The segment models – our final model is in Table 27 – support the assertion that most of the variables in the study are significant. Some variables that correlate with accidents (e.g., commercial traffic percentage T) are omitted because they are not as significant as competing variables. However, the chief variables – exposure, lane and shoulder width, Roadside Hazard Rating and driveway density, and the alignment variables – are all represented. Differences between States appear to be genuine and are captured by the variable STATE. When we pass to the negative binomial and the extended negative binomial, the coefficient estimates are reapportioned somewhat as overdispersion and localized vertical and horizontal measures make their contribution to the variation in accident counts.

With regard to intersections, the final models are presented in Table 35. Minnesota data are taken as fundamental because the Washington intersection data are non-random and less reliable. Furthermore, the criteria for significance are relaxed so that "best guess" coefficients for alignment design variables can be presented. The effects of number of driveways, Roadside Hazard Rating, the angle variables, and channelization show notable variation between the three-legged intersections and the four-legged. Number of driveways has unexpected sign (negative) on three-leggeds in both States. Roadside Hazard Rating has unexpected sign (negative) on four-leggeds in both States. The acute/obtuse angle variable HAU behaves as expected on four-leggeds but not on three-leggeds, but another angle variable, deviation DEV from 90°, is more significant on four-leggeds. The presence of major road turning lanes increases accidents on three-leggeds but decreases them on four-leggeds. In the final models of Table 35 number of driveways (wrong sign) is omitted from the three-legged intersections, while Roadside Hazard Rating (wrong sign) and right turn lanes (insignificant) are omitted from the four-legged intersections.

Some noteworthy differences also appear between the Minnesota and Washington models, for example, the insignificance of Roadside Hazard Rating in Minnesota segments (due perhaps in part to less variation), the anomalous sign of lane width in Washington segments (perhaps related to design differences), differences in the commercial traffic percentage variable T between the two States, and insignificance of most variables on the Washington three-legged intersections.

The combined segment model (Table 27) and the Minnesota intersection models (Table 35) exhibit the effects of the chief variables, while minimizing anomalies found in some variables and in Washington intersection data.

6. VALIDATION AND FURTHER ANALYSIS

This chapter is devoted to miscellaneous analytical tasks relevant to possible uses of the models:

- Validation tests are performed to measure the predictive efficacy of the leading models. The Minnesota models are tested against Minnesota data from a later time period (1990-1993) on the same segments and intersections. They are also tested against Washington data, and the Washington segment model and the combined segment model are tested on Minnesota data from 1985-89 and 1990-93.

- The relative explanatory value of different groups of variables in the final models (Tables 27 and 35) is assessed by means of the Log-Likelihood R-squared introduced in Chapter 5.

- Scaled residuals (observed accident counts minus predicted mean accident counts divided by estimated standard error) are compared graphically with leading variables to check for systematic trends that might contradict the assumed model form or suggest model refinements.

VALIDATION

Validation Techniques

The chi-square statistic χ^2 provides a rough validation measure. More precisely, use is made here of a concocted χ^2, called χ_c^2, that applies to both the Poisson and the negative binomial distribution:

$$\chi_c^2 = \sum_{i=1}^{N} \frac{(y_i - \hat{y}_i)^2}{\hat{y}_i + K(\hat{y}_i)^2} ,$$

where

y_i	=	observed accident count at site number i
$\hat{y}_i$	=	predicted mean accident count at site number i (according to the model)
K	=	overdispersion parameter of the model
N	=	sample size to which the model is applied.

A more refined approach is to compute the z-score of the concocted statistic χ_c^2. If the null hypothesis that the model is valid is true, it can be shown that the expected value of χ_c^2 is the sample size N and its variance is given by:

$$V(\chi_c^2) = 2N(1+3K) + \sum_{i=1}^{N} \frac{1}{\hat{y}_i(1 + K\hat{y}_i)}.$$

Then the z-score of χ_c^2 is

$$Z = \frac{\chi_c^2 - N}{\sqrt{V(\chi_c^2)}} ,$$

and this statistic is approximately normal.

Also computed are the mean absolute deviation (MAD) and the mean absolute scaled deviation (MASD):

$$MAD = \frac{1}{N}\sum_{i=1}^{N} |y_i - \hat{y}_i|$$

$$MASD = \frac{1}{N}\sum_{i=1}^{N} \frac{|y_i - \hat{y}_i|}{\sqrt{\hat{y}_i + K(\hat{y}_i)^2}} .$$

These are two additional measures of the predictive power of the model.

Minnesota Models versus Later Minnesota Data

Highway Safety Information System data became available during the course of this study for the years 1990-1993 in Minnesota. These data included accident counts, traffic, shoulder widths, lane widths, and speeds for 392 segments (out of the 619 in the original sample), and accident, traffic, and speed data for 365 three-legged intersections (out of the original 389) and 309 four-legged intersections (out of the original 327). The sample sizes for the second time period are smaller because sites with major changes (for example, segments that had changed length) or for which accident counts were not available were omitted. The new values of the highway variables were applied to the leading models and the predicted mean accident counts were compared with actual accident counts to test how the models performed. Variables such as number of driveways, Roadside Hazard Rating, and alignment were not revised for the new data sets. The values of these variables were obtained from photologs for 1985-89 and original construction plans. Updated values were not available, and it was assumed that few changes had occurred.

Table 41 shows the results of applying the Minnesota models from Tables 26 and 35 to the 1990-93 Minnesota data. The first model is an extended negative binomial model for segments with an overdispersion parameter K = .2722, the second and third models are negative binomial models with K = .4811 and .2055, respectively. The critical value $\chi^2_{95\%}$ has been listed for comparison purposes. The segment data fit the model quite well, while the three-legged and four-legged intersections fail to fall within the 95% critical value. If we adopted as null hypotheses that the segments, the three-legged intersections, and the four-legged intersections were drawn from intersections with mean accident counts given by the models, we would reject these hypotheses for the intersections and fail to reject for the segments.

TABLE 41. Validation of Minnesota Models with 1990-1993 Minnesota Data

	Sample size of data used in modeling	Sample size N of validation data	χ_c^2	Critical value $\chi^2_{95\%}$	Z-score of χ_c^2	Mean Abs. Dev. MAD	Mean Abs. Scaled Dev. MASD
MN Segment Model (Table 26)	619	392	304.6	439	-1.94	1.17	0.71
MN 3-legged Intersection Model (Table 35)	389	365	464.1	410	1.94	1.02	0.73
MN 4-legged Intersection Model (Table 35)	327	309 (308*)	386.6 (343.3)	351	2.05 (0.91)	1.28 (1.15)	0.85 (0.83)

* One outlier removed

Nonetheless, in other respects the fits are reasonably good, not only for the segments but also for the intersections, with small mean absolute and absolute scaled deviations. The four-legged intersections improve dramatically when one outlier is removed, an intersection with 51 accidents in 1990-1993.

The objection may be made that accidents in the new time period are correlated with accidents in the old time period, and that the validation sample is not independent of the sample used to derive the model. The effect of this might be to generate predicted accident counts for the new time period similar to those in the old time period, but with the dependency on highway variables not receiving a genuinely independent test. Indeed, the overfitting of the segment data suggests this possibility.

Minnesota Models versus Washington Data

Table 42 below shows validation results when the Minnesota models of Table 41 are applied to the Washington segments and intersections. In this case there is no danger of correlation and the validation data serve as an independent sample.

TABLE 42. Validation of Minnesota Models with 1993-1995 Washington Data

	Sample size of data used in modeling	Sample size N of validation data	χ_c^2	Critical value $\chi^2_{95\%}$	Z-score of χ_c^2	Mean Abs. Dev. MAD	Mean Abs. Scaled Dev. MASD
MN Segment Model (Table 26)	619	712	991.8	775	4.69	1.52	0.85
MN 3-legged Intersection Model (Table 35)	389	181	141.4	213	-1.14	1.17	0.74
MN 4-legged Intersection Model (Table 35)	327	90	188.0	113	5.22	2.68	1.13

Table 42 shows a marked difference between the segment and four-legged Minnesota models and the corresponding Washington data with respect to χ_c^2. The MAD and the MASD look somewhat better. The three-legged model looks relatively good, but it should be recalled that this model has the largest overdispersion parameter (K = .4811 for the three-leggeds versus K = .2722 for the segments and K = .2055 for the four-leggeds). The large overdispersion parameter indicates more unexplained variation than in the other models, and also has the effect of increasing the denominator in χ_c^2 and MASD.

In the case of the segments one explanation of the large z-score of χ_c^2 is the difference in overall accident rate (accidents per million vehicle-miles) between Minnesota and Washington. In Table 43 a comparison is shown of three different ways of applying the Minnesota segment model to the Washington data:

i) the model is used as is;

ii) the predicted mean is taken to be that in i), multiplied by the ratio (1.0228/.6656) of the accident rate (accidents per million vehicle-miles) in Washington to the accident rate in Minnesota; or

iii) the predicted mean is taken to be that in i) multiplied by that factor which gives the maximum likelihood estimate when the predicted mean in i) is used as an offset.

TABLE 43. Validation of Adjusted MN Segment Model with 1993-1995 WA Data

	Sample size of data used in modeling	Sample size N of validation data	χ_c^2	Critical value $\chi^2_{95\%}$	Z-score of χ_c^2	Mean Abs. Dev. MAD	Mean Abs. Scaled Dev. MASD
MN Segment Model (Table 26) without adjustment	619	712	991.8	775	4.69	1.52	0.85
MN Segment Model (Table 26), mult. by 1.0228/.6656	619	712	630.3	775	-1.45	2.07	0.77
MN Segment Model (Table 26) mult. by exp(.0914)	619	712	869.8	775	2.69	1.57	0.81

Table 43 shows that multipliers lead to better fits. An argument in favor of the maximum likelihood multiplier, exp(.0914), is that the ratio of the overall accident rates, 1.0228/.6656 = exp(.430), does not measure the effect of variables besides exposure observation by observation and that differences between the two States in these other variables may already be represented in the model. Method iii) introduces the intercept giving the maximum likelihood fit after the model has accounted for other variables to the extent possible.

Table 43 calls attention to the important question of how a model developed for one or more States in some time period should be applied to other States and/or other time periods. A multiplier such as the ratio of accident rates or the maximum likelihood intercept can be applied, or even one tailored to minimize χ_c^2 or MAD or MASD. The choice of multiplier in general depends on the quantity being optimized. Thus, for example, to obtain a value for χ_c^2 as close as possible to zero in Table 43, a multiplier intermediate between exp(.0914) and exp(.430) might be used.

Washington and Combined Segment Models versus Minnesota Data

Table 44, similar to Table 43, can be generated by applying a Washington State segment model to the Minnesota data. The extended negative binomial model for Washington State from Table 26 is applied to the 1985-1989 Minnesota data with and without a multiplier in Table 44. The ratio of accident rates, .6656/1.0228 = exp(-.430), yields the largest z-score for χ_c^2, while the maximum likelihood intercept, exp(-.2108), yields the z-score closest to zero.

TABLE 44. Validation of Adjusted WA Segment Model with 1985-1989 MN Data

	Sample size of data used in modeling	Sample size N of validation data	χ_c^2	Critical value $\chi^2_{95\%}$	Z-score of χ_c^2	Mean Abs. Dev. MAD	Mean Abs. Scaled Dev. MASD
WA Segment Model (Table 26) without adjustment	712	619	513.2	678	-1.96	1.75	0.72
WA Segment Model (Table 26), mult. by .6656/1.0228	712	619	900.3	678	4.93	1.66	0.88
WA Segment Model (Table 26) mult. by exp(-.2108)	712	619	645.0	678	0.47	1.65	0.78

The combined extended negative binomial model for segments (Table 27) can be applied to the segment data for Minnesota and Washington individually and, as expected, yields z-scores for χ_c^2

close to zero (.926 on Minnesota data, -.0577 on Washington data). When applied to the 1990-1993 Minnesota data (with STATE = 0) it yields the results in Table 45. The accident rate for the 1990-1993 Minnesota segments is .5509 accidents per million vehicle-miles, whereas for the combined Minnesota-Washington data set, used in the modeling, the rate is .8070 accidents per million vehicle-miles.

The data used for validation in Table 45 are not independent of those used in modeling since some of the segments are the same. Nonetheless, it is of interest to note that adjustments may be appropriate when a model is applied to a new time period. Table 45 shows that adjustments that increase likelihood may have variable effects on χ_c^2, MAD, and MASD.

TABLE 45. Validation of Combined Segment Model with 1990-1993 MN Data

	Sample size of data used in modeling	Sample size N of validation data	χ_c^2	Critical value $\chi^2_{95\%}$	Z-score of χ_c^2	Mean Abs. Dev. MAD	Mean Abs. Scaled Dev. MASD
Combined Segment Model (Table 27) without adjustment	1331	392	296.1	439	-2.09	1.20	0.71
Combined Segment Model (Table 27), mult. by .5509/.8070	1331	392	495.0	439	2.09	1.12	0.85
Combined. Segment Model (Table 27) mult. by exp(.0938)	1331	392	273.4	439	-2.62	1.26	0.69

EXPLANATORY VALUE OF FINAL MODELS

One way to assess the explanatory power of models is to examine the coefficient of determination R^2 and see how it changes as one adds variables to the model. In Tables 46 and 47 and Figures 6

TABLE 46. Accident Variation by Groups of Covariates, Final Segment Model

Combined Extended Negative Binomial Model (Table 27)	Log-Likelihood Coefficient of Determination(%)
Randomness	45.20
Exposure	26.81
State	2.63
Lane Width, Shoulder Width	2.33
Roadside Hazard Rating, Driveway Density	1.38
Alignment (DEG{i}, V{i}, GR{i})	1.95
Unexplained	19.70
TOTAL	100.00

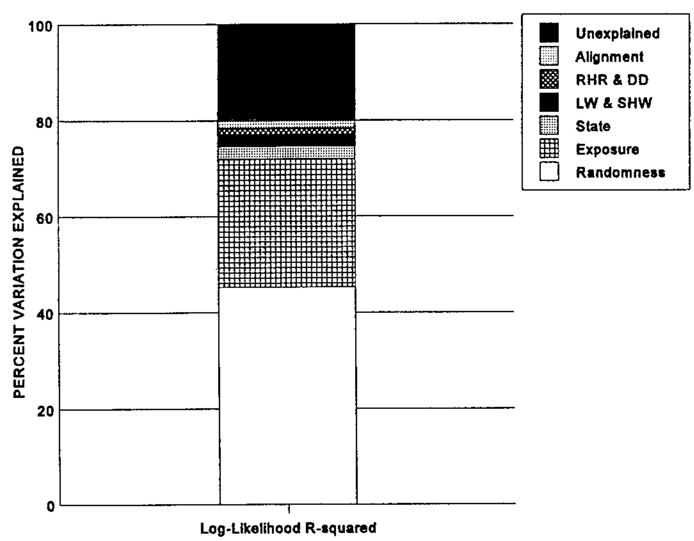


FIGURE 6. GRAPH OF ACCIDENT VARIATION BY GROUPS OF COVARIATES, Final Segment Model

and 7, this is done for three of the models – the combined segment model of Table 27, and the Minnesota three-legged and four-legged models of Table 35. Because all of these models are of

TABLE 47. Accident Variation by Groups of Covariates, Final Intersection Models

Minnesota Intersection Models (Table 35)	Log-Likelihood Coefficient of Determination (%)	
	three-legged	four-legged
Randomness	53.96	59.19
Exposure (ADT1, ADT2)	27.12	27.99
Design (All other variables)	1.78	2.06
Unexplained	17.14	10.76
TOTAL	100.00	100.00

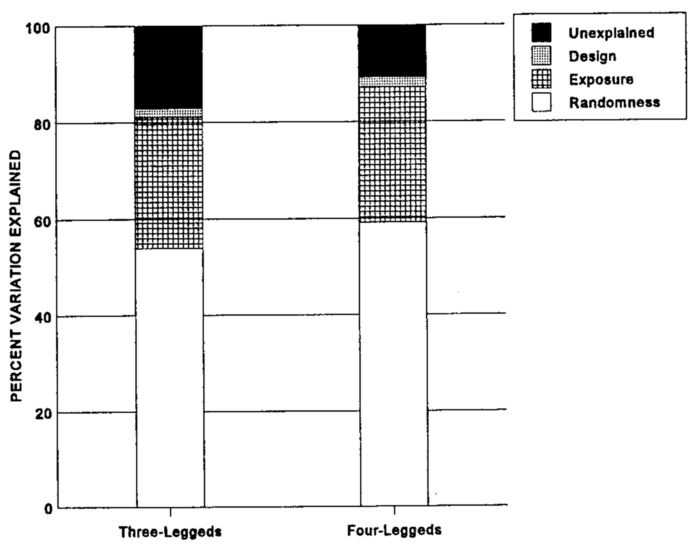


FIGURE 7. GRAPH OF ACCIDENT VARIATION BY GROUPS OF COVARIATES, Final Intersection Models

negative binomial type, we use the Log-Likelihood R-squared proposed by Fridstrøm et al. (1995). With respect to this measure, negative binomial randomness is represented by $1 - P_D^2$. The contribution of other factors is represented by R_D^2 for the first variable when a model with that variable present is used, and then the increment in R_D^2 for each additional variable as it is added to the model. Finally the unexplained portion of variation is $P_D^2 - R_D^2$, where R_D^2 is the R-squared value obtained when all variables are present.

Although the Log-Likelihood R-squared is not the only way to compare explanatory values, it is a reasonable way to do so for negative binomial models (and we presume for their extended negative binomial counterparts). The tables and figures indicate that the portion of mean accident counts explained by variables other than exposure and ADT is small.

CUMULATIVE SCALED RESIDUALS

Figures 8 through 15 below show cumulative scaled residual plots for the extended negative binomial model (combined segments, Table 27) and for negative binomial models (Minnesota three-legged and four-legged intersections, Table 35). The cumulative scaled residuals are plotted against leading explanatory variables. For an explanatory variable x, a plot is made of j versus

$$\sum_{i\,:\,x_i \le j} \frac{y_i - \hat{y}_i}{\sqrt{\hat{y}_i + K(\hat{y}_i)^2}} ,$$

where j runs through the values of x. Each term, a scaled residual, should be approximately unbiased. However, if the sum depends in some regular way on j, then the model may have missed some systematic effects (e.g., quadratic dependency). If there is no systematic effect and the terms are otherwise independent, the expected value of the sum is approximately zero, and its standard deviation is approximately the square root of the number of observations for which $x \le j$. For the segments this means a standard deviation not in excess of $\sqrt{1331} \approx 36.5$ and for the intersections one not in excess of $\sqrt{389} \approx 19.7$ (three-legged) or $\sqrt{327} \approx 18.1$ (four-legged). The cumulative scaled residuals should represent the net distance traveled after each step in a random walk that ends at the sum of the scaled residuals for the entire data set.

For the segments (Figures 8, 9, 10, and 11) the overall sum of the scaled residuals is about -8, for the three-legged intersections (Figures 12 and 13) the sum is about -2, and for the four-legged intersections (Figures 14 and 15) the sum is about +1. Thus the segment graphs and the three-legged graphs should end below the horizontal axis, while the four-legged graphs should end above.

Table 48 summarizes the residual behavior.

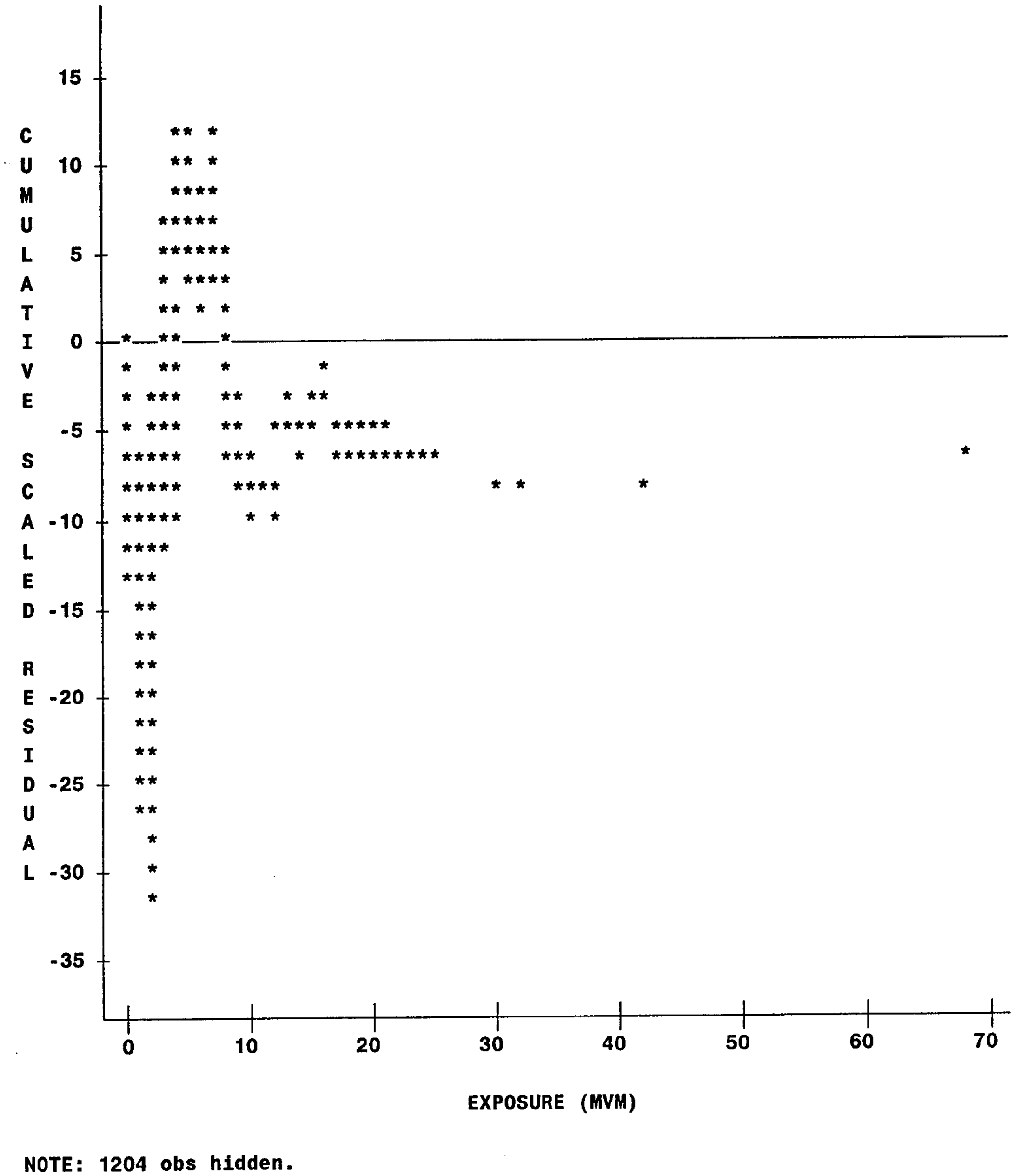


FIGURE 8. CUMULATIVE SCALED RESIDUAL VERSUS EXPOSURE (MVM)

The segment model overpredicts (predicted mean number of accidents higher than actual number) at the low end of exposure. The cumulative scaled residual varies from -32 to +12.

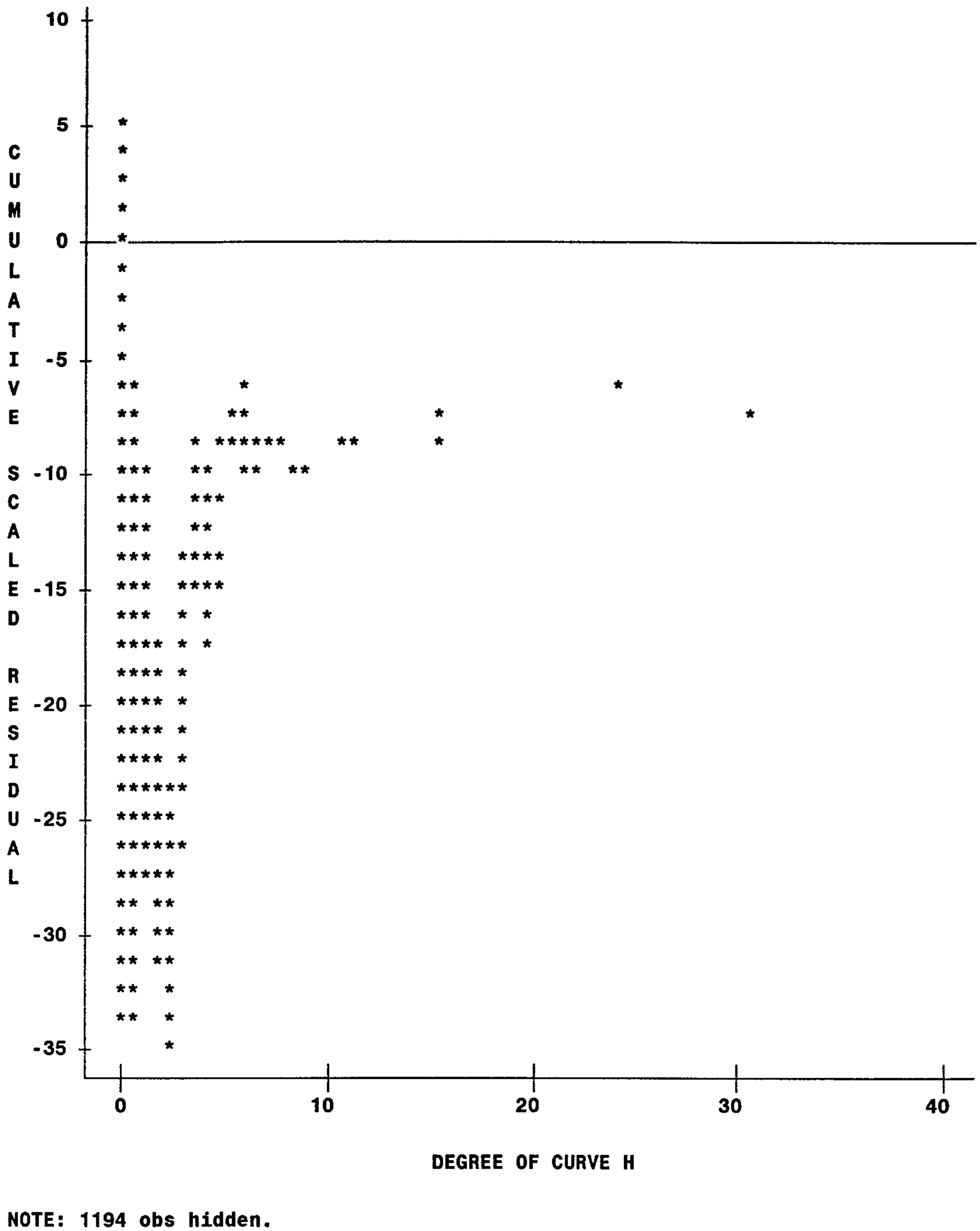


FIGURE 9. CUMULATIVE SCALED RESIDUAL VERSUS DEGREE OF CURVE H

Overprediction occurs on segments without horizontal curves. The cumulative scaled residual varies from -36 to +7.

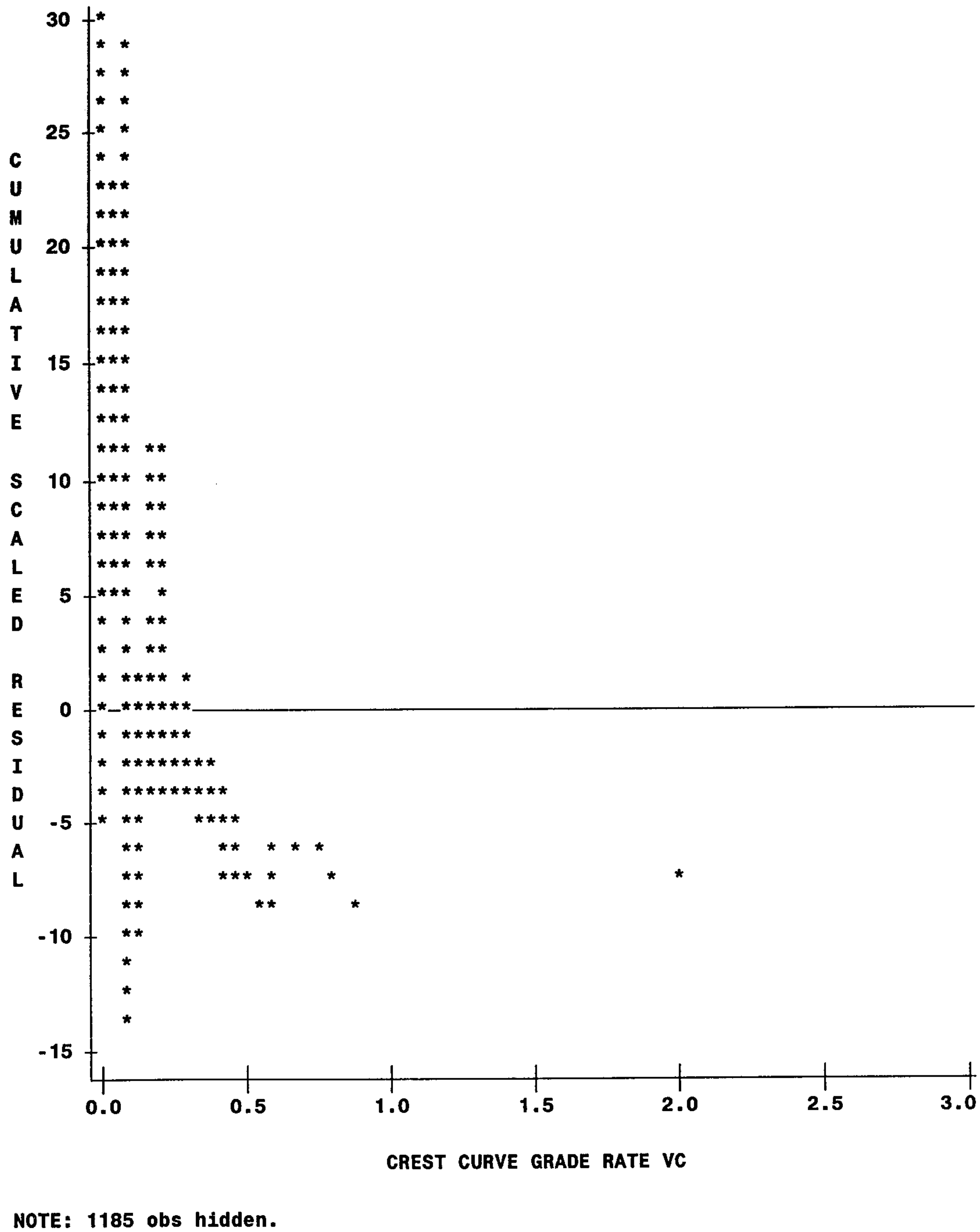


FIGURE 10. CUMULATIVE SCALED RESIDUAL VERSUS CREST CURVE GRADE RATE VC

The segment model underpredicts on segments without crest curves. The cumulative scaled residual varies from -13 to +30 .

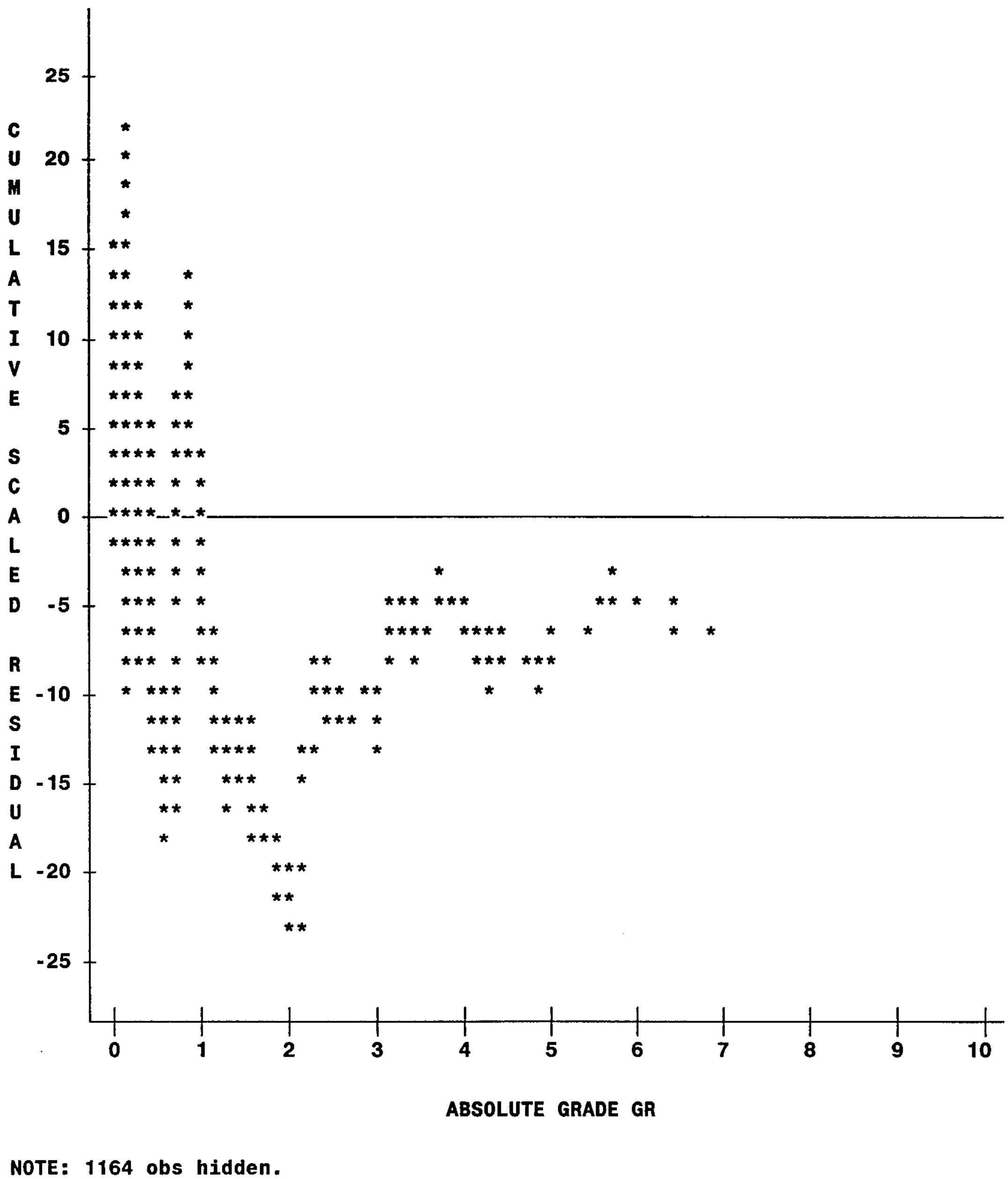


FIGURE 11. CUMULATIVE SCALED RESIDUAL VERSUS ABSOLUTE GRADE GR

The cumulative scaled residual varies from -24 to + 22.

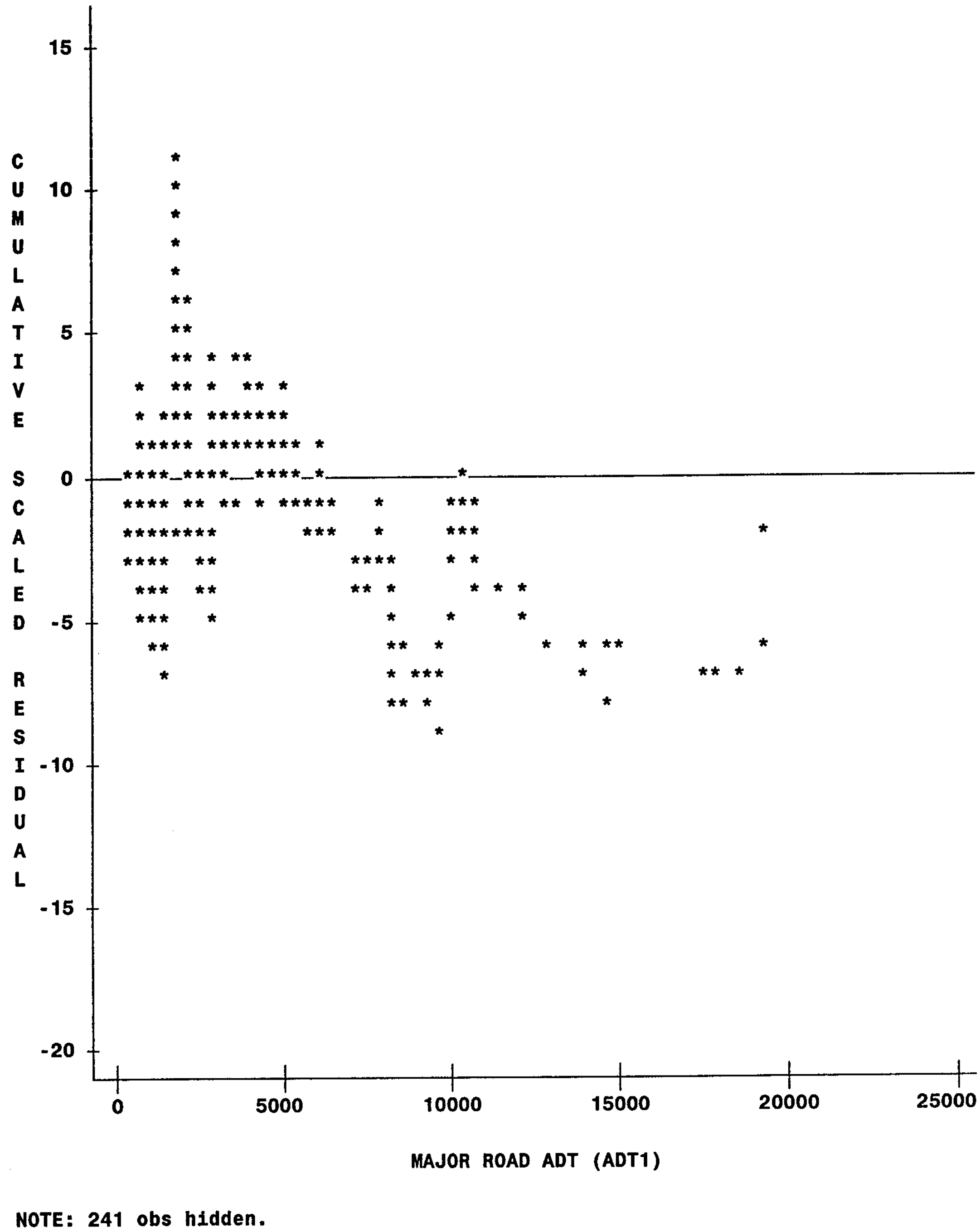


FIGURE 12. CUMULATIVE SCALED RESIDUAL VERSUS ADT1, 3-Legged Intersection.

The cumulative scaled residual varies from -9 to +11 .

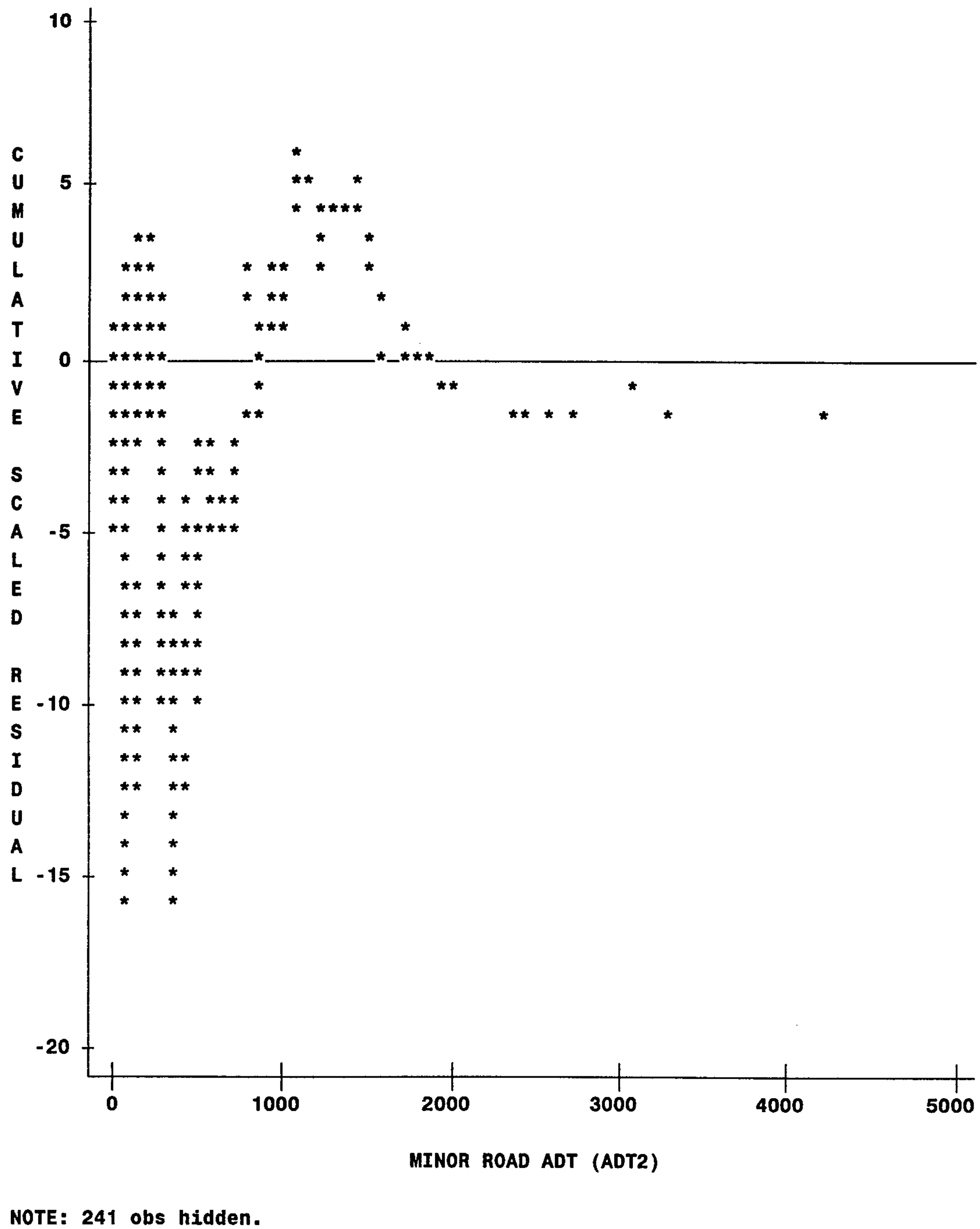
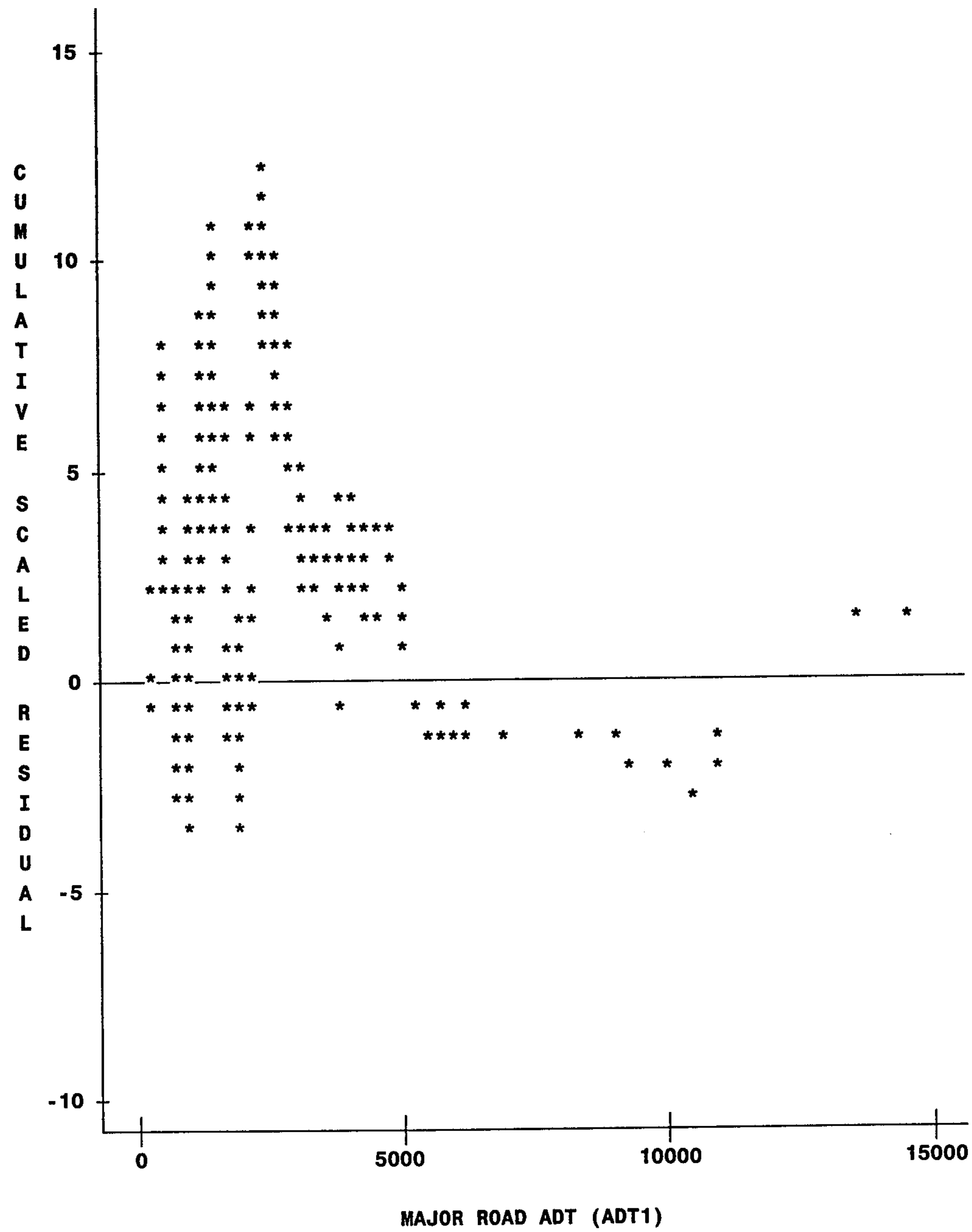


FIGURE 13. CUMULATIVE SCALED RESIDUAL VERSUS ADT2, 3-Legged Intersections

The cumulative scaled residual varies from -16 to +7.



NOTE: 180 obs hidden.

FIGURE 14. CUMULATIVE SCALED RESIDUAL VERSUS ADT1, 4-Legged Intersections

The cumulative scaled residual varies from -4 to +12.

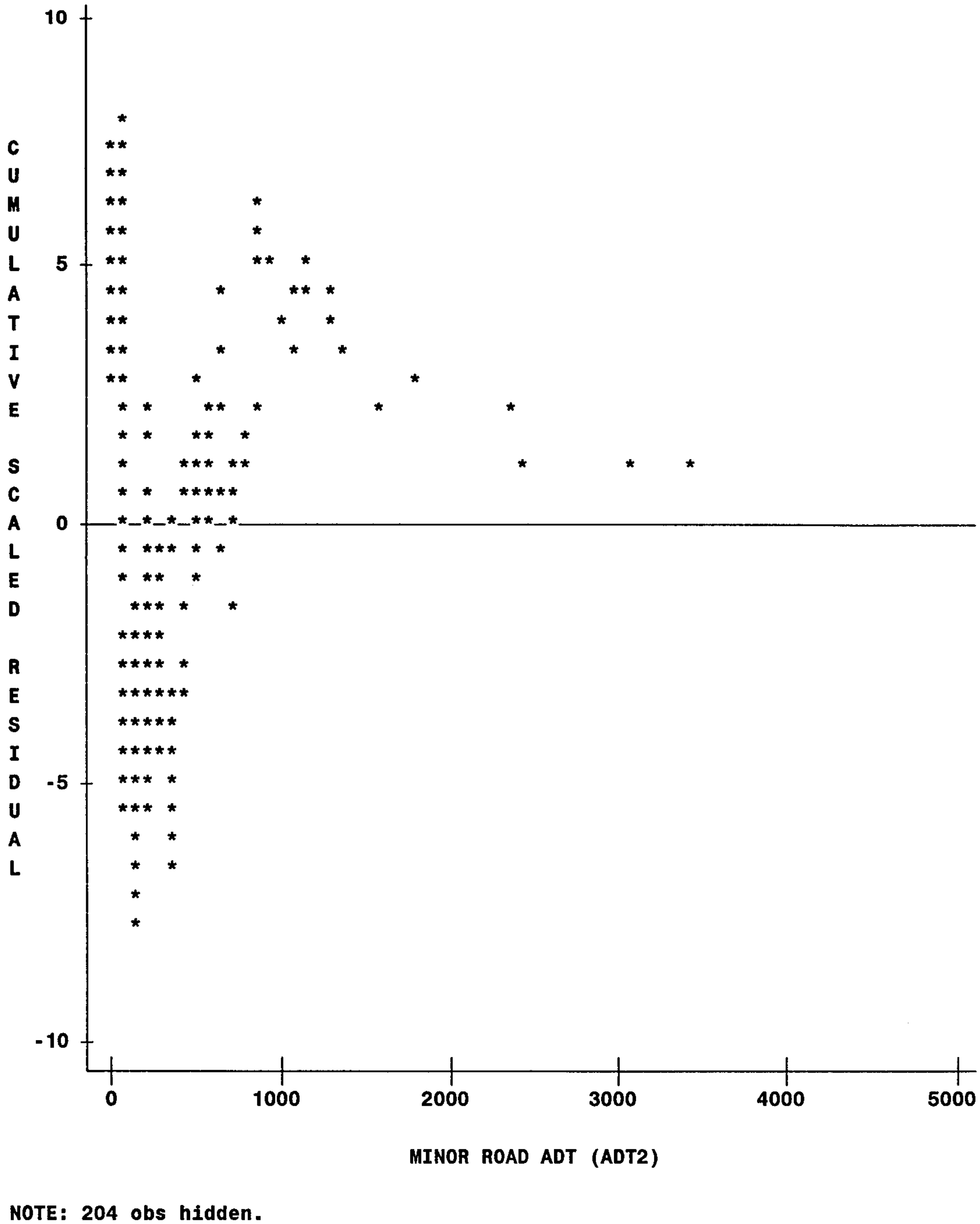


FIGURE 15. CUMULATIVE SCALED RESIDUAL VERSUS ADT2,
4-Legged Intersections

The cumulative scaled residual varies from -8 to +8.

TABLE 48. Cumulative Scaled Residuals versus Increasing Value of Variables, Final Models

	Highway Variable	Range of Cumulative Scaled Residual
1,331 Segments Combined Segment Model (Table 27)	Exposure	-32 to +12
	Degree of Curve H	-36 to +7
	Crest Curve Grade Rate VC	-13 to + 30
	Absolute Grade GR	-24 to + 22
MN 3-leggeds, 389 intersections (Table 35 model)	ADT1	-9 to +11
	ADT2	-16 to +7
MN 4-leggeds, 327 intersections (Table 35 model)	ADT1	-4 to +12
	ADT2	-8 to +8

$\sqrt{1331} \approx 36.5$, $\sqrt{389} \approx 19.7$, $\sqrt{327} \approx 18.1$

Despite the indications of overprediction or underprediction in some regimes in the segment model, which might lead one to develop separate models in different regimes (e.g., one model for low exposure, one for medium exposure, and one for high), the graphs are generally consistent with random walks. In particular the ranges shown in Table 48 above are reasonable. In a random walk, as mentioned, the n-th step or observation on average will take one a distance of less than $\pm(n)^{1/2}$ units from the origin. In addition it is not at all uncommon to stay on one side of zero (above or below) for many steps in succession. Negative binomial models never predict zero values for the dependent variable (in our case numbers of accidents). Thus at low values of highway variables (presumed to be associated with fewer accidents), when the true number of accidents is zero, the negative binomial predicts a positive number and hence must overpredict at least somewhat.

SUMMARY

Validation based on a chi-square statistic χ_c^2, mean absolute deviation MAD, and mean absolute scaled deviation MASD suggests that the models have some predictive power. The Minnesota models behave well on the later Minnesota data (Table 41): the segment model is even underdispersed. This does not constitute a real test, though, since the data sets are dependent so that accidents in the later time period might be expected to correlate well with accidents on the same segment in the earlier time period (and the latter are the basis for the model). A better test is to validate models from one State with data from the other. On Washington data (Table 42) the Minnesota models give small values for MAD and MASD, although the Washington four-legged sample gives somewhat large values. The Washington segment model also gives small values of MAD and MASD on Minnesota data (Table 44). To get a small value of χ_c^2, one adjusts the intercept term of the model to account for a difference in accident experience between the States. Inspection of Tables 43 and 44 shows that the multiplier that makes χ_c^2 smallest for the Minnesota segment model applied to Washington data is approximately 1.35, while the best multiplier for the Washington segment model applied to the Minnesota data is on the order of 0.85. The product of these numbers is approximately 1.0, as is reasonable.

As assessed by the Log-Likelihood R-squared, the explanatory power of the highway variables is rather limited. Exposure and ADT account for about 27% of the variation. For the segments a total of 5.7% of the variation is accounted for by other highway variables (while STATE accounts for 2.6%). For the three-legged intersections, all highway variables other than ADT account for only 1.8% (perhaps in part because of the large overdispersion parameter in the three-legged model), while for the four-leggeds the other variables account for 2.1%. See Tables 46 and 47, and Figures 6 and 7.

Although the cumulative scaled residual graphs for the segments suggest some differences in regimes, the graphs in Figures 8 through 15 are generally consistent with the model forms in Tables 27 and 35. Different models applied when some of the highway variables are confined to subsets of their full range (first quartile, second quartile, etc.) might yield better fits, but if a single overarching model is wanted for each of the three classes of data, the final models in Tables 27 and 35 are plausible candidates (with adjustments for different States and times).

7. CONCLUSIONS

We present the final models of this study in the form of equations and make a few remarks about their significance. Appendix 2 gives the equations in metric form.

The final models proposed in this study are the following:

I. Segments of two-lane rural roads (Table 27)

Extended Negative Binomial Model with K = .306

$$\hat{y} =$$
$$\begin{aligned} & \boldsymbol{EXPO} \times \exp(0.641) \\ & \times \exp(-0.0846\boldsymbol{LW} - 0.0591\boldsymbol{SHW} + 0.0668\boldsymbol{RHR} + 0.0084\boldsymbol{DD} + 0.139\boldsymbol{STATE}) \\ & \times (\textstyle\sum_i \boldsymbol{WH}\{i\} \exp(0.0450\boldsymbol{DEG}\{i\})) \\ & \times (\textstyle\sum_j \boldsymbol{WV}\{j\} \exp(0.465\boldsymbol{V}\{j\})) \\ & \times (\textstyle\sum_k \boldsymbol{WG}\{k\} \exp(0.105\boldsymbol{GR}\{k\})) \end{aligned}$$

where

$\hat{y}$ = predicted mean number of non-intersection accidents on the segment
EXPO = traffic exposure in millions of vehicle-miles
LW = lane width in feet
SHW = average of left and right shoulder widths in feet
RHR = average Roadside Hazard Rating along segment
DD = driveway density in driveways per mile
STATE = 0 for Minnesota, 1 for Washington
DEG{i} = degree of curve in degrees per hundred feet of the i-th horizontal curve that overlaps the segment
WH{i} = fraction of the total segment length occupied by the i-th horizontal curve
V{j} = absolute change in grade in percent per hundred feet of the j-th vertical crest curve that overlaps the segment
WV{j} = the fraction of the total segment length occupied by the j-th vertical crest curve
GR{k} = absolute grade in percent of the k-th uniform grade section that overlaps the segment
WG{k} = fraction of the total segment length occupied by the k-th uniform grade section.

NOTE: Each set of weights WH{i}, WV{j}, and WG{k} separately must sum to 1. To ensure this, usually it is necessary to insert one artificial horizontal curve with DEG = 0, one artificial crest with V = 0, and one artificial straightaway with GR = 0, each one having whatever weight is needed to make the sum equal 1.

II. Three-legged intersections of two-lane rural roads, stop-controlled on the minor road (Table 35)

Negative Binomial Model with K = .481

$$\hat{y} =$$

$$\boldsymbol{NUMBER\ OF\ YEARS \times (ADT1)^{.805} \times (ADT2)^{.504} \times \exp(-13.0)}$$
$$\boldsymbol{\times \exp(0.0339HI + 0.290VCI + 0.0285SPDI)}$$
$$\boldsymbol{\times \exp(0.173RHRI + 0.267RT + 0.0045HAU)}$$

where the variables are:

$\hat{y}$ = predicted mean number of intersection or intersection-related accidents within 250 feet of the intersection center
ADT1 = average two-way major road traffic in vehicles per day
ADT2 = average two-way minor road traffic in vehicles per day
HI = sum of degree of curve in degrees per hundred feet for each horizontal curve on major road any portion of which is within 250 feet of the intersection center, divided by the number of such curves
VCI = sum of absolute change of grade in percent per hundred feet for each crest curve (incoming signed grade larger than outgoing signed grade) on major road any portion of which is within 250 feet of the intersection center, divided by the number of such curves
SPDI = average posted speed in miles per hour on major road in vicinity of the intersection
RHRI = average Roadside Hazard Rating within 250 feet of intersection center along major road
RT = 0 if no right turn lane on major road, 1 if right turn lane exists on major road
HAU = angle in degrees between increasing direction of major road and minor road minus 90 degree, multiplied by 1 if minor road is to right or by -1 if minor road is to left.

III. Four-legged intersections of two-lane rural roads, stop-controlled on the minor road (Table 35)

Negative Binomial Model with K = .205

$$\hat{y} =$$

$$NUMBER\ OF\ YEARS \times (ADT1)^{.603} \times (ADT2)^{.609} \times \exp(-10.4) \times \exp(0.0449HI + 0.289VCI + 0.0187SPDI + 0.124ND - 0.0049HAU)$$

where the variables are:

$\hat{y}$ = predicted mean number of intersection or intersection-related accidents within 250 feet of the intersection center
ADT1 = average two-way major road traffic in vehicles per day
ADT2 = average two-way minor road traffic in vehicles per day
HI = sum of degree of curve in degrees per hundred feet for each horizontal curve on major road any portion of which is within 250 feet of the intersection center, divided by the number of such curves
VCI = sum of absolute change of grade in percent per hundred feet for each crest curve (incoming signed grade larger than outgoing signed grade) on major road any portion of which is within 250 feet of the intersection center, divided by the number of such curves
SPDI = average posted speed in miles per hour on major road in vicinity of the intersection
ND = number of driveways within 250 feet of the intersection center along major road
HAU = angle in degrees between increasing direction of major road and right leg of minor road minus angle in degrees between increasing direction of major road and left leg of minor road.

These models yield the Accident Reduction Factors shown in Table 49 below. Recall that the Accident Reduction Factor is the percentage decrease in mean predicted accident count when a variable is increased by one unit, all other variables being held fixed. A negative value signifies that accidents increase by that percentage when the variable is increased by one unit.

TABLE 49. Accident Reduction Factors for the Final Models

Segment Model (Table 27)		3-Legged Intersection Model (Table 35)		4-Legged Intersection Model (Table 35)	
LW	+8.1%				
SHW	+5.7%				
RHR	-6.9%	RHRI	-18.8%		
DD	-0.84%			ND	-13.1%
DEG	-4.6%	HI	-3.4%	HI	-4.6%
V	-59.2%	VCI	-33.7%	VCI	-33.4%
GR	-11.0%				
		HAU	-0.5%	HAU	+.5%

The Accident Reduction Factors for DD and ND are roughly comparable. Since DD = $ND \times 5280 \div 500$, the coefficient 0.0084 of DD in the segment model (Table 27) translates into a coefficient 0.0887 of ND and an Accident Reduction Factor of -9.3% for an intersection model, as compared with -13.1% in the actual four-legged intersection model (Tables 35 and 49).

The ultimate use of models such as these is to aid the highway designer to improve highway safety and to determine what design measures will do this most effectively. The coefficients proposed for each of the models – in Tables 27 and 35 and in the equations above – are directly translatable into predicted accident counts and Accident Reduction Factors. Even if the models considered here were taken to be definitive, each coefficient has an estimated standard deviation or standard error (shown in Tables 27 through 35), and there is no reason to believe that the estimated coefficients are known to much greater accuracy than one standard deviation. For a normal random variable about 68% of measured values lie within one standard deviation of the mean. In addition there are numerous uncertainties that cannot be quantified in the highway variables. Variables such as ADT are crude averages over time, and some variables are incorrect for unknown causes (new construction without plans to confirm the change, data entry errors in one of the multiple data bases from which the data are obtained, inaccuracies in location of accidents, mileposts, alignments, etc.).

One informal way to estimate the error in a coefficient is to examine alternative models and note how coefficients vary from model to model. As well as referring to the literature for models obtained by other investigators, one may compare the different models in this study in Tables 21 through 37. Although there is some stability in coefficients as one passes from Poisson to negative binomial to extended negative binomial, there is less as one passes from one State to another, or from all accidents to injury accidents.

Of great importance for the practical utility of models such as the ones presented here is the issue of how to adapt them to different States and regions and/or different time epochs. In general what is needed is a multiplier that can be applied to a standard model to adjust it to a different State or region (for example, New England versus the Great Plains) and/or a different era (1999 versus 2001-2005), to circumstances in which drivers, vehicles, law enforcement, and demographics may differ from those under which the standard model was developed. Engineering judgment together with historical data from different States and eras can be used to develop multipliers. Alternatively, a small recent sample of accidents in a region can be compared with predictions from the standard model and an adjustment factor derived from the sample. Yet another approach is the Empirical Bayesian one: combine past data on a particular segment or intersection with a standard model of negative binomial type as discussed in Hauer et al. (1988).

Although the segment model developed here summarizes data from two reasonably diverse States (and two epochs), the intersection models are based on Minnesota alone. In Table 42 they have only partial success when applied to Washington State. Moreover, the design variables (e.g., Roadside Hazard Rating, number of driveways, channelization, and intersection angle) behave in unexpected ways as one moves from three-legged intersections to four-legged ones. These peculiarities, as well as the relatively high accident rates at intersections, suggest that intersection studies should continue as a highway safety research priority.

REFERENCES

Agresti, A., Analysis of Ordinal Categorical Data, Wiley Publication, 1984.

Bared, J.G., and Lum, H., Safety Evaluation of Intersections Design Elements (Pilot Study), Transportation Research Board Conference Presentation, Washington, D.C., 1992.

Ben-Akiva, M., and Lerman, S.R., Discrete Choice Analysis - Theory and Application to Travel Demand, MIT Press, Cambridge, Massachusetts, 1985.

Cirillo, J.A., "Safety Effectiveness of Highway Design Features, Volume I: Access Control," FHWA-RD-91-044, Federal Highway Administration, Washington, D.C., 1992.

Consul, P.C., and Famoye, F., "Generalized Poisson Regression Model," Communication Statistics - Theory Methodology, 21(1): 89-109, 1992.

David, N., and Norman, J.R., "Motor Vehicle Accidents in Relation to Geometric and Traffic Features of Highway Intersection," Volume II, FHWA-RD-76-129, Federal Highway Administration, Washington, D.C., 1975.

Dean, C., and Lawless, J.F., "Tests for Detecting Overdispersion in Poisson Regression Models," Journal of the American Statistical Association, 84 (406): 467-472, 1989.

Designing Safer Roads: Practices for Resurfacing, Restoration, and Rehabilitation, Transportation Research Board, SR 214, Washington, D.C., 1986.

Durth, W., "Risk Analysis in Highway Engineering", Proceedings of Strategic Highway Research Program and Traffic Safety on Two Continents, VTI Report 351A, Gothenburg, Sweden, 1989.

Fee, J.A., Beatty, R.L., Dietz, S.K., Kaufman, S.F., and Yates, S.F., "Interstate System Accident Research Study 1," U.S. Government Printing Office, Washington, D.C., 1970.

Fridstrøm, L., Ifver, J., Ingebrigsten, S., Kulmala, R., and Thomsen L. K., "Measuring the Contribution of Randomness, Exposure, Weather, and Daylight to the Variation in the Road Accident Counts," Accident Analysis and Prevention, 27(1): 1-20, 1995.

Garber, N.J., and Srinivasan, R., "Risk Assessment of Elderly Drivers at Intersections: Statistical Modeling," Transportation Research Board, TRR 1325, Washington, D.C., 1991.

Graham, J.L., and Harwood, D.W., Effectiveness of Clear Recovery Zones, National Cooperative Highway Research Program, Report 247, Transportation Research Board, National Research Council, Washington, D.C., 1982.

Hakkert, A.-S., and Mahalel, D., "Estimating the Number of Accidents at Intersections from a Knowledge of the Traffic Flows on the Approaches," Accident Analysis and Prevention, 10: 69-79, 1978.

Harwood, D.W., Mason, J.M., and Graham, J.L., "Conceptual Plan for an Interactive Highway Safety Design Model," FHWA-RD-93-122, Federal Highway Administration, Washington, D.C., 1994.

Hauer, E., "On Two Uses of Exposure," Paper Presented at the Transportation Research Board Annual Meeting, Washington, D.C., 1994.

Hauer E., Ng, J. C.N., and Lovell, J., "Estimation of Safety at Signalized Intersections," Transportation Research Board, TRR 1185, Washington, D.C., 1988.

Highway Capacity Manual, Transportation Research Board Special Report 209, Third Edition, 1994.

Hosmer, D.W., and Lemeshow, S., Applied Logistic Regression, Wiley-Interscience Publication, 1989.

Roy Jorgensen Associates, Inc. "Cost and Safety Effectiveness of Highway Design Elements", National Cooperative Highway Research Program, Report 197, Transportation Research Board, National Research Council, Washington, D.C., 1978.

King, G.F., and Goldblatt, R.B., "Relationship of Accident Patterns to Type of Intersection Control," Transportation Research Board, TRR 540, Washington, D.C., 1975.

Kuciemba, S.R, and Cirillo, J.A., "Safety Effectiveness of Highway Design Features, Volume V: Intersections," FHWA-RD-91-048, Federal Highway Administration, Washington, D.C., 1992.

Kulmala, R., "Safety at Three- and Four-Arm Junctions: Development and Application of Accident Prediction Models," VTT Publication 233, Technical Research Centre of Finland, Espoo, 1995.

Kulmala, R., and Roine, M., "Accident Prediction Models for Two-Lane Roads in Finland," Technical Research Centre of Finland, Traffic Safety Theory and Research Methods, Session 4, Statistical Analysis and Models, Amsterdam, 1988.

Kuo-Liang, T., and Chin-Lung, Y., "A Predictive Accident Model for Two-Lane Rural Highways in Taiwan," Republic of China, Traffic Safety Theory and Research Methods, Session 4, Statistical Analysis and Models, Amsterdam, 1988.

Kvalseth, T.O., "Cautionary Note About R^2," The American Statistician, American Statistical Association, 39(4): 279-285, 1985.

Lau, M.Y.-K., and May, A.D., "Accident Prediction Model Development for Unsignalized Intersections: Final Report," University of California at Berkeley, Institute of Transportation Studies, Report UCB-ITS-RR-89-12, Berkeley, California, 1989.

Lau, M.Y.-K., and May, A.D., "Accident Prediction Model Development: Signalized Intersections, Final Report," University of California at Berkeley, Institute of Transportation Studies, Report UCB-ITS-RR-88-7, Berkeley, California, 1988.

Lawless, J.F., "Negative Binomial and Mixed Poisson Regression," The Canadian Journal of Statistics, 15(3): 209-225, 1987.

Leong, W.H.J., "Relationship Between Accidents and Traffic Volumes at Urban Intersections," Journal of Australian Road Research Board, 5(3): 72-90, 1973.

Lovell, J., and Hauer, E., "The Safety Effect of Conversion to All-Way Stop Control," Transportation Research Board, TRR 1068, Washington, D.C., 1986.

Luyanda, F., Smith R.W., Padron, M., Resto, P., Gutierrez, J., and Fernandez, L., "Multivariate Statistical Analysis of Highway Accident and Highway Conditions," University of Puerto Rico, Mayaguez School of Engineering Research Center, Report DOT-RSPA-DMA-50/84/9, Puerto Rico, 1983.

Mak, K.K., "Effect of Bridge Width on Safety," State of the Art Report 6, Relationship Between Safety and Key Highway Features – A Synthesis of Prior Research, Transportation Research Board, National Research Council, Washington, D.C., 1987.

Manual of Uniform Traffic Control Devices for Streets and Highways, U.S. Department of Transportation, Federal Highway Administration, Washington, D.C., 1988.

McDonald, J.W., "Relationship Between Number of Accidents and Traffic Volumes at Divided Highway Intersections," National Research Council, Highway Research Board Report 74: 7-17, Washington, D.C., 1966.

McGee, H.W., Hughes, W.E., and Daily, H., "Effect of Highway Standards on Safety," National Cooperative Highway Research Program, Report 374, Transportation Research Board, National Research Council, Washington, D.C., 1995.

McGuirk, W.W.; and Staterly, G.T., Jr., "Evaluation of Factors Influencing Driveway Accidents," Transportation Research Board, TRR 601, Washington, D.C., 1976.

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

Miaou, S.-P., and Lum, H., "Modeling Vehicle Accidents and Highway Geometric Design Relationships," Accident Analysis and Prevention, 25(6): 689-709, 1993.

Miaou, S.-P., Hu, P.S., Wright, T., Davis, S.C., and Rathi, A.K., "Development of Relationship between Truck Accidents and Geometric Design: Phase I," FHWA-RD-91-124, Federal Highway Administration, Washington, D.C., 1993.

Moore, W.L. Jr., and Humphreys, J.B., "Sight Distance Obstructions on Private Property at Urban Intersections," Transportation Research Board, TRR 541, Washington, D.C., 1975.

Naclerio, M.T., Kruger, P, and May, A.D., "Accident Prediction Models for Signalized and Unsignalized Intersections. Addendum," University of California at Berkeley, Institute of Transportation Studies, Report UCB-ITS-RR-89-17, Berkeley, California, 1989.

Persaud, B.N., "Safety Migration, the Influence of Traffic Volumes, and Other Issues in Evaluating Safety Effectiveness - Some Findings on Conversion of Intersections to Multiway Stop Control," Transportation Research Board, TRR 1068, Washington, D.C., 1986.

A Policy on Geometric Design of Highways and Streets, American Association of State Highway and Transportation Officials (AASHTO), Washington, D.C., 1994.

Polus, A., "The Relationship of Overall Geometric Characteristics to the Safety Level of Rural Highways," Traffic Quarterly, 34(4), 1980.

Roadside Design Guide, American Association of State Highway and Transportation Officials, Washington, D.C., 1988.

Rumar, K., "The Role of Perceptual and Cognitive Filters in Observed Behavior," in Human Behavior and Traffic Safety, Evans, L. and Sching, R.C. eds., Plenum Press, pp. 151-170, New York, 1985.

SAS Technical Report P243, SAS/STAT Software: The GENMOD Procedure, Release 6.09, SAS Institute Inc., Cary, North Carolina, 1993.

Shankar, V., Mannering, F., and Barfield, W., "Effect of Roadway Geometrics and Environmental Factors on Rural Freeway Accident Frequencies," Accident Analysis and Prevention, 27(3): 371-389, 1995.

Smith, Murray, "Neural Networks for Statistical Modeling," Van Nostrand, New York, 1993.

Surti, V.H. , "Accident Exposure for At-Grade Intersections," Traffic Engineering, 36 (3): 26-27 and 53, 1965.

Tessmer, J.M., “Rural and Urban Crashes: A Comparative Analysis,” Technical Report DOT-HS-808-450, National Highway Traffic Safety Administration, United States Department of Transportation, Washington, D.C., 1996.

Twomey, J.M., Heckman, M.L., and Hayward, J.C., "Safety Effectiveness of Highway Design Features - INTERCHANGES," Volume IV, FHWA-RD-91-047, Federal Highway Administration, Washington, D.C., 1992.

Urbanik, T., II, Hinshaw, W., and Fambro, D.B., "Safety Effects of Limited Sight Distance on Crest Vertical Curves," Transportation Research Board, TRR 1208, Washington, D.C., 1989.

Vogt, A., and Bared, J.G., “Accident Models for Two-Lane Rural Segments and Intersections,” Transportation Research Board, TRR 1635, Washington, D.C., 1998, to appear.

Wedderburn, C.W.M., "Quasi-likelihood Functions, Generalized Linear Models, and the Gauss-Newton Method," Biometrika, 54: 439-447, 1974.

Wu, Y.S., "Effect of Clear Vision Right-of Way on Traffic Accidents at Urban and Rural Signalized Intersections," Report TSD-228-73, Department of State Highways, Michigan, 1973.

Zegeer, C.V., Twomey, J.M., Heckman, M.L., and Hayward, J.C., "Safety Effectiveness of Highway Design Features, Volume II: Alignment," FHWA-RD-91-045, Federal Highway Administration, Washington, D.C., 1992.

Zegeer, C.V., and Council, F.M., "Safety Effectiveness of Highway Design Features – Volume III: Cross Sections," FHWA-RD-91-046, Federal Highway Administration, Washington, D.C., 1992.

Zegeer, C.V., Hummer, J., Reinfurt, D., Herf, L., and Hunter, W., "Safety Cost-Effectiveness of Incremental Changes in Cross-Section Design – Informational Guide ," FHWA-RD-87-094, Federal Highway Administration, Washington, D.C., 1987.

Zegeer, C.V., Hummer, J., Reinfurt, D., Herf, L., and Hunter, W., "Safety Effects of Cross-Section Design for Two-Lane Roads," FHWA-RD-87-008, Federal Highway Administration, Washington, D.C., 1986.

Zegeer, C.V., Stewart, R., Reinfurt, D., Council, F., Neuman, T., Hamilton, E., Miller, T., and Hunter, W., "Cost Effective Geometric Improvements for Safety Upgrading of Horizontal Curves," FHWA-RD-90-021, Federal Highway Administration, Washington, D.C., 1991.

APPENDIX 1 - STATISTICS ON THE MINNESOTA POPULATIONS

Percentage of Accidents versus Accident and Vehicle Variables for Three-legged and Four-legged Intersections and Segments (Minnesota two-lane rural roads, 1985-1989)

VARIABLE	THREE-LEGGED INTERSECTIONS	FOUR-LEGGED INTERSECTIONS	SEGMENTS
SEVERITY			
FATAL	1.6	2.8	2.1
INJURY	6.0	8.3	5.4
NON-INCAP INJURY	17.2	18.5	15.0
POSSIBLE INJURY	19.1	17.9	14.1
PROPERTY DAMAGE	56.2	52.4	63.5
ACCIDENT TYPE			
REAR END	31.8	19.3	13.2
HEAD ON	3.7	1.8	6.4
RIGHT ANGLE	19.4	44.8	10.8
SIDESWIPE	9.4	9.1	5.0
SIDESWIPE OPP.	2.6	2.0	5.4
RAN OFF ROAD R.	6.5	3.8	17.6
RAN OFF ROAD L.	3.7	2.6	13.9
RIGHT TURN	0.9	0.5	0.1
LEFT TURN	8.3	6.7	1.5
OTHER	11.3	7.6	19.4
UNKNOWN	24	1.8	6.9

VARIABLE	THREE-LEGGED INTERSECTIONS	FOUR-LEGGED INTERSECTIONS	SEGMENTS
LIGHT CONDITION			
DAYLIGHT	64.5	74.3	46.8
DAWN	2.4	1.7	3.7
DUSK	2.8	3.1	3.6
DARK ST. L. ON	9.9	.3.6	3.0
DARK ST. L. OFF	0.4	0.4	1.3
DARK NO ST. L.	19.6	16.8	40.9
OTHER	0.0	0.0	0.1
UNKNOWN	0.3	0.2	0.6
ROAD SURFACE CONDITION			
DRY	71.6	76.4	65.6
WET	15.3	12.1	9.9
SNOW/SLUSH	3.3	2.6	5.8
ICE/PACK SNOW	9.0	8.0	17.2
MUDDY	0.0	0.0	0.1
DEBRIS	0.1	0.0	0.0
OILY	0.0	0.0	0.0
OTHER	0.4	0.5	0.7
UNKNOWN	0.3	0.2	0.7

VARIABLE	THREE-LEGGED INTERSECTIONS	FOUR-LEGGED INTERSECTIONS	SEGMENTS
PHYSICAL CONDITION			
NORMAL	82.6	87.1	74.8
UNDER INFLUENCE	6.7	4.2	6.8
DRINKING	4.1	3.6	6.4
USING DRUGS	0.1	0.1	0.1
ASLEEP	1.4	0.6	3.5
FATIGUED	0.3	0.2	1.1
ILL	0.3	0.1	0.4
HANDICAP	0.1	0.1	0.1
OTHER	0.6	0.4	0.7
UNKNOWN	4.0	3.4	6.1
TYPE OF VEHICLE			
AUTO.	72.3	69.6	70.4
PICKUP/VAN	17.2	21.1	19.7
TRUCKS	5.4	3.4	6.5
MOTORCYCLES	2.0	1.0	1.8
EMERG. VEHICLE	0.3	0.2	0.3
BUS/MOTOR HOME	0.3	0.4	0.5
BIKER/PED.	0.1	0.0	0.0
OTHER EQUIP.	0.3	0.1	0.5

949 three-legged intersections and 1,440 accidents
1,156 four-legged intersections and 2,028 accidents
3,308 segments and 8,083 accidents

Average Daily Traffic
(Minnesota two-lane rural road segments, 1985-1989)

AVERAGE DAILY TRAFFIC	NO. OF SEGMENTS	TOTAL SEGMENT LENGTH (miles)	TOTAL NO. OF ACCIDENTS (non-intersection)
51-500	178	400.54	194
501-1000	563	910.34	832
1001-1500	546	619.90	875
1501-2000	511	480.53	962
2001-3000	676	597.78	1821
3001-5000	537	372.40	1551
5001-25000	297	172.21	1848
TOTAL	3,308	3,553.70	8,083

Commercial Traffic Percentage
(Minnesota two-lane rural road segments, 1985-1989)

COMMERCIAL TRAFFIC PERCENTAGE	NO. OF SEGMENTS	TOTAL SEGMENT LENGTH (miles)	TOTAL NO. OF ACCIDENTS (non-intersection)
0-5.0	334	288.83	1382
5.1-10.0	1336	1308.50	3602
10.1-15.0	1053	1192.15	2054
15.1-20.0	425	513.84	799
20.1-30.0	160	250.38	246
TOTAL	3,308	3,553.70	8,083

Shoulder Width
(Minnesota two-lane rural road segments, 1985-1989)

SHOULDER WIDTH (feet)	NO. OF SEGMENTS	TOTAL SEGMENT LENGTH (miles)	TOTAL NO. OF ACCIDENTS (non-intersection)
0-3	471	830.04	1255
3-5	393	566.10	1159
5-7	522	598.39	1354
7-8	959	847.09	2623
8-10	922	696.98	1667
10-12	41	15.10	25
TOTAL	3,308	3,553.70	8,083

Lane Width
(Minnesota two-lane rural road segments, 1985-1989)

LANE WIDTH (feet)	NO. OF SEGMENTS	TOTAL SEGMENT LENGTH (miles)	TOTAL NO. OF ACCIDENTS (non-intersection)
9	1	0.58	5
10	92	117.96	371
11	254	279.40	982
11.5	5	6.71	3
12	2956	3149.06	6722
TOTAL	3,308	3,553.71	8,083

Shoulder Type
(Minnesota two-lane rural road segments, 1985-1989)

SHOULDER TYPE	NO. OF SEGMENTS	TOTAL SEGMENT LENGTH (miles)	TOTAL NO. OF ACCIDENTS (non-intersection)
GRAV & DIRT SHLD	1520	2065.17	3494
COMPOSIT SHLD	227	300.44	432
PAVED SHLD	1454	1102.48	3891
TOTAL	3,201*	3,459.09	7,817*

*3,308 segments with 8,083 non-intersection accidents were studied, but the constraint that shoulder type remain the same from left to right and throughout the time period 1985-1989 reduces these to 3,203 segments. Of these two had no shoulders, yielding the numbers shown above.

APPENDIX 2 - FINAL MODELS IN METRIC UNITS

The metric versions of the final models are:

I. Segments of two-lane rural roads (Table 27 in metric form)

Extended Negative Binomial Model with K = .306

$$
\begin{aligned}
\hat{y} &= \\
&\boldsymbol{EXPO_m \times \exp(0.165)} \\
&\boldsymbol{\times \exp(-0.278 LW_m - 0.194 SHW_m + 0.0668 RHR + 0.0135 DD_m + 0.139 STATE)} \\
&\boldsymbol{\times (\textstyle\sum_i WH\{i\} \exp(0.0137 DEG_m\{i\}))} \\
&\boldsymbol{\times (\textstyle\sum_j WV\{j\} \exp(0.142 V_m\{j\}))} \\
&\boldsymbol{\times (\textstyle\sum_k WG\{k\} \exp(0.105 GR\{k\}))}
\end{aligned}
$$

where

$\hat{y}$ = predicted mean number of non-intersection accidents on the segment
$EXPO_m$ = traffic exposure in millions of vehicle-kilometers
LW_m = lane width in meters
SHW_m = average of left and right shoulder widths in meters
RHR = average Roadside Hazard Rating along segment
DD_m = driveway density in driveways per kilometer
STATE = 0 for Minnesota, 1 for Washington
$DEG_m\{i\}$ = degree of curve in degrees per hundred meters of the i-th horizontal curve that overlaps the segment
WH{i} = fraction of the total segment length occupied by the i-th horizontal curve
$V_m\{j\}$ = absolute change in grade in percent per hundred meters of the j-th vertical crest curve that overlaps the segment
WV{j} = the fraction of the total segment length occupied by the j-th vertical crest curve
GR{k} = absolute grade in percent of the k-th uniform grade section that overlaps the segment
WG{k} = fraction of the total segment length occupied by the k-th uniform grade section.

NOTE: Each set of weights WH{i}, WV{j}, and WG{k} separately must sum to 1. To ensure this, usually it is necessary to insert one artificial horizontal curve with DEG = 0, one artificial crest with V = 0, and one artificial straightaway with GR = 0, each one having whatever weight is needed to make the sum equal 1.

II. Three-legged intersections of two-lane rural roads, stop-controlled on the minor road (Table 35 in metric form)

Negative Binomial Model with K = .481

$$\hat{y} =$$

$$\mathbf{\mathit{NUMBER\ OF\ YEARS} \times (\mathit{ADT1})^{.805} \times (\mathit{ADT2})^{.504} \times exp(-13.0)}$$
$$\mathbf{\times\ exp(0.0103\mathit{HI_m} + 0.0884\mathit{VCI_m} + 0.0177\mathit{SPDI_m})}$$
$$\mathbf{\times\ exp(0.173\mathit{RHRI} + 0.267\mathit{RT} + 0.0045\mathit{HAU})}$$

where the variables are:

$\hat{y}$ = predicted mean number of intersection or intersection-related accidents within 76 meters of intersection center
ADT1 = average two-way major road traffic in vehicles per day
ADT2 = average two-way minor road traffic in vehicles per day
HI_m = sum of degree of curve in degrees per hundred meters for each horizontal curve on major road any portion of which is within 76 meters of the intersection center, divided by the number of such curves
VCI_m = sum of absolute change of grade in percent per hundred meters for each crest curve (incoming signed grade larger than outgoing signed grade) on major road any portion of which is within 76 meters of the intersection center, divided by the number of such curves
$SPDI_m$ = average posted speed in kilometers per hour on major road in vicinity of the intersection
RHRI = average Roadside Hazard Rating within 76 meters of intersection center along major road
RT = 0 if no right turn lane on major road, 1 if right turn lane exists on major road
HAU = angle in degrees between increasing direction of major road and minor road minus 90 degree, multiplied by 1 if minor road is to right or by -1 if minor road is to left.

III. Four-legged intersections of two-lane rural roads, stop-controlled on the minor road (Table 35 in metric form)

Negative Binomial Model with K = .205

$$\hat{y} =$$

$$\textbf{\textit{NUMBER OF YEARS}} \times (\textbf{\textit{ADT1}})^{.603} \times (\textbf{\textit{ADT2}})^{.609} \times \exp(-10.4)$$
$$\times \exp(0.0137\textbf{\textit{HI}}_m + 0.0879\textbf{\textit{VCI}}_m + 0.0116\textbf{\textit{SPDI}}_m + 0.124\textbf{\textit{ND}} - 0.0049\textbf{\textit{HAU}})$$

where the variables are:

$\hat{y}$ = predicted mean number of intersection or intersection-related accidents within 76 meters of the intersection center
ADT1 = average two-way major road traffic in vehicles per day
ADT2 = average two-way minor road traffic in vehicles per day
HI_m = sum of degree of curve in degrees per hundred meters for each horizontal curve on major road any portion of which is within 76 meters of the intersection center, divided by the number of such curves
VCI_m = sum of absolute change of grade in percent per hundred meters for each crest curve (incoming signed grade larger than outgoing signed grade) on major road any portion of which is within 76 meters of the intersection center, divided by the number of such curves
$SPDI_m$ = average posted speed in kilometers per hour on major road in vicinity of the intersection
ND = number of driveways within 76 meters of the intersection center along major road
HAU = angle in degrees between increasing direction of major road and right leg of minor road minus angle in degrees between increasing direction of major road and left leg of minor road.