Crash Models for Rural Intersections: Four-Lane by Two-Lane Stop-Controlled and Two-Lane by Two-Lane Signalized

PUBLICATION NO. FHWA-RD-99-128

OCTOBER 1999

PB2000-102367

U.S. Department of Transportation
Federal Highway Administration

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

FOREWORD

This report provides direct input into the Accident Analysis Module (AAM) of the Interactive Highway Safety Design Model. The AAM is a tool that highway engineers can use to evaluate the impacts of highway design elements in project planning and preliminary design. Three crash models were developed relating crashes to three types of rural intersections. These types are: (1) three-legged intersections with major four-lane roads and minor two-lane roads that are stop-controlled, (2) four-legged intersections with major four-lane roads and minor two-lane roads that are stop-controlled, and (3) signalized intersections with both major and minor two-lane roads.

Elaborate sets of data were acquired from State data sources (Michigan and California) and collected in the field. The final data sets consist of 84 sites of the three-legged intersections, 72 sites of the four-legged intersections, and 49 sites of the signalized intersections. Negative binomial models — variants of Poisson models that allow for overdispersion — were developed for each of the three data sets. Significant variables included major and minor road traffic; peak major and minor left-turn percentage; peak truck percentage; number of driveways; and channelization, intersection median widths, vertical alignment, and, in the case of signalized intersections, the presence or absence of protected left-turn phases. Separate models were developed for crashes resulting in injuries and total crashes.

Michael F. Trentacoste

Michael F. Trentacoste
Director, Office of Safety
Research and Development

NOTICE

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
FHWA-RD-99-128		

4. Title and Subtitle	5. Report Date
CRASH MODELS FOR RURAL INTERSECTIONS: FOUR-LANE BY TWO-LANE STOP-CONTROLLED AND TWO-LANE BY TWO-LANE SIGNALIZED	October 1999
	6. Performing Organization Code
7. Author(s)	8. Performing Organization Report No.
Andrew Vogt	
9. Performing Organization Name and Address	10. Work Unit No. (TRAIS)
Pragmatics, Inc. 7926 Jones Branch Drive, Suite 711 McLean, Virginia 22102	3A5A
	11. Contract or Grant No.
	DTFH61-96-C-00076
12. Sponsoring Agency Name and Address	13. Type of Report and Period Covered
Office of Safety R&D Federal Highway Administration 6300 Georgetown Pike McLean, Virginia 22101-2296	Final Report September 1996 - May 1999
	14. Sponsoring Agency Code

15. Supplementary Notes

Contracting Officer's Technical Representative (COTR): Joe G. Bared, HRDS

16. Abstract

This report describes the collection, analysis, and modeling of crash and roadway data for intersections on rural roads in California and Michigan for the years 1993-1995. Three classes of intersections are considered: (1) three-legged intersections with major road four-lane and minor leg two-lane stop-controlled, (2) four-legged intersections with major road four-lane and minor legs two-lane stop-controlled, and (3) signalized intersections with both major and minor roads two-lane. Data were acquired from the Highway Safety Information System, State and Federal photologs, and field work at all intersections. The field work included morning and evening traffic counts by movement and vehicle type as well as alignment measurements out to 800 feet (244 meters) along the major road. The final data sets consist of 84 three-legged intersections, 72 four-legged intersections, and 49 signalized intersections. Other variables collected include Roadside Hazard Rating, number of driveways, channelization, intersection angles, and speed limits.

Negative binomial models — variants of Poisson models that allow for overdispersion — are developed for each of the three data sets. Significant variables include major and minor road traffic, peak major and minor road left-turning percentage, number of driveways, channelization, median widths, vertical alignment, and, in the case of the signalized intersections, the presence or absence of protected left-turn phases and peak truck percentage. Models are developed for all crashes within 250 feet (76 meters) of the intersection center, for intersection-related crashes within 250 feet (76 meters), and for injury crashes. For injury crashes, intersection angle and minor road posted speed are significant. Models of crashes at signalized intersections by approach flows are also investigated, and other model forms are proposed for future consideration.

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

19. Security Classification (of this report)	20. Security Classification (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	182	

Form DOT F 1700.7 Reproduction of completed page authorized

ACKNOWLEDGMENTS

At the direction of FHWA, the word "crash" and its variants have been substituted for "accident" and its variants except in special contexts.

The author expresses appreciation to staff members of the Federal Highway Administration's Turner-Fairbank Highway Research Center (especially Dr. Joe Bared), the Highway Safety Information System, the California and Michigan Departments of Transportation, and Pragmatics, Inc. (especially Highway Engineer Jian Song). Thanks also to Ezra Hauer, to Tim Neuman and Kevin Slack of CH2M HILL, and to engineering students at Sacramento State, Michigan State, and Lansing Community College.

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

Page

1. INTRODUCTION 1

2. LITERATURE REVIEW 3
   CRASHES AND TRAFFIC 3
   VARIABLES BESIDES TRAFFIC 7
      Channelization 7
      Sight Distance 8
      Horizontal and Vertical Alignment 8
      Intersection Angle 9
      Median Width, Surface Width, and Shoulder Width 10
      Signal Characteristics 10
      Lighting 10
      Roadside Conditions 11
      Truck Percentage 11
      Speed 12
      Weather 12
   MODEL FORMS AND FUNCTIONAL FORMS 12
      State of the Art 12
      Relevant Literature 14
      Empirical Bayesian Methods 15
      Alternative Functional Forms 16
      Special Studies 17
   CONCLUSIONS 18

3. DATA COLLECTION 21
   THE POPULATIONS 21
      The States and the Three Data Classes 21
      Pilot Studies 23
   SAMPLE SELECTION 25
   DATA COLLECTED 28
      Highway Safety Information System (HSIS) Data 28
      Traffic-Count Variables 28
      Roadside Variables 32
      Channelization and Intersection Geometry 33
      Sight Distances 33
      Horizontal Alignment 34
      Vertical Alignment 34
      Other Variables 35

TABLE OF CONTENTS
(continued)

Page

DATA LIMITATIONS 35
HSIS Data 35
Field Data 35

4. ANALYSIS 39
NEW VARIABLES 39
Crash Variables 39
ADT Variables 41
Variables Derived From Traffic Counts 42
Intersection Angle Variables 45
Sight Distances 48
Horizontal Alignment 49
Vertical Alignment 49
Miscellaneous Variables 51
UNIVARIATE STATISTICS 52
Crash Data Versus Intersection Class and State 65
BIVARIATE STATISTICS 67
Crashes Versus Other Variables 67
ADT and State Versus Other Variables 76
Correlations Between Intersection Variables 85
Correlations for Single-Vehicle and Multiple-Vehicle Crashes at Signalized Intersections 90
Turning Percentage Variables 90
Crashes Versus ADT 96
A.M. Versus P.M. Truck Percentages 101
CONCLUSIONS 101

5. MODELING 105
THEORY 105
Modeling 105
P-Values and Goodness of Fit 107
Model Building 110
MODELS FOR THREE-LEGGED INTERSECTIONS 111
MODELS FOR FOUR-LEGGED INTERSECTIONS 115
MODELS FOR THE SIGNALIZED INTERSECTIONS 120
Negative Binomial Models 120
Flow Models 126
RESIDUAL ANALYSIS 129

TABLE OF CONTENTS
(continued)

Page

6. CONCLUSIONS 145
THE MAIN MODELS 145
Three-Legged Intersections 145
Four-Legged Intersections 147
Signalized Intersections 149
EXPLANATORY VALUE OF MAIN MODELS 151
ACCIDENT REDUCTION FACTORS 155
SUMMARY 157

APPENDIX. DATA FROM PILOT STUDY PHASE OF DATA COLLECTION 161

REFERENCES 167

LIST OF FIGURES

Figure		Page
1.	Intersection Diagram Showing Leg Numbers	32
2.	Intersection Angle Geometries	46
3.	Examples of Calculation of the Angle Variable HAU	47
4.	Vertical Curve	50
5.	Crashes Versus Major Road ADT, Three-Legged Sample	97
6.	Crashes Versus Minor Road ADT, Three-Legged Sample	98
7.	Crashes Versus Major Road ADT, Four-Legged Sample	99
8.	Crashes Versus Minor Road ADT, Four-Legged Sample	99
9.	Crashes Versus Major Road ADT, Signalized Sample	100
10.	Crashes Versus Minor Road ADT, Signalized Sample	101
11.	Cumulative Scaled Residual Versus ADT1 for Three-Legged Intersections, TOTACCI Main Model of Table 29	131
12.	Cumulative Scaled Residual Versus ADT2 for Three-Legged Intersections, TOTACCI Main Model of Table 29	132
13.	Cumulative Scaled Residual Versus MEDWIDTH1 for Three-Legged Intersections, TOTACCI Main Model of Table 29	133
14.	Cumulative Scaled Residual Versus NODRWY1 for Three-Legged Intersections, TOTACCI Main Model of Table 29	134
15.	Cumulative Scaled Residual Versus ADT1 for Four-Legged Intersections, TOTACCI Main Model of Table 33	135
16.	Cumulative Scaled Residual Versus ADT2 for Four-Legged Intersections, TOTACCI Main Model of Table 33	136
17.	Cumulative Scaled Residual Versus PK%LEFT1 for Four-Legged Intersections, TOTACCI Main Model of Table 33	137
18.	Cumulative Scaled Residual Versus ADT1 for Signalized Intersections, TOTACCI Main Model of Table 36	138
19.	Cumulative Scaled Residual Versus ADT2 for Signalized Intersections, TOTACCI Main Model of Table 36	139
20.	Cumulative Scaled Residual Versus PK%LEFT2 for Signalized Intersections, TOTACCI Main Model of Table 36	140
21.	Cumulative Scaled Residual Versus VEICOM for Signalized Intersections, TOTACCI Main Model of Table 36	141
22.	Cumulative Scaled Residual Versus PK%TRUCK for Signalized Intersections, TOTACCI Main Model of Table 36	142
23.	Explanation of Variation of TOTACC and TOTACCI by Groups of Covariates, Main Negative Binomial Models for Three-Legged Intersections, Log-Likelihood R-Squared	152

LIST OF FIGURES
(continued)

Figure **Page**

24. Explanation of Variation of TOTACC and TOTACCI by Groups of Covariates, Main Negative Binomial Models for Four-Legged Intersections, Log-Likelihood R-Squared 154

25. Explanation of Variation of TOTACC and TOTACCI by Groups of Covariates, Main Negative Binomial Models for Signalized Intersections, Log-Likelihood R-Squared 155

A-1. Posted Speed Versus Operating Speed 162

A-2. Daytime Speed Versus 24-Hour Speed 163

A-3. Peak Truck Percentage Versus 24-Hour Truck Percentage 164

A-4. Crash Locations and Relationships at a Three-Legged Intersection 165

LIST OF TABLES

Table Page

1. Frequency of Signalized Rural Two-Lane Intersections in Four States 21
2. Pilot Study Intersections in California and Michigan 23
3. Samples as Proportions of Nominal Populations 27
4. Variables Collected in the Study 29
5. Summary Statistics: 84 Three-Legged Rural Intersections 53
6. Summary Statistics: 72 Four-Legged Rural Intersections 57
7. Summary Statistics: 49 Signalized Rural Intersections 61
8. Mean Number of Crashes per Intersection by Crash Variable and Intersection Class 65
9. Total Number of Crashes by Severity, State, and Intersection Class 66
10. Percentage of Intersections and Crashes in California for Each Intersection Class .. 66
11. Correlation Coefficients and P-Values for Crashes Versus Other Variables, Three-Legged Intersections 68
12. Correlation Coefficients and P-Values for Crashes Versus Other Variables, Four-Legged Intersections 70
13. Correlation Coefficients and P-Values for Crashes Versus Other Variables, Signalized Intersections 72
14. Correlates of TOTACC 77
15. Correlation Coefficients and P-Values for ADT and STATE Versus Intersection Variables, Three-Legged Intersections 79
16. Correlation Coefficients and P-Values for ADT and STATE Versus Intersection Variables, Four-Legged Intersections 81
17. Correlation Coefficients and P-Values for ADT and STATE Versus IntersectionVariables, Signalized Intersections 83
18. Correlations Between Intersection Variables in the Three-Legged Sample 86
19. Correlations Between Intersection Variables in the Four-Legged Sample 87
20. Correlations Between Intersection Variables in the Signalized Sample 88
21. Correlation Coefficients and P-Values for Single-Vehicle and Multiple-Vehicle Crashes Versus Signalized Intersection Variables 91
22. Correlation Coefficients and P-Values for Peak Turning Percentage Variables 94
23. Correlation Coefficients and P-Values for Crashes Versus Peak Turning Percentage Variables 95
24. Crashes Versus Grouped Major and Minor Road ADT, Three-Legged Sample 97
25. Crashes Versus Grouped Major and Minor Road ADT, Four-Legged Sample 98
26. Crashes Versus Grouped Major and Minor Road ADT, Signalized Sample 100
27. A.M. and P.M. Truck Percentages by State 102
28. Negative Binomial Models for Crashes per Year (TOTACC), Three-Legged Intersections 111
29. Negative Binomial Models for Crashes per Year (TOTACCI), Three-Legged Intersections 112

LIST OF TABLES
(continued)

Table Page

30. Negative Binomial Models for Crashes per Year (INJACC), Three-Legged Intersections 113
31. Negative Binomial Models for Crashes per Year (INJACCI), Three-Legged Intersections 114
32. Negative Binomial Models for Crashes per Year (TOTACC), Four-Legged Intersections 116
33. Negative Binomial Models for Crashes per Year (TOTACCI), Four-Legged Intersections 117
34. Negative Binomial Models for Crashes per Year (INJACC and INJACCI), Four-Legged Intersections 118
35. Negative Binomial Models for Crashes per Year (TOTACC), Signalized Intersections 122
36. Negative Binomial Models for Crashes per Year (TOTACCI), Signalized Intersections 123
37. Negative Binomial Models for Crashes per Year (INJACC, INJACCI), Signalized Intersections 124
38. Cumulative Scaled Residuals Versus Increasing Value of Intersection Variables, TOTACCI Main Models 143
39. Explanation of Variation in Total Crashes by Groups of Covariates, Main Three-Legged Intersection Models 151
40. Explanation of Variation in Total Crashes by Groups of Covariates, Main Four-Legged Intersection Models 153
41. Explanation of Variation in Total Crashes by Groups of Covariates, Main Signalized Intersection Models 153
42. Accident Reduction Factors for the Main Models 156

INDEX OF VARIABLES

Variable	Page
ABSGRD1	51
ABSGRD2	51
ACCTYPE	40
ADT	3
ADT1	41
ADT2	41
ADTM	29
AM%TRUCK	43
ANALYS	40
ANGLEi	30
DEGHi	30
Δgi	49
DEV	45
DIRi	30
DRINCMP	30
EBEG	29
E_HR	29
F_i	45
GBi	31
GEi	31
GRADE1	31
HAU	45
HBi	30
HEi	30
HEICOM	49
HEI-1	49
HEI-2	49
HICOM	49
HI-1	49
HI-2	49
INJACC	41
INJACCI	41
Li	49
LIGHT	31
LTLN1	30
LTLN1S	51
LTLN2	30
MBEG	29
MEDTYPE	30
MEDWIDTH1	30
MEDWIDTH1_m	145
MISCACT1	31
M_HR	29
NODRWY1	51
NODRWY2	51
PK%LEFT	43
PK%LEFT1	43
PK%LEFT2	43
PK%THRU1	43
PK%THRU2	43
PK%TRUCK	43
PK%TURN	43
PM%TRUCK	43
POSTADVi	31
PRODFADJ	45
PRODFOPP	45
PROT_LT	31
RAWEPCij	29
RAWETRij	29
RAWMPCij	29
RAWMTRij	29
RSD1	49
RSD2	49
RSDL1	49
RSDL2	49
RSDR	49
RTLN1	30
RTLN2	30
SDi	30
SDLi	30
SDRi	30
SIG_TYPE	31
SPDLIMi	31
SPD1	51
SPD2	51
STATE	52

INDEX OF VARIABLES
(continued)

Variable	Page
SUMF	45
terrain	31
TOTACC	31, 39
TOTACCI	39
TOTACCM	41
TOTACCS	41
VBi	31
VCEICOM	50
VCEI-1	50
VCICOM	50
VCI-1	50
VEH_INVOL	40
VEi	31
VEICOM	50
$VEICOM_m$	150
VEI-1	50
V_i	49
VICOM	50
VI-1	50

1. INTRODUCTION

This study develops crash models for:

- Rural three-legged and four-legged intersections on four-lane highways, stop-controlled on the minor legs.
- Signalized rural intersections of two-lane roads.

An earlier study,[1] of which this may be regarded as a continuation, treats segments of two-lane rural roads and rural three- and four-legged intersections of two-lane roads, stop-controlled on the minor legs. The two studies together consider the chief geometries on two-lane roads — segments, intersections with minor road stop-controlled, and signalized intersections. In addition, this study branches out by passing from intersections on two-lane roads to ones on four-lane roads.

A major intended use of crash models such as the ones developed here is in the Accident Analysis component of the Interactive Highway Safety Design Model (IHSDM).[2] The IHSDM is a proposed set of interactive computer programs that will allow highway designers to examine the safety consequences of various design alternatives. These programs will assess how proposed designs relate to driver expectations, vehicle and driver capabilities, traffic flows, and established design principles.

The Accident Analysis component, or Accident Analysis Module, is intended to estimate, in quantitative terms, the safety effects — crash frequencies and severities — that may result from different designs. In addition to driver and vehicle variables, safety is influenced by the volume and movement of traffic. It is also influenced by such design features as channelization, horizontal and vertical curves, sight distances, and roadside conditions. The module was tentatively envisioned (op. cit.) to have four parts, dealing respectively with segment crashes, intersection crashes, interchange ramp crashes, and roadside crashes. The safety consequences of a particular design would be the sum of the contributions of each part. A model would be developed for each type of crash and the models would be combined to yield an overall picture of design consequences.

---

[1] A. Vogt and J.G. Bared, *Accident Models for Two-Lane Rural Roads: Segments and Intersections*, Report No. FHWA-RD-98-133, Federal Highway Administration, McLean, Va., 1998; and A. Vogt and J.G. Bared, "Accident Models for Two-Lane Rural Segments and Intersections," *Transportation Research Record* 1635: 19-29, 1998.

[2] J.A. Reagan, "The Interactive Highway Safety Design Model: Designing for Safety by Analyzing Road Geometrics," *Public Roads*: 37-43, Summer 1994.

H. Lum and J. Reagan, "Interactive Highway Safety Design Model: Accident Predictive Module," FHWA Draft 8-22-94.

The goal of the present study is to assess the combined and relative effects of highway variables on intersection crashes for the classes of intersections noted above. The method used, by now a well-established method, is that of generalized linear models based on a negative binomial distribution. Crashes are thought of as discrete rare events, the number of crashes at an intersection being a random variable of the Poisson type with overdispersion. The mean number of crashes is an exponential function of a linear combination of intersection variables and the variance in crash counts depends on the mean, as well as on an overdispersion parameter representing factors not included in the model.

In Chapter 2, literature on modeling of intersection crashes is reviewed. In Chapters 3 and 4, the data collection and preliminary analysis are described, and in Chapter 5, the models are presented and evaluated. A final chapter, Chapter 6, summarizes the results of this study.

2. LITERATURE REVIEW

In this chapter, representative studies are reviewed that relate intersection crashes to highway variables. The chief highway variables are the Average Daily Traffic (ADT) on the intersecting roads, but closer analysis indicates an important role for traffic movements as they pertain to different crash types. Most studies recognize that other variables, such as sight distances and channelization, also affect safety, and some studies that consider these other variables are discussed below. In addition, a number of studies are reviewed that examine the issue of the appropriate model form and/or functional form for mean number of crashes. Studies that deal with special issues, such as underreporting of crashes and crash location, are also noted.

This review is not meant to be exhaustive. Further review of the literature and many additional references may be found in the articles cited here. Of particular value for its up-to-dateness is the MRI Report (1997).[3] Our interest is rural intersections and, where possible, we shall emphasize studies in rural settings.

The chapter closes with a few overall conclusions.

CRASHES AND TRAFFIC

Many studies have been devoted to the relationship between crashes and traffic.

A 1953 study by McDonald[4] in California of intersections on divided highways, stop-controlled on the minor legs, represents crashes per year in graphical form as a function of major and minor road incoming daily traffic. A total of 150 three-legged and four-legged intersections on U.S. 99 and U.S. 40 were treated together and a dependency of the form:

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

was found where N is the number of crashes per year, V_d is entering major road Average Daily

[3] Midwest Research Institute, *Critical Reviews of Intersection Safety Studies Task K Resource Paper*, MRI Report, Contract No. DTFH61-96-C-00055, MRI Project No. 4584-09, Kansas City, Mo., 1997.

[4] J.W. McDonald, "Relation Between Number of Accidents and Traffic Volume at Divided-Highway Intersections," *Highway Research Board Bulletin 74, Traffic-Accident Studies*, pp. 7-17, National Academy of Sciences, National Research Council, Washington, D.C., 1953.

Traffic (ADT), and V_c is entering minor road ADT. This study advocates crashes per year rather than crashes per million entering vehicles as a measure of intersection safety, and emphasizes that crash experience at an individual intersection is a variable, while N is the mean for an aggregate of intersections with the given volumes. Median widths, channelization, and number of lanes at sample intersections were not explicitly noted. The study concludes that crashes are more sensitive to minor road volumes. Of interest is that the minor road ADT in this study was based on weekday 24-hour mechanical traffic counts at most sites and may be more accurate than that in other studies.

Another study in California, by Webb[5] in 1955, examines two-phase signalized intersections and arrives at the equations:

$$
\begin{aligned}
N_U &= 0.000189(ADT1)^{0.55}(ADT2)^{0.55} \\
N_S &= 0.00389(ADT1)^{0.45}(ADT2)^{0.38} \\
N_R &= 0.00703(ADT1)^{0.51}(ADT2)^{0.29}
\end{aligned}
$$

where N_U, N_S, and N_R, respectively, are the number of crashes per year at urban, semi-urban, and rural two-phase intersections, and ADT1 and ADT2 are major and minor road two-way average daily traffic counts (units have been adjusted from the original study). The three categories were determined by speed limits: 25 mph (40.2 km/h) was regarded as urban; more than 25 mph (40.2 km/h) but less than 45 mph (72.4 km/h) as semi-urban; and 45 mph (72.4 km/h) or more as rural. Intersections having unusual features were eliminated, and the resulting sample sizes were 23, 60, and 14 intersections for urban, semi-urban, and rural, respectively. Some of those that remained were on four-lane divided highways. Rear-end crashes on the minor road, a county road, were omitted, and the author notes that this may, in part, be responsible for the decreasing power of minor road ADT as one moves from urban to rural and from lower to higher major road speeds. The author also notes that intersection geometry, roadside development, and sight distance are influential causal factors for crashes. Hauer and Persaud (1996, p. 84)[6] find Webb's equation for N_R the most plausible among available studies.

---

[5] G.M. Webb, "The Relation Between Accidents and Traffic Volumes at Signalized Intersections," *Institute of Transportation Engineers Proceedings*, Technical Session No. 3B, pp. 149-167, 1955.

[6] E. Hauer and B. Persaud, *Safety Analysis of Roadway Geometry and Ancillary Features*, Transportation Association of Canada, Ottawa, 1996.

Yet another California study, David and Norman (1975),[7] considers crash factors at San Francisco Bay Area intersections, but only at intersections with at least two crashes in the time period 1971-1973. This study includes numerous tabular presentations of crash counts for ranges of crash factors. Crashes were classified by severity and by traffic conflicts and movements. Let us call the conflict/movement categories "Typical" and "Other." The study includes a linear regression model for the number of "Typical" intersection crashes per 3 years. The chief factors in the model in decreasing order of importance (as measured by R-squared statistics), along with the sign of their effect, are:

+	A measure of traffic volume based on "Typical" conflict/turning movement.
+	Number of "Other" crashes in time period.
+	Number of U-turn restrictions.
-	Number of right-turn lanes.
-	Number of lanes on major road.
+	Stop-controlled versus signalized (0 versus 1).
+	Width of minor road.
-	Number of divided streets.
-	Number of left-turn lanes.

This model (David and Norman, 1975, p. 105) was based on 82 intersections for which directional ADT data were available. David and Norman note, as does Webb, that introduction of left-turn lanes at signalized intersections without conversion of two-phase signals into three or more phases tends to increase crash counts. For a sample of 558 intersections, the percentage of nighttime crashes was usually 20 to 30%, with no notable variation when lighting was present. Possibly, the percentage of crashes at the lighted intersections would have been higher if they had not been lighted.

Hakkert and Mahalel (1978)[8] observe that more than 50% of crashes occur at intersections. They analyze four-legged intersections in terms of 24 crossing or merging pairs of traffic flows (vehicles per unit time). For each pair, they calculate the product of the two flows and sum over all 24 pairs to obtain a traffic flow index x. For urban and interurban intersections in Israel, they obtain a Poisson-type model of the form:

$$N = A + Bx$$

---

[7] N.A. David and J.R. Norman, *Motor Vehicle Accidents in Relation to Geometric and Traffic Features of Highway Intersections, Volume II - Research Report*, Report No. FHWA-RD-76-129, Federal Highway Administration and National Highway Traffic Safety Administration, Washington, D.C., 1975.

[8] A.S. Hakkert and D. Mahalel, "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.

where N is the mean number of crashes per unit time at the intersection and A and B are suitable positive constants. The crashes were injury or fatality crashes, the roads a mix of two-lane and four-lane, and the intersections a mix of signalized and non-signalized. Traffic flows for the modeling were based on 16-hour weekday counts. The presence of the constant term A is taken as evidence that for small values of x, other factors come into play.

Pickering, Hall, and Grimmer (1986)[9] consider crashes at three-legged intersections of two-lane roads. They report that in 1983, one-third of injury crashes occurred at intersections, and 45% of these were at tee intersections. Their basic model is a Poisson model, with mean number of crashes per unit time N of the form:

$$N = K(ADT1 \times ADT2)^p$$

where p is approximately 0.5. They consider such issues as how far a crash is from the intersection, presence or absence of islands and channelization, and the dependence of crashes on pairs of traffic flows. For different crash types, products of the relevant flows tended to be most significant, but the model above performed respectably when all types of crashes were summed. Motorcycles and bicycles were involved in a disproportionate number of crashes relative to their percentage of the flow. Operating speeds of vehicles were significant, but depending on the type of crash, higher speeds did not always lead to more frequent crashes.

A study of Hauer, Ng, and Lovell (1988),[10] based on 145 signalized intersections in Toronto, considers 15 different crash patterns and develops negative binomial models for each pattern of the forms:

$$N = K \times F^a$$
$$N = K \times F_1^a \times F_2^b$$

depending on whether one flow F or two flows F_1 and F_2 are involved, with $a, b \geq 0$. Here N is the mean number of crashes of the given pattern on the population of all intersections having these flows. Crashes are weekday daytime crashes involving two vehicles. The number of lanes on the roads and the channelization are not noted This study is notable for, among other things, its very thoughtful explication of assumptions underlying the use of the negative binomial model.

---

[9] D. Pickering, R.D. Hall, and M. Grimmer, *Accidents at Rural T-Junctions*, Research Report 65, Transport and Road Research Laboratory, Department of Transport, Crowthorne, Berkshire, United Kingdom, 1986.

[10] E. Hauer, J.C.N. Ng, and J. Lovell, "Estimation of Safety at Signalized Intersections," *Transportation Research Record* 1185: 48-61, 1988.

Bonneson and McCoy (1993)[11] develop a negative binomial model of the form:

$$N = K \times (ADT1)^{0.256}(ADT2)^{0.831}$$

Here N is the mean number of crashes. The overdispersion parameter for this model is 4.0, which is rather large. A total of 125 non-urban four-legged intersections from Minnesota were considered in the study, 17 of which had four-lane major roads with substantial medians. All crashes occurring within 500 feet (153 meters) of the intersection were included.

VARIABLES BESIDES TRAFFIC

The primary importance of traffic as an explanatory factor for intersection crashes relative to other highway variables has long been acknowledged, and recent studies do not contradict this observation. The study of Bauer and Harwood (1996)[12] concludes that highway variables other than traffic have only a slight influence on crashes. A review, described by Bauer and Harwood, of hard-copy crash reports at eight urban intersections found that "only 5 to 14% of the accidents had causes that appeared to be related to geometric design features of the intersections." The report of Vogt and Bared (Vogt and Bared, 1998, p. 137), which develops crash models for three-legged and four-legged intersections of rural two-lane roads, attributes about 2% explanatory value to design variables as compared with 27% to ADT.

Nonetheless, designs aimed at improving safety will always be in demand, and attempts to quantify design effect are entirely proper. Design variables that have received special attention in connection with intersection crashes include: channelization, sight distances, horizontal and vertical alignment, intersection angle, median width, and signal characteristics. Also noted below are the effects of truck percentage in the traffic stream, speed, and weather.

Channelization

It is generally thought that right-turn and left-turn lanes on major and/or minor roads contribute to intersection safety. The model of David and Norman (1975) mentioned earlier indicates that left- and right-turn lanes reduce crashes. They also list left-turn storage lanes as one of six "demonstrably accident-related" intersection design features, but they find that opposing left-turn lanes without multi-phasing or at stop-controlled intersections increase crashes. They suggest raised lane markers to help drivers define their lateral location and multi-phasing at signalized intersections. The

[11] J.A. Bonneson and P.T. McCoy, "Estimation of Safety at Two-Way Stop-Controlled Intersections on Rural Highways," *Transportation Research Record* 1401: 83-89, 1993.

[12] K.M. Bauer and D. Harwood, *Statistical Models of At-Grade Intersection Accidents*, Report No. FHWA-RD-96-125, Federal Highway Administration, McLean, Va., 1996.

summary of Kuciemba and Cirillo (1992)[13] mentions channelization, along with sight distance improvement, as a safety factor for intersections where turning traffic is high. Use of lane dividers is recommended in urban settings, while left-turn lanes in rural areas are expected to reduce passing crashes. The study of Bauer and Harwood (1996) finds that left-turn lanes lower crashes, although curbed dividers may not be more effective than painted ones. A study of McCoy, Hoppe, and Dvorak (1985)[14] points out that left-turn lanes may be more necessary in the absence of paved shoulders or when truck percentages are high. The study of Pickering, Hall, and Grimmer (1986) finds channelization, including islands, to be significant for certain crash types, but not for total crashes. Garber and Srinivasan (1991)[15] in a study of elderly drivers conclude that left-turn lanes (and protected phasing) would have special benefits for the elderly because of their proclivity for crashes with opposing traffic.

Sight Distance

Intersection sight distances are an intuitively evident safety consideration at intersections. They are noted as such by David and Norman (1975) and in the summary of Kuciemba and Cirillo (1992). A study of Hanna, Flynn, and Tyler (1976)[16] notes that sight distances on all approaches, for both non-signalized and signalized intersections, affect crash rates in the expected way. Bared and Lum (1992)[17] also find that sight distances are shorter at high-crash intersections.

Horizontal and Vertical Alignment

Horizontal and vertical alignment are, of course, related to sight distances. Horizontal curves, in particular, are associated with high crash rates. Their effects on roadway crashes are noted in the

---

[13] S.R. Kuciemba and J.A. Cirillo, *Safety Effectiveness of Highway Design Features, Volume V - Intersections*, Report No. FHWA-RD-91-048, Federal Highway Administration, Washington, D.C., 1992.

[14] P.T. McCoy, W.J. Hoppe, and D.V. Dvorak, "Benefit-Cost Evaluation of Left-Turn Lanes on Uncontrolled Approaches of Rural Intersections (Abridgement)," *Transportation Research Record* 1025: 40-43, 1985.

[15] N.J. Garber and R. Srinivasan, "Risk Assessment of Elderly Drivers at Intersections: Statistical Modeling," *Transportation Research Record* 1325: 17-22, 1991.

[16] J.T. Hanna, T.E. Flynn, and W.L. Tyler, "Characteristics of Intersection Accidents in Rural Municipalities," *Transportation Research Record* 601: 79-82, 1976.

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

report of McGee, Hughes, and Daily (1995)[18] and the references cited therein, as well as in the study of truck crashes by Miaou, Hu, Wright, Davis, and Rathi (1993)[19]; the paper of Shankar, Mannering, and Barfield (1995)[20]; and the paper of Vogt and Bared (1998). This paper and the FHWA report of Vogt and Bared (1998) also exhibit intersection crash models for three-legged and four-legged intersections of two-lane roads in which the average degree of curve for nearby horizontal curves and the average grade change per 100 feet (30.1 meters) for nearby crest curves are represented. These curves are required to be on the major road, with some portion within 250 feet (76 meters) of the intersection center. The minor roads are stop-controlled. Although the alignment variables are not particularly significant (with P-values on the order of 0.30), they correlate reasonably well with crash counts, especially on the four-legged intersections.

One oddity on the subject of alignments is the finding of Hanna et al. (1976) that steep grades tend to decrease intersection crash counts. Grades different from zero appear to increase crash counts on segments according to Miaou et al. (1993), Shankar et al. (1995), and Vogt and Bared (1998).

Intersection Angle

Right-angled intersections are encouraged in design. A study of McCoy, Tripi, and Bonneson (1994)[21] indicates that severely skewed intersections have higher crash experience. However, Bared and Lum (1992) find right-angled intersections more dangerous than mildly skewed ones. This is also supported by Bauer and Harwood (1996) for urban signalized intersections and by Vogt and Bared (1998) for rural stop-controlled intersections of two-lane roads. A study of Kulmala (1995)[22] suggests that when major road turning traffic that must cross the opposing major road lane(s) turns

---

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

[19] S.-P. Miaou, P.S. Hu, T. Wright, S.C. Davis, and A.K. Rathi, *Development of Relationship Between Truck Accidents and Geometric Design: Phase I*, Report No. FHWA-RD-91-124, Federal Highway Administration, McLean, Va., 1993.

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

[21] P.T. McCoy, E.J. Tripi, and J.A. Bonneson, *Guidelines for Realignment of Skewed Intersections*, Nebraska Department of Roads Research Project Number RES1 (0099) P471, 1994.

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

through an angle from $0°$ to $90°$, fewer crashes occur than when the turning angle is from $90°$ to $180°$. This is presumably because traffic exiting from the major road has better sight of oncoming major road traffic for small angles. The intersection models of Vogt and Bared support this conclusion in the case of four-legged intersections, but not in the case of three-legged ones.

Median Width, Surface Width, and Shoulder Width

Wider medians are generally associated with fewer crashes on divided highways. See the study of Knuiman, Council, and Reinfurt (1993).[23] At intersections, a median region allows a zone of protection for turning traffic (although if the zone is too wide, it converts one intersection into two). Harwood et al. (1995)[24] find that increased median widths are associated with fewer crashes at rural unsignalized intersections, but with more crashes at suburban signalized intersections.

Bauer and Harwood (1996) find that increased lane widths and increased shoulder widths lower the probability of serious crashes and/or multiple-vehicle crashes at urban non-signalized intersections.

Signal Characteristics

King and Goldblatt (1975)[25] discuss the important issue of whether signalization decreases crashes. Their study and some others have found no significant decrease, but rather a change in the relative frequencies of crash types (from right-angle to rear-end). The commonly accepted view is that at high-volume intersections, signalization is beneficial, but that at low-volume ones, it may not be.

With regard to phasing, David and Norman (1975) indicate that protected left turns are beneficial. For the elderly, this is supported by Garber and Srinivasan (1991), who also propose a longer amber light. Bauer and Harwood (1996) likewise find a beneficial effect for multi-phase, rather than two-phase, signaling in their modeling of urban intersections, as well as for actuated signals versus pre-timed ones.

Lighting

Bauer and Harwood (1996) find that the absence of lighting contributed significantly to the number

---

[23] M.W. Knuiman, F.M. Council, and D.W. Reinfurt, "Association of Median Width and Highway Accident Rates," *Transportation Research Record* 1401: 70-82, 1993.

[24] D.W. Harwood, M.T. Pietrucha, M.D. Woolridge, R.E. Brydia, and K. Fitzpatrick, *Median Intersection Design*, National Cooperative Highway Research Program Report 375, Transportation Research Board, National Research Council, National Academy Press, Washington, D.C., 1995.

[25] G.F. King and R.B. Goldblatt, "Relationship of Accident Patterns to Type of Intersection Control," *Transportation Research Record* 540: 1-12, 1975.

of injury crashes at rural three-legged and four-legged intersections. A study by Blower, Campbell, and Green (1993)[26] indicates that truck crashes in Michigan are more frequent at night and in rural settings; the combination of the two is deemed to imply less lighting. See also the study of Elvik (1995).[27]

Roadside Conditions

Vogt and Bared (1998) find that roadside hazards, as measured by the Roadside Hazard Rating of Zeeger et al. (1987), contribute to crashes on three-legged intersections, while driveway density near the intersection center contributes to crashes on four-legged intersections.

The Roadside Hazard Rating is a whole number from 1 to 7 (with 1 representing perfectly flat and unobstructed roadsides, the least hazardous case) that evaluates sideslope, clear zone, and distance to the nearest hard object. In the Vogt-Bared study, the value is a subjective average along the major road within ± 250 feet (76.2 meters) of the intersection center. Although it is reasonable that nearby driveways might make an intersection more dangerous, the Vogt-Bared results are based on Minnesota data and it was not possible to eliminate driveway crashes explicitly from the data set.

Truck Percentage

David and Norman (1975) note the safety-relatedness of bus routing and zones, of clearly visible street name signs, and of raised markers and striping to indicate turning lanes and to remind the driver of intersection control features. Their study is primarily urban, but the routing of buses and the placement of bus zones can be thought of as the equivalent of truck traffic and truck turning percentages. Not only are trucks more difficult to maneuver and potentially more likely to cause serious crashes, but they are also obstacles that interfere with the line of sight of drivers (including the truck driver making a turn).

Blower, Campbell, and Green (1993) find that significant causative factors for truck crashes are: rural environment, nighttime, and road type "other" (versus "major arterial" or "limited access"). Furthermore, bobtail trucks (no tractor) are more crash-prone than single or double tractors. McCoy, Hoppe, and Dvorak (1985), as noted, favor left-turn lanes when truck percentages are high.

Miaou et al. (1993) and the Vogt-Bared (1998) FHWA report find that a higher percentage of truck traffic is associated, respectively, with fewer truck crashes and fewer crashes on rural roads. Miaou et al. (1993, p. 62) suggest that perhaps "for a constant vehicle density, as percent trucks increases, the frequency of lane changing and overtaking movements by cars decreases."

---

[26] D. Blower, K.L. Campbell, and P.E. Green, "Accident Rates for Heavy Truck-Tractors in Michigan," *Accident Analysis and Prevention* 25 (3): 307-321, 1993.

[27] R. Elvik, "Meta-Analysis of Evaluations of Public Lighting as Accident Countermeasures," *Transportation Research Record* 1485: 112-123, 1995.

Speed

Bauer and Harwood (1996) find that crash rates increase with increasing design speed on four-legged rural intersections. Vogt and Bared (1998) find the same for posted speeds on rural three-legged and four-legged intersections. Pickering, Hall, and Grimmer (1986) observe that higher operating speeds at three-legged intersections are associated with more right-turn crashes, but with fewer crashes of other types.

Weather

Bad weather is recognized as a contributing factor to crashes. Shankar, Mannering, and Barfield (1995) call attention to the interaction of extreme weather and extreme alignment. Miaou et al. (1993) note the relevance of weather to truck crashes. Fridstrøm et al. (1995)[28] in a study of Scandinavian roadway crashes find weather significant, although bad weather does not always increase crashes. Vogt and Bared (1998), using a regional, but not particularly local weather variable in Minnesota, find that weather conditions do not have a strong effect on crashes.

MODEL FORMS AND FUNCTIONAL FORMS

State of the Art

In recent years, a consensus has formed in favor of modeling crashes as discrete, rare, independent events. In a static environment, such events can be characterized by their mean number λ per unit time and are simply represented by a Poisson random variable, i.e., the probability that y crashes will be observed per unit time is:

$$P(Y = y) = e^{-\lambda} \frac{\lambda^y}{y!}$$

where y = 0, 1, 2, ... To proceed further, one analyzes the mean λ in terms of familiar variables that characterize or partially characterize the crash location (in our case, an intersection). Thus, one assumes that

$$\lambda = f(\{x_i\},\{\beta_j\})$$

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

that is, λ is taken to be a function of suitable variables $x_0, x_1, ..., x_m$ pertaining to the intersection. This function is also assumed to depend on parameters β_j that are independent of the intersection. The form of the function f is up to the modeler except that it is required not to yield negative values. At different intersections, the variables x_i may take different values, so different intersections may have different mean crash counts λ.

A commonly used functional form is the generalized linear one:

$$\lambda = \exp\,(\beta_0 x_0 + \beta_1 x_1 + ... + \beta_m x_m) = \exp\,(\textstyle\sum_{i=0}^{m} \beta_i x_i) \qquad (2.1)$$

This form guarantees a non-negative integer value for the mean number of crashes per unit time. A major attraction of the form is that it is possible to estimate the coefficients β_i from data using methods originated by Nelder and Wedderburn (1972)[29] and implemented by the software packages SAS and LIMDEP. If the first variable x_0 is taken to be identically equal to 1, the combination in equation (2.1) includes a constant term β_0, sometimes called the intercept term. Another advantage is easy comparability with existing models since the form $\lambda = \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2)$ can easily be converted to the multiplicative form $\lambda = K(y_1)^{\beta 1}\,(y_2)^{\beta 2}$, where $K = \exp(\beta_0)$, $y_1 = \exp(x_1)$, and $y_2 = \exp(x_2)$. The multiplicative form is common in earlier studies.

The model form equation (2.1) is based on the assumptions that crashes are independent events, that suitable input variables x_i are discoverable taking fixed values at the intersection on some appropriate time scale, and that the functional form in equation (2.1) is superior to other possible forms. It is useful to act as if these assumptions are approximately true, in part because they yield an analytically tractable generalized linear model and in part because they have proved their worth elsewhere in biology and economics.

A refinement of this approach, described in Hauer, Ng, and Lovell (1988), is to acknowledge that the mean for a particular intersection is unknowable and to consider an imaginary population of intersections all having the same values for the variables x_i and having means that are grouped around the value λ in equation (2.1). The variance of the crash counts of the intersections in this population depends on further assumptions, but can be taken to have the form:

$$\lambda + K\,\lambda^2 \qquad (2.2)$$

where K is a parameter, applicable to the entire population but independent of the particular intersection, called the overdispersion parameter. The variance of crash counts has two components, the first due to Poisson variation and the second due to differences among members of the

[29] J.A. Nelder and R.W. Wedderburn, "Generalized Linear Models," *Journal of the Royal Statistical Society*, Series A, 135(3): 370-384, 1972.

population, the latter perhaps due to omitted variables. Dean and Lawless (1989)[30] propose that the mean of individual intersections in the population is equal to a multiplier times the value λ in equation (2.1), and that the multiplier is a continuous positive random variable with mean 1 and variance K having the same distribution at each intersection. From this, they derive the overall variance (2.2). The number of crashes Y per unit time at individual intersections is distributed according to a compound Poisson distribution: Y given the intersection mean is a Poisson variable, but the intersection mean itself is a variable. It is customary to assume that this variable obeys a gamma distribution on each population and hence that Y obeys a negative binomial distribution.

$$P(Y = y) = (\frac{1}{y!})(\frac{\Gamma(y+(1/K))}{\Gamma(1/K)})(\frac{K\lambda}{1+K\lambda})^y(1+K\lambda)^{-1/K}$$

With the assumptions that λ is given by equation (2.1) and that K is independent of $\{x_i\}$, it is possible to estimate the parameters $\{\beta_j\}$ and K in LIMDEP and SAS by maximum likelihood methods. When prior crash experience is known at a particular intersection, along with the variables x_i, the negative binomial form makes it possible to revise the estimated crash count for a new time period by empirical Bayesian methods. See the discussion on p. 15 below.

Relevant Literature

Many of the studies alluded to earlier in this chapter have used Poisson and negative binomial models. Hakkert and Mahalel (1977) use a Poisson model with some refinements to study intersection crashes. Pickering, Hall, and Grimmer (1986), in their study of tee intersections, use a Poisson model along with the generalized linear model technique (and the software packages GENSTAT and GLM). Maycock and Hall (1984),[31] studying roundabouts, and Hauer, Ng, and Lovell (1988), studying urban intersections, employ the negative binomial technique. A sampling of other studies that have used negative binomial models includes: Miaou et al. (1993) - truck roadway crashes; Bonneson and McCoy (1993) - rural intersection crashes; Knuiman, Council, and Reinfurt (1993) - divided highway crashes; Fridstrøm et al. (1995) - roadway crashes; Poch and Mannering (1996)[32] - urban intersection crashes; Bauer and Harwood (1996) - intersection crashes;

---

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

[31] G. Maycock and R.D. Hall, *Accidents at 4-Arm Roundabouts*, Laboratory Report 1120, Transport and Road Research Laboratory, Department of Transport, Crowthorne, Berkshire, United Kingdom, 1984.

[32] M. Poch and F. Mannering, "Negative Binomial Analysis of Intersection-Accident Frequencies," *Journal of Transportation Engineering* 122 (2): 105-113, 1996.

and Vogt and Bared (1998) - rural segment and rural intersection crashes.

Miaou et al. (1993), Bauer and Harwood (1996), and Vogt and Bared (1998) make use of both Poisson and negative binomial models. Miaou and Lum (1993)[33] compare two linear regression models and two Poisson models, prefer the latter, and indicate that the negative binomial or "double Poisson" may be even better. Miaou (1994)[34] compares Poisson models and negative binomial models and indicates that both kinds of models have their place, with negative binomial to be preferred if the data are sufficiently overdispersed.

Empirical Bayesian Methods

Hauer, Ng, and Lovell (1996, p. 56) note that the negative binomial model permits past information about an intersection to be incorporated into modeling with relative ease. The essential idea is that intersections in the imaginary population with identical values of $\{x_i\}$ have their mean grouped around the value λ in equation (2.1), but past experience at an intersection gives some indication of where in this grouping the intersection mean is likely to be. If an intersection has had A crashes in the past T time units, then the grand mean λ and the crash count variance $\lambda + K\lambda^2$ are no longer applicable. Instead, for the sub-population with the given crash experience, crash counts still obey a negative binomial distribution, but the appropriate grand mean is:

$$\lambda_{new} = \frac{\lambda(1 + AK)}{1 + K\lambda T} \tag{2.3}$$

and the total variance of crash counts on members of this sub-population is:

$$\lambda_{new} + K_{new}(\lambda_{new})^2$$

where

$$K_{new} = \frac{K}{1 + AK} \tag{2.4}$$

The overdispersion parameter decreases in equation (2.4) if A > 0, and the grand mean increases or

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

[34] S.-P. Miaou, "The Relationship Between Truck Accidents and Geometric Design of Road Sections: Poisson Versus Negative Binomial Regressions," *Accident Analysis and Prevention* 26 (4): 471-482, 1994.

decreases in equation (2.3) depending on whether the crash experience is above average ($A > \lambda T$) or not.

Further discussion of this methodology is to be found in Hauer, Terry, and Griffith (1994),[35] Pendleton (1996),[36] Hauer and Persaud (1996), and the book of Hauer (1997).[37]

Hauer's book explores a variety of issues that relate to the use of crash models. His chief point is that if the goal is increased safety, cross-sectional studies are inadequate by themselves. Before-and-after studies are needed, and the effect of "regression to the mean" must be taken into account. This can be done with suitable models, based in part on cross-sectional studies, for reference populations that incorporate year-by-year crash data. Methods for predicting future trends are offered, along with ways to compare the safety of treated and untreated intersections in light of the models and crash history.

Alternative Functional Forms

Hakkert and Mahalel (1978) use a traffic flow index and a "sum of products" approach to modeling intersection crashes. Hauer, Ng, and Lovell (1988) analyze crashes by patterns and have a model for each approach pattern. Thus, it is desirable to have enough data by pattern to build separate models for each. Then the mean count for each type of crash can be summed to obtain an overall mean.

Miaou (1994) considers, in addition to Poisson and negative binomial models, zero-inflated Poisson (ZIP) models. These are Poisson models adjusted by increasing the probability of zero crashes (and rescaling the remaining probabilities so that the sum is still one). Miaou concludes that these are useful when there is underreporting of crashes, so that some locations have undeserved zero crash counts.

Bauer and Harwood (1996) do Poisson and negative binomial modeling, but they also exhibit a lognormal model where the log of the number of crashes is regarded as a normal variable with mean μ and variance σ^2. Log μ is assumed to be a linear function of intersection variables, while the variance is constant. They find this model useful for classes of high crash intersections (where few intersections have zero crashes in the time period under consideration).

---

[35] E. Hauer, D. Terry, and M.S. Griffith, "Effect of Resurfacing on Safety of Two-Lane Rural Roads in New York State," *Transportation Research Record* 1467: 30-37, 1994.

[36] O. Pendleton, *Evaluation of Accident Methodology*, Report No. FHWA-RD-96-039, Federal Highway Administration, McLean, Va., 1996.

[37] E. Hauer, *Observational Before-After Studies in Road Safety*, Pergamon Press, Oxford, U.K., 1997.

Lau and May (1988)[38] use Classification and Regression Trees (CART) to study intersection crashes. Data are divided into classes by binary trees of multiple levels until terminal nodes are reached (ones from which little further improvement can be made). A split is based on dividing a sample into two sub-samples so that the combined weighted variance of the two strata is a minimum for the residual crash count (left over from the previous split). This method seems to be applicable when most variables are categorical rather than continuous. Predicted crash counts under this approach may be modified on the basis of individual intersection histories.

Joksch and Kostyniuk (1998)[39] apply smoothing techniques to study the relationship between intersection crashes and major and minor road ADT. They consider crashes by type at stop-controlled and signalized intersections. After some data smoothing, surfaces are developed to represent crash as a function of major and minor road ADT for each class of intersections. They find that the crash surface for urban signalized intersections in California contains a "ridge": for reasonably large major road volumes, as minor road ADT increases, crash counts rise to a maximum near 20,000 vehicles per day and then decrease for higher minor road traffic. Figure 23 (op. cit., p. 76) also shows a plateau and perhaps a ridge as major road ADT increases.

Special Studies

Pickering, Hall, and Grimmer (1986) study intersection crashes within 20 meters of rural tee intersections and within 100 meters of these intersections. They find that crashes from 20 to 100 meters away are three or four times as common as crashes on segments of similar length. Far from the intersection center, head-on crashes are more frequent; close to the center, turning crashes dominate. They raise the delicate issue of what an intersection-related crash really is.

Hauer and Hakkert (1988)[40] estimate that fatal crash counts are accurate to within 5%, serious injury crash counts to within 20%, and minor injury counts to within 50%. Reporting varies with the driver, the location, and the time. The count of fatalities can also vary with the quality and timeliness of medical attention, even with progress in medicine. Property damage crashes have threshold reporting requirements and are subject to inflation as repair costs rise. These considerations and similar ones are important caveats for modelers.

---

[38] M.Y.-K. Lau and A.D. May, *Accident Prediction Model Development: Signalized Intersections*, Research Report UCB-ITS-RR-88-7, Institute of Transportation Studies, University of California, Berkeley, Ca., 1988.

[39] H.C. Joksch and L.P. Kostyniuk, *Modeling Intersection Accident Counts and Traffic Volume*, Report No. FHWA-RD-98-096, Federal Highway Administration, McLean, Va., 1998.

[40] E. Hauer and A.S. Hakkert, "Extent and Some Implications of Incomplete Accident Reporting," *Transportation Research Record* 1185: 1-10, 1988.

The statistical abstract of Tessmer (1996)[41] reports that from 1975 to 1993, there were more than 420,000 fatal crashes in rural areas versus about 300,000 fatal ones in urban areas in the Fatal Accident Reporting System (FARS), despite fewer vehicle-miles driven (14.2 trillion versus 19.7 trillion (22.9 trillion versus 31.7 trillion vehicle-kilometers)). Also noted was the rural time delay in receiving medical attention. About 77% of the rural fatal crashes involved trucks versus about 62% of urban fatal crashes. A higher percentage of single-vehicle fatal crashes, and a lower percentage of multiple-vehicle crashes, occurred in rural settings than in urban settings in sampled States.

CONCLUSIONS

The issues in model development include: model form, choice of variables, and interpretation.

Models of the Poisson and negative binomial types, with mean a generalized linear function of covariates, have the dual virtues of being tractable computationally with present software and of capturing the discrete, random, non-negative integer character of crash counts. The log-linearity in these models also permits equations of traditional multiplicative types, and hence easy comparison with the results of earlier studies.

Although coefficients in both the Poisson and negative binomial types tend to be similar, the negative binomial has additional advantages. The presence of an overdispersion factor offers a way to account for omitted variables (the larger this parameter is, the more important such variables are). It also offers the possibility of combining the given model with empirical data from the past at a given intersection to obtain Bayesian refinements of the model predictions.

With regard to choice of variables, there is an infinity of possibilities, although resources are finite. Most of the variables discussed above are collected in this study, with the exception of weather. These variables further proliferate through mathematical transformations, e.g., composite measures of horizontal and vertical alignment near an intersection, or sight distance averages, or estimates of daily traffic by incoming and outgoing intersection leg. Transformations are suggested by past practice and common sense, but new combinations are always possible. In the analysis of the sample data in Chapter 4, correlations between crashes and variables are examined. These correlations, and successive ones found between residuals and variables, serve to select the variables used in the models. The selection should also be influenced by engineering judgment so that variables found to be important in the literature, or considered so by designers, receive full consideration.

Finally, there is the question of model interpretation. The studies above note that many factors influence crashes. However, a quantitative agreement on their relative importance has not been

---

[41] J.M.Tessmer, *Rural and Urban Crashes: A Comparative Analysis*, Report No. DOT-HS-808-450, U.S. Department of Transportation, National Highway Traffic Safety Administration Technical Report, Washington, D.C., 1996.

achieved. What a model can do, chiefly, is to summarize sample data. It can indicate which variables are most important with regard to the crashes on the sample intersections. Because of collinearity (i.e., two or more variables that are strongly dependent through design or coincidence), there is no guarantee that for variables present in the model, causation has been established. A model selects the variables that look "best" on the given data, and related variables may thereby be omitted. It is thus wise to identify families or clusters of variables that are related and tentatively view these families as the causal factors. Since families overlap, this task is not simple.

Using a model that summarizes to predict is best done with even more engineering judgment. The model summarizes a data set, but there are sampling and non-sampling errors in the data. Often what one wishes to predict has new or different factors influencing it. One is dealing with a moving target. Thus, judgment and some flexibility are in order.

3. DATA COLLECTION

The data collected in this study come from two primary sources: Highway Safety Information System (HSIS) files for California and Michigan, and field visits to the intersections made by Pragmatics personnel.

This chapter discusses the populations from which the data were selected, sample selection, data collection techniques, and data limitations.

THE POPULATIONS

The States and the Three Data Classes

An issue of early importance for this study was the selection of States in which to carry out the sampling. HSIS has extensive files for eight States - California, Illinois, Maine, Michigan, Minnesota, North Carolina, Utah, and Washington. File formats and contents vary from State to State. For three of the States - California, Michigan, and Minnesota - separate HSIS intersection files exist, while for another three - North Carolina, Utah, and Washington - there is no HSIS intersection information. Maine has node-and-link files (intersection-and-segment); Illinois treats intersections as segments of zero length. California gives details about signal characteristics; Illinois gives details about medians. Neither Illinois nor Michigan has minor road ADT available, except for cases in Michigan where the minor road, like the major road, is a State road.

The three intersection classes in this study were originally intended to be signalized three- and four-legged rural intersections of two-lane roads, along with four-legged rural intersections of a four-lane road with a two-lane stop-controlled minor road. However, examination of data bases for California, Michigan, and Minnesota indicated that there were very few signalized three-legged rural intersections of two-lane roads. The same indication came from information on three State routes in Washington. See Table 1. "Other" refers almost exclusively to stop-controlled on the minor road,

TABLE 1. Frequency of Signalized Rural Two-Lane Intersections in Four States

Rural two-lane roads	Three-legged intersections		Four-legged intersections	
	signalized	other	signalized	other
California - 1995	14 (0.2%)	6126 (99.8%)	35 (1.9%)	1832 (98.1%)
Michigan - 1994	16 (0.2%)	6513 (99.8%)	158 (4.1%)	3722 (95.9%)
Minnesota - 1992	4 (0.3%)	1307 (99.7%)	11 (0.7%)	1591 (99.3%)
Washington* - ca. 1993	2 (0.3%)	645 (99.7%)	10 (5%)	190 (95%)

*Routes 002, 009, 101 only.

but includes a few cases of flashers, and stop-controls on the major road. Of chief concern is not the low percentages, but rather the low absolute numbers, which might make acquisition of samples of adequate sizes difficult. Thus, these intersections were replaced by three-legged rural intersections with four-lane major roads and two-lane stop-controlled minor legs.

Table 1 reveals a similar, but less drastic, shortage of four-legged signalized rural intersections of two-lane roads. California has relatively few of them, especially for such a large State. On the other hand, Michigan appears from this table to have an adequate number for sampling.

In order to gain useful variety in the analysis, California and Michigan were chosen for the modeling effort, with the possibility, if resources permitted, of addition of a third State later.

Constraints imposed on the populations from which the samples were chosen were as follows:

1. Three-legged rural intersections, major road four-lane, minor leg two-lane stop-controlled: median width less than or equal to 36 feet (11 meters) on major road, all approaches two-way, stop-controlled on minor leg only.

2. Four-legged rural intersections, major road four-lane, minor legs two-lane stop-controlled: median width less than or equal to 36 feet (11 meters) on major road, all approaches two-way, stop-controlled on minor legs only.

3. Four-legged rural signalized intersections, major and minor roads two-lane: all approaches two-way.

Implementing these constraints was not completely straightforward. The California (CA) and Michigan (MI) HSIS intersection files had no information on whether intersections were rural or urban, nor on median widths, while MI's intersection file had no information on number of lanes. To obtain these items, the intersections were linked with segments in the CA and MI Roadlog files where such information was available.

For CA, a Roadlog variable entitled RU_IO was available to indicate whether the segment was rural, urbanized, or urban and inside a city or outside a city. For this study, we elected to use those marked as "rural, outside city" and did not include those that were rural, but inside or partly inside a city. The numbers for CA in Table 1 would have increased by only a small amount if other rural categories were added. For MI, a Roadlog variable entitled RURURB, with three rural categories (rural, rural dense small city, and rural small city boundary), was available. In the case of Michigan, all three categories were allowed. Roughly 50% of the Michigan intersections fell under "rural" and roughly 50% under "rural dense small city," and very few fell in the third category.

An intersection in CA or MI was considered rural if a neighboring segment was rural according to the classification above. In addition, in the case of Michigan, since the intersection file did not include a lane count, the major road was assumed to be either two-lane or four-lane, depending on how the segments adjacent to the intersection were described in the Michigan Roadlog file.

Pilot Studies

Pilot studies were conducted from bases in Sacramento, CA, and Lansing, MI, in March and May 1997, respectively, with a view to visiting all intersections sufficiently close to the State capitals as resources permitted.

In the case of California, 115 intersections in Districts 3, 4, 10, and the northern part of 6 (within approximately 250 miles (402 kilometers) of Sacramento) were qualified for membership in the populations on the basis of HSIS data. In Michigan, the pilot study concentrated on signalized intersections, and 66 such intersections were identified in Districts 5, 6, 7, 8, and 9 from the HSIS data base.

In both States, photologs were examined for all such intersections. If the photolog indicated that the intersection was not rural (e.g., curb parking, significant urban build-up for several blocks) or the lane count was incorrect or the signalization (several flashers were found that had been listed as fully signalized) or there was an adjacent intersection within 500 feet (152.4 meters) on the major road, then the intersection was eliminated. Thereafter, site visits were made to most of the intersections, additional intersections were eliminated by the site visit, and data were collected at the remaining ones. Even among those for which data were collected, in some cases, it was unclear whether they should be considered rural or urban. Table 2 indicates the disposition of the pilot study samples.

TABLE 2. Pilot Study Intersections in California and Michigan

	California	Michigan
sample units - three-legged	28	
sample units - four-legged	27	
sample units - signalized	10	23
Y intersections	6	
disqualified from photologs	25	13
disqualified from visits	13	17
too remote/isolated	6	13
Total	115	66

Table 2 reveals some difficulties that were to affect the entire study. Photologs did not match what was in the HSIS data base in a fair number of cases, and site visits revealed that changes not shown in the photologs had also taken place. This was particularly true of the Michigan signalized intersections. The Y intersections, three-legged intersections with two legs diverging from the third, were included in the pilot study, but it was later decided to eliminate them from the full data collection in part because of their relative rarity. A few intersections in both States were excluded from visits on the grounds that they were too remote or isolated.

The issue of how to handle remote and/or isolated intersection sites is a rather delicate one since it relates to both resource consumption and sample integrity. Rural intersections can be few and far between. To conserve resources, it is advisable to select intersections that are in close proximity to one another and to a suitable base of operation where junior highway engineers can be recruited for field work. With Sacramento or Lansing as a base, there were numbers of intersections each of which would require an overnight trip for two people, with driving time to and from and downtime between morning and evening traffic counts (if the site was not disqualified). While distances in California are well-known, it is less well-known that the distance from Lansing, Michigan, to the farthest reach of Michigan's Upper Peninsula, 550 miles (885 kilometers), is greater than Lansing's distance to New York City, about 500 miles (805 kilometers). Paradoxically, the most classically rural intersections, ones without suburban or small town features, are likely to be far from each other and far from suitable bases of operation and thus require disproportionate resources to visit. If intersections are close to each other, within a few miles, so that a team can visit several in the same day, the independence of the sample may be jeopardized. If they are close to a central point, such as a major city or the State capital, they are likely to be less rural and to be in transition.

During the pilot studies, in addition to examination of photologs and field work, construction plans and aerial photographs were reviewed, and the possibility of obtaining crash reports was investigated. Aerial photographs, photologs, and some (but not all) horizontal construction plans were available in the Traffic Operation Office at Caltrans headquarters in Sacramento. More complete computerized vertical and horizontal plans were not available, since the computer application that accessed them was undergoing major repair and renovation. At a later date, this system was running, but some plans were found to be missing and others were difficult to locate. District Offices in California, 11 in all, also have construction plans and hard-copy crash reports, but these offices are understaffed and the Project Team was told that retrieval would take much time. Caltrans personnel did indicate that the HSIS crash file for California would have numerous variables from which crash details could be reconstructed. Michigan had aerial photographs for many intersections in Southern Michigan taken in the years from 1972 to 1988, and some negatives for photos from prior years. Michigan also had a library of construction plans (maps, microfilms, and hanging files, depending on the year), although a fire in 1955 had destroyed some plans and others were misfiled. Road segments will have as many as 50 jobs and corresponding plans. For a minor job, the plan will not show the alignments of the road. The Project Team was told that, in Michigan, confidentiality laws make crash reports difficult to obtain since preliminary deletions by State employees are required. Although Michigan photologs were one cycle more recent than those of FHWA in McLean, Virginia, numerous discrepancies were found among the HSIS files, the photologs, and site visit observations.

In both the California and Michigan pilot studies, field work was done at all intersections. The typical routine was a morning site visit to hand-count traffic (on specially designed tally sheets). Thereafter, radar guns were used to determine operating speeds for samples of vehicles approaching the intersection on each leg. These measurements were made at discreetly placed locations before vehicles began to slow for the intersection. Speeds would be determined only for the lead car in a platoon, and the angle between the radar path and the direction of vehicle travel was noted to permit calculation of true travel speed. Typically, 25 measurements would be made if the leg traffic was adequate. If the traffic was light, as many measurements as a 15- to 20-minute stay would permit would be made. Measuring wheels were used to pace off sight distances. Other intersection features and geometry were recorded, as well as signal characteristics at signalized intersections. In the late afternoon, a second traffic count would be done. In the case of Michigan, where only signalized intersections were visited, computerized plate counters were also used to measure minor leg traffic. The plate counters were nailed to the minor road at mid-day and left there for 24 hours. They were recovered on a subsequent visit to the intersection and unwrapped. Data were downloaded from them and they were recharged and rewrapped for the next count. Three people were required for placement of the plate counters since traffic had to be disrupted. For all site visits, permits were required from District Offices, and safety precautions, including wearing of hardhats and orange vests, and placement of cones and signs, were taken.

Pilot study data were subsequently used to prepare some small special studies. Three kinds of speed data were compared: posted speeds obtained by inspection along intersection legs, operating speeds measured by radar guns, and speeds recorded by the plate counters. The plate counters also permitted a determination of 24-hour truck percentages, and these could be compared with observed peak-hour truck percentages from the manual traffic counts. To assess the "intersection-relatedness" of crashes, a review was also undertaken on the area of influence of an intersection for a few pilot study intersections. The results of these investigations are reported in the appendix to this report.

SAMPLE SELECTION

After both pilot studies were completed, the studies were assessed and plans were made for the subsequent main data collection effort. The chief decisions made were to restrict attention to tee intersections and omit Y intersections,[42] to measure horizontal and vertical alignments at each intersection rather than attempt to extract this information from plans or photos, to discontinue the minor leg plate counts, and to follow an informal sample selection plan.

---

[42] A three-legged intersection is a T intersection (or tee) when "two of the three intersection legs form a through road and the angle of intersection is not acute"; it is a Y intersection (or wye) when "all three intersection legs have a through character or the intersection angle with the third intersection leg is small." These definitions are taken from p. 836 of *A Policy on Geometric Design of Highways and Intersections* (also known as the "Green Book"), American Association of State Highway and Transportation Officials, Washington, D.C., 1994.

Three-legged intersections for the main data collection effort were restricted to T intersections because of the relative scarcity of Y intersections and in the interest of sample homogeneity. Sample homogeneity contributes to successful modeling by removing variables that will not be modeled. However, such homogeneity can only be achieved to a limited extent. Too many restrictions (e.g., requiring that all intersections have lighting, that they all have medians of a certain type, that they have ADT in a certain narrow range, etc.) can be counterproductive. There may be too few intersections meeting all the constraints to be useful for modeling, and data collection from all of them to maximize sample size may be too expensive if they are geographically dispersed. The distinction between T and Y intersections is of recognized importance, T intersections are favored by intersection designers, and restricting the sample to T intersections was judged to pose no problem.

Collection of alignment data during the field work and discontinuation of minor leg plate counts were undertaken for reasons of economy.

In both Michigan and California, the availability and accessibility of plans showing recorded alignments were in doubt. Plans and photos were sought for sub-samples of the pilot study samples in both States, and for roughly 30% of the intersections in the sub-samples, no information could be found. Since morning and afternoon traffic counts were to be done at each intersection, acquisition of alignment data at midday did not seem to be unduly burdensome for field workers. At intersections visited in the pilot studies, alignment data had not been collected, but revisits during other field work could be done without hardship.

With respect to the plate counters, the pilot study revealed that they provided good data, but that they were resource-intensive and that the data were not essential to the overall effort. The plate counters are HISTAR units that detect changes in the magnetic field above the roadway. The associated computer data are generated and printed with NU-METRICS software. Information available includes: counts of incoming vehicles by type, counts of occupants per vehicle, vehicle speeds, weather conditions (temperature and precipitation), and gaps between vehicle arrival times. During the Michigan pilot study, the weather variables did not seem reliable, and the counters did not work properly on occasion. At two intersections, they were placed on major roads for which ADT was available. The count at one of these roads was 4,400 vehicles per day versus 6,400 vehicles per day according to HSIS files. The difference is that the plate count data is for one day in 1997 and the HSIS data is a State estimate for 1993-1995. The plate counters do not determine turning movements, and manual counts still have to be done at each intersection to obtain these. The plate counts, as already noted, require two visits spaced 24 hours apart with adequate personnel to ensure safety during placement and removal. Thus, to conserve resources, they were omitted in the main data collection effort. For Michigan intersections where minor road ADT was not available, major road ADT plus 1997 peak-hour traffic counts, in particular ratios of traffic by movement, were used to estimate minor road ADT. This method, while making use of 1997 data to estimate minor road ADT for earlier years, is arguably more reliable than using an absolute 1997 count.

Both pilot studies revealed that field work at rural intersections is time-consuming. Rural intersections suitable for the study, especially signalized ones, tended to be few and far between, as noted earlier. In some cases, overnight lodgings were required both before and after in order for Project Team members to get to a site at an adequately early hour and remain there until an adequately late hour. During the Michigan pilot study, a number of intersections thought to be in the population on the basis of HSIS files and Michigan photologs were found to be unqualified at a site visit, primarily because they were not at all rural.

An informal sampling plan, as follows, was developed. Complete lists of intersections whose HSIS records satisfied the constraints had been developed. All intersections within approximately a 3- to 4-hour drive from Sacramento or Los Angeles or Lansing were automatically included in the sample, together with a few other selected intersections at farther distances. Photologs were reviewed for all of these, some were disqualified as a result, and with the exception of those that had been in the pilot study and a few especially remote ones in California, all of the remaining ones were pre-visited prior to the field work. The purpose of the pre-visit was to ascertain whether each intersection was, in fact, qualified — no legs or medians closed, no offsets, no additional lanes or legs, number of lanes unchanging out to ±800 feet (243.8 meters), no urbanization, with signalization or signage as advertised. In addition, a large number of intersections were eliminated because they were too close to other intersections of the same type and were likely to have strongly correlated data values. This was especially true for the three-legged and four-legged non-signalized intersections. In both States, such intersections tended to be grouped on a relatively small number of highways and tended to be placed along these highways in close sequence.

Pre-visits by senior Project Team members were found to be very useful since field workers would not spend unnecessary time at unqualified intersections and the senior members of the team could make experienced judgments about the appropriateness of intersections.

The final samples, including pilot study observations, are shown in Table 3.

TABLE 3. Samples as Proportions of Nominal Populations

	CA	MI	Total
3-legged	60/302 (19.9%)	24/93 (25.8%)	84/395 (21.3%)
4-legged	54/150 (36%)	18/49 (36.7%)	72/199 (36.2%)
Signalized	18/27 (66.7%)	31/100 (31%)	49/127 (38.6%)

The first number in Table 3 is the sample size and the second is the nominal population size in the State. These numbers are adjusted from Table 1 by elimination of duplicate observations and photolog reviews, but the denominators include numerous dependent intersections and, especially in the Michigan signalized case, intersections that are no longer rural. The denominators also include remote intersections that were not visited for lack of resources. It should be kept in mind that some of these would have been disqualified if they had been visited.

The informal sample selection method raises the issue of representativeness. It should be noted that crash data were not consulted in selecting the samples, but that there was some tendency to favor larger ADT intersections or ones with more irregular alignment when, for example, only one of two nearby intersections could be chosen because of dependence. In addition, many, but not all, of the most remote intersections in both States were omitted from the samples.

DATA COLLECTED

The data collected in this study and the sources are shown in Table 4.

Highway Safety Information System (HSIS) Data

Average Daily Traffic (ADT) data and crash data were extracted from HSIS files.

ADT data were extracted from HSIS Intersection and Roadlog files. For California, major and minor road ADT were available in HSIS intersection files for the years 1993, 1994, and 1995. For Michigan, ADT data were available in HSIS Roadlog files for segments of State roads, although 1993 data were unavailable and had to be interpolated from 1992 data.

HSIS crash variables for 1993, 1994, and 1995 were consulted. These include Accident Location variables, Accident Number, Accident Severity, Accident Type, Number of Vehicles, Vehicle Motion Prior to Accident (MISCACT1). All variables, but the last, are in the HSIS Accident file for the State. The last is in the HSIS Vehicle file.

Traffic-Count Variables

For all intersections in the study, field counts were done on traffic during morning and evening hours. Due to limited resources, the counts were not done at a fixed time, but were typically done in the morning for about 45 minutes between 7:00 a.m. and 9:30 a.m. and in the afternoon between 3:30 p.m. and 6:00 p.m. The counts were done on non-holiday weekdays.

In a few cases in California, no traffic was seen emanating from the minor road during the hours of visitation. This happened at two three-legged intersections and at two four-legged intersections, all of them in California on Route 395. The first two intersections had incoming traffic, but the last two had no traffic, incoming or outgoing. These intersections are in high-altitude regions near Mono Lake and Independence, the counts were made in the fall of 1997, and traffic may have been reduced for seasonal reasons.

TABLE 4. Variables Collected in the Study

	Variable	Meaning	Units	Source
Identifiers	cnty_rte	California (CA) route identifier		HSIS
	cntl_sec	Michigan (MI) route identifier		HSIS
	milepost	intersection center milepost along route	miles	HSIS
Traffic 1993-95 (MI 1992 instead of MI 1993)	ADT1	CA average daily traffic on major road at intersection by year	vehicles per day	HSIS
	ADT2	CA average daily traffic on minor road at intersection by year	vehicles per day	HSIS
	ADTM	MI average daily traffic on adjacent segment of State road by year	vehicles per day	HSIS
Peak Traffic	RAWMPCij	no. of cars traveling from leg i to leg j in morning count period	vehicles	Field
	RAWMTRij	no. of trucks traveling from leg i to leg j in morning count period	vehicles	Field
	M_HR	duration of morning count period	hours	Field
	MBEG	start time of morning count period	clock-hours	Field
	RAWEPCij	no. of cars traveling from leg i to leg j in evening count period	vehicles	Field
	RAWETRij	no. of trucks traveling from leg i to leg j in evening count period	vehicles	Field
	E_HR	duration of evening count period	hours	Field
	EBEG	start time of evening count period	clock-hours	Field
Roadside	HAZRAT	Roadside Hazard Rating within ±250 ft of intersection center on major road	1, 2, 3, 4, 5, 6, 7	Field
	NODRWYR1	no. of residential driveways within ±250 ft of intersection on major road	0, 1, ...	Field
	NODRWYC1	no. of commercial driveways within ±250 ft of intersection on major road	0, 1, ...	Field
	NODRWYR2 NODRWYC2	minor road counterparts for signalized intersections only	0, 1, ...	Field

1 mi = 1.61 km, 1 ft = 0.305 m

TABLE 4. Variables Collected in the Study (continued)

	Variable	Meaning	Units	Source
Channel-ization	LTLN1	no. of left-turn lanes on major road	0, 1, or 2	Field
	RTLN1	no. of right-turn lanes on major road	0, 1, or 2	Field
	LTLN2	no. of left-turn lanes on minor road	0, 1, or 2	Field
	RTLN2	no. of right-turn lanes on minor road	0, 1, or 2	Field
Intersection Geometry	MEDWIDTH1	median width on major road	feet	Field
	MEDTYPE	median type of major road	none, curbed, painted, other	Field
	DRINCMP	direction of increasing mileposts along major road	E for east, N for north	Field
	ANGLEi	angle between increas. dir. of major road and left leg (i =1) or right leg (i=2)	degrees	Field
Sight Distances	SDi	longitudinal sight distance along leg i of major road, i = 1 or 2, or of minor road (signalized int. only), i = 3 or 4	feet	Field
	SDLi	left sight distance along leg i of minor road, i = 3 or 4, or of major road (signalized int. only), leg = 1 or 2	feet	Field
	SDRi	right sight distance along leg i of minor road, i = 3 or 4	feet	Field
Horizontal alignment on major road (and minor road of signalized inter-sections)	HBi	beginning point of curve no. i (if any portion of curve is within ±800 ft of intersection center along major road)	feet ± from intersection center	Plans
	HEi	end point of curve no. i	feet ±	Plans
	DEGHi	degree of curve, curve no. i	degrees per hundred feet	Field
	DIRi	direction along increasing direction of major/minor road of curve no. i	L for left, R for right	Field

1 ft = 0.305 m

TABLE 4. Variables Collected in the Study (continued)

	Variable	Meaning	Units	Source
Vertical alignment on major road (and minor road of signalized inter-sections)	VBi	beginning point of curve no. i (if any portion of curve is within ±800 ft of intersection center along major road)	feet ± from intersection center	Field
	VEi	end point of curve no.	feet ±	Field
	GBi	grade prior to curve no. i	%	Field
	GEi	grade after curve no. i	%	Field
	GRADE1	grade of major road, if only one	%	Field
Speed Limits	SPDLIMi	posted regulatory speed on leg i, if seen	mph	Field
	POSTADVi	posted advisory speed on leg i, if seen	mph	Field
Signali-zation	SIG_TYPE	signal type - pre-timed, actuated, or semi-actuated		Field
	PROT_LT	protected left turn - multiphasing; 1 for yes, 0 for no	0, 1	Field
Miscel-laneous	LIGHT	1 if lighting is present, 0 if not	0, 1	Field
	terrain	flat, rolling, or mountainous		Field
Crash data 1993-1995	TOTACC	no. of crashes occurring at intersection or within ±250 feet of intersection on major road during 1993-95	0, 1, ...	HSIS
	fatal, injury, propdam	no. of fatal, injury, property damage only crashes, respectively	0, 1, ...	HSIS
	head-on, sideswipe, rear end, broadside, hit object, overturned, pedestrian, other	no. of head-on, sideswipe, rear end, broadside, hit object, overturned, pedestrian, or other crashes, respectively	0, 1, ...	HSIS
	MISCACT1	movement of vehicle in crash prior to crash (left turn, etc.)		HSIS

1 mph = 1.61 km/h, 1 ft = 0.305 m

During the counts, the number of passenger vehicles and the number of trucks entering and leaving the intersection were recorded, along with the incoming and outgoing legs. The beginning and ending times of the counts were also recorded. A typical duration was 45 minutes. When the data were processed later at Pragmatics, Inc., all counts were converted to hourly counts. Intersection legs were identified by leg numbers, in the clockwise order 1, 3, 2, 4 shown in Figure 1. Legs numbered 1 and 2 are on the major road; from leg 1 to leg 2 is the increasing milepost direction. Legs 3 and/or 4 are on the minor road, with leg 3 to the left of the major road's increasing direction. For traffic going from leg number i to leg number j, the morning counts were M_PCij and M_TRij in vehicles per hour, while the evening counts were E_PCij and E_TRij. The distinction between passenger vehicles and commercial vehicles/trucks was based on the number of tires. A commercial vehicle was taken to be any vehicle with more than four tires, and included cars with trailers. This almost always meant a vehicle with more than two axles.

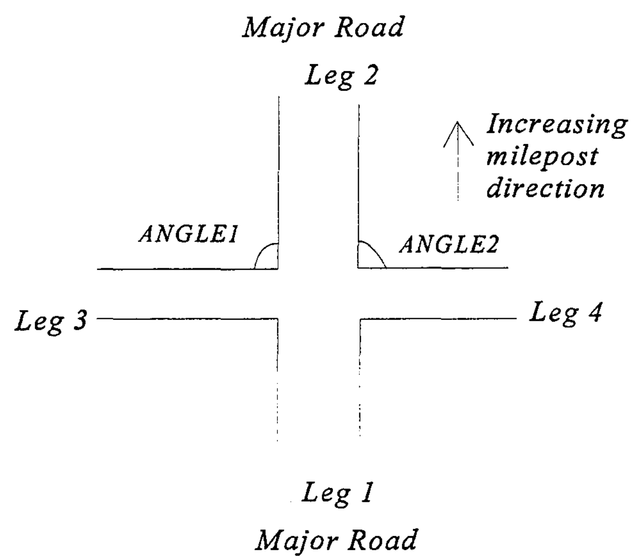


FIGURE 1. Intersection Diagram Showing Leg Numbers

Roadside Variables

During the field work, the roadside variables Number of Driveways and Roadside Hazard Rating (HAZRAT) were collected by inspection.

The number of driveways within 250 feet (76 meters) of the intersection center was counted along the major road. Residential and commercial driveways were counted separately. A gas station with two entranceways would be counted as having two commercial driveways. For signalized intersections, the number of driveways was also counted on the minor road out to 250 feet (76 meters).

HAZRAT is a variable devised by Zegeer et al. (1987)[43] that is an amalgam of sideslope, clear zone, and distance to nearest hard object. It takes whole number values from 1 to 7, with 7 being the most hazardous. Field workers were provided with images of typical roadsides with different ratings, and at site visits, they made estimates of the average rating of the major road within 250 feet (76 meters) of the intersection center.

Channelization and Intersection Geometry

At each intersection, field workers recorded left- and right-turning lanes on all approaches, median widths and characteristics, and intersection angles. At a three-legged intersection, the number of left-turn lanes on the major road, or right-turn lanes, is always 0 or 1, and likewise on the minor road. At a four-legged intersection, signalized or not, the number of left-turn lanes on each road, or right-turn lanes, is 0, 1, or 2. The measured intersection angles, ANGLE1 and ANGLE2, are between legs 2 and 3 and between legs 2 and 4, respectively. See Figure 1. In California, intersections are squared up by policy, i.e., although the basic angle between the major and minor roads may be substantially different from 90 degrees, the minor road will curve sharply within a few car lengths of the intersection to create a right angle. Field workers were instructed to record the large-scale angle of the approach when very sharp curves of this type were present.

Sight Distances

Sight distances were estimated longitudinally on the major road and left and right on each minor leg. At three-legged intersections, the longitudinal sight distance was only measured in one direction, e.g., if the third leg was leg 3 in Figure 1, then the sight distance from leg 1 to leg 2 was measured, but not from leg 2 to leg 1. Likewise, at signalized intersections, left sight distances were not measured. For the signalized intersections, longitudinal and left sight distances were estimated on all legs. When a protected left turn exists from leg 1 to leg 3, one may argue that longitudinal sight distance from leg 1 to leg 2 is unimportant.

The Green Book (1994, p. 702) recommends that left and right intersection sight distances from the minor road be measured at 6 meters (20 feet) from the edge of the traveled way. At many intersections, this yields very little sight distance, and only a foolhardy driver would decide to enter the intersection from this location. An alternative standard is 3 meters (10 feet) from the edge of the traveled way, approximately the location of a seated driver prior to entering the intersection. The latter standard has apparently been adopted by many States, and is the one that was used in measurements here. For longitudinal sight distance (along the major road from one lane to the opposing lane), measurements were made from the edge of the traveled way of the minor road in the leftmost incoming lane of the major road. The driver's eye was assumed to be at a height of 1070 millimeters (3.5 feet) and the object viewed was assumed to have a height of 1300 millimeters (4.25

---

[43] C.V. Zegeer, J. Hummer, D. Reinfuhrt, L. Herf, and W. Hunter, *Safety Cost-Effectiveness of Incremental Changes in Cross-Section Design — Information Guide*, Report No. FHWA-RD-87-094, Federal Highway Administration, Washington, D.C., 1987.

feet). See the Green Book, pp. 136-7.

Sight distances, if they were sufficiently short, were paced off with a measuring wheel to record the distance. If they were many hundreds of feet long, they were estimated with a range finder. The latter is an optical device with two light paths from the distant object to the eyepiece. Then a dial is turned until the two images of the object merge, and a distance can be read from the dial.

Horizontal Alignment

Horizontal curves were recorded for the major road and, in the case of signalized intersections, for the minor road. A segment from 800 feet (244 meters) before the intersection to 800 feet after the intersection was determined, and any horizontal curve that overlapped this segment was included. For each such horizontal curve, the beginning and end points were noted, along with the direction of curvature and the degree of curve. Measuring wheels and chalk were used to determine beginning points and endpoints. Degree of curve was measured by marking off a straight line distance, typically 100 feet (30.5 meters), between two points at the edge of the traveled way, and calculating the perpendicular distance at the midpoint to the edge of the traveled way. The degree of curve, in degrees per 100 feet (30.5 meters), is then calculated from the formula:

$$DEGH = \frac{18000 \times 8 \times H}{\pi \times (4H^2 + L^2)}$$

where L is the length of the straight line in feet and H is the perpendicular distance in feet. (The metric equivalent is $DEGH_m = DEG/0.305$ in degrees per 100 meters.) No adjustment was made for the roadway width. Even on a four-lane road, an adjustment that replaces the edge of the traveled way by the centerline of the road would typically change the value by no more than a few percent.

Vertical Alignment

As with horizontal curves, vertical curves were recorded that overlapped a segment out to ± 800 feet (244 meters) from the intersection center along the major road and, for signalized intersections, along the minor road. Beginning points and endpoints of each vertical curve were determined with measuring wheels and chalk. Then, incoming and outgoing grades were estimated at the beginning and end of each curve. Grades were considered positive if they were uphill in the direction from leg 1 to leg 2, the increasing direction of the major road, or from leg 3 to leg 4 along the minor road. For any intersection that had no vertical curves, a unique grade, GRADE1, was reported.

Grades were measured in one of two ways. An optical level and a measuring rod were sometimes used. A distance of 25 feet (7.6 meters) or so would be paced off along the edge of the traveled way. A marked height at that distance would be compared with the corresponding height on a measuring rod determined by sighting the optical level horizontally. The difference in height divided by the horizontal distance yields the slope. An alternative method was to place a 4-foot (1.2-meter) level along the roadway (or along a flat board on the roadway) and record the slope directly from a display.

Other Variables

Posted advisory and regulatory speeds were recorded for each leg when seen within a few thousand feet, existence of lighting at the intersection was noted, and a qualitative measure of terrain (flat, rolling, or mountainous) was also noted. At signalized intersections, it was noted whether the signal appeared to be pre-timed, actuated, or semi-actuated. Protected left turns on the major road were also noted, but no record was made of which pairs of legs had such protection. A reasonable assumption is that the left-turn movement from the major road leg with the highest volume, either leg 1 to leg 3 or leg 2 to leg 4, had such protection when PROT_LT equals 1 and no left turns were protected when PROT_LT equals 0.

DATA LIMITATIONS

HSIS Data

The HSIS variables are ADT and crash data for the years 1993, 1994, and 1995.

California ADT data are determined systematically and regularly on State roads through 400 permanent continuous operation count stations and another 1,700 permanent stations that are used once every 3 years. Intersection major road ADT is based on the segment ADT. Minor road ADT is generally estimated rather than counted, is done by the Districts, and is thought to be older and of lesser quality. Michigan has about 120 permanent count stations, not all on State roads, and attempts to do counts on each State road once every 3 years. It does not have ADT for minor roads unless they are State roads.

Crash data for both States are subject to the limitations noted in the study of Hakkert and Hauer (1988). Many Michigan property damage only crashes, and some injury crashes, are reported by the driver without an officer at the scene. Not only are there issues of underreporting and classification for both States, but there is also the question of crash location. Some Michigan observers think that crash locations are often incorrect, and mention examples where a crash was attributed not just to the wrong milepost, but to the wrong intersection.

Field Data

The traffic data collected in the field during this study have obvious limitations. They were collected on a single weekday in a particular season of the year and during a short time period in nominal peak morning and evening hours. Field workers reported that in different locations, the traffic volumes might be especially high early or late in the morning or evening, depending on such factors as the presence of a manufacturing plant versus a shopping center nearby. The true definition of peak hour varies from location to location, while this study had to follow visitation timetables based on available resources. In California, some rural intersections had relatively low traffic, reflecting likely seasonal variations at resorts and camping areas such as Lake Tahoe. The site visits were conducted in late fall and early spring at some of these locations. No attempt was made to adjust the data to take into account such variability. Yet another limitation is that these data were collected in 1997 for use in modeling 1993-1995 events.

Variables such as HAZRAT, number of driveways, channelization, angle, speed limits, sight distances, and horizontal and vertical alignments were also measured in 1997 and are presumed to be valid for the earlier time period. These items, however, tend to be much more stable than traffic movements, and temporal variation is not thought to be a significant source of error.

HAZRAT is a subjective rating of roadside hazards. The measure is supposed to average the hazards alongside the major road within ±250 feet (76 meters) of the intersection. Typically, two experienced observers will agree on a value or differ by 1, e.g., one observer may assign a 3 and the other a 2.

Sight distances, as noted, were measured with a measuring wheel or a range finder. Because of the limitations of the range finder and some subjectivity about when an object becomes visible (seeing something versus recognizing what it is), sight distances are likely to be accurate to within roughly 10%. For the purposes of this study, sight distances in excess of 1800 feet (550 meters) or more were not distinguished, and any sight distance thought to be in excess of 2000 feet (610 meters) was generally marked as 2000 feet. A sight distance of 1600 feet (488 meters) would be noticeably smaller, and absolute accuracy would improve as sight distances decrease.

Horizontal and vertical curves present unique difficulties. For many rural roads, the line of a highway is quite irregular when examined on a small scale. Potholes, bumps, and other small irregularities due to the lay of the land or due to wear caused by traffic and weather are often present. Field workers were asked to idealize roadways by smoothing road lines out to scales of several vehicle lengths. Decisions about where a curve begins and ends are thus to some extent arbitrary, particularly for curves of large radius or small grade. Beginning points and endpoints as judged by two different observers might differ by as much as 20 feet (6 meters), while degree of curve might vary by 5% or more. Michigan was relatively flat, with many grades less than 1%. Vertical grades of less than 1% were probably measured to no greater accuracy than ±0.25%, so that a grade listed as 0.5% might be 0.25% or 0.75%. A much larger grade, say 5%, would be accurate to within ±0.5%. Differences in successive grades accompanying a vertical curve would have about the same accuracy since the observers would be sensitive to the change of grade.

Perhaps the greatest limitation of the data is that they do not reflect the special circumstances of each intersection. When individuals are classified by such conventional (and imperfect) measures as age, height, weight, sex, IQ, race, etc., sometimes the most important and most relevant points are missed. Site visits reveal that the intersections in this study are quite diverse, with very individual personalities. Significant items that would not appear in a data base are quite common, e.g., a nearby amusement park, beach turn-offs along Lake Huron, canyon roads off of Pacific coastal highways, sideroads into California deserts, resort areas such as Lake Tahoe and Squaw Valley, small towns where a two-lane road flares out to four lanes for a few thousand feet or where a two-lane road abruptly arrives at a single signalized intersection, or rural intersections along heavily trafficked commuter highways connecting big cities to rural homesites.

Signalized rural intersections, in particular, are in transition. The signal is often in place because of increasing local development and increased minor road traffic. With increasing traffic come more businesses and residences, and soon a very rural area becomes a small town and a small town becomes a city.

Analysis and modeling are bound to be inexact because the population under study is a moving target, and qualitative changes can overtake the quantitative ones, bringing unforeseen variables into prominence.

4. ANALYSIS

The analysis consists of developing a variety of new variables derived from the variables collected, determining the statistics for new and old variables singly and jointly, determining correlations between variables (especially between crash counts and other variables), and studying the chief relationships found.

Of particular interest is the relationship between crash counts and traffic. Without question, average daily traffic (ADT) on all approaches is a significant (and usually the most significant) predictor of crashes. Not only does greater traffic imply greater numbers of crash-prone drivers, even with the percentage of crash-prone drivers assumed to be independent of traffic or increasing with traffic, but for multiple-vehicle crashes at intersections, an adequate amount of traffic is a necessary condition for a crash.

Successive sections of this chapter treat new variables, univariate statistics, bivariate statistics and correlations, and the relationship between crash counts and traffic.

NEW VARIABLES

The chief classes of variables in this study are: crash variables, traffic variables, intersection geometric variables, roadside variables, alignment variables, and sight distances. The intersection geometric variables concern medians, channelization, and intersection angle. Alignment variables and sight distance variables, which pertain to the roadway as far out as 800 feet (244 meters) to several thousand feet from the intersection center, are treated separately.

Crash Variables

The chief crash variable is TOTACC. This is the total number of crashes occurring at the intersection in the years 1993, 1994, and 1995. Any crash occurring at the intersection or within 250 feet (76 meters) of the intersection center along the major road is included in this number. Crashes occurring along the minor road near the intersection are recorded as being at the intersection (if within 100 feet (30.5 meters) of the intersection center in Michigan, if within 250 feet (76 meters) in California). One exception to this is when the minor road is a State road (the major road is always a State road). This happens for some signalized intersections. In such cases, Accident files for the minor road were also consulted and all crashes within 250 feet (76 meters) of the intersection center along the minor State road were included.

A second crash variable is TOTACCI. For this variable, criteria proposed by Bellomo-McGee, Inc. (BMI) were used to restrict the crashes to ones considered intersection-related. Michigan's HSIS Accident file has a variable called Highway Area Type that indicates whether a crash occurred in the vicinity of an intersection. This perhaps could have been used to establish intersection-relatedness. However, California has no similar variable. Indeed, an important modeling issue is to establish criteria for intersection-related crashes that are uniform from State to State. A set of criteria with this aim were prepared by Warren Hughes and A.J. Nedzesky of BMI, with the assistance of Forrest Council, and were submitted to FHWA in a memo dated March 26, 1998. The BMI criteria are the following: (1) crashes must occur within 250 feet (76 meters) of the intersection center and (2) they must be (a) vehicle-pedestrian crashes; (b) crashes in which one vehicle involved in the crash is making a left turn, right turn, or U-turn prior to the crash; or (c) multiple-vehicle crashes in which the accident type is either sideswipe, rear end, or broadside/angle.

Applying these criteria in California and Michigan was not completely straightforward. Minor road crashes could sometimes only be obtained out to a lesser distance, as noted above, because of the recording methods of the States. The California data base is silent on whether crashes, including turning crashes, may or may not involve driveways, while Michigan has separate categories for some crashes involving driveways (e.g., "angle driveway"). For accident type, California uses the term "broadside," while Michigan uses the terms "angle straight" and "angle turn." California does not distinguish between "sideswipe same" and "sideswipe opposite," whereas Michigan does. The precise criteria used in the two States, apart from location as specified in TOTACC, were:

CALIFORNIA

Some vehicle in the crash had MISCACT1 (Motion preceding collision) equal to "making right turn," "making left turn," or "making U turn";

or

ACCTYPE (Type of collision) was "Auto-pedestrian";

or

VEH_INVOL (Motor vehicles involved with) was "Pedestrian";

or

VEH_INVOL was "Other motor vehicle" or "Motor vehicle on other roadway," and ACCTYPE was "Sideswipe" or "Rear end" or "Broadside."

MICHIGAN

ANALYS (Accident analysis) was "Motor vehicle/motor vehicle," and ACCTYPE (Accident type) was "Head-on" or "Sideswipe opposite," and MISCACT1 (Driver intent) for some vehicle in the crash was "Make right turn," "Make left turn," or "Make U turn";

or

ANALYS was "Auto-pedestrian";

or

ANALYS was "Motor vehicle/motor vehicle," and ACCTYPE was "Angle straight," "Rear end," "Angle turn," "Sideswipe same," "Rear end left turn," "Rear end right turn," "Head-on left turn," "Dual left turn," or "Dual right turn."

From these comparisons, it is evident that the problem of uniformity among States also arises when multiple data fields are used to ascertain whether an crash is intersection-related. The data fields and associated definitions do not always match up precisely.

Yet another problem is that the BMI criteria were developed for use with two-lane rural roads. The present study, in part, concerns four-lane rural roads and it is not clear that the same criteria should be used for them. Observers have also raised the issue of whether different criteria should be used for signalized versus non-signalized intersections of two-lane rural roads.

Four other crash variables were developed for this study. Their definitions are given below:

$$'NJACC = Aall\ accidents\ with\ fatalities,\ injuries,\ or\ possible\ injuries\ counted\ in\ TOTACC$$

$$INJACCI = All\ accidents\ with\ fatalities,\ injuries,\ or\ possible\ injuries\ counted\ in\ TOTACC$$

$$TOTACCS = All\ single\text{-}vehicle\ accidents\ counted\ in\ TOTACC$$

$$TOTACCM = All\ multiple\text{-}vehicle\ accidents\ counted\ in\ TOTACC$$

The first two variables, INJACC and INJACCI, exclude crashes in which only property damage occurred, but include all others. In California, one of the severity categories is "Complaint of pain." In the time period 1993 through 1995, the reporting threshold for property damage only crashes was $400 in Michigan and $500 in California. The last two variables, TOTACCS and TOTACCM, were determined for the signalized intersections only, and were used in some of the modeling to relate crashes to traffic flows by leg.

ADT Variables

Two average daily traffic variables were used in this study - ADT1 and ADT2. ADT1 is estimated average daily two-way traffic on the major road measured in vehicles per day (vpd) in the vicinity of the intersection for the 3 years 1993, 1994, and 1995. ADT2 is the estimated average daily two-way traffic for the minor road in this period.

For California, ADT1 and ADT2 were obtained by taking annual figures provided in the HSIS intersection files, summing them, and dividing by three.

For Michigan, ADT data were not available in the HSIS intersection file. However, ADT data were available in the HSIS Roadlog file for State roads in the years 1992, 1994, and 1995. The values of ADT for these years were interpolated to obtain a value for the year 1993, and the values for 1993, 1994, and 1995 were averaged. These estimated ADT values were for segments of roads. Then, ADT on segments of the major road adjacent to intersections in the study were averaged to yield ADT1. In some cases (about 20% of the Michigan intersections), the minor road was also a State road, and ADT2 could be obtained in the same way. In all other cases, ADT2 was estimated on the basis of morning and evening traffic counts done by Pragmatics, Inc. (see below). An average morning-hour traffic count (incoming plus outgoing) was determined for each leg, converted into a fraction of all incoming and outgoing traffic; the same was done for evening traffic, and the two fractions were averaged. Then this fraction was applied to the known estimated ADT on the two legs of the major road to obtain an estimated ADT for each minor leg. ADT2 was this value if there was only one minor leg, and it was the average of the values for the two minor legs otherwise. This method has two evident defects: it only represents peak-hour ADT, and a sample of such at that, and it was done in the year 1997 rather than the study years. Nonetheless, it probably has the correct order of magnitude and may well be as reliable as other minor road ADT estimates in the HSIS files.

In the case of the signalized intersections, the decision about which of two two-lane roads at a four-legged intersection is major and which is minor was based on ADT. The one with the higher ADT is defined to be the major road, and the other the minor road. In general, the major road is the State road, but in Michigan, sometimes both roads are State roads and thus the ADT criterion is used to declare one of them to be the major road. There are three cases, two in Michigan and one in California, where the State road has a lower ADT than the other road, a county or local road. In these three cases, the other road is taken to be the major road, its ADT is ADT1, and its legs are taken to be legs 1 and 2.

Variables Derived From Traffic Counts

Traffic count data were converted into hourly form so that for each ordered pair of approaches (i,j), an estimated number of vehicles per hour was given traveling from leg i to leg j. This was calculated for passenger vehicles and trucks separately and for a morning and evening hour separately.

$$
\begin{aligned}
M\_PCij &= \frac{RAWMPCij}{M\_HR} \\
M\_TRij &= \frac{RAWMTRij}{M\_HR} \\
E\_PCij &= \frac{RAWEPCij}{E\_HR} \\
E\_TRij &= \frac{RAWETRij}{E\_HR}
\end{aligned}
$$

A rather large variety of variables can be derived from such quantities. For the present study, selected variables shown below were developed.

Commercial or truck percentage was measured by three variables, AM%TRUCK, PM%TRUCK, and PK%TRUCK, representing the morning, evening, and combined morning and evening percentages of truck traffic passing through the intersection. These are defined as follows:

$$AM\%TRUCK = \frac{\sum_{all\ pairs\ (i,j)} M\_TRij}{\sum_{all\ pairs\ (i,j)} (M\_TRij + M\_PCij)} \times 100$$

$$PM\%TRUCK = \frac{\sum_{all\ pairs\ (i,j)} E\_TRij}{\sum_{all\ pairs\ (i,j)} (E\_TRij + E\_PCij)} \times 100$$

$$PK\%TRUCK = \frac{\sum_{all\ pairs\ (i,j)} (M\_TRij + E\_TRij)}{\sum_{all\ pairs\ (i,j)} (M\_TRij + M\_PCij + E\_TRij + E\_PCij)} \times 100$$

The sums are over all ordered pairs of legs (i,j), $i \neq j$. Notice that PK%TRUCK is not necessarily the average of the other two variables. It is rather a weighted combination of the two, weighted by the fractions of the overall traffic in morning and evening, respectively.

Turning percentages were calculated along the major road, the minor road, and combined by methods similar to the above. Define the auxiliary variables Mij and Eij by:

$$Mij = M\_PCij + M\_TRij$$

$$Eij = E\_PCij + E\_TRij$$

summing passenger and commercial vehicle flows to get total vehicle flows. Then, the variables PK%TURN, PK%LEFT, PK%THRU1, PK%LEFT1, PK%THRU2, PK%LEFT2 are given by:

$$PK\%TURN = \frac{\sum_{all\ pairs\ (i,j)\ except\ (1,2),(2,1),(3,4),\ and\ (4,3)} (Mij + Eij)}{\sum_{all\ pairs\ (i,j)} (Mij + Eij)} \times 100$$

$$PK\%LEFT = \frac{\sum_{(i,j)=(1,3),(4,1),(2,4),\ or\ (3,4)} Mij + Eij}{\sum_{all\ pairs\ (i,j)} (Mij + Eij)} \times 100$$

$$PK\%THRU1 = \frac{\sum_{(i,j)=(1,2)\ or\ (2,1)} (Mij + Eij)}{\sum_{all\ pairs\ (i,j)\ with\ i=1\ or\ 2} (Mij + Eij)} \times 100$$

$$PK\%LEFT1 = \frac{\sum_{(i,j)=(1,3)\ or\ (2,4)} (Mij + Eij)}{\sum_{all\ pairs\ (i,j)\ with\ i=1\ or\ 2} (Mij + Eij)} \times 100$$

$$PK\%THRU2 = \frac{\sum_{(i,j)=(3,4)\ or\ (4,3)} (Mij + Eij)}{\sum_{all\ pairs\ (i,j)\ with\ i=3\ or\ 4} (Mij + Eij)} \times 100$$

$$PK\%LEFT2 = \frac{\sum_{(i,j)=(4,1)\ or\ (3,2)} (Mij + Eij)}{\sum_{all\ pairs\ (i,j)\ with\ i=3\ or\ 4} (Mij + Eij)} \times 100$$

In case the intersection is three-legged, traffic flows to and from one of legs 3 and 4 will always be zero and, in particular, PK%THRU2 is zero. Three more variables that might be considered are:

$$PK\%RIGHT = (PK\%TURN - PK\%THRU)$$

$$PK\%RIGHT1 = (100 - PK\%LEFT1 - PK\%THRU1)$$

$$PK\%RIGHT2 = (100 - PK\%LEFT2 - PK\%THRU2)$$

In connection with the modeling of the signalized intersections, variables were developed to estimate the incoming traffic on each leg. These variables were based on the ADT information and the peak-hour traffic flows. They are:

$$F_1 = \frac{\sum_{j=2,3,4} (M1j + E1j)}{1/2(\sum_{i=1,2} (\sum_{j\neq i} (Mij + Eij + Mji + Eji)))} \times \frac{ADT1}{1000}$$

$$F_2 = \frac{\sum_{j=1,3,4} (M2j + E2j)}{1/2(\sum_{i=1,2} (\sum_{j\neq i} (Mij + Eij + Mji + Eji)))} \times \frac{ADT1}{1000}$$

$$F_3 = \frac{\sum_{j=1,2,4} (M3j + E3j)}{1/2(\sum_{i=3,4} (\sum_{j\neq i} (Mij + Eij + Mji + Eji)))} \times \frac{ADT2}{1000}$$

$$F_4 = \frac{\sum_{j=1,2,3} (M4j + E4j)}{1/2(\sum_{i=3,4} (\sum_{j\neq i} (Mij + Eij + Mji + Eji)))} \times \frac{ADT2}{1000}$$

where the units are thousands of vehicles per day and F_i is the estimated number of thousands of vehicles per day entering the intersection along leg number i (cf. Figure 1). Three other variables derived from the F_i's were also considered:

$$PRODFADJ = F_1F_4 + F_4F_2 + F_2F_3 + F_3F_1$$

$$PRODFOPP = F_1F_2 + F_3F_4$$

$$SUMF = F_1 + F_2 + F_3 + F_4$$

The first variable PRODFADJ is a variable representing the interaction of adjacent legs, the second PRODFOPP does the same for opposite legs, and the third SUMF is the sum of all the flows.

Intersection Angle Variables

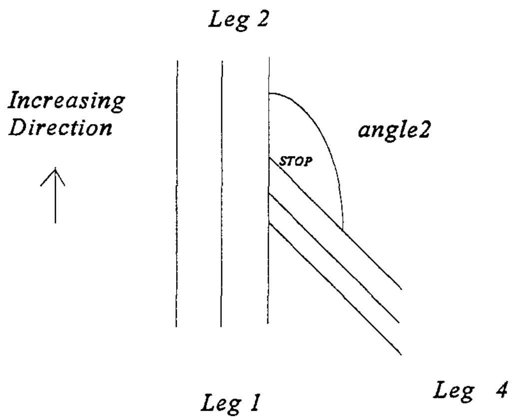
An angle variable DEV, representing the average deviation from 90°, is defined by:

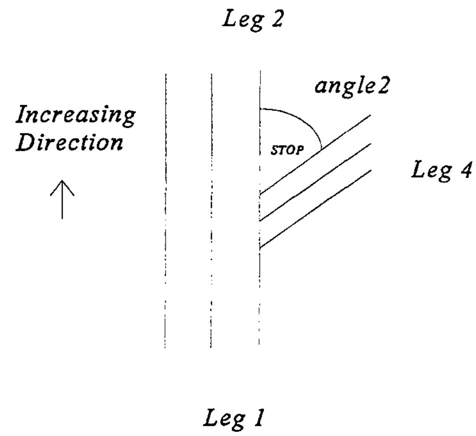
$$DEV = \begin{cases} |angle1 - 90| & \textit{if intersection is three-legged with third leg left (leg 3)} \\ |angle2 - 90| & \textit{if intersection is three-legged with third leg right (leg 4)} \\ \dfrac{|angle1 - 90| + |angle2 - 90|}{2} & \textit{if intersection is four-legged} \end{cases}$$

Another angle variable considered in this study, suggested by E. Hauer, is HAU:

$$HAU = \begin{cases} angle2 - 90 & \textit{if the third leg is to the right (leg 4)} \\ & \textit{at a three-legged intersection} \\ 90 - angle1 & \textit{if the third leg is to the left (leg 3)} \\ & \textit{at a three-legged intersection} \\ \dfrac{angle2 - angle1}{2} & \textit{at a four-legged intersection} \end{cases}$$

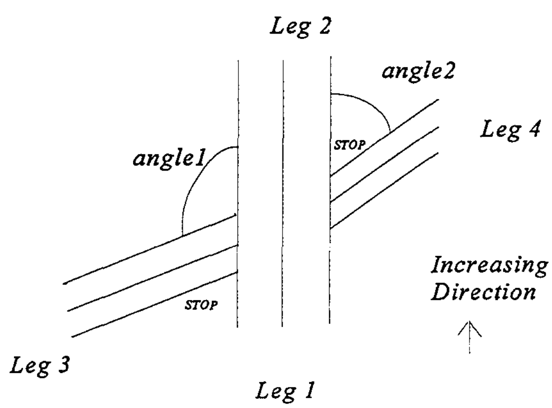
The variable HAU is a signed variable. See Figures 2 and 3. For a three-legged intersection with the angle to the right of the increasing direction, HAU is positive when the angle is larger than 90°, as in 2(a), and HAU is negative when the angle is smaller than 90°, as in 2(b). If the angle is to the left of the increasing direction (see Figure 3), 180° minus the angle becomes the new angle and HAU is defined as ((180 - angle) - 90) = (90 - angle), as above. For a four-legged intersection, as in 2(c), it is the average of the two three-legged values (and thus 90° cancels out). Figure 4 illustrates the calculation of HAU in a variety of cases. Kulmala (1995) proposes that turns from the





(a) Three-legged intersection, angle larger than 90°

(b) Three-legged intersection, angle smaller than 90°



(c) Four-legged intersection

FIGURE 2. Intersection Angle Geometries

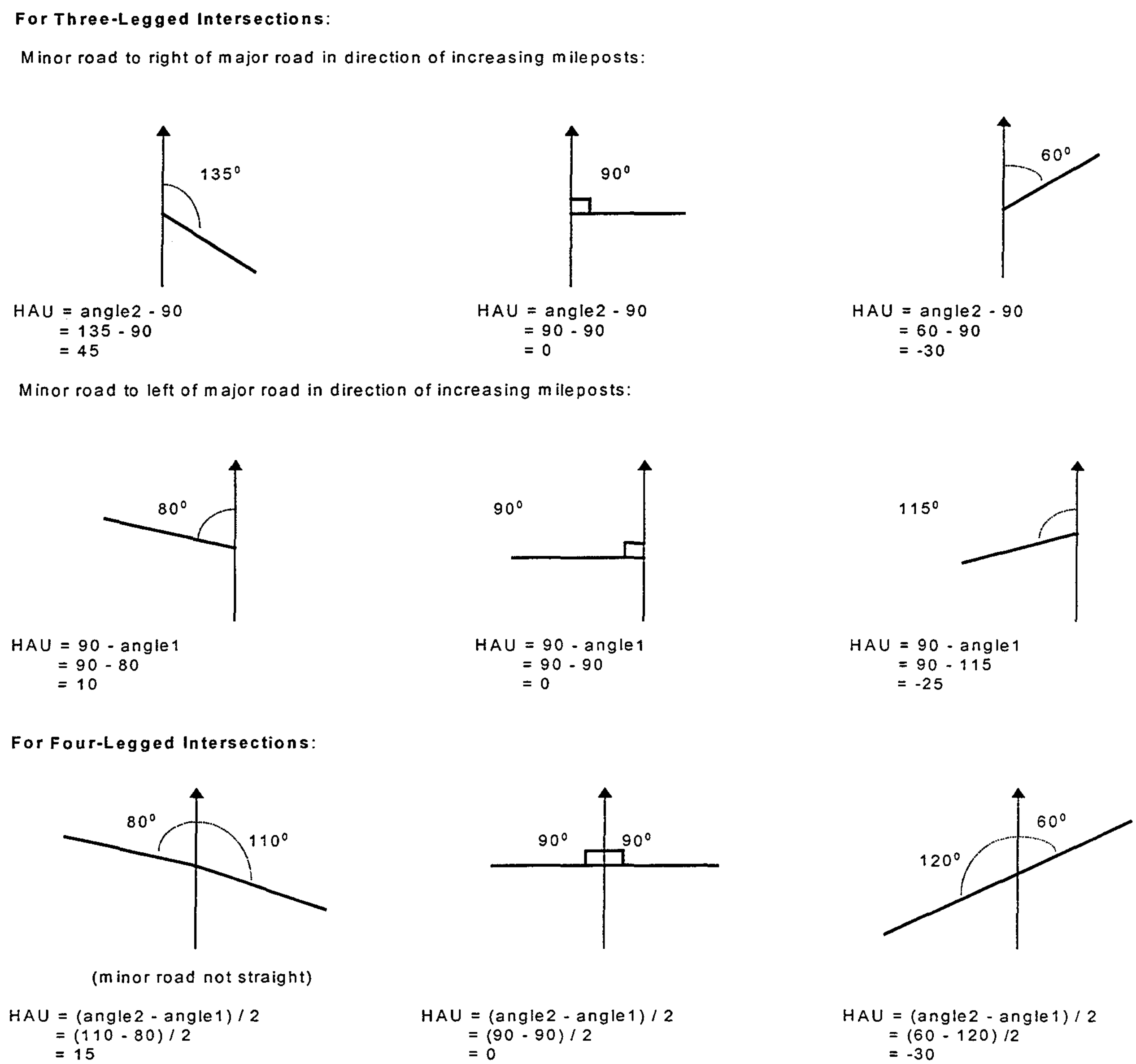


FIGURE 3. Examples of Calculation of the Angle Variable HAU

far lane of the major road may be less crash prone in situation 2(a) than in situation 2(b), so that positive values of HAU correspond to fewer crashes.

Sight Distances

To represent sight distances for the modeling, reciprocals were chosen. Large values of the reciprocals corresponded to poor sight distances, and small ones corresponded to lengthy sight distances, and thus crashes might be expected to increase with increasing values of the reciprocals.

$$RSD1 = \begin{cases} \dfrac{1}{SD1} & \textit{if intersection is three-legged with minor leg being leg } 3 \\[2ex] \dfrac{1}{SD2} & \textit{if intersection is three-legged with minor leg being leg } 4 \\[2ex] (\dfrac{1}{2})(\dfrac{1}{SD1} + \dfrac{1}{SD2}) & \textit{if intersection has four legs;} \end{cases}$$

$$RSDL2 = \begin{cases} \dfrac{1}{SDL3} & \textit{if intersection is three-legged with minor leg being leg } 3 \\[2ex] \dfrac{1}{SDL4} & \textit{if intersection is three-legged with minor leg being leg } 4 \\[2ex] (\dfrac{1}{2})(\dfrac{1}{SDL3} + \dfrac{1}{SDL4}) & \textit{if intersection has four legs;} \end{cases}$$

$$RSDR2 = \begin{cases} \dfrac{1}{SDR3} & \textit{if intersection is three-legged with minor leg being leg } 3 \\[2ex] \dfrac{1}{SDR4} & \textit{if intersection is three-legged with minor leg being leg } 4 \\[2ex] (\dfrac{1}{2})(\dfrac{1}{SDR3} + \dfrac{1}{SDR4}) & \textit{if intersection has four legs;} \end{cases}$$

$$RSDL1 = (\frac{1}{2})(\frac{1}{SDL1} + \frac{1}{SDL2}) \ \textit{if intersection has four legs;}$$

$$RSD2 = (\frac{1}{2})(\frac{1}{SD3} + \frac{1}{SD4}) \ \textit{for a four-legged signalized intersection.}$$

The variables are RSD1, RSDL2, RSDR2, RSDL1, and RSD2.

Horizontal Alignment

Variables used to represent composite horizontal curvature are the same as those used by Vogt and Bared (1998), except that 764 feet (233 meters) has been replaced by 800 feet (244 meters):

$$HI\text{-}1 = \frac{\sum_i DEGHi}{Number\ of\ horizontal\ curves\ overlapping\ intersection\ center\ \pm 250\ feet}$$

$$HEI\text{-}1 = \frac{\sum_j DEGHj}{Number\ of\ horizontal\ curves\ overlapping\ intersection\ center\ \pm 800\ feet}$$

where the sum is over the corresponding curves along the major road. HI-1 and HEI-1 (E for extended) are the unweighted averages of the degrees of curvature of the corresponding curves. Similar quantities for the minor road, in the case of signalized intersections, are denoted by HI-2 and HEI-2. These are combined with the major road variables to generate two more variables HICOM and HEICOM:

$$HICOM = (\frac{1}{2})(HI\text{-}1 + HI\text{-}2)$$

$$HEICOM = (\frac{1}{2})(HEI\text{-}1 + HEI\text{-}2)$$

to be used in the modeling of the signalized intersections.

Vertical Alignment

Vertical alignment variables likewise are taken from Vogt and Bared (1998).

A basic variable associated with each vertical curve is V_i:

$$V_i = \frac{|GBi - GEi|}{length\ Li\ of\ i\text{-}th\ vertical\ curve\ in\ hundreds\ of\ feet}$$

with units of percent per 100 feet (30.5 meters), where the numerator is the absolute change of grade $\Delta gi = |GBi - GEi|$ and $Li = (VEi - VBi)/100$.

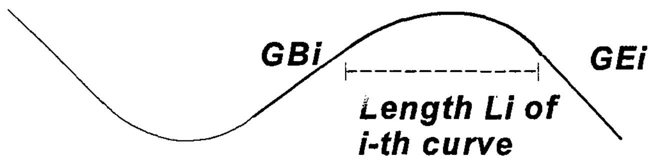


FIGURE 4. Vertical Curve

Four vertical variables VCI-1, VCEI-1, VI-1, and VEI-1 were considered:

$$VCI\text{-}1 = \frac{\sum_i Vi}{\textit{Number of vertical crest curves overlapping intersection center} \pm 250 \textit{ feet}}$$

$$VCEI\text{-}1 = \frac{\sum_i Vi}{\textit{Number of vertical crest curves overlapping intersection center} \pm 800 \textit{ feet}}$$

$$VI\text{-}1 = \frac{\sum_i Vi}{\textit{Number of vertical curves overlapping intersection center} \pm 250 \textit{ feet}}$$

$$VEI\text{-}1 = \frac{\sum_i Vi}{\textit{Number of vertical curves overlapping intersection center} \pm 800 \textit{ feet}}$$

These sums are over the stipulated vertical curves along the major road. For signalized intersections, similar variables with the suffix 2 rather than 1 were also employed for the minor road, as well as the combined variables VCICOM, VCEICOM, VICOM, and VEICOM:

$$VCICOM = (\frac{1}{2})(VCI\text{-}1 + VCI\text{-}2)$$

$$VCEICOM = (\frac{1}{2})(VCEI\text{-}1 + VCEI\text{-}2)$$

$$VICOM = (\frac{1}{2})(VI\text{-}1 + VI\text{-}2)$$

$$VEICOM = (\frac{1}{2})(VEI\text{-}1 + VEI\text{-}2)$$

Recall that crest curves are vertical curves for which the grade decreases (positive to negative, positive to less positive, negative to more negative).

Another variable developed pertaining to vertical alignment is ABSGRD1. If only one grade was seen on the major road, ABSGRD1 was the absolute value of this grade. If more than one grade was seen in the vicinity of the intersection on the major road, absolute values were computed of all grades seen at the beginnings and endings of those vertical curves that overlapped the segment of the major road within ±800 feet (244 meters) of the intersection center. These absolute values were then averaged (e.g., if six grades occur corresponding to three vertical curves, their absolute values were summed and divided by six), without regard to where they occurred (in some cases more than 800 feet (244 meters) from the intersection) or the distance for which the grade remained constant. A similar variable ABSGRD2 was also developed for the minor road of signalized intersections.

Miscellaneous Variables

Driveway variables were combined to yield NODRWY1 as follows:

$$NODRWY1 = NODRWYR1 + NODRWYC1$$

and a similar combination, NODRWY2, was used for minor road driveways at signalized intersections.

Median widths varied between legs of the major road in 18 out of 84 three-legged intersections, 18 out of 72 four-legged intersections, and 1 out of 49 signalized intersections (most of the signalized intersections had no median). Thus, the median width variable here, MEDWIDTH1, is the average of the median widths of the two legs, leg 1 and leg 2, of the major road.

Speed limit variables, SPD1 and SPD2, with values in miles per hour were assigned to the major and minor road. On the major road, SPD1 was the average of the posted speeds on Legs 1 and 2 or the unique value seen if a posted speed was seen on only one of these legs. The same rule was applied for the minor road to get SPD2. In some cases, no posted speed limit was seen on the leg or legs of the minor road. In this case, SPD2 was assigned the default value 35.

During the modeling, it became convenient to introduce the channelization variable LTLN1S:

$$LTLN1S = \begin{array}{ll} 1 & \text{if } LTLN1 \text{ is } 1 \text{ or } 2 \\ 0 & \text{if } LTLN1 \text{ is } 0 \end{array}$$

Yet another numerical variable was devised to denote the State:

$$STATE = \begin{cases} 0 & \textit{if the intersection is in California} \\ 1 & \textit{if the intersection is in Michigan} \end{cases}$$

The STATE variable can be used to study whether crash experience at the various intersections is due in part to differences between the States. Such factors as driver behavior and/or crash reporting practices may be significantly different between the two States.

UNIVARIATE STATISTICS

A summary of the data obtained is shown in Tables 5, 6, and 7. The first item that strikes the eye is that the mean number of crashes per intersection, no matter how they are measured, is highest at signalized intersections, moderate at four-legged ones, and lowest at three-legged ones.

There are a number of other ways in which the intersection classes differ. The signalized intersections have much higher minor road ADT and much higher turning percentages than the other two classes. The signalized intersections tend to have more turning lanes on both major and minor legs, and lower speed limits on the major road as well as higher ones on the minor road. There is more lighting on the signalized intersection, a moderate amount on the four-legged intersection, and the least on the three-legged intersection. Likewise, the general terrain is flattest on the signalized intersection, less so on the four-legged intersection, and least on the three-legged intersection. This is due at least in part to the fact that two-thirds of the signalized intersections are in Michigan, while only 25% of the other intersections are, and Michigan is a relatively flat State. The three intersection classes are similar in other ways. Peak Truck Percentages at the three classes of intersections are from 9 to 11% on average. There are two or three driveways per intersection on average, and the average value of HAZRAT is from 2.2 to 2.5. Sight distances are comparable, except that signalized intersections have a lower average sight distance left on the minor road. The signalized intersections have even lower sight distances left on the major road than on the minor road. This suggests that woods, buildings, and other obstacles are not cleared away from the minor road to the extent that they are from the major road.

Horizontal and vertical alignments are generally similar. Fewer of the signalized intersections, primarily in Michigan as mentioned, have horizontal curves and fewer have vertical curves. Although the average value of HEI-1 varies substantially among the intersection classes, this average is strongly influenced by a few intersections with sharp turns. The average grade of signalized intersections is a bit lower than the average for the nonsignalized intersections, and the minor road has a higher average grade than the major road. This phenomenon was also noted in the three-legged and four-legged intersections, although no measurements were made. Frequently, the minor legs leading to an intersection on a four-lane road have fairly steep grades as they are brought up or down to conform with the level of the major road.

TABLE 5. Summary Statistics: 84 Three-Legged Rural Intersections
Major road four-lane, minor leg stop-controlled, California and Michigan, 1993-1995

Variable and Abbreviation	Min.	Max.	Median	Mean	Freq.	%Zero
No. of Crashes TOTACC	0	19	2	3.88	326	21.4
No. of Injury Crashes INJACC	0	11	1	1.61	220 (67.5%)	38.1
No. of Intersection-Type Crashes TOTACCI	0	13	1	2.62	135 (41.4%)	34.5
No. of Intersection-Type Injury Crashes INJACCI	0	9	1	1.21	102 (31.3%)	48.8
Average Daily Traffic on Major Road ADT1, vpd	2,367	33,058	12,050	12,870		
Average Daily Traffic on Minor Road ADT2, vpd	15	3,001	349	596		
Peak Truck Percentage PK%TRUCK	1.18	28.16	7.79	9.15		
Peak Turning Percentage PK%TURN	0.26	53.09	4.28	6.68		
Peak Left-Turn Percentage PK%LEFT	0.13	25.97	2.16	3.29		
Peak Through Percentage on Major Road PK%THRU1	63.26	100.00	97.98	96.44		
Peak Left-Turn Percentage on Major Road PK%LEFT1	0.00	21.29	0.69	1.49		13.1
Peak Left-Turn Percentage on Minor Leg PK%LEFT2	0.00	100.00	60.99	56.64		7.1

TABLE 5. Summary Statistics: 84 Three-Legged Rural Intersections (continued)
Major road four-lane, minor leg stop-controlled, California and Michigan, 1993-1995

Variable and Abbreviation	Min.	Max.	Median	Mean	Freq.	%Zero
Roadside Hazard Rating HAZRAT	1	7	2	2.52		
1					16 (19.0%)	
2					37 (44.0%)	
3					13 (15.5%)	
4					10 (11.9%)	
5					6 (7.1%)	
6					1 (1.2%)	
7					1 (1.2%)	
No. of Res. Driveways on Major Road NODRWYR1	0	7	0	1.17	98	56.0
No. of Comm. Driveways on Major Road NODRWYC1	0	14	0	1.93	162	57.1
No. of Driveways on Major Road NODRWY1	0	15	1	3.10	259	42.9
Left-Turn Lane on Major Road LTLN1	0	1	1	0.54		46.4
0 = no					39 (46.4%)	
1 = yes					45 (53.6%)	
Right-Turn Lane on Major Road RTLN1	0	1	0	0.19		81.0
0 = no					68 (81.0%)	
1 = yes					16 (19.0%)	
Left-Turn Lane on Minor Road LTLN2	0	1	0	3.57		96.4
0 = no					81 (96.4%)	
1 = yes					3 (3.6%)	
Right-Turn Lane on Minor Road RTLN2	0	1	0	11.90		88.1
0 = no					74 (88.1%)	
1 = yes					10 (11.9%)	

TABLE 5. Summary Statistics: 84 Three-Legged Rural Intersections (continued)
Major road four-lane, minor leg stop-controlled, California and Michigan, 1993-1995

Variable and Abbreviation	Min.	Max.	Median	Mean	Freq.	%Zero
Median Width on Major Road MEDWIDTH1, feet	0	36	0	3.74		53.6
Median Type on Major Road MEDTYPE No Median Painted Curbed Other (Guardrail, Mixed, etc.)					 45 (53.6%) 23 (27.4%) 9 (10.7%) 7 (8.3%)	
Angle Variable HAU, degrees	-45	55	0	-0.36		83.3
Longitudinal Sight Distance on Major Road SD1, feet	500	2000+	2000+	1543+		
Left-Side Sight Distance on Minor Road SDL2, feet	45	2000+	1470	1399+		
Right-Side Sight Distance on Minor Road SDR2, feet	80	2000+	1375	1388+		
Degree of Curve HEI-1= $(1/n)\sum$ DEGHi, deg/100 ft	0	26.6	0	2.47		52.4
Curve Grade Rate VEI-1= $(1/m)\sum (\lvert\Delta gi\rvert/Li)$, %/100 ft	0	6.71	0.04	0.89		50.0
Crest Grade Rate VCEI-1 = $(1/m)\sum (\lvert\Delta gi\rvert/Li)$, %/100 ft	0	11.0	0	0.65		59.5
Average Absolute Grade on Major Road ABSGRD1, %	0	5.85	0.65	1.11		25.0

1 ft = 0.305 m

TABLE 5. Summary Statistics: 84 Three-Legged Rural Intersections (continued)
Major road four-lane, minor leg stop-controlled, California and Michigan, 1993-1995

Variable and Abbreviation		Min.	Max.	Median	Mean	Freq.	%Zero
Speed Limit on Major Road SPD1, mph		30	65	55	50.4		
Speed Limit on Minor Road SPD2, mph		15	35	35	31.5		
Light at Intersection LIGHT	0 = no 1 = yes					52 (61.9%) 32 (38.1%)	
Terrain	Flat Rolling Mountainous					48 (57.1%) 29 (34.5%) 7 (8.3%)	
STATE	0 = CA 1 = MI					60 (71.4%) 24 (28.6%)	

1 mph = 1.61 km/h

TABLE 6. Summary Statistics: 72 Four-Legged Rural Intersections

Major road four-lane, minor legs stop-controlled, California and Michigan, 1993-1995

Variable and Abbreviation	Min.	Max.	Median	Mean	Freq.	%Zero
No. of Crashes TOTACC	0	38	3.5	5.53	398	12.5
No. of Injury Crashes INJACC	0	20	2	2.64	190 (47.7%)	25.0
No. of Intersection-Type Crashes TOTACCI	0	27	2	4.13	297 (74.6%)	22.2
No. of Intersection-Type Injury Crashes INJACCI	0	19	1	2.19	158 (39.7%)	36.1
Average Daily Traffic on Major Road ADT1, vpd	3,350	73,000	11,166	13,018		
Average Daily Traffic on Minor Road ADT2, vpd	21	2,018	410	559		
Peak Truck Percentage PK%TRUCK	1.70	37.24	8.36	10.95		
Peak Turning Percentage PK%TURN	0.00	48.52	6.56	9.47		2.8
Peak Left-Turn Percentage PK%LEFT	0.00	25.26	6.56	9.47		2.8
Peak Through Percentage on Major Road PK%THRU1	67.77	100.0	96.51	94.41		
Peak Left-Turn Percentage on Major Road PK%LEFT1	0.00	13.96	1.51	2.78		5.6
Peak Through Percentage on Minor Road PK%THRU2	0.00	68.1	12.0	16.37		17.1
Peak Left-Turn Percentage on Minor Road PK%LEFT2	0.00	100.00	37.5	40.58		5.7

TABLE 6. Summary Statistics: 72 Four-Legged Rural Intersections (continued)
Major road four-lane, minor legs stop-controlled, California and Michigan, 1993-1995

Variable and Abbreviation	Min.	Max.	Median	Mean	Freq.	%Zero
Roadside Hazard Rating HAZRAT	1	6	2	2.19		
1					21 (29.2%)	
2					29 (40.3%)	
3					12 (16.7%)	
4					8 (11.1%)	
5					1 (1.4%)	
6					1 (1.4%)	
No. of Res. Driveways on Major Road NODRWYR1	0	7	0	1.04	75	66.7
No. of Comm. Driveways on Major Road NODRWYC1	0	12	0	0.88	63	66.7
Number of Driveways on Major Road NODRWY1	0	15	0	1.92	138	54.2
Left-Turn Lanes on Major Road LTLN1	0	2	2	1.33		30.6
0					22 (30.6%)	
1					4 (5.5%)	
2					46 (63.9%)	
Right-Turn Lanes on Major Road RTLN1	0	2	0	0.65		62.5
0					45 (62.5%)	
1					7 (9.7%)	
2					20 (27.8%)	
Left-Turn Lanes on Minor Road LTLN2	0	1	0	0.028		97.2
0					70 (97.2%)	
1					2 (2.8%)	
2					0 (0.0%)	
Right-Turn Lanes on Minor Road RTLN2	0	2	0	0.61		62.5
0					45 (62.5%)	
1					10 (13.9%)	
2					17 (23.6%)	

TABLE 6. Summary Statistics: 72 Four-Legged Rural Intersections (continued)
Major road four-lane, minor legs stop-controlled, California and Michigan, 1993-1995

Variable and Abbreviation	Min.	Max.	Median	Mean	Freq.	%Zero
Median Width on Major Road MEDWIDTH1, feet	0	36	2	3.78		43.1
Median Type on Major Road MEDTYPE No Median Painted Curbed Other					 31 (43.1%) 17 (23.1%) 22 (30.6%) 2 (2.8%)	
Angle Variable HAU, degrees	-20	30	0	0.868		77.8
Longitudinal Sight Distance on Major Road SD1, feet	400	2000+	1500	1430+		
Left-Side Sight Distance on Minor Road SDL2, feet	324	2000+	1438	1358+		
Right-Side Sight Distance on Minor Road SDR2, feet	215	2000+	1430	1377+		
Degree of Curve HEI-1 = $(1/n)\sum$ DEGHi, deg/100 ft	0	233.3	0	5.01		56.9
Curve Grade Rate VEI-1 = $(1/m)\sum (\lvert\Delta gi\rvert/Li)$, %/100 ft	0	12.5	0	0.70		61.1
Crest Grade Rate VCEI-1 = $(1/m)\sum (\lvert\Delta gi\rvert/Li)$, %/100 ft	0	12.5	0	0.50		75.0
Average Absolute Grade on Major Road ABSGRD1, %	0	5.8	0.4	0.98		38.9

1 ft = 0.305 m

TABLE 6. Summary Statistics: 72 Four-Legged Rural Intersections (continued)
Major road four-lane, minor legs stop-controlled, California and Michigan, 1993-1995

Variable and Abbreviation		**Min.**	**Max.**	**Median**	**Mean**	**Freq.**	**%Zero**
Speed Limit on Major Road SPD1, mph		25	65	55	53.68		
Speed Limit on Minor Road SPD2, mph		25	50	35	33.35		
Light at Intersection LIGHT	0 = no 1 = yes					40 (55.6%) 32 (44.4%)	
Terrain	Flat Rolling Mountainous					49 (68.1%) 14 (19.4%) 9 (12.5%)	
STATE	0 = CA 1 = MI					54 (75.0%) 18 (25.0%)	

1 mph = 1.61 km/h

TABLE 7. Summary Statistics: 49 Signalized Rural Intersections
Four-legged intersections of two-lane roads, California and Michigan, 1993-1995

Variable and Abbreviation	Min.	Max.	Median	Mean	Freq.	%Zero
No. of Crashes TOTACC	2	48	21	20.8	1017	0.0
No. of Injury Crashes INJACC	0	25	7	7.47	366 (36.0%)	4.1
No. of Intersection-Type Crashes TOTACCI	1	37	17	16.1	790 (77.7%)	0.0
No. of Intersection-Type Injury Crashes INJACCI	0	21	6	6.14	301 (29.6%)	4.1
Average Daily Traffic on Major Road ADT1, vpd	4,917	25,133	8,900	10,491		
Average Daily Traffic on Minor Road ADT2, vpd	940	12,478	3,670	4,367		
Peak Truck Percentage PK%TRUCK	2.69	45.43	7.71	8.96		
Peak Turning Percentage PK%TURN	7.07	72.66	34.48	35.64		
Peak Left-Turn Percentage PK%LEFT	4.20	37.07	17.97	18.17		
Peak Through Percentage on Major Road PK%THRU1	18.01	96.73	73.77	71.19		
Peak Left-Turn Percentage on Major Road PK%LEFT1	1.78	36.67	12.99	14.71		
Peak Through Percentage on Minor Road PK%THRU2	8.45	84.09	41.97	43.90		
Peak Left-Turn Percentage on Minor Road PK%LEFT2	2.50	75.73	24.88	28.69		

TABLE 7. Summary Statistics: 49 Signalized Rural Intersections (continued)
Four-legged intersections of two-lane roads, California and Michigan, 1993-1995

Variable and Abbreviation	Min.	Max.	Median	Mean	Freq.	%Zero
Roadside Hazard Rating HAZRAT	1	6	2	2.35		
1					10 (20.4%)	
2					20 (40.8%)	
3					14 (28.6%)	
4					3 (6.1%)	
5					1 (2.0%)	
6					1 (2.0%)	
No. of Res. Driveways on Major Road NODRWYR1	0	6	0	0.67	33	71.4
No. of Comm. Driveways on Major Road NODRWYC1	0	11	2	2.35	115	32.7
No. of Driveways on Major Road NODRWY1	0	15	3	3.02	148	28.6
No. of Res. Driveways on Minor Road NODRWYR2	0	8	0	0.94	46	65.3
No. of Comm. Driveways on Minor Road NODRWYC2	0	11	3	2.24	110	22.4
No. of Driveways on Minor Road NODRWY2	0	11	3	3.18	156	12.2
Left-Turn Lanes on Major Road LTLN1	0	2	2	1.69		14.3
0					7 (14.3%)	
1					1 (2.0%)	
2					41 (83.7%)	
Right-Turn Lanes on Major Road RTLN1	0	2	1	0.98		42.9
0					21 (42.9%)	
1					8 (16.3%)	
2					20 (40.8%)	

TABLE 7. Summary Statistics: 49 Signalized Rural Intersections (continued)
Four-legged intersections of two-lane roads, California and Michigan, 1993-1995

Variable and Abbreviation	Min.	Max.	Median	Mean	Freq.	%Zero
Left-Turn Lanes on Minor Road LTLN2	0	2	2	1.24		34.7
0					17 (34.7%)	
1					3 (6.1%)	
2					29 (59.2%)	
Right-Turn Lanes on Minor Road RTLN2	0	2	0	0.73		53.1
0					26 (53.1%)	
1					10 (20.4%)	
2					13 (26.5%)	
Median Width on Major Road MEDWIDTH1, feet	0	6.5	0	0.58		87.8
Median Type on Major Road MEDTYPE						
No Median						43 (87.8%)
Painted						1 (2.0%)
Mixed						5 (10.2%)
Angle variable HAU, degrees	-45	40	0	0.102		67.35
Longitudinal Sight Distance on Major Road SD1, feet	267	2000+	1538	1454+		
Left-Side Sight Distance on Major Road SDL1, feet	186	2000+	612	833+		
Longitudinal Sight Distance on Minor Road SD2, feet	390	2000+	1333	1406+		
Left-Side Sight Distance on Minor Road SDL2, feet	253	2000+	825	1007+		

1 ft = 0.305 m

TABLE 7. Summary Statistics: 49 Signalized Rural Intersections (continued)
Four-legged intersections of two-lane roads, California and Michigan, 1993-1995

Variable and Abbreviation		Min.	Max.	Median	Mean	Freq.	%Zero
Degree of Curve HEI-1= $(1/n)\sum$ DEGi, deg/100 ft		0	94.9	0	4.61		57.1
Curve Grade Rate VEI-1= $(1/m)\sum (\lvert\Delta gi\rvert/Li)$, %/100 ft		0	5.98	0.67	1.26		38.8
Crest Grade Rate VCEI-1= $(1/m)\sum (\lvert\Delta gi\rvert/Li)$, %/100 ft		0	6.88	0	1.01		51.0
Degree of Curve HEI-2 = $(1/n)\sum$ DEGHi, deg/100 ft		0	36.41	0	2.27		73.57
Curve Grade Rate VEI-2 = $(1/m)\sum (\lvert\Delta gi\rvert/Li)$, %/100 ft		0	11.97	1.35	2.24		24.5
Crest Grade Rate VCEI-2 = $(1/m)\sum (\lvert\Delta gi\rvert/Li)$, %/100 ft		0	12.13	1	1.88		32.7
Average Absolute Grade on Major Road ABSGRD1, %		0	3.45	0.73	0.83		28.6
Average Absolute Grade on Minor Road ABSGRD2, %		0	5.3	0.71	1.00		18.4
Speed Limit on Major Road SPD1, mph		30	65	55	48.7		
Speed Limit on Minor Road SPD2, mph		25	55	45	43.8		
Protected Left Turn PROT_LT	0 = no 1 = yes					28 (57.1%) 21 (42.9%)	57.1
Signal Type SIG_TYPE	Pre-Timed Actuated Semi-Actuated					22 (44.9%) 21 (42.9%) 6 (12.2%)	
Light at Intersection LIGHT	0 = no 1 = yes					10 (20.4%) 39 (79.6%)	
Terrain	Flat Rolling Mountainous					36 (73.5%) 11 (22.4%) 2 (4.1%)	
STATE	0 = CA 1 = MI					18 (36.7%) 31 (63.3%)	

1 ft = 0.305 m, 1 mph = 1.61 km/h

Crash Data Versus Intersection Class and State

Table 8 is an extract from Tables 5, 6, and 7, comparing the mean number of crashes per intersection for the three intersection classes. It indicates that four-legged intersections have from 1.42 to 1.81 times as many crashes as three-legged intersections. The higher ratio comes into effect as the crash severity and the intersection-relatedness increase. This is consistent with the rough rule of thumb that a four-legged intersection behaves like a pair of three-legged intersections, with a consequent crash ratio of 2. Note that average major and minor road ADT's, ADT1 and ADT2, in Tables 5 and 6 for three-legged and four-legged intersections, respectively, are very nearly equal, and thus that the comparison of three-legged and four-legged intersections is justifiable.

TABLE 8. Mean Number of Crashes per Intersection by Crash Variable and Intersection Class

	three-legged	four-legged	signalized
TOTACC	3.88	5.53= 1.42×3.88	20.8=3.76×5.53
TOTACCI	2.62	4.13=1.58×2.62	16.1=3.90×4.13
INJACC	1.61	2.64=1.64×1.61	7.47=2.83×2.64
INJACCI	1.21	2.19=1.81×1.21	6.14=2.80×2.19

With regard to the signalized intersections, Table 8 indicates that they have from 3.90 to 2.80 times as many crashes as four-legged intersections. These two intersection classes have in common four-leggedness, but otherwise are quite different (lanes, control, and ADT). Nonetheless, it appears that intersection-relatedness, i.e., all crashes versus those satisfying the BMI criteria (see p. 40), has a negligible effect on the crash ratio, but that the fraction of serious crashes is lower at signalized intersections than it is at the four-legged intersections.

Table 9 provides a decomposition of crashes by severity and State for the three intersection classes. With respect to State, it indicates that Michigan crashes tend to be less severe than California crashes for all classes, regardless of intersection-relatedness. Regardless of State, signalized intersections have the lowest percentage of serious crashes and four-legged intersections have the highest percentage. Intersection-related crashes (TOTACCI) have a slightly higher tendency to be serious than all crashes (TOTACC) for both States and all three intersection classes.

The data in Table 9 are represented in another way in Table 10. Table 10 indicates that California is underrepresented in crashes in both the four-legged and signalized intersection samples and partly underrepresented in the three-legged intersection sample. It also shows that such under-representation decreases for serious crashes and that for the three-legged intersections, California is overrepresented in serious crashes. The modeling later in this report will attempt to sort out whether the dependence on STATE reflects a difference in highway variables between the States.

TABLE 9. Total Number of Crashes by Severity, State, and Intersection Class

	TOTACC		TOTACCI	
	CA Three-legged	MI	CA Three-legged	MI
Property damage only	106 (50.7%)	85 (72.6%)	73 (47.7%)	45 (67.2%)
Injury	101 (48.3%)	32 (27.4%)	78 (51.0%)	22 (32.8%)
Fatal	2 (1.0%)	0	2 (1.3%)	0
Total	209	117	153	67
	CA Four-legged	MI	CA Four-legged	MI
Property damage only	95 (41.5%)	113 (66.9%)	73 (38.6%)	66 (61.1%)
Injury	129 (56.3%)	55 (32.5%)	112 (59.3%)	41 (38.0%)
Fatal	5 (2.2%)	1 (0.6%)	4 (2.1%)	1 (0.9%)
Total	229	169	189	108
	CA Signalized	MI	CA Signalized	MI
Property damage only	159 (58.2%)	492 (66.1%)	143 (57.4%)	346 (64.0%)
Injury	112 (41.0%)	247 (33.2%)	104 (41.8%)	190 (35.1%)
Fatal	2 (0.7%)	5 (0.7%)	2 (0.8%)	5 (0.9%)
Total	273	744	249	541

TABLE 10. Percentage of Intersections and Crashes in California for Each Intersection Class

	% of int.	% of TOTACC	% of TOTACCI	% of INJACC	% of INJACCI
3-legged	72	64.1	69.5	76.3	78.4
4-legged	75	57.5	63.6	70.5	72
Signalized	36.7	26.8	31.5	31.1	35.2

BIVARIATE STATISTICS

To prepare for model development, it is appropriate to ask what variables correlate strongly with crash counts and to note the mutual correlations of highway variables with one another.

In the tables that follow, correlation coefficients between variables are shown, along with P-values, for each of the three data sets. Recall that the P-value is the estimated probability that the measured correlation coefficient would be at least as far from 0 as it is found to be if the true correlation coefficient for the population from which the sample is drawn is zero. A small P-value indicates that a correlation is significant, a large one indicates that no particular significance can be attached to it. The correlation coefficient summarizes the sample: if it is positive, the variables compared tend to increase together in the sample; if it is negative, they tend to decrease together. If the correlation coefficient is far from zero and its P-value is small, the sample is unlikely to have been drawn from a population where the true correlation is zero; if the correlation coefficient is close to zero and its P-value is large, the sample resembles a sample drawn randomly from a population whose overall correlation coefficient is zero.

Other cautions should be offered in the interpretation of correlation coefficients. If a variable correlates strongly with, say, number of crashes, it may be that the variable is not in itself influential, but that it happens to correlate strongly with another variable that is influential. Likewise, if a variable seems to have a weak correlation with the number of crashes, it may be in part because the influence of the variable is masked by the presence of other more influential variables. The point of modeling is to determine the leading influences and then discover secondary influences, e.g., crashes may be strongly dependent on ADT, but after ADT is properly taken into account, the residual, the portion of crash count that cannot be expressed in terms of ADT, may be strongly correlated with another variable.

Crashes Versus Other Variables

Tables 11, 12, and 13 exhibit correlation coefficients and P-values between crash counts and other variables for the three data sets.

Table 11 exhibits the correlations between intersection crashes and highway variables for the three-legged intersections. Major and minor road ADT's correlate positively with crashes, as expected. Peak turning percentages also correlate with crashes, both positively and negatively. Since these turning percentages correlate with each other, it is not immediately clear what the chief influences are. While HAZRAT is insignificant, number of driveways correlates positively with crashes and median width correlates negatively; neither result is unexpected. The angle variables HAU and DEV are both significant, with HAU more so than DEV. The sign, however, is not what the Kulmala (1995) study suggests, but it is consistent with the work of Vogt and Bared (1998) for three-legged intersections. Sight distance is not significant, although minor road sight distance left is marginally significant. Both left and right turns from the minor road are affected by sight distance left. The horizontal variable HEI-1 and the vertical variables VI-1 and VEI-1 are significant. LIGHT and

TABLE 11. Correlation Coefficients and P-Values for Crashes Versus Other Variables, Three-Legged Intersections

84 rural intersections, major road 4-lane, minor leg stop-controlled, CA and MI, 1993-95

Highway Variable	TOTACC		INJACC		TOTACCI		INJACCI	
	Corr.	P-value	Corr.	P-value	Corr.	P-value	Corr.	P-value
ADT1	0.3623	0.0007	0.3383	0.0016	0.3810	0.0003	0.3223	0.0028
ADT2	0.5009	0.0001	0.3780	0.0004	0.5007	0.0001	0.4315	0.0001
PK%TRUCK	-0.2502	0.0217	-0.1540	0.1620	-0.2662	0.0144	-0.1596	0.1470
PK%TURN	0.2574	0.0181	0.2362	0.0305	0.3113	0.0039	0.2811	0.0096
PK%LEFT	0.2323	0.0335	0.2142	0.0504	0.2834	0.0090	0.2574	0.0181
PK%THRU1	-0.2170	0.0474	-0.1745	0.1123	-0.2819	0.0094	-0.2242	0.0403
PK%LEFT1	0.2786	0.0103	0.2612	0.0164	0.3098	0.0041	0.2884	0.0078
PK%LEFT2	-0.2096	0.0588	-0.1628	0.1440	-0.1900	0.0873	-0.1446	0.1950
HAZRAT	-0.0720	0.5150	0.0449	0.6850	-0.0419	0.7050	0.0595	0.5907
NODRWY1	0.3888	0.0003	0.1591	0.1484	0.4132	0.0001	0.1876	0.0874
LTLN1	-0.1753	0.1106	0.0190	0.8635	-0.1347	0.2218	-0.0086	0.9382
RTLN1	-0.1203	0.2757	0.0041	0.9704	-0.0717	0.5168	-0.0242	0.8267
LTLN2	0.1691	0.1241	0.1579	0.1515	0.1563	0.1556	0.1564	0.1553
RTLN2	0.1552	0.1586	0.1210	0.2728	0.1519	0.1677	0.1411	0.2005
MEDWIDTH1	-0.2557	0.0189	-0.1252	0.2566	-0.2259	0.0388	-0.1223	0.2679
HAU	0.2871	0.0081	0.3817	0.0003	0.2265	0.0383	0.3753	0.0004
DEV	0.1743	0.1127	0.2422	0.0264	0.1332	0.2269	0.2401	0.0278
RSD1	0.0775	0.4836	0.0778	0.4818	0.1126	0.3079	0.0736	0.5061
RSDL2	0.1597	0.1467	0.1264	0.2520	0.0908	0.4116	0.1143	0.3006
RSDR2	0.0684	0.5366	0.1006	0.3625	0.0626	0.5717	0.0861	0.4361

TABLE 11. Correlation Coefficients and P-Values for Crashes Versus Other Variables, Three-Legged Intersections (continued)

84 rural intersections, major road 4-lane, minor leg stop-controlled, CA and MI, 1993-95

Highway Variable	TOTACC		INJACC		TOTACCI		INJACCI	
	Corr.	P-value	Corr.	P-value	Corr.	P-value	Corr.	P-value
HI-1	0.0552	0.6181	0.0834	0.4507	0.0489	0.6590	0.0753	0.4958
HEI-1	0.2366	0.0303	0.1786	0.1041	0.1946	0.0761	0.1676	0.1275
VI-1	0.1742	0.1131	0.1614	0.1426	0.2437	0.0255	0.2287	0.0364
VEI-1	0.1673	0.1283	0.1530	0.1647	0.2208	0.0436	0.2060	0.0601
VCI-1	0.0251	0.8210	0.0513	0.6429	0.0637	0.5647	0.0676	0.5410
VCEI-1	0.1321	0.2308	0.1234	0.2633	0.1774	0.1065	0.1922	0.0799
ABSGRD1	0.0099	0.9288	0.1158	0.2942	0.0492	0.6567	0.0931	0.3997
SPD1	-0.3688	0.0006	-0.1314	0.2334	-0.3509	0.0011	-0.1591	0.1483
SPD2	-0.1133	0.3047	0.0174	0.8753	-0.0208	0.8513	0.0664	0.5483
LIGHT	0.3290	0.0022	0.2163	0.0481	0.3242	0.0026	0.2078	0.0579
STATE	0.1459	0.1853	-0.0823	0.4568	0.0327	0.7680	-0.1054	0.3402

major road speed (SPD1) correlate positively and negatively, respectively, with crashes, but they also correlate positively and negatively, respectively, with minor road ADT (cf. Table 15), and this may be an example of one variable representing another. The same applies to Peak Truck Percentage, which correlates negatively with both crashes and ADT (Tables 11 and 15). The variable STATE does not seem to play an important role in three-legged intersection crashes.

In Table 12, similar correlations are found between crashes on four-legged intersections and highway variables. ADT1 is a bit less significant than in the three-legged case. Peak turning percentages correlate with crashes, but the minor road turning percentages are less significant. HAZRAT remains insignificant, but now it is joined by number of driveways and median width, which are also insignificant. The typical Hazard Rating and number of driveways at four-legged intersections are

TABLE 12. Correlation Coefficients and P-Values for Crashes Versus Other Variables, Four-Legged Intersections

72 rural intersections, major road 4-lane, minor legs stop-controlled, CA and MI, 1993-95

Highway Variable	TOTACC		INJACC		TOTACCI		INJACCI	
	Corr.	P-value	Corr.	P-value	Corr.	P-value	Corr.	P-value
ADT1	0.1519	0.2027	0.3088	0.0083	0.1642	0.1682	0.2705	0.0216
ADT2	0.4801	0.0001	0.3123	0.0076	0.4612	0.0001	0.2945	0.0120
PK%TRUCK	-0.3035	0.0096	-0.3154	0.0070	-0.2932	0.0124	-0.3003	0.0104
PK%TURN	0.3225	0.0057	0.1651	0.1659	0.3400	0.0035	0.1810	0.1282
PK%LEFT	0.3117	0.0077	0.1598	0.1799	0.3258	0.0052	0.1745	0.1426
PK%THRU1	-0.3022	0.0099	-0.1457	0.2219	-0.3263	0.0052	-0.1647	0.1668
PK%LEFT1	0.3532	0.0023	0.2020	0.0889	0.3794	0.0010	0.2190	0.0645
PK%THRU2	0.1688	0.1625	0.0813	0.5033	0.2013	0.0948	0.1081	0.3729
PK%LEFT2	-0.1021	0.4003	-0.0883	0.4674	-0.1088	0.3702	-0.0961	0.4288
HAZRAT	-0.1663	0.1628	-0.1452	0.2237	-0.1367	0.2521	-0.1294	0.2789
NODRWY1	0.1780	0.1346	0.0389	0.7455	0.1702	0.1528	0.0132	0.9121
LTLN1	-0.2904	0.0133	-0.0662	0.5809	-0.1828	0.1244	-0.0127	0.9156
RTLN1	-0.1910	0.1080	-0.0525	0.6612	-0.1352	0.2574	-0.0450	0.7076
LTLN2	0.1689	0.1562	0.1723	0.1478	0.2181	0.0657	0.2016	0.0895
RTLN2	-0.0998	0.4042	-0.0056	0.9631	-0.1006	0.4007	-0.0132	0.9124
MEDWIDTH1	-0.1579	0.1852	0.0102	0.9324	-0.1172	0.3270	0.0289	0.8093
HAU	0.0101	0.9330	-0.0572	0.6333	-0.0413	0.7307	-0.0940	0.4320
DEV	0.0599	0.6174	0.1381	0.2473	0.0416	0.7289	0.1117	0.3500
RSD1	0.0884	0.4604	0.0095	0.9369	0.0619	0.6054	0.0168	0.8889
RSDL2	0.1278	0.2850	0.0110	0.9270	0.0846	0.4800	-0.0004	0.9971
RSDR2	0.3314	0.0045	0.2060	0.0826	0.3420	0.0033	0.2068	0.0814

TABLE 12. Correlation Coefficients and P-Values for Crashes Versus Other Variables, Four-Legged Intersections (continued)

72 rural intersections, major road 4-lane, minor legs stop-controlled, CA and MI, 1993-95

Highway Variable	TOTACC		INJACC		TOTACCI		INJACCI	
	Corr.	P-value	Corr.	P-value	Corr.	P-value	Corr.	P-value
HI-1	0.0396	0.7411	-0.0740	0.5366	-0.0139	0.9081	-0.0487	0.6848
HEI-1	-0.0423	0.7240	-0.0481	0.6880	-0.0829	0.4890	-0.0762	0.5249
VI-1	-0.0414	0.7298	0.0147	0.9025	-0.0453	0.7057	0.0081	0.9459
VEI-1	-0.0087	0.9421	0.0096	0.9360	-0.0181	0.8804	0.0033	0.9781
VCI-1	-0.0323	0.7879	-0.0018	0.9880	-0.0477	0.6908	-0.0041	0.9725
VCEI-1	0.0281	0.8145	0.0330	0.7831	0.0075	0.9499	0.0125	0.9171
ABSGRD1	-0.0177	0.8826	-0.0332	0.7822	-0.0012	0.9918	-0.0140	0.9073
SPD1	-0.2753	0.0193	-0.0306	0.7988	-0.2477	0.0359	-0.0007	0.9957
SPD2	-0.0778	0.5158	0.2541	0.0312	-0.0006	0.9963	0.2742	0.0197
LIGHT	0.0393	0.7430	-0.0377	0.7533	-0.0105	0.9303	-0.0633	0.5976
STATE	0.3441	0.0031	0.0827	0.4898	0.2032	0.0869	0.0251	0.8340

slightly less than they are at three-legged intersections, and this perhaps is relevant. However, median width, on average, is as high at four-legged intersections as at three-legged intersections, with a lower percentage of zero medians at four-legged intersections. Four-legged geometries, perhaps, lessen the safety effect of medians.

As with the three-legged intersections, major road turning lanes tend to decrease the number of crashes (or are insignificant for injury crashes), while minor road turning lanes increase the number of crashes or are insignificant. In the three-legged case, minor road turning lanes correlate strongly with minor road ADT, but this is not true for four-legged intersections. Peak truck percentage still correlates negatively with crashes and positively with ADT (Table 16), but LIGHT, which there is more of on the four-legged intersections, is now insignificant.

Neither angle variable HAU or DEV is significant on the four-legged intersections. Perhaps this is

TABLE 13. Correlation Coefficients and P-Values for Crashes Versus Other Variables, Signalized Intersections

49 signalized 4-legged rural intersections, 2-lane by 2-lane roads, CA and MI, 1993-95

Highway Variable	TOTACC		INJACC		TOTACCI		INJACCI	
	Corr.	P-value	Corr.	P-value	Corr.	P-value	Corr.	P-value
ADT1	0.0166	0.9099	0.0330	0.8219	0.0686	0.6393	0.0537	0.7138
ADT2	0.4490	0.0012	0.1020	0.4857	0.3873	0.0060	0.0392	0.7893
PK%TRUCK	0.2675	0.0631	0.4431	0.0014	0.2760	0.0549	0.4308	0.0020
PK%TURN	0.2110	0.1457	0.0147	0.9202	0.1496	0.3049	-0.0642	0.6615
PK%LEFT	0.2175	0.1333	0.0022	0.9879	0.1489	0.3071	-0.0801	0.5845
PK%THRU1	-0.2693	0.0614	-0.0660	0.6524	-0.2472	0.0868	-0.0086	0.9533
PK%LEFT1	0.3557	0.0121	0.1521	0.2967	0.3507	0.0135	0.1450	0.3203
PK%THRU2	0.1482	0.3096	0.1176	0.4210	0.1996	0.1692	0.1686	0.2468
PK%LEFT2	-0.3230	0.0236	-0.2526	0.0800	-0.3629	0.0104	-0.3101	0.0301
HAZRAT	0.0136	0.9260	0.0890	0.5433	0.0631	0.6667	0.1462	0.3163
NODRWY1	0.4005	0.0044	0.1823	0.2099	0.3641	0.0101	0.1021	0.4852
NODRWY2	0.0255	0.8618	0.0179	0.9028	0.0331	0.8212	0.0014	0.9924
LTLN1	-0.2046	0.1584	-0.0058	0.9683	-0.1022	0.4849	0.1088	0.4569
RTLN1	-0.1107	0.4490	-0.0728	0.6194	-0.1085	0.4582	-0.0824	0.5737
LTLN2	-0.1755	0.2277	-0.0760	0.6037	-0.1838	0.2062	-0.0688	0.6387
RTLN2	0.2425	0.0932	0.1363	0.3504	0.2216	0.1260	0.1301	0.3730
MEDWIDTH1	-0.0394	0.7882	-0.0216	0.8827	0.0190	0.8968	0.0401	0.7843
HAU	-0.0070	0.9610	0.0079	0.9571	0.0535	0.7153	0.0533	0.7163
DEV	-0.0587	0.6886	-0.1496	0.3051	-0.0874	0.5504	-0.1639	0.2605

TABLE 13. Correlation Coefficients and P-Values for Crashes Versus Other Variables, Signalized Intersections (continued)

49 signalized 4-legged rural intersections, 2-lane by 2-lane roads, CA and MI, 1993-95

Highway Variable	TOTACC		INJACC		TOTACCI		INJACCI	
	Corr.	P-value	Corr.	P-value	Corr.	P-value	Corr.	P-value
RSD1	-0.1129	0.4401	-0.0403	0.7832	-0.0984	0.5013	-0.0473	0.7468
RSDL1	-0.2085	0.1505	-0.1310	0.3695	-0.2115	0.1446	-0.1613	0.2681
RSD2	0.0079	0.9571	-0.0165	0.9104	0.0197	0.8929	-0.0421	0.7739
RSDL2	-0.0615	0.6749	-0.0829	0.5713	-0.0662	0.6514	-0.1134	0.4379
HI-1	-0.2232	0.1232	-0.1936	0.1825	-0.2398	0.0970	-0.1815	0.2120
HEI-1	-0.0152	0.9177	-0.0892	0.5421	-0.0651	0.6567	-0.1457	0.3178
HI-2	-0.2391	0.0980	-0.2039	0.1601	-0.2230	0.1236	-0.1867	0.1990
HEI-2	-0.1749	0.2295	-0.1540	0.2907	-0.1487	0.3079	-0.1363	0.3503
HICOM	-0.3268	0.0219	-0.2815	0.0501	-0.3317	0.0199	-0.2613	0.0697
HEICOM	-0.0817	0.5766	-0.1434	0.3258	-0.1186	0.4169	-0.1897	0.1918
VI-1	0.0634	0.6654	0.0277	0.8504	0.0113	0.9386	0.0519	0.7230
VEI-1	0.2196	0.1294	0.1316	0.3674	0.1631	0.2627	0.0891	0.5429
VCI-1	0.1942	0.1811	0.1029	0.4818	0.0782	0.5933	0.0302	0.8368
VCEI-1	0.0465	0.7511	0.0069	0.9627	0.0069	0.9626	-0.0549	0.7082
VI-2	0.1356	0.3531	0.0931	0.5246	0.1388	0.3417	0.1295	0.3752
VEI-2	0.1486	0.3081	0.1038	0.4778	0.1524	0.2957	0.1353	0.3541
VCI-2	0.1065	0.4663	0.0355	0.8086	0.0988	0.4993	0.0729	0.6187
VCEI-2	0.1472	0.3127	0.0875	0.5501	0.1466	0.3147	0.1217	0.4050
VICOM	0.1417	0.3316	0.0903	0.5372	0.1214	0.4061	0.1315	0.3677
VEICOM	0.2188	0.1310	0.1437	0.3245	0.1985	0.1715	0.1530	0.2938
VCICOM	0.1633	0.2621	0.0676	0.6442	0.1163	0.4263	0.0762	0.6026
VCEICOM	0.1534	0.2927	0.0815	0.5779	0.1345	0.3570	0.0834	0.5687

TABLE 13. Correlation Coefficients and P-Values for Crashes Versus Other Variables, Signalized Intersections (continued)

49 signalized 4-legged rural intersections, 2-lane by 2-lane roads, CA and MI, 1993-95

Highway Variable	TOTACC		INJACC		TOTACCI		INJACCI	
	Corr.	P-value	Corr.	P-value	Corr.	P-value	Corr.	P-value
ABSGRD1	0.0328	0.8228	-0.0365	0.8032	0.0269	0.8545	-0.0445	0.7614
ABSGRD2	-0.0822	0.5744	-0.0316	0.8294	-0.1005	0.4920	-0.0461	0.7530
SPD1	-0.1201	0.4111	0.1354	0.3538	-0.0744	0.6112	0.2006	0.1670
SPD2	0.0246	0.8668	0.2031	0.1616	0.0960	0.5118	0.2816	0.0499
PROT_LT	-0.2925	0.0414	-0.0767	0.6006	-0.1242	0.3951	0.0307	0.8340
LIGHT	-0.1336	0.3601	-0.0670	0.6473	-0.0619	0.6729	-0.0827	0.5723
STATE	0.3690	0.0091	0.1817	0.2115	0.1977	0.1732	0.0481	0.7429

because they are less variable on the four-legged intersections than on the three-legged intersections, with standard deviations on the three-legged intersections being about twice what they are on the four-legged intersections. Minor road sight distance right is significant on the four-legged intersections, an indication that left-turn and through traffic on the minor road may have a greater tendency toward crashes than right-turn traffic. All remaining alignment variables, including grade, are insignificant on the four-legged intersections. STATE appears to be significant for four-legged intersections, and this is consistent with Table 10.

Correlation coefficients of crashes with other variables for the signalized intersections are shown in Table 12. Remarkably, ADT1 is insignificant, and ADT2 is insignificant for injury crashes. This is perhaps due to the relatively small sample size and the presence of a variety of other influential factors. Peak Truck Percentage, which negatively correlates with ADT (cf. Table 16), although weakly, has a strong positive correlation with crashes. Peak turning percentages have some significant correlations, positive and negative, with crashes, and they will be examined more closely later in this chapter. HAZRAT, channelization, median width, and the angle variables are generally insignificant. Median widths are mostly zero, but HAU and DEV, the angle variables, are about as variable as in the three-legged intersections and still have a negligible effect. Sight distances and horizontal alignment are generally insignificant with the wrong sign. This indicates that when other factors are ignored, shorter sight distance and more horizontal curvature lead to fewer crashes. On the other hand, vertical alignment, although generally insignificant, has the right sign: other factors ignored, crashes rise with more grade change per unit distance. Speeds, though generally insignificant, seem to correlate positively with injury crashes. The existence of a protected left turn, which correlates positively with major road ADT, correlates negatively with crashes, as one might expect. This may, in part, account for the poor showing of ADT1. Finally, LIGHT is insignificant, but STATE shows a positive correlation with crashes.

As a general rule, correlations are similar for TOTACC and TOTACCI, and for INJACC and INJACCI. However, there are significant differences as one passes from all crashes to serious crashes (from TOTACC to INJACC, or from TOTACCI to INJACCI). Items that stand out include the following:

- ADT1 and ADT2 are both significant at three-legged and four-legged intersections, with ADT2 generally more significant; but at the signalized intersections, neither is significant except ADT2 with TOTACC and TOTACCI.

- PK%TRUCK correlates negatively with crashes of all types at three-legged and four-legged intersections and positively at signalized intersections.

- Peak turning percentage variables correlate strongly with crashes of all types at three-legged and four-legged intersections, and with TOTACC and TOTACCI at signalized intersections.

- NODRWY1 correlates positively with TOTACC and TOTACCI at all intersection types, but correlates insignificantly with INJACC and INJACCI at four-legged and signalized intersections.

- MEDWIDTH1 correlates negatively with TOTACC and TOTACCI at three-legged and four-legged intersections, but insignificantly with INJACC and INJACCI.

- Channelization variables correlate less significantly with INJACC and INJACCI than with TOTACC and TOTACCI and sometimes have correlation coefficients of unexpected sign.

- HAU and DEV correlate strongly with all crash types at three-legged intersections.

- Sight distance variables generally have insignificant correlation, except for RSDR2 at four-legged intersections, which correlates positively with all crash types.

- Horizontal alignment variables have insignificant correlation and/or correlation coefficients with unexpected sign, except for HEI-1 at three-legged intersections, while HICOM correlates negatively with all crash types at signalized intersections (fewer crashes at signalized intersections with major or minor road horizontal curves out to 250 feet (76 meters)).

- Vertical alignment variables are insignificant and/or have correlation of unexpected sign, except for VI-1 and VEI-1 at three-legged intersections and VEICOM at signalized intersections, the effect being stronger for TOTACC and TOTACCI than for INJACC and INJACCI.

- SPD1 correlates negatively with TOTACC and TOTACCI at three-legged and four-legged intersections; SPD2 correlates positively with INJACC and INJACCI at four-legged and signalized intersections.

- LIGHT correlates positively with all crash types at three-legged intersections.

- STATE correlates positively with TOTACC and TOTACCI at four-legged and signalized intersections.

- PROT_LT correlates negatively with TOTACC, but not significantly with INJACC and INJACCI for signalized intersections.

Information pertaining to TOTACC is summarized in Table 14. Features not already mentioned that are related to TOTACC include:

- ADT1 has lessened significance as one passes from three-legged to four-legged to signalized intersections.

- LIGHT correlates positively with TOTACC at three-legged and four-legged intersections (perhaps because lights are placed at high crash locations).

- LTLN1 correlates negatively with TOTACC on all three data sets.

- At three-legged intersections, HEI-1 and RSDL2 correlate positively with TOTACC.

- At four-legged intersections, horizontal and vertical variables have correlation coefficients of mixed signs with TOTACC, while all sight distances have coefficients of appropriate signs, with RSDR2's being significant.

- At signalized intersections, vertical variables have positive correlation with TOTACC, horizontal variables have negative correlation, and sight distance variables have mixed correlation.

ADT and State Versus Other Variables

It is generally recognized that ADT is the most important explanatory variable in modeling crashes. It is therefore appropriate to make a special effort to determine when other variables are correlated with ADT so that one can begin to distinguish effects that are properly due to these variables apart

TABLE 14. Correlates of TOTACC

84 Three-legged Intersections		
Positive correlates	Negative correlates	Insignificant correlates
ADT1*, ADT2* PK%TURN*, PK%LEFT* PK%LEFT1* PK%RIGHT2* (82 int.) NODRWY1*, HAU*, DEV* HEI-1*, LIGHT* LTLN2, RTLN2, RSDL2 VI-1, VEI-1, STATE	PK%TRUCK* PK%THRU1* PK%LEFT2* (82 int.) MEDWIDTH1* SPD1* LTLN1	PK%RIGHT1 HAZRAT (neg) RTLN1 (neg) RSD1, RSDR2 HI-1 VCI-1, VCEI-1 ABSGRD1 SPD2 (neg)
72 Four-legged Intersections		
Positive correlates	Negative correlates	Insignificant correlates
ADT2* PK%TURN* PK%LEFT*, PK%LEFT1* PK%RIGHT1* RSDR2*, STATE* PK%THRU2 (70 int.) NODRWY1, LTLN2 ADT1 (P-value = 0.2027)	PK%TRUCK* PK%THRU1* LTLN1* SPD1* HAZRAT RTLN1 MEDWIDTH1	PK%LEFT2, PK%RIGHT2 (both neg., 70 int.) RTLN2 (neg), HAU, DEV RSD1, RSDL2, HI-1 HEI-1 (neg), VI-1 (neg) VEI-1 (neg), VCI-1 (neg) VCEI-1, ABSGRD1 (neg) SPD2 (neg), LIGHT
49 Signalized Intersections		
Positive correlates	Negative correlates	Insignificant correlates
ADT2*, PK%TRUCK* PK%LEFT1* NODRWY1*, NODRWYCOM* F_3*, F_4*, STATE* RTLN2*, PK%TURN PK%LEFT, VCI-1,VEI-1 VEICOM	PK%THRU1* PK%LEFT2* PROT_LT* HI-2*, HICOM*, HI-1 LTLN1, RSDL1	ADT1, F_1, F_2, PK%RIGHT1 PK%THRU2, PK%RIGHT2 HAZRAT, NODRWY2 LTLN1, RTLN1 (neg) MEDWIDTH1 (neg) HAU, DEV (both neg) RSD1, RSDL2 (both neg), RSD2 HEI-1, HEI-2 (both neg) HEICOM (neg) VI-1, VCEI-1 VI-2, VCI-2, VEI-2, VCEI-2, VICOM, VCICOM, VCEICOM ABSGRD1, ABSGRD2 (neg) SPD1 (neg), SPD2, LIGHT (neg)

"Insignificant" means P-value in excess of 0.20, "*" means P-value less than 0.10

from their relationship to ADT. Likewise, in a multi-State study, it is desirable to have a sense of variables that correlate with the choice of State. This offers some guidance as to whether the State variable is genuinely relevant or whether it is a stand-in for other collected variables that may differ from one State to another. Tables 15, 16, and 17 exhibit correlations between ADT1, ADT2, and STATE versus other highway variables for our three classes of intersections.

Here we call attention to those variables that correlate with ADT and STATE in all three data sets (Tables 15, 16, and 17). Variables that correlate with major road ADT (ADT1), minor road ADT (ADT2), and STATE are:

<u>Positive Correlates with ADT1</u>:
PK%THRU, PK%THRU1, RSD1
<u>Negative Correlates with ADT1</u>:
PK%TRUCK, PK%TURN, PK%LEFT, PK%RIGHT1

<u>Positive Correlates with ADT2</u>:
STATE, PK%TURN, PK%LEFT, PK%LEFT1, PK%RIGHT1, NODRWY1, HEI-1
<u>Negative Correlates with ADT2</u>:
PK%TRUCK, PK%THRU1, HAZRAT?, MEDWIDTH1, SPD1

<u>Positive Correlates with STATE</u>:
ADT2, PK%TURN, PK%LEFT, NODRWY1, VCI-1?
<u>Negative Correlates with STATE</u>:
HAZRAT, LTLN1, RTLN1, MEDWIDTH1, SPD1, SPD2

The criteria for inclusion in this list are that the sign of the correlation coefficient is constant and that the P-value is less than 0.10 in at least two of the three intersection classes. Two exceptions are noted with question marks. Items that stand out from this catalogue include:

- Lower truck percentages on heavily traveled roads (trucks avoid these roads, or passenger cars favor them).

- Higher ADT2, more driveways, lower HAZRAT, and lower speeds in Michigan (presumably because it is less rural than California).

- Narrower medians and less major road channelization in Michigan (presumably reflecting differences in highway design principles from Michigan to California).

- Shorter longitudinal sight distances for higher major road ADT (somewhat baffling - possibly due to both occurring more often in California).

Another noteworthy feature of Tables 15, 16, and 17 is the relationship of LIGHT to ADT2 and STATE. On the three-legged and four-legged intersections, LIGHT and ADT2 are strongly

TABLE 15. Correlation Coefficients and P-Values for ADT and STATE Versus Intersection Variables, Three-Legged Intersections

84 rural intersections, major road 4-lane, minor leg stop-controlled, CA and MI, 1993-95

Highway Variable	ADT1		ADT2		STATE	
	Corr.	P-value	Corr.	P-value	Corr.	P-value
ADT1	1.0000	0.0000	0.1612	0.1429	-0.1156	0.2951
ADT2	0.1612	0.1429	1.0000	0.0000	0.2240	0.0406
STATE	-0.1156	0.2951	0.2240	0.0406	1.0000	0.0000
PK%TRUCK	-0.2349	0.0315	-0.2211	0.0433	-0.0993	0.3686
PK%TURN	-0.1079	0.3286	0.6842	0.0001	0.0251	0.8208
PK%LEFT	-0.1319	0.2317	0.6658	0.0001	0.0530	0.6323
PK%THRU1	0.1024	0.3540	-0.6183	0.0001	0.0213	0.8477
PK%LEFT1	-0.0353	0.7500	0.6404	0.0001	0.0132	0.9052
PK%LEFT2	-0.2709	0.0138	-0.1145	0.3058	-0.0380	0.7345
HAZRAT	0.1405	0.2025	-0.1416	0.1990	-0.4795	0.0001
NODRWY1	0.1347	0.2217	0.2166	0.0478	0.2425	0.0262
LTLN1	0.2027	0.0644	-0.1127	0.3076	-0.6794	0.0001
RTLN1	0.2585	0.0176	-0.0218	0.8442	-0.3067	0.0045
LTLN2	0.0195	0.8601	0.4336	0.0001	-0.1217	0.2701
RTLN2	0.0311	0.7786	0.2513	0.0211	0.1744	0.1127
MEDWIDTH1	0.0251	0.8211	-0.2267	0.0381	-0.3923	0.0002
HAU	-0.0164	0.8823	0.1250	0.2574	0.2042	0.0624
DEV	0.0992	0.3691	0.0418	0.7056	-0.0654	0.5545

TABLE 15 . Correlation Coefficients and P-Values for ADT and STATE Versus Intersection Variables, Three-Legged Intersections (continued)

84 rural intersections, major road 4-lane, minor leg stop-controlled, CA and MI, 1993-95

Highway Variable	ADT1		ADT2		STATE	
	Corr.	P-value	Corr.	P-value	Corr.	P-value
RSD1	0.2673	0.0140	0.0576	0.6030	0.0280	0.8003
RSDL2	0.1149	0.2998	-0.0424	0.7020	-0.0923	0.4038
RSDR2	0.1339	0.2248	0.0034	0.9755	-0.0818	0.4597
HI-1	0.0765	0.4892	0.0214	0.8472	-0.0258	0.8160
HEI-1	0.1326	0.2294	0.0347	0.7540	0.1134	0.3043
VI-1	0.2868	0.0082	0.0772	0.4852	-0.0484	0.6623
VEI-1	0.2501	0.0218	0.0509	0.6455	-0.0471	0.6706
VCI-1	-0.0203	0.8545	-0.0719	0.5159	0.1620	0.1410
VCEI-1	0.1607	0.1442	0.0854	0.4401	0.0467	0.6733
ABSGRD1	0.1299	0.2389	-0.0680	0.5387	-0.3052	0.0048
SPD1	-0.0703	0.5250	-0.2895	0.0076	-0.4397	0.0001
SPD2	0.0375	0.7348	-0.1394	0.2061	-0.7916	0.0001
LIGHT	0.0917	0.4070	0.3625	0.0007	0.3178	0.0032

TABLE 16. Correlation Coefficients and P-Values for ADT and STATE Versus Intersection Variables, Four-Legged Intersections

72 rural intersections, major road 4-lane, minor legs stop-controlled, CA and MI, 1993-95

Highway Variable	ADT1		ADT2		STATE	
	Corr.	P-value	Corr.	P-value	Corr.	P-value
ADT1	1.0000	0.0000	-0.1083	0.3653	-0.1436	0.2288
ADT2	-0.1083	0.3653	1.0000	0.0000	0.4082	0.0004
STATE	-0.1436	0.2288	0.4082	0.0004	1.0000	0.0000
PK%TRUCK	-0.2673	0.0232	-0.2044	0.0850	-0.2459	0.0374
PK%TURN	-0.3284	0.0049	0.6402	0.0001	0.2795	0.0174
PK%LEFT	-0.3205	0.0061	0.5921	0.0001	0.2622	0.0261
PK%THRU1	0.3087	0.0083	-0.6207	0.0001	-0.2240	0.0586
PK%LEFT1	-0.2754	0.0192	0.5777	0.0001	0.2677	0.0230
PK%THRU2	-0.3957	0.0007	0.3468	0.0033	0.0117	0.9231
PK%LEFT2	0.2937	0.0136	-0.0896	0.4609	-0.0982	0.4186
HAZRAT	0.1181	0.3230	-0.2264	0.0558	-0.3059	0.0090
NODRWY1	-0.0582	0.6272	0.2336	0.0483	0.3567	0.0021
LTLN1	0.0548	0.6474	-0.2563	0.0297	-0.8433	0.0001
RTLN1	0.1089	0.3623	-0.0734	0.5403	-0.4261	0.0002
LTLN2	-0.0736	0.5389	0.0935	0.4349	-0.0976	0.4148
RTLN2	0.0991	0.4077	-0.0642	0.5920	-0.0761	0.5250
MEDWIDTH1	0.2571	0.0292	-0.2597	0.0276	-0.3968	0.0006
HAU	-0.0431	0.7195	-0.0592	0.6214	0.0206	0.8636
DEV	-0.0687	0.5663	0.0417	0.7282	0.0542	0.6514

TABLE 16. Correlation Coefficients and P-Values for ADT and STATE Versus Intersection Variables, Four-Legged Intersections (continued)

72 rural intersections, major road 4-lane, minor legs stop-controlled, CA and MI, 1993-95

Highway Variable	ADT1		ADT2		STATE	
	Corr.	P-value	Corr.	P-value	Corr.	P-value
RSD1	0.0798	0.5054	-0.0628	0.6003	-0.0706	0.5557
RSDL2	0.0589	0.6233	0.0066	0.9560	-0.0072	0.9523
RSDR2	-0.0233	0.8458	0.2311	0.0508	-0.0565	0.6372
HI-1	0.0037	0.9754	-0.0549	0.6469	-0.0881	0.4620
HEI-1	0.0080	0.9472	0.3428	0.0032	0.2339	0.0587
VI-1	-0.0115	0.9237	-0.1108	0.3540	0.1860	0.1178
VEI-1	-0.0132	0.9122	-0.1220	0.3075	0.1794	0.1316
VCI-1	-0.0741	0.5365	-0.0976	0.4147	0.2322	0.0497
VCEI-1	-0.0215	0.8575	-0.0958	0.4233	0.2107	0.0757
ABSGRD1	0.0926	0.4392	-0.2053	0.0837	-0.2760	0.0190
SPD1	0.2020	0.0888	-0.3133	0.0074	-0.4738	0.0001
SPD2	0.0858	0.4738	-0.0523	0.6627	-0.5648	0.0001
LIGHT	-0.1626	0.1725	0.2560	0.0300	0.3873	0.0008

TABLE 17. Correlation Coefficients and P-Values for ADT and STATE Versus Intersection Variables, Signalized Intersections

49 signalized 4-legged rural intersections, 2-lane by 2-lane roads, CA and MI, 1993-95

Highway Variable	ADT1		ADT2		STATE	
	Corr.	P-value	Corr.	P-value	Corr.	P-value
ADT1	1.0000	0.0000	0.1965	0.1759	-0.4544	0.0010
ADT2	0.1965	0.1759	1.0000	0.0000	0.2397	0.0972
STATE	-0.4544	0.0010	0.2397	0.0972	1.0000	0.0000
PK%TRUCK	-0.2051	0.1575	-0.1001	0.4938	0.1836	0.2067
PK%TURN	-0.2818	0.0498	0.4554	0.0010	0.3116	0.0293
PK%LEFT	-0.2630	0.0679	0.4940	0.0003	0.2893	0.0438
PK%THRU1	0.3358	0.0183	-0.5271	0.0001	-0.1819	0.2111
PK%LEFT 1	-0.1856	0.2018	0.5179	0.0001	0.0997	0.4955
PK%THRU2	-0.3224	0.0239	0.1868	0.1988	-0.0342	0.8157
PK%LEFT2	0.1800	0.2158	-0.1472	0.3127	-0.0214	0.8839
HAZRAT	0.2309	0.1105	-0.1939	0.1818	-0.3096	0.0304
NODRWY1	-0.0642	0.6611	0.3133	0.0284	0.3613	0.0108
NODRWY2	0.1781	0.2209	0.1905	0.1899	-0.0278	0.8495
LTLN1	0.1756	0.2276	-0.0651	0.6569	-0.3305	0.0204
RTLN1	-0.1709	0.2404	-0.0648	0.6583	-0.0633	0.6657
LTLN2	-0.3372	0.0178	0.0284	0.8463	0.1088	0.4570
RTLN2	-0.0858	0.5578	0.3733	0.0083	0.2100	0.1476
MEDWIDTH1	-0.0155	0.9159	-0.1377	0.3456	-0.3992	0.0045
HAU	0.1417	0.3313	-0.1621	0.2659	-0.2000	0.1682
DEV	-0.1103	0.4504	-0.0192	0.8957	0.0573	0.6956

TABLE 17. Correlation Coefficients and P-Values for ADT and STATE Versus Intersection Variables, Signalized Intersections (continued)

49 signalized 4-legged rural intersections, 2-lane by 2-lane roads, CA and MI, 1993-95

Highway Variable	ADT1		ADT2		STATE	
	Corr.	P-value	Corr.	P-value	Corr.	P-value
RSD1	0.5043	0.0002	0.0011	0.9940	-0.1830	0.2081
RSDL1	0.3642	0.0101	-0.1004	0.4927	-0.1342	0.3578
RSD2	0.3954	0.0049	0.1536	0.2921	-0.0701	0.6320
RSDL2	0.0701	0.6325	-0.0805	0.5827	-0.0702	0.6317
HI-1	-0.1852	0.2027	-0.0010	0.9944	0.1156	0.4289
HEI-1	0.0018	0.9903	0.3390	0.0172	0.1476	0.3115
HI-2	0.1542	0.2900	-0.0449	0.7596	-0.2588	0.0726
HEI-2	0.1706	0.2412	-0.0373	0.7994	-0.2991	0.0368
HICOM	-0.0688	0.6386	-0.0266	0.8560	-0.0503	0.7314
HEICOM	0.0675	0.6448	0.3044	0.0334	0.0234	0.8730
VI-1	-0.0187	0.8987	-0.0131	0.9287	0.0289	0.8439
VEI-1	-0.0722	0.6221	0.1015	0.4875	0.1496	0.3048
VCI-1	-0.0985	0.5009	0.0558	0.7036	0.2056	0.1565
VCEI-1	0.1259	0.3889	0.0336	0.8187	0.0880	0.5478
VI-2	-0.1754	0.2281	-0.1649	0.2575	-0.1547	0.2886
VEI-2	-0.1287	0.3783	-0.1801	0.2156	-0.1431	0.3267
VCI-2	-0.1837	0.2064	-0.1725	0.2359	-0.1489	0.3073
VCEI-2	-0.1569	0.2818	-0.1358	0.3523	-0.1089	0.4566
VICOM	-0.1553	0.2866	-0.1441	0.3233	-0.1170	0.4234
VEICOM	-0.1403	0.3361	-0.1123	0.4424	-0.0605	0.6794
VCICOM	-0.1999	0.1685	-0.1364	0.3500	-0.0633	0.6658
VCEICOM	-0.0820	0.5754	-0.1059	0.4688	-0.0566	0.6992

TABLE 17. Correlation Coefficients and P-Values for ADT and STATE Versus Intersection Variables, Signalized Intersections (continued)

49 signalized 4-legged rural intersections, 2-lane by 2-lane roads, CA and MI, 1993-95

Highway Variable	ADT1		ADT2		STATE	
	Corr.	P-value	Corr.	P-value	Corr.	P-value
ABSGRD1	-0.0795	0.5873	-0.0396	0.7871	-0.0446	0.7611
ABSGRD2	-0.0075	0.9590	-0.1583	0.2774	-0.0037	0.9800
SPD1	-0.1053	0.4713	-0.3336	0.0192	-0.1853	0.2025
SPD2	-0.1776	0.2222	-0.1849	0.2034	-0.0871	0.5520
PROT_LT	0.4829	0.0004	-0.0023	0.9875	-0.7943	0.0001
LIGHT	0.2200	0.1288	-0.0421	0.7739	-0.2808	0.0506

correlated: if the minor road ADT is high, there will tend to be lighting. At the signalized intersections, 80% of which have lighting (see Table 7), there is no correlation with minor road ADT. On the three-legged and four-legged intersections, LIGHT and STATE are positively correlated. As noted, Michigan tends to be less rural and to have more minor road ADT, and hence more lighting. But LIGHT negatively correlates with STATE on signalized intersections, an indication that California signalized intersections are more likely to have lighting.

Correlations Between Intersection Variables

Tables 18, 19, and 20 show correlations between pairs of intersection variables within the three data sets. Only those correlations are shown for which P-values are less than 0.10. In addition, rather than exhibit all peak turning percentage, channelization, alignment, and sight distance variables, we only show representative variables from each of these classes.

Items of special note in these tables that have not already been mentioned include the following:

- Wider medians, left-turn lanes on the major road, and fewer major road driveways tend to go together in the three-legged and four-legged samples.

TABLE 18. Correlations Between Intersection Variables in the Three-Legged Sample

VARIABLE	POSITIVE CORRELATES	NEGATIVE CORRELATES
ADT1	LTLN1, RTLN1, RSD1, VI-1, VEI-1	PK%TRUCK, PK%LEFT2
ADT2	PK%TURN, PK%LEFT, PK%LEFT1, NODRWY1, LTLN2, RTLN2, LIGHT, STATE	PK%TRUCK, PK%THRU1, MEDWIDTH1, SPD1
STATE CA = 0, MI = 1	ADT2, NODRWY1, HAU, LIGHT	HAZRAT, LTLN1, RTLN1, MEDWIDTH1, ABSGRD1, SPD1, SPD2
PK%TRUCK	MEDWIDTH1, HAU, SPD1	ADT1, ADT2, NODRWY1, RSD1, RSDR2, HEI-1, VI-1, VEI-1, LIGHT
PK%TURN	ADT2, PK%LEFT, PK%LEFT1, LTLN2, RTLN2, LIGHT	PK%THRU1
SPD1	PK%TRUCK, HAZRAT, LTLN1, RTLN1, MEDWIDTH1, SPD2	ADT2, NODRWY1, RSD1, RSDR2, HI-1, HEI-1, VI-1, VEI-1, VCEI-1, LIGHT, STATE
SPD2	HAZRAT, LTLN1, RTLN1, MEDWIDTH1, ABSGRD1, SPD1	LIGHT, STATE
HAZRAT	LTLN1, HI-1, ABSGRD1, SPD1, SPD2	NODRWY1, LIGHT, STATE
NODRWY1	ADT2, RSD1, RSDR2, HI-1, HEI-1, VEI-1, LIGHT, STATE	PK%TRUCK, HAZRAT, LTLN1, RTLN1, MEDWIDTH1, SPD1
MEDWIDTH1	PK%TRUCK, LTLN1, RTLN1, SPD1, SPD2	ADT2, NODRWY1, RSD1, HI-1, HEI-1, LIGHT
LTLN1	ADT1, HAZRAT, RTLN1, MEDWIDTH1, SPD1, SPD2, STATE	NODRWY1, HI-1, HEI-1, VCI-1, LIGHT
HAU	PK%TRUCK, LIGHT, STATE	RSD1, HI-1, ABSGRD1
DEV	RSD1, HEI-1	
RSD1 Reciprocal Sight Distance	ADT1, NODRWY1, DEV, HI-1, HEI-1, VI-1, VEI-1, ABSGRD1, LIGHT	PK%TRUCK, MEDWIDTH1, HAU, SPD1
HEI-1 Horizontal out to 800 ft	NODRWY1, DEV, RSD1, RSDL2, RSDR2, HI-1, VEI-1	PK%TRUCK, LTLN1, MEDWIDTH1, SPD1
VEI-1 Vertical out to 800 ft	ADT1, NODRWY1, RSD1, RSDR2, HI-1, HEI-1, VI-1, VCI-1, VCEI-1, ABSGRD1	PK%TRUCK, SPD1
ABSGRD1	HAZRAT, RTLN1, RSD1, RSDR2, HI-1, VI-1, VEI-1, VCEI-1, SPD2	HAU, STATE
LIGHT No = 0, Yes = 1	ADT2, PK%TURN, PK%LEFT, NODRWY1, LTLN2, HAU, RSD1, STATE	PK%TRUCK, PK%THRU1, HAZRAT, LTLN1, MEDWIDTH1, SPD1, SPD2

1 ft = 0.305 m

TABLE 19. Correlations Between Intersection Variables in the Four-Legged Sample

VARIABLE	POSITIVE CORRELATES	NEGATIVE CORRELATES
ADT1	PK%THRU1, PK%LEFT2, MEDWIDTH1, SPD1	PK%TRUCK, PK%TURN, PK%LEFT, PK%LEFT1, PK%THRU2
ADT2	PK%TURN, PK%LEFT, PK%LEFT1, PK%THRU2, NODRWY1, RSDR2, HEI-1, LIGHT, STATE	PK%TRUCK, PK%THRU1, HAZRAT, LTLN1, MEDWIDTH1, ABSGRD1, SPD1
STATE CA = 0, MI = 1	ADT2, PK%TURN, PK%LEFT, PK%LEFT1, NODRWY1, HEI-1, VCI-1, VCEI-1, LIGHT	PK%TRUCK, PK%THRU1, HAZRAT, LTLN1, RTLN1, MEDWIDTH1, ABSGRD1, SPD1, SPD2
PK%TRUCK	PK%THRU2, LTLN1, RTLN1, RTLN2, SPD1	ADT1, ADT2, PK%TURN, PK%LEFT, PK%LEFT1, HAZRAT, NODRWY1, DEV, RSD1, RSDL2, RSDR2, LIGHT, STATE
PK%TURN	ADT2, PK%LEFT, PK%LEFT1, PK%THRU2, NODRWY1, LIGHT, STATE	ADT1, PK%TRUCK, PK%THRU1, LTLN1, RTLN1, RTLN2, MEDWIDTH1, SPD1
SPD1	ADT1, PK%TRUCK, PK%THRU1, LTLN1, RTLN1, RTLN2, MEDWIDTH1, SPD2	ADT2, PK%TURN, PK%LEFT, PK%LEFT1, NODRWY1, RSDL2, RSDR2, HEI-1, LIGHT, STATE
SPD2	LTLN1, RTLN1, DEV, SPD1	NODRWY1, HEI-1, VI-1, VEI-1, VCI-1, VCEI-1, STATE
HAZRAT	LTLN1, DEV, RSD1, RSDL2, RSDR2, HI-1, ABSGRD1	ADT2, PK%TRUCK, PK%THRU2, RTLN1, RTLN2, STATE
NODRWY1	ADT2, PK%TURN, PK%LEFT, PK%LEFT1, RSD1, RSDL2, RSDR2, HEI-1, LIGHT, STATE	PK%TRUCK, PK%THRU1, LTLN1, RTLN1, RTLN2, MEDWIDTH1, SPD1, SPD2
MEDWIDTH1	ADT1, PK%THRU1, LTLN1, LTLN2, SPD1	ADT2, PK%TURN, PK%LEFT, PK%LEFT1, NODRWY1, LIGHT, STATE
LTLN1	PK%TRUCK, HAZRAT, RTLN1, MEDWIDTH1, SPD1, SPD2	ADT2, PK%TURN, PK%LEFT, PK%LEFT1, NODRWY1, VCI-1, VCEI-1, LIGHT, STATE
HAU	DEV, ABSGRD1	
DEV	HAZRAT, HAU, RSDR2, ABSGRD1, SPD2	PK%TRUCK, RTLN1, RTLN2
RSD, Reciprocal Sight Distance	HAZRAT, NODRWY1, RSDL2, RSDR2, HI-1, ABSGRD1	PK%TRUCK, PK%THRU2, RTLN1
HEI-1, Horizontal out to 800 ft	ADT2, NODRWY1, STATE	SPD1, SPD2
VEI-1, Vertical out to 800 ft	RSDL2, VI-1, VCI-1, VCEI-1, ABSGRD1	RTLN2, SPD2
ABSGRD1	HAZRAT, HAU, DEV, RSD1, RSDL2, RSDR2, HI-1, VEI-1	ADT2, PK%THRU2, RTLN1, RTLN2, STATE
LIGHT No = 0, Yes = 1	ADT2, PK%TURN, PK%LEFT, PK%LEFT1, NODRWY1, STATE	PK%TRUCK, PK%THRU1, LTLN1, RTLN1, RTLN2, MEDWIDTH1, SPD1

1 ft = 0.305 m

TABLE 20. Correlations Between Intersection Variables in the Signalized Sample

VARIABLE	POSITIVE CORRELATES	NEGATIVE CORRELATES
ADT1	PK%THRU1, RSD1, RSDL1, RSD2, PROT_LT	PK%TURN, PK%LEFT, PK%THRU2, LTLN2, STATE
ADT2	PK%TURN, PK%LEFT, PK%LEFT1, NODRWY1, RTLN2, HEI-1, HEICOM, STATE	PK%THRU1, SPD1
STATE CA = 0, MI = 1	ADT2, PK%TURN, PK%LEFT, NODRWY1	ADT1, HAZRAT, LTLN1, MEDWIDTH1, HI-2, HEI-2, PROT_LT, LIGHT
PK%TRUCK	SPD2	
PK%TURN	ADT2, PK%LEFT, PK%LEFT1, NODRWY1, HEI-1, HEICOM, VEI-1, VEICOM, VCI-2, VCEI-2, VCEICOM, ABSGRD1	ADT2, PK%THRU1, PK%THRU2, SPD1
SPD1	RTLN1, RTLN2, SPD2	ADT2, PK%TURN, PK%LEFT, NODRWY1, NODRWY2, RSD1, RSDL1, RSDL2, HEI-1, HEICOM, VCEI-1, LIGHT
SPD2	PK%TRUCK, PK%THRU2, LTLN1, RTLN2, SPD1	PK%LEFT2, NODRWY1, NODRWY2, RSD1, RSDL1, RSD2, RSDL2, HEICOM, LIGHT
HAZRAT	RSD1, VCEI-1	NODRWY2, RTLN1, VCI-1, STATE
NODRWY1	ADT2, PK%TURN, PK%LEFT, NODRWY2, RSD2, RSDL2, STATE	PK%THRU1, HI-2, PROT_LT, SPD1, SPD2
NODRWY2	PK%LEFT2, NODRWY1, RSDL2, LIGHT	PK%THRU2, HAZRAT, SPD1, SPD2
MEDWIDTH1	VI-2, VEI-2, VCI-2, VCEI-2, VCICOM, VCEICOM, PROT_LT	RTLN2, HAU, STATE
LTLN1	LTLN2, PROT_LT, SPD2	HEI-1, VCI-1, STATE
HAU	MEDWIDTH1	HI-1, HICOM
DEV	HI-1, HICOM	
RSD1, Reciprocal Sight Distance along Major Road	ADT1, HAZRAT, RSDL1, RSD2, RSDL2, HI-1, HEI-1, HEICOM, VI-1, VEI-1, VCEI-1, ABSGRD1, PROT_LT, VEI-1, ABSGRD1, LIGHT	RTLN1, SPD1, SPD2
RSD2, Recip. Sight Dist along Minor Rd	ADT1, PK%LEFT2, NODRWY1, RSD1, RSDL1, HI-2, HEI-2, ABSGRD1, ABSGRD2	PK%THRU2, SPD2
HEICOM Horizontal out to 800 ft, All legs	ADT2, PK%TURN, PK%LEFT, PK%LEFT1, RSD1, RSDL2, HI-1, HI-2, HICOM, HEI-1, HEI-2, VEI-1, VCEI-1, ABSGRD1	PK%THRU1, SPD1, SPD2
VEICOM, Vertical out to 800 ft, All legs	PK%TURN, VI-1, VI-2, VICOM, VEI-1, VEI-2, VCI-1, VCI-2, VCICOM, VCEI-1, VCEI-2, VCEICOM, ABSGRD1, ABSGRD2	PK%THRU

1 ft = 0.305 m

TABLE 20. Correlations Between Intersection Variables in the Signalized Sample (continued)

VARIABLE	POSITIVE CORRELATES	NEGATIVE CORRELATES
ABSGRD1 Major Road	PK%TURN, PK%LEFT, RSD1, RSD2, RSDL2, HI-1, HICOM, VI-1, VICOM, VEI-1, VEICOM, VCI-1, VCEI-1, VCEICOM, ABSGRD2	PK%THRU1, MEDWIDTH1
ABSGRD2 Minor Road	RSD2, VI-2, VICOM, VEI-2, VEICOM, VCI-2, VCICOM, VCEI-2, VCEICOM, ABSGRD1	
LIGHT No = 0, Yes = 1	PK%LEFT2, NODRWY2, RSDL1, RSDL2	PK%LEFT1, RTLN2, SPD1, SPD2, STATE
PROT_LT No = 0, Yes = 1	ADT1, LTLN1, MEDWIDTH1, RSD1, RSDL1, HEI-2	NODRWY1, STATE

- Major road speeds tend to be higher when major road channelization is present and when medians are wider in the three-legged and four-legged samples.

- Major road speeds tend to be lower when minor road ADT is higher, when there are more major road driveways, when sight distance is restricted, when lighting is present, or when horizontal or vertical curves are present. This happens in all three data sets.

- Lighting is more likely to be present when minor road ADT is high, when the peak turning percentages are high, or when the number of major road driveways is high. This applies to the three-legged and four-legged samples.

- At signalized intersections, protected left turns are more likely to occur in California than in Michigan (17 out of 18 CA signalized intersections have protected left turns, while only 4 out of 31 MI signalized intersections do).

A couple of anomalies are evident from the tables. In Table 19, for the four-legged intersections, a negative correlation exists between the presence of a left-turn lane on the major road and peak turning percentages, including major road left turns. When a higher fraction of the traffic is turning, it is less likely that there is a turning lane. It may be that the motive for installing turning lanes is more to prevent disruption of through traffic than to assist turning drivers. Another oddity, this time in Table 20 for signalized intersections, is the negative correlation between HAZRAT and VCI-1, accompanied by a positive correlation between HAZRAT and VCEI-1. HAZRAT is measured out to 250 feet (76 meters) as is VCI-1, while VCEI-1 is measured out to 800 feet (244 meters). A total of 37 out of 49 signalized intersections have VCI equal to zero, while 25 out of 49 have VCEI equal to zero. The two highest hazard ratings occur at intersections with VCI equal to zero, but with VCEI equal to 6.0 and 3.46 (average VCEI is 0.952), and this contributes to the anomalous correlation.

Correlations for Single-Vehicle and Multiple-Vehicle Crashes at Signalized Intersections

For the signalized intersections, an attempt was made to analyze single-vehicle crashes and multiple-vehicle crashes separately and to relate them to various flow patterns derived from the traffic data. The variables TOTACCS and TOTACCM, representing a decomposition of TOTACC into single-vehicle crashes and multiple-vehicle crashes, were compared with the intersection variables and with the flow variables F_1, F_2, F_3, F_4, PRODFADJ, PRODFOPP, and SUMF. The correlation coefficients and P-values are shown in Table 21.

Conclusions that can be drawn from Table 21 with regards to the signalized sample are:

- Single-vehicle crashes show a slight negative correlation with major road ADT and major road flows.

- Multiple-vehicle crashes are strongly correlated with minor road flows and with the interaction variable for adjacent legs, as well as with peak truck percentage and left-turn percentage on the major road.

- HAZRAT's correlation coefficient has the correct sign for single-vehicle crashes, but is insignificant, as are the driveway variables.

- Horizontal alignment variables are negatively correlated with both kinds of crashes (one may speculate that horizontal alignment causes drivers to exert extra caution at signalized intersections), and protected left turns reduce both kinds of crashes.

- Minor road vertical alignment contributes to single-vehicle crashes, and lighting reduces these crashes significantly.

The correlation of both kinds of crashes with the STATE variable has already been noted, i.e., Michigan is overrepresented in crashes. However, since STATE has a strong negative correlation with PROT_LT, it is not clear which of these two variables has the dominant influence.

Turning Percentage Variables

Intersection crashes are naturally related to turning percentages at intersections. However, sorting out the relative importance of left turns versus right turns and turns from the major road versus the

TABLE 21. Correlation Coefficients and P-Values for Single-Vehicle and Multiple-Vehicle Crashes Versus Signalized Intersection Variables

49 signalized 4-legged rural intersections, 2-lane by 2-lane roads, CA and MI, 1993-95

Highway Variable	TOTACCS		TOTACCM	
	Corr.	P-value	Corr.	P-value
ADT1	-0.1175	0.4213	0.0386	0.7923
ADT2	0.1682	0.2480	0.4545	0.0010
F_1	-0.1140	0.4365	0.0303	0.8365
F_2	-0.0722	0.6222	0.0809	0.5807
F_3	0.1650	0.2571	0.3048	0.0332
F_4	0.0584	0.6904	0.4461	0.0013
PRODFADJ	0.1018	0.4865	0.3931	0.0052
PRODFOPP	-0.0588	0.6882	0.1543	0.2899
SUMF	-0.0201	0.8907	0.2349	0.1043
PK%TRUCK	0.1853	0.2025	0.2558	0.0761
PK%TURN	0.1132	0.4386	0.2075	0.1526
PK%LEFT	0.1188	0.4161	0.2136	0.1406
PK%THRU1	-0.0793	0.5882	-0.2763	0.0546
PK%LEFT 1	0.1450	0.3203	0.3579	0.0116
PK%THRU2	-0.0374	0.7986	0.1664	0.2533
PK%LEFT2	-0.1490	0.3069	-0.3220	0.0240
HAZRAT	0.1001	0.4938	-0.0030	0.9838
NODRWY1	0.1245	0.3940	0.4098	0.0035
NODRWY2	-0.1132	0.4386	0.0475	0.7459

TABLE 21. Correlation Coefficients and P-Values for Single-Vehicle and Multiple-Vehicle Crashes Versus Signalized Intersection Variables (continued)

49 signalized 4-legged rural intersections, 2-lane by 2-lane roads, CA and MI, 1993-95

Highway Variable	TOTACCS		TOTACCM	
	Corr.	P-value	Corr.	P-value
LTLN1	-0.0669	0.6478	-0.2088	0.1499
RTLN1	0.0683	0.6410	-0.1314	0.3682
LTLN2	-0.0727	0.6194	-0.1764	0.2253
RTLN2	0.2169	0.1345	0.2232	0.1231
MEDWIDTH1	-0.0724	0.6209	-0.0297	0.8395
HAU	-0.0737	0.6149	0.0054	0.9704
DEV	-0.0708	0.6289	-0.0508	0.7288
RSD1	-0.1429	0.3275	-0.0965	0.5095
RSDL1	-0.1760	0.2265	-0.1938	0.1821
RSD2	-0.1453	0.3191	0.0341	0.8159
RSDL2	-0.2095	0.1485	-0.0293	0.8415
HI-1	-0.1041	0.4766	-0.2223	0.1248
HEI-1	-0.1815	0.2121	0.0156	0.9150
HI-2	-0.1812	0.2128	-0.2258	0.1187
HEI-2	-0.1750	0.2292	-0.1577	0.2792
HICOM	-0.1924	0.1853	-0.3184	0.0258
HEICOM	-0.2382	0.0993	-0.0461	0.7529

TABLE 21. Correlation Coefficients and P-Values for Single-Vehicle and Multiple-Vehicle Crashes Versus Signalized Intersection Variables (continued)

49 signalized 4-legged rural intersections, 2-lane by 2-lane roads, CA and MI, 1993-95

Highway Variable	TOTACCS		TOTACCM	
	Corr.	P-value	Corr.	P-value
VI-1	-0.0295	0.8403	0.0735	0.6156
VEI-1	-0.0421	0.7741	0.2442	0.0908
VCI-1	-0.0302	0.8367	0.2148	0.1384
VCEI-1	-0.1738	0.2324	0.0808	0.5811
VI-2	0.3434	0.0157	0.0856	0.5587
VEI-2	0.3269	0.0219	0.1026	0.4829
VCI-2	0.3238	0.0232	0.0578	0.6934
VCEI-2	0.3051	0.0331	0.1050	0.4729
VICOM	0.2749	0.0559	0.1043	0.4759
VEICOM	0.2629	0.0680	0.1895	0.1922
VCICOM	0.2818	0.0498	0.1264	0.3867
VCEICOM	0.1924	0.1853	0.1315	0.3680
ABSGRD1	-0.0753	0.6071	0.0487	0.7398
ABSGRD2	0.1030	0.4812	-0.1068	0.4652
SPD1	0.0794	0.5875	-0.1435	0.3253
SPD2	0.1588	0.2758	-0.0015	0.9919
PROT-LT	-0.3996	0.0045	-0.2449	0.0899
LIGHT	-0.4359	0.0017	-0.0672	0.6464
STATE	0.3356	0.0184	0.3387	0.0173

minor road, as well as the direction of the effect in each case, is not easy since the turning percentage variables are strongly related to one another. In Tables 22 and 23, some of the relevant correlation coefficients are presented.

Table 22 shows the correlation coefficients for the various turning percentages. It supports the conventional wisdom, although not perfectly. PK%LEFT1 correlates positively with PK%RIGHT2

TABLE 22. Correlation Coefficients and P-Values for Peak Turning Percentage Variables

Variable Pair		3-legged Corr.	3-legged P-value	4-legged Corr.	4-legged P-value	signalized Corr.	signalized P-value
PK%LEFT1 vs.	PK%THRU1	-0.8853	0.0001	-0.8964	0.0001	-0.7744	0.0001
	PK%RIGHT1	0.5588	0.0001	0.6519	0.0001	0.2101	0.1473
	PK%LEFT2	-0.2891	0.0084	-0.1704	0.1584	-0.4724	0.0006
	PK%THRU2			0.2642	0.0271	0.1307	0.3709
	PK%RIGHT2	0.2891	0.0084	-0.0109	0.9290	0.3477	0.0144
PK%THRU1 vs.	PK%RIGHT1	-0.8803	0.0001	-0.9205	0.0001	-0.7813	0.0001
	PK%LEFT2	0.0165	0.8829	0.0248	0.8385	0.0716	0.6248
	PK%THRU2			-0.2937	0.0136	-0.0995	0.4965
	PK%RIGHT2	-0.0165	0.8829	0.1621	0.1801	0.0491	0.7378
PK%RIGHT1 vs.	PK%LEFT2	0.2673	0.0152	0.1077	0.3750	0.3554	0.0122
	PK%THRU2			0.2686	0.0245	0.0248	0.8657
	PK%RIGHT2	-0.2673	0.0152	-0.2670	0.0255	-0.4189	0.0027
PK%LEFT2 vs.	PK%THRU2			-0.1966	0.1028	-0.6536	0.0001
	PK%RIGHT2	-1.0000	0.0001	-0.7874	0.0001	-0.2543	0.0779
PK%THRU2 vs.	PK%RIGHT2			-0.4495	0.0001	-0.5657	0.0001

(exception: the four-legged intersections where the correlation is negligible) and negatively with PK%LEFT2. Likewise, PK%RIGHT1 correlates positively with PK%LEFT2 and negatively with PK%RIGHT2. The positive correlations are expected since the corresponding flows are reversals of one another. The negative correlations result at least in part from the fact that PK%LEFT2 and PK%RIGHT2 are negatively correlated with one another. The four-legged intersections are less regular than the other two intersection classes. These correlations are, of course, based on rough information since peak hours in the morning and the afternoon were selected in a crude manner and there is no reason why flows should reverse in any precise way (even if peak hours were selected with great care).

Table 23, extracted in part from Tables 11, 12, and 13, shows the relationship between the crash variables and the turning percentages for the three classes of intersections. What immediately strikes the eye is that PK%LEFT1 is positively correlated with all types of crashes at all types of intersections, while PK%LEFT2 (or for that matter PK%THRU1) is negatively correlated with all types of crashes at all types of intersections. Since PK%LEFT2 = 100 - PK%THRU2 - PK%RIGHT2, what is being said is that the sum of PK%THRU2 and PK%RIGHT2 is positively correlated with crashes. The last two columns of Table 23 confirm this. In general, both PK%THRU2 and PK%RIGHT2 are positively correlated with crashes; in cases where one of them

TABLE 23. Correlation Coefficients and P-Values for Crashes Versus Peak Turning Percentage Variables

	PK%LEFT1	PK%THRU1	PK%RIGHT1	PK%LEFT2	PK%THRU2	PK%RIGHT2
TOTACC						
3-legged	0.2786, 0.0103	-0.2170, 0.0474	0.1027, 0.3525	-0.2096, 0.0588		0.2096, 0.0588
4-legged	0.3532, 0.0023	-0.3022, 0.0099	0.2055, 0.0833	-0.1021, 0.4003	0.1688, 0.1625	-0.0131, 0.9144
signalized	0.3557, 0.0121	-0.2693, 0.0614	0.0652, 0.6565	-0.3230, 0.0236	0.1482, 0.3096	0.1626, 0.2643
TOTACCI						
3-legged	0.3098, 0.0041	-0.2819, 0.0094	0.1867, 0.0890	-0.1900, 0.0873		0.1900, 0.0873
4-legged	0.3794, 0.0010	-0.3263, 0.0052	0.2235, 0.0590	-0.1088, 0.3702	0.2013, 0.0948	-0.0275, 0.8215
signalized	0.3507, 0.0135	-0.2472, 0.0868	0.0361, 0.8056	-0.3629, 0.0104	0.1996, 0.1692	0.1403, 0.3362
INJACC						
3-legged	0.2612, 0.0164	-0.1745, 0.1123	0.0448, 0.6861	-0.1628, 0.1440		0.1628, 0.1440
4-legged	0.2020, 0.0889	-0.1457, 0.2219	0.0712, 0.5521	-0.0883, 0.4674	0.0813, 0.5033	0.0293, 0.8098
signalized	0.1521, 0.2967	-0.0660, 0.6524	-0.0481, 0.7427	-0.2526, 0.0800	0.1176, 0.4210	0.1249, 0.3925
INJACCI						
3-legged	0.2884, 0.0078	-0.2242, 0.0403	0.1056, 0.3389	-0.1446, 0.1950		0.1446, 0.1950
4-legged	0.2190, 0.0645	-0.1647, 0.1668	0.0886, 0.4591	-0.0961, 0.4288	0.1081, 0.3729	0.0196, 0.8723
signalized	0.1450, 0.3203	-0.0086, 0.9533	-0.1298, 0.3742	-0.3101, 0.0301	0.1686, 0.2468	0.1224, 0.4022
TOTACCS						
signalized	0.1450, 0.3203	-0.0793, 0.5882	-0.0205, 0.8887	-0.1490, 0.3069	-0.0374, 0.7986	0.2101, 0.1473
TOTACCM						
signalized	0.3579, 0.0116	-0.2763, 0.0546	0.0739, 0.6140	-0.3220, 0.0240	0.1664, 0.2533	0.1383, 0.3434

is not (and in those cases, the correlation is insignificant), the other one is still positively correlated with crashes. Right turns from the minor road, including right turns on red, are certainly occasions for crashes. It might be argued that drivers turning left from the minor road are more vigilant than drivers turning right and are at less risk than drivers going through (between legs 3 and 4). Drivers turning left from the major road, at least at the three-legged and four-legged intersections, must be concerned about both opposing traffic and traffic behind them, whereas drivers turning left from a stop-controlled minor road have less risk from traffic behind them. According to Table 22, PK%LEFT2 correlates negatively with PK%LEFT1. This suggests that the negative correlation of crashes with PK%LEFT2 may, in part, be a consequence of the positive correlation of crashes with PK%LEFT1.

Crashes Versus ADT

Examination of correlation coefficients shows that ADT unsupported by other variables, especially ADT1, plays a smaller role as one passes from three-legged to four-legged to signalized intersections. To understand this phenomenon better, we examine grouped data in the manner of Hauer et al. (1988). For each of the three data sets, intersections were divided into four groups by increasing major road ADT with an effort to equalize the number of crashes in each group to the extent possible. Likewise, intersections were divided into four groups by increasing minor road ADT with roughly equal crash counts in each group. Then, 16 cells were defined by means of the grouping. In each cell, the number of intersections was counted, along with the number of crashes (TOTACC) at its intersections during 1993-1995 and the ratio (the average number of crashes per intersection). The numbers obtained are shown in Tables 24, 25, and 26. In addition, marginal counts were made for the major road ADT groups and the minor road ADT groups of the same variables (number of intersections, number of crashes, and average number of crashes per intersection).

It is evident from the tables that some cells were empty or sparsely occupied. For example, in Table 24, there are no intersections in the highest quartile for major road ADT and the second highest quartile for minor road ADT. There are also two empty cells in Table 25. If the cells were uniformly occupied, the average number in each cell would be $84/16 = 5.25$, $72/16 = 4.5$, and $49/16 \approx 3.1$ in Tables 24, 25, and 26, respectively.

In Figures 5, 7, and 9, the marginal distributions with respect to major road ADT are plotted, and in Figures 6, 8, and 10, those with respect to minor road ADT are plotted. The horizontal variable in each case is the median ADT of the group, and the vertical variable is the average number of crashes per intersection in the group. The number of crashes per intersection generally appears to increase with increasing minor road ADT, with allowances made for noise due to the smallness of the sample sizes. The number of crashes per intersection versus major road ADT shows a similar but more erratic trend, except for the signalized intersections (Figure 9). The plot for the latter shows very little change in the crashes per intersection as major road ADT is varied. Note the scale.

TABLE 24. Crashes Versus Grouped Major and Minor Road ADT, Three-Legged Sample

3-legged, 4-lane by 2-lane, stop-controlled rural intersections, CA and MI, 1993-95

No. of Intersections No. of Crashes Crashes/Intersection		ADT 1 2,367 to 11,916	 11,917 to 15,167	 15,168 to 17,378	 17,379 to 33,058	
ADT 2	15 - 250	22 18 0.82	5 21 4.20	2 7 3.50	6 33 5.50	35 79 2.26
	251 - 820	11 30 2.73	6 20 3.33	1 6 6.00	7 31 4.43	25 87 3.48
	821 - 1,270	6 23 3.83	4 19 4.75	3 33 11.00	0	13 75 5.77
	1,271 - 3,001	2 9 4.50	4 29 7.25	2 23 11.50	3 24 8.00	11 85 7.73
		41 80 1.95	19 89 4.68	8 69 8.63	16 88 5.50	84 326 3.88

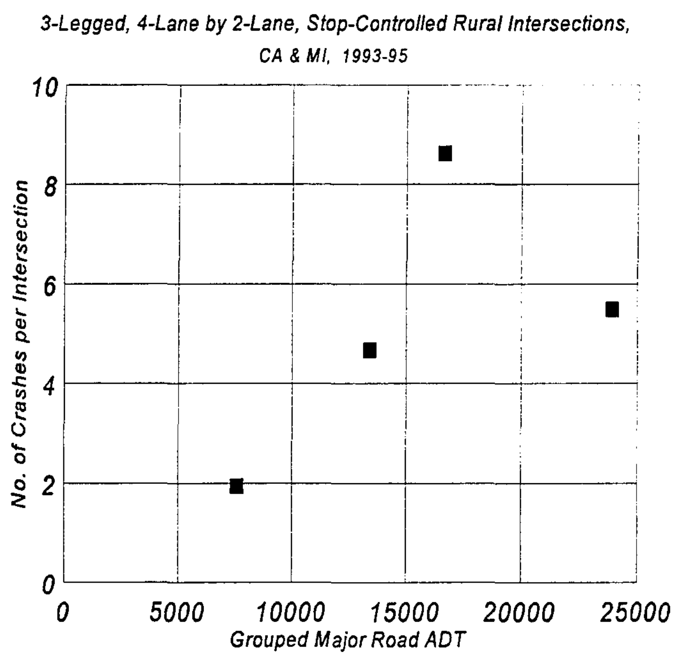


FIGURE 5. Crashes Versus Major Road ADT, Three-Legged Sample

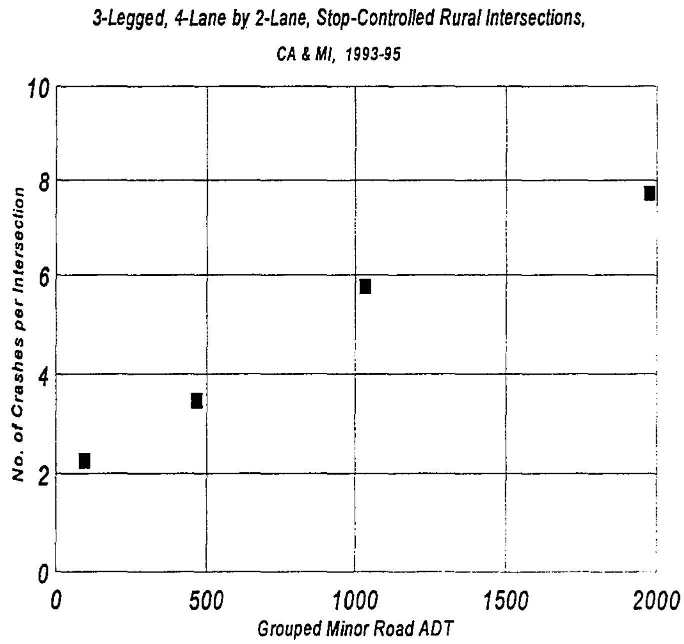


FIGURE 6. Crashes Versus Minor Road ADT, Three-Legged Sample

TABLE 25. Crashes Versus Grouped Major and Minor Road ADT, Four-Legged Sample
4-legged, 4-lane by 2-lane, stop-controlled rural intersections, CA and MI, 1993-95

No. of Intersections No. of Crashes Crashes/Intersection		**A D T 1**				
		3,350 to 7,684	7,685 to 12,000	12,001 to 19,332	19,333 to 73,000	
A	21 - 340	9 21 2.33	15 36 2.40	4 17 4.25	6 24 4.00	34 98 2.88
D	341 - 800	3 6 2.00	8 31 3.88	6 44 7.33	4 17 4.25	21 98 4.67
T	801 - 1,051	0	3 17 5.56	1 18 18.00	3 62 20.67	7 97 13.86
2	1,052 - 2,018	6 72 12.00	2 17 8.50	2 16 8.00	0	10 105 10.50
		18 99 5.50	28 101 3.61	13 95 7.31	13 103 7.92	72 398 5.53

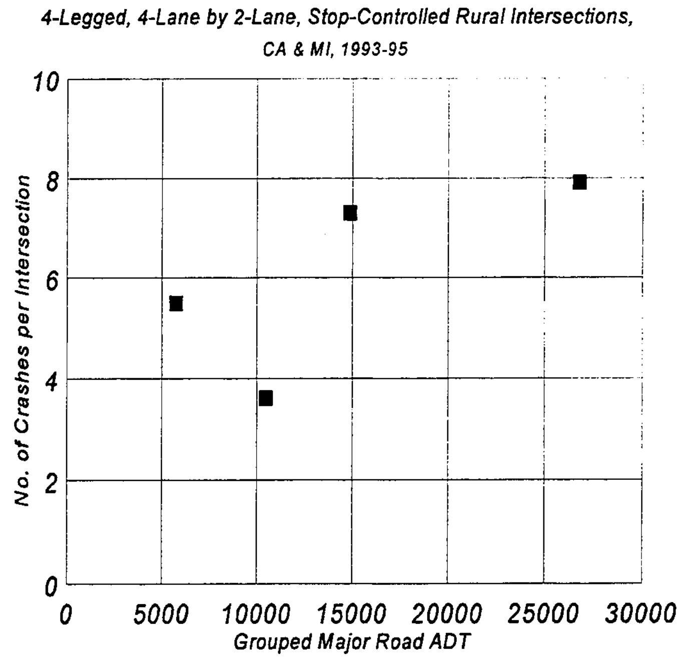


FIGURE 7. Crashes Versus Major Road ADT, Four-Legged Sample

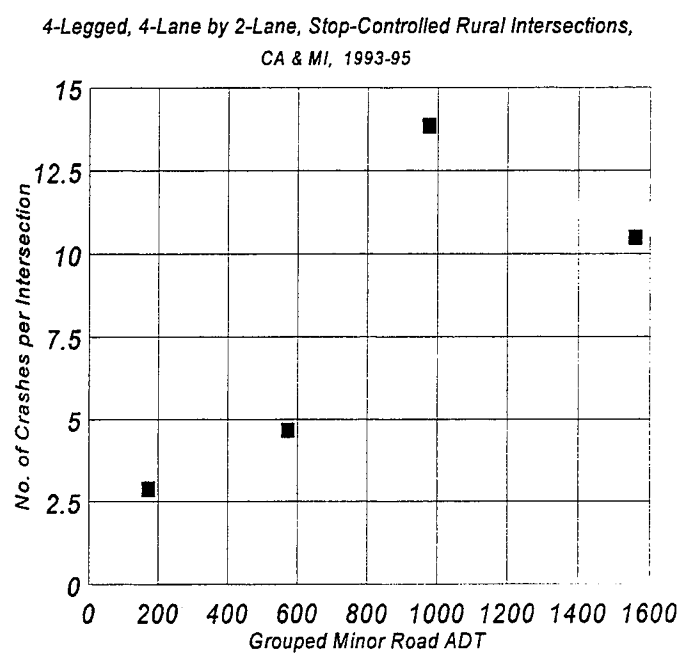


FIGURE 8. Crashes Versus Minor Road ADT, Four-Legged Sample

TABLE 26. Crashes Versus Grouped Major and Minor Road ADT, Signalized Sample

Signalized, 2-lane by 2-lane, 4-legged rural intersections, CA and MI, 1993-95

No. of Intersections No. of Crashes Crashes/Intersection		A D T 1 4,647 to 7,580	 7,581 to 8,833	 8,834 to 12,825	 12,826 to 25,133	
A	940 - 3,003	4 70 17.50	3 65 21.67	1 17 17.00	7 98 14.00	15 250 16.67
D	3,004 - 4,192	4 68 17.00	3 69 23.00	4 119 29.75	2 13 6.50	13 269 20.69
T	4,193 - 5,450	4 76 19.00	4 78 19.50	4 79 19.75	1 25 25.00	13 258 19.85
2	5,451 - 12,478	1 31 31.00	1 30 30.00	3 60 20.00	3 119 39.67	8 240 30.00
		13 245 18.85	11 242 22.00	12 275 22.92	13 255 19.62	49 1017 20.76

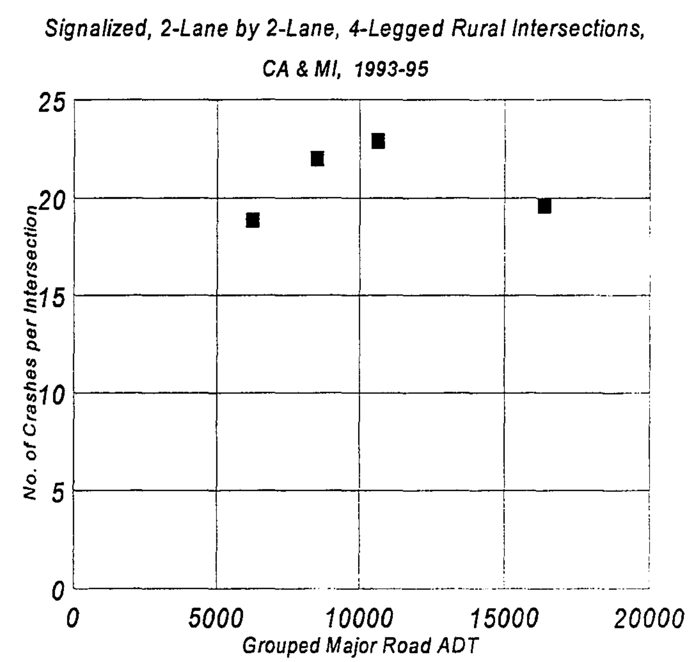


FIGURE 9. Crashes Versus Major Road ADT, Signalized Sample

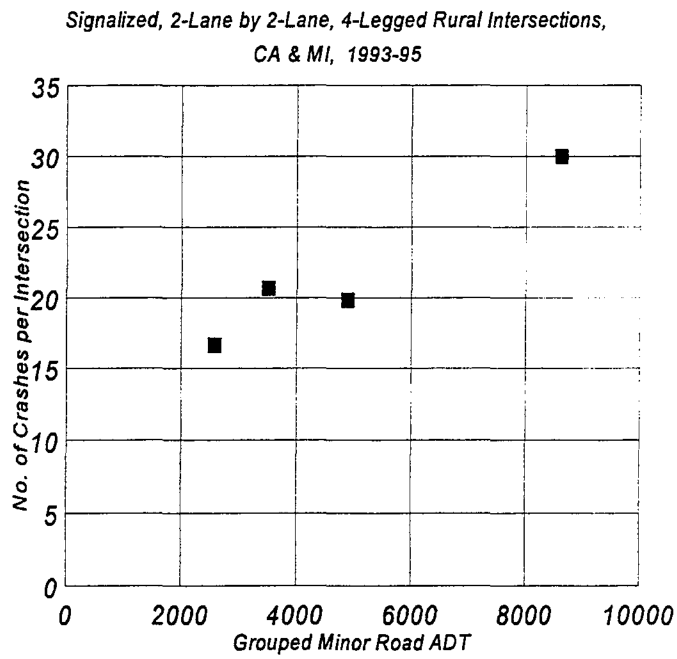


FIGURE 10. Crashes Versus Minor Road ADT, Signalized Sample

A.M. Versus P.M. Truck Percentages

The large amount of traffic movement data collected for this report permits a variety of special studies. Table 27 is one illustrative example. For related items, see the appendix.

Table 27 indicates that the truck percentage is somewhat variable, and that in the morning, the truck percentage is higher than in the evening (except for the Michigan signalized intersections). Miaou et al. (1993) recommend that future studies include a time-of-day variable in estimating truck percentages.

CONCLUSIONS

This chapter began with the development of variables for analysis and modeling. A variety of variables were constructed relating to crash counts, ADT, peak-hour truck traffic, turning percentages, geometry, channelization, alignments, and driveway counts. A variable for State was defined, underscoring the possibility that in different regions and/or epochs, crash experience may be quantitatively distinct despite similar values for intersection variables.

TABLE 27. A.M. and P.M. Truck Percentages by State

	3-Legged Intersections		
	California (60)	Michigan (24)	CA & MI (84)
AM%TRUCK	10.11	10.31	10.17
PM%TRUCK	9.19	6.98	8.56
PK%TRUCK	9.52	8.21	9.15
	4-Legged Intersections		
	California (54)	Michigan (18)	CA & MI (72)
AM%TRUCK	13.76	8.94	12.56
PM%TRUCK	10.98	7.11	10.01
PK%TRUCK	11.98	7.83	10.95
	Signalized Intersections		
	California (18)	Michigan (31)	CA & MI (49)
AM%TRUCK	8.34	9.70	9.20
PM%TRUCK	6.62	10.16	8.86
PK%TRUCK	7.36	9.89	8.96

Then, in Tables 5, 6, and 7, a summary of univariate statistics for these variables on the three data sets was given. More crashes occur at signalized intersections than at four-legged intersections, and more occur at four-legged than at three-legged intersections (cf. Table 8). Crashes tend to be more severe in California (Table 9), but more frequent in Michigan (Table 10). While this may, in part, be attributable to systematic differences in intersection variables between the two States (cf. Tables 15, 16, and 17), it is a reminder that the STATE variable may make an independent contribution.

The chapter also examines correlations between pairs of variables. This includes crashes versus other variables (Tables 11, 12, 13, and 14), ADT and STATE versus other intersection variables (Tables 15, 16, and 17), single-vehicle and multiple-vehicle crashes versus other signalized intersection variables (Table 21), and turning percentage variables (Tables 22 and 23). The most striking finding is the relevant insignificance of major road ADT in relation to the signalized intersection crashes (see especially Figure 9). Another finding of importance is the negative correlation between minor road left-turn percentage and crashes present for all three intersection classes (Table 23). This, of course, implies a positive correlation between crashes and the sum of minor road through and right-turn percentages. Given the range of minor road left-turn percentages within and among the three data sets (Tables 5, 6, and 7), this seems especially significant. Since the hazards that a left-turning vehicle faces are greater than those that a right-turning vehicle faces, the possibility exists that drivers making left turns from the minor road exercise more care than other drivers approaching the intersection from a minor leg. Perhaps more relevant is the fact that since left-turn percentages from the minor road correlate negatively with right-turn percentages from the minor road, they also correlate negatively with left-turn percentages from the major road. As minor road left turns increase, major road left turns decrease, and the net effect of the two opposite changes is to reduce crashes.

An issue that will affect the modeling is the multivariate relationships, especially the relationship among crashes and pairs of highway variables. Thus, for the signalized intersections, the relative insignificance of crashes versus major road ADT may indicate the effect of a third variable that correlates with ADT. Again, for the signalized data, the effect of STATE on crash counts may be confounded with that of other variables such as LIGHT, PROT_LT, HAZRAT, NODRWY1, MEDWIDTH1, and even ADT1, all of which strongly correlate with STATE. The general strategy will be to see which variable has the chief effect, in accordance with common sense, and, thereafter, to determine which remaining variable, if any, has a significant effect on the residual, i.e., the portion of the crash count not predicted by the chief variable.

5. MODELING

In this chapter, we use the sample data to develop generalized linear models of the Poisson/negative binomial type for the mean number of crashes per unit time at an intersection in terms of the intersection variables discussed in earlier chapters. These models summarize the data collected. It is hoped that they have predictive value for other data sets from the same intersection classes.

The chapter begins with a review of some of the theoretical aspects of model building and measurement of goodness of fit. Thereafter, models are built for each of the three classes of intersections. This is done for each of the four crash variables — TOTACC, TOTACCI, INJACC, and INJACCI. We study how these variables can be represented in terms of major and minor road ADT, and then we add variables with the aim of improving the fit and discovering design elements that might affect safety.

Separate models are also developed for TOTACCS and TOTACCM in the case of the signalized intersections. These models use only the flows F_1, F_2, F_3, and F_4 as explanatory variables.

Finally, the main models for TOTACCI are subjected to residual analysis to uncover systematic shortcomings.

THEORY

Modeling

We shall use a negative binomial model with mean a generalized linear function of intersection variables. Thus,

$$\mu_i = \exp(\beta_1 + \textstyle\sum_{j=2}^{p} x_{ij}\beta_j) \tag{5.1}$$

where μ_i is the mean number of crashes to be expected at intersection number i in a given time period; $x_{i2}, ..., x_{im}$, are the values of the intersection variables at this intersection during that time period ($x_{i1} \equiv 1$ corresponds to the intercept term); and $\beta_1, ..., \beta_p$ are coefficients to be estimated by the modeling. More sensitively, one might say that μ_i is the grand mean of crashes to be expected at a hypothetical population of intersections having the same values as intersection number i for the intersection variables considered. Variables not included in the model account for differences in the expected number of crashes among members of this population, and these differences are described by the term overdispersion. See Hauer et al. (1988). The variance $(\sigma_i)^2$ of the number of crashes in this population under the model is:

$$(\sigma_i)^2 = \mu_i + K(\mu_i)^2$$

where K is the overdispersion parameter. The second term on the right side of this equation represents the variation in means among different members of the population existing even when all intersections have the same value for the considered intersection variables. In principle, K could also depend on these intersection variables, but for simplicity, that possibility is ignored.

Under the negative binomial model, the probability of y_i crashes at intersection number i is given by:

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

When 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(β, K) for the negative binomial distribution. The likelihood function is the probability that the values y_1, ..., y_N would be observed for intersections number 1 through N. If crash counts are independent at the different intersections, the likelihood is:

$$\prod_{i=1}^{N} P(y_i)$$

and application of the logarithm yields the log-likelihood function:

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

(5.2)

Here, $\beta = (\beta_1, ..., \beta_p)$ is the vector of coefficients, y_i is now taken to be the observed crash count at intersection no. i, and μ_i is given by equation (5.1). The values of β and K that maximize the function $L(\beta, K)$ in (5.2) are the estimated coefficient vector $\hat{\beta}$ and the estimated overdispersion parameter $\hat{K}$. The estimated value of μ_i obtained by substituting $\hat{\beta}$ and $\hat{K}$ for β and K in equation (5.1) is denoted by $\hat{y}_i$. For convenience, the same letters will often be used for both the parameters and their estimated values, i.e., carets (^) will be omitted in references to β and $\hat{K}$.

P-Values and Goodness of Fit

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. Even if the P-value is large, the parameter estimate has some value since the null hypothesis that the parameter is zero is a somewhat arbitrary starting point and the estimate obtained is the one dictated by the data. A large P-value lowers our confidence in the estimate and indicates that even if the basic model form is correct, the true coefficient may be quite different from the one estimated. One may expect the true coefficient to be within one or two estimated standard errors of the estimated coefficient.

Goodness-of-fit measures associated with Poisson-type models have been introduced and reviewed by Fridstrøm et al. (1995) and Miaou (1996). For the modeling, we shall use three measures of goodness-of-fit.

One measure is the ordinary R-squared, or coefficient of determination, used in linear regression models:

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \tag{5.3}$$

where

y_i = observed crash count for intersection no. i,

$\bar{y}$ = average crash count for the sample, and

$\hat{y}_I$ = estimated mean crash count for intersection no. i.

This measure is used because of its great familiarity. In case a model with no variables is used, i.e., in equation (5.1), $\mu_i \equiv \exp(\beta_0)$ so that there is only a constant or intercept term in the linear expression, the maximum likelihood estimate for β_0 can be shown to yield

$$\hat{y}_i \equiv \bar{y}$$

and hence, R^2 equals zero. This model is called the zero model. At the other extreme, it might, in principle, happen that

$$\hat{y}_i = y_i$$

for each i, and hence, R^2 equals 1. The value of R^2 is always less than or equal to 1 by definition. It is greater than or equal to zero since maximum likelihood guarantees a result at least as good as the zero model.

Fridstrøm et al. (1995) have pointed out that in Poisson or negative binomial models, R^2 is very unlikely to equal 1 since a Poisson-type variable takes a variety of values other than its mean, and y_i is unlikely to equal the estimated mean $\hat{y}_i$ for each i in a sample of any appreciable size. They have proposed taking a ratio of R^2 to its largest expected value P^2 under a best fit as a measure of goodness-of-fit.

A form of this that they recommend for negative binomial models is the log-likelihood R-squared, based on the deviance D^m of the model. The deviance of a model m is:

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

where

$$L^f = \sum_{i=1}^{N} y_i \log(y_i) - y_i - \log(y_i!)$$

is the log-likelihood that would be achieved if the model did give a perfect fit ($\mu_i = y_i$ for each i, and $K = 0$). Such a model is called the full or saturated model by Fridstrøm et al. L^m is the log-likelihood, as in (5.2), 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 of parameters. The number of parameters is (p + 1), where p is the number of explanatory variables in the model plus the intercept, and the extra 1 is for the overdispersion parameter.

Fridstrøm et al. propose the following measures:

$$R_D^2 = 1 - \left(\frac{\frac{D^m}{N-p-1}}{\frac{D^0}{N-2}}\right) \tag{5.4}$$

$$P_D^2 = 1 - \left(\frac{\frac{D_E^m}{N-p}}{\frac{D^0}{N-2}}\right) \tag{5.5}$$

$$R_{PD}^2 = \frac{R_D^2}{P_D^2} \tag{5.6}$$

Here D^0 is the deviance of a model with only two parameters — the constant term (intercept) and the overdispersion parameter; p 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 where 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.

A third measure of goodness-of-fit, proposed by Miaou (1996), is based explicitly on the overdispersion parameter:

$$R_K^2 = 1 - \frac{K}{K_{max}} \tag{5.7}$$

Here, K is the overdispersion parameter estimated in the model, and K_{max} is the overdispersion parameter estimated in the zero model. 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.

Model Building

Adopting Miaou's parameter as a measure of goodness-of-fit is equivalent to taking the over-dispersion parameter as such a measure. A smaller overdispersion parameter signifies a better fit. Such improvement may occur because explanatory variables have been discovered or because the number of independent variables is large relative to the sample size.

Akaike has proposed a criterion for judging models, and a corrected version of this, applicable to small samples, has been developed by Hurvich and Tsai. See Miaou (1996, Chapter 4) for a discussion. This statistic, in a form relevant to negative binomial models, is:

$$CAIC_{NB} = -2L(\hat{\beta},\hat{K}) + 2(p + 1) + \frac{2(p + 1)(p + 2)}{N - p - 2} \tag{5.8}$$

where N is the sample size and p is the number of parameters in the model (excluding the overdispersion parameter). Models with smaller values of $CAIC_{NB}$ are deemed to be better fits. This measure involves a trade-off between increased probability and a penalty for adding parameters on small data sets. If $N = \infty$, the last term is dropped and the uncorrected Akaike criterion results. Even without the last term, the criterion includes a penalty in the second term for adding parameters.

The model building described in subsequent sections of this chapter is guided by certain principles. Intersection variables of known importance, namely ADT1 and ADT2, should be included in the model. Other variables with understandable interpretations, i.e., those presented in the previous chapter (some of which were developed in the course of the modeling), are added to the model provided they satisfy some combination of the criteria below:

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

- Among variables that measure strongly overlapping properties, at most one, will be used.

- Examination of residuals $y_i - \hat{y}_i$ under a predecessor model not including the variable should indicate that the variable is strongly correlated with the residual.

- Inclusion of the variable should lead to reductions of the overdispersion parameter and $CAIC_{NB}$, increases in the R-squared values, and respectable P-values for the estimated coefficient of the variable to the extent possible.

These criteria are guidelines rather than precise and strict requirements, since model-building is an art rather than a science.

MODELS FOR THREE-LEGGED INTERSECTIONS

Tables 28, 29, 30, and 31 show negative binomial models of crashes in terms of intersection variables for the three-legged intersections.

TABLE 28. Negative Binomial Models for Crashes per Year (TOTACC), Three-Legged Intersections

Estimated regression coefficients (estimated standard error and P-value in parentheses).

Variables	ADT only	Main Model	Variant
Intercept	-12.9243 (2.3682, 0.0001)	-12.2196 (2.3575, 0.0001)	-12.2577 (2.3626, 0.0001)
Log of ADT1	1.1989 (0.2477, 0.0001)	1.1479 (0.2527, 0.0001)	1.1778 (0.2517, 0.0001)
Log of ADT2	0.3027 (0.0892, 0.0007)	0.2624 (0.0866, 0.0024)	0.2034 (0.1032, 0.0487)
MEDWIDTH1 (in feet), major road		-0.0546 (0.0249, 0.0285)	-0.0551 (0.0246, 0.0254)
NODRWY1, driveways to 250 ft, major road		0.0391 (0.0239, 0.1023)	0.0414 (0.0245, 0.0912)
PK%LEFT1 major road			0.0544 (0.0471, 0.2479)
N, p	84, 3	84, 5	84, 6
K R_K^2	0.5256 (0.1366, 0.0001) 0.5158	0.3893 (0.1160, 0.0008) 0.6413	0.3658 (0.1095,0.0008) 0.6630
R^2	0.2294	0.4351	0.4473
R_D^2, P_D^2 R_{PD}^2	0.1821, 0.5628 0.3237	0.2247, 0.5589 0.4021	0.2275, 0.5524 0.4119
$CAIC_{NB}$	381.930	373.887	373.742

1 ft = 0.305 m

Table 28 indicates that the regression coefficient for (the log of) major road ADT is about four to five times that for minor road ADT. Among the next most significant variables, as measured by residuals after use of the ADT-only model, are NODRWY1, MEDWIDTH1, and SPD1. A second

TABLE 29. Negative Binomial Models for Crashes per Year (TOTACCI), Three-Legged Intersections

Estimated regression coefficients (estimated standard error and P-value in parentheses).

Variables	ADT only	Main Model	Variant 1	Variant 2	Variant 3
Intercept	-16.1636 (3.4655, 0.0001)	-15.4661 (3.4685, 0.0001)	-16.6179 (3.3126, 0.0001)	-15.7008 (3.3955, 0.0001)	-13.6339 (3.0516, 0.0001)
Log of ADT1	1.5023 (0.3507, 0.0001)	1.4331 (0.3608, 0.0001)	1.6117 (0.3541, 0.0001)	1.4962 (0.3530, 0.0001)	1.1954 (0.3109, 0.0001)
Log of ADT2	0.2904 (0.1001, 0.0037)	0.2686 (0.0988, 0.0065)	0.1276 (0.1283, 0.3199)	0.1801 (0.1187, 0.1294)	0.2646 (0.1014, 0.0091)
MEDWIDTH1 ft		-0.0612 (0.0360, 0.0888)	-0.0687 (0.0384, 0.0738)	-0.0607 (0.0340, 0.0739)	
NODRWY1 major road		0.0560 (0.0289, 0.0525)	0.0552 (0.0290, 0.0565)	0.0597 (0.0283, 0.0350)	0.0903 (0.0266, 0.0007)
PK%TURN			0.0401 (0.0215, 0.0617)		
PK%LEFT1, major road				0.0764 (0.0665, 0.2509)	
VEI-1, vertical out to 800 ft, major road					0.1180 (0.0700, 0.0919)
HAU angle					0.0197 (0.0174, 0.2591)
N, p	84, 3	84, 5	84, 6	84, 6	84, 6
K R_K^2	0.7332 (0.2068, 0.0004) 0.5139	0.5118 (0.1719, 0.0029) 0.6607	0.4195 (0.1478, 0.0046) 0.7219	0.4416 (0.1513, 0.0035) 0.7072	0.4416 (0.1642, 0.0072) 0.7072
R^2	0.1666	0.4452	0.4644	0.4757	0.4287
R_D^2, P_D^2 R_{PD}^2	0.1731, 0.5322 0.3253	0.2233, 0.5374 0.4155	0.2491, 0.5356 0.4652	0.2371, 0.5313 0.4462	0.2278, 0.5196 0.4384
$CAIC_{NB}$	326.278	317.747	313.620	315.800	317.480

1 ft = 0.305 m

TABLE 30. Negative Binomial Models for Crashes per Year (INJACC), Three-Legged Intersections

Estimated regression coefficients (estimated standard error and P-value in parentheses).

Variables	ADT only	Variant 1	Variant 2
Intercept	**-13.1685 (3.0319, 0.0001)**	**-12.3246 (2.8076, 0.0001)**	**-11.0061 (2.6937,0 .0001)**
Log of ADT1	**1.2028 (0.3082, 0.0001)**	**1.1436 (0.2763, 0.0001)**	**0.9526 (0.2843, 0.0008)**
Log of ADT2	**0.1925 (0.0931, 0.0388)**	**0.1357 (0.1029, 0.1872)**	**0.1499 (0.0916, 0.1018)**
HAU angle		**0.0230 (0.0131, 0.0790)**	**0.0289 (0.0105, 0.0061)**
NODRWY1, driveways out to 250 ft, major road			**0.0481 (0.0262, 0.0664)**
ABSGRD1, average grade, major road			**0.1838 (0.1130, 0.1038)**
N, p	**84, 3**	**84, 4**	**84, 6**
K **R_K^2**	**0.5649 (0.2032, 0.0055)** **0.4535**	**0.3787 (0.1792, 0.0346)** **0.6336**	**0.2588 (0.1848, 0.1613)** **0.7496**
R^2	**0.1400**	**0.3755**	**0.4505**
R_D^2, P_D^2 **R_{PD}^2**	**0.1437, 0.4039** **0.3558**	**0.1841, 0.3966** **0.4644**	**0.2036, 0.3837** **0.5306**
$CAIC_{NB}$	**274.653**	**269.275**	**268.081**

1 ft = 0.305 m

tier of significant variables includes PK%TRUCK and LTLN1. All of these variables correlate with NODRWY1 and MEDWIDTH1 (see Table 18), and when the latter two variables are added, the others become much less significant. However, it is also true that NODRWY1 and MEDWIDTH1 correlate strongly with each other (correlation coefficient -0.37654 and P-value 0.0004). Nonetheless, we keep them both because they seem to have separate effects in the main model of Table 28. When we consider residuals for the main model, the angle variables DEV and HAU show positive correlation, as do turning percentage variables. If we add an angle variable and a turning percentage variable to the model, the overdispersion parameter K reduces to about 0.29, but the P-

TABLE 31. Negative Binomial Models for Crashes per Year (INJACCI), Three-Legged Intersections

Estimated regression coefficients (estimated standard error and P-value in parentheses).

Variables	ADT only	Variant 1	Variant 2
Intercept	-14.6858 (4.0902, 0.0004)	-13.9216 (3.7706, 0.0003)	-12.4996 (3.5376, 0.0004)
Log of ADT1	1.3145 (0.4202, 0.0018)	1.2616 (0.3810, 0.0009)	1.0701 (0.3691, 0.0037)
Log of ADT2	0.2179 (0.1076, 0.0429)	0.1629 (0.1097, 0.1373)	0.1657 (0.1019, 0.1038)
HAU angle		0.0253 (0.0205, 0.2179)	0.0319 (0.0138, 0.0205)
NODRWY1 driveways out to 250 ft, major rd			0.0487 (0.0302, 0.1068)
VEI-1 vertical out to 800 ft			0.1555 (0.1075, 0.1479)
N, p	84, 3	84, 4	84, 6
K R^2_K	0.7219 (0.2846, .0112) 0.4725	0.4857 (0.2401, 0.0431) 0.6451	0.3295 (0.2723, 0.2263) 0.7592
R^2	0.1470	0.3674	0.4119
R^2_D, P^2_D R^2_{PD}	0.1375, 0.3848 0.3573	0.1786, 0.3816 0.4680	0.2084, 0.3774 0.5522
$CAIC_{NB}$	240.718	235.734	233.492

1 ft = 0.305 m

values for the angle variable range from 0.30 to 0.39. We have retained only one variant model in Table 28. Note that inclusion of PK%LEFT1 in the variant reduces the coefficient of the log of ADT2, not unexpectedly, since these variables are correlated.

In Table 29, similar models are shown for TOTACCI. With crashes restricted to those that are intersection-related, the effect of ADT1 becomes stronger. Two variant models show the slightly different effects of two turning percentage variables: one has smaller P-values and a larger R^2, the other has smaller $CAIC_{NB}$ and the other two R-squared measures are larger; they give differing magnitudes to the minor road coefficient. A third variant shows that vertical alignment VEI-1 and the angle variable HAU in a suitable combination, but not without each other, have some explanatory value in place of MEDWIDTH1.

When we turn to serious crashes (INJACC and INJACCI) in Tables 30 and 31, the importance of ADT1 relative to ADT2 continues to increase. In addition, MEDWIDTH1 ceases to be significant and HAU becomes an important variable. Recall from Figures 2 and 3 that the sign of HAU is positive when a driver turning from the major road across traffic need only turn through a small angle. According to these models, this increases crashes. This suggests that perhaps the more relevant movement is turning from the minor road. A driver turning right from the minor road may have the illusion of easy access, but inadequate visibility for traffic on the major road traveling in the same direction, while a driver turning left will have poor visibility of the traffic that must be crossed. Only 17% of the three-legged intersections had HAU different from zero (cf. Table 5), and the ones with HAU higher than zero had more injury crashes than average and the ones with HAU lower than zero had fewer crashes. Recall from Table 11 that HAU has a strong positive correlation with all crash types. With MEDWIDTH1 removed and HAU added, NODRWY1 and one of the two vertical alignment variables VEI-1 or ABSGRD1 also contribute to injury crashes in the other models shown in Tables 30 and 31.

Variables not included in these models, such as STATE, sight distances, and HEI-1, had very insignificant P-values after inclusion of the variables shown in the tables.

The three-legged models have the following general features:

- TOTACC and TOTACCI models are similar, INJACC and INJACCI models are similar.
- For all four crash variables, ADT1, ADT2, and NODRWY1 are influential.
- For TOTACC and TOTACCI, MEDWIDTH1 and turning percentage are influential.
- For INJACC and INJACCI, the angle variable HAU and vertical alignment are influential, and, to some extent, this is also true for TOTACCI.

MODELS FOR FOUR-LEGGED INTERSECTIONS

The models for the four-legged intersections are exhibited in Tables 32, 33, and 34.

Table 32 shows models for TOTACC. In the absence of other variables, minor road ADT appears to be more influential than major road ADT. When other variables are added, in particular, turning and through percentages, ADT1 becomes much more influential than ADT2. The variables that correlate most strongly with the residuals of the ADT-only model are RSDR2, PK%LEFT1, LTLN1S, and STATE, in order. However, when these variables are added to the models, the ones that are most significant are PK%LEFT1 and LTLN1S. Both of them correlate strongly with

TABLE 32. Negative Binomial Models for Crashes per Year (TOTACC), Four-Legged Intersections

Estimated regression coefficients (estimated standard error and P-value in parentheses).

Variables	ADT only	Main Model	Variant 1	Variant 2	Variant 3
Intercept	-6.9352 (2.3767, 0.0035)	-9.46311 (2.5991, 0.0003)	-10.1902 (3.3126, 0.0021)	-10.9526 (2.5907, 0.0001)	-9.7859 (2.6032, 0.0002)
Log of ADT1	0.4683 (0.2330, 0.0444)	0.8503 (0.2779, 0.0022)	0.8873 (0.2653, 0.0008)	1.0382 (0.2870, 0.0003)	0.8894 (0.2873, 0.0020)
Log of ADT2	0.5135 (0.0896, 0.0001)	0.3294 (0.1255, 0.0087)	0.2924 (0.1316, 0.0263)	0.2206 (0.1219, 0.0704)	0.2845 (0.1375, 0.0385)
PK%LEFT1, major road		0.1100 (0.0412, 0.0076)	0.2976 (0.1393, 0.0326)	0.1054 (0.0372, 0.0046)	
LTLN1S (0 or 1)		-0.4841 (0.2311, 0.0362)		-0.6607 (0.2347, 0.0049)	-0.5471 (0.2445, 0.0252)
PK%LEFT1 squared			-0.0131 (0.0094, 0.1643)		
PK%THRU2 minor road				0.0220 (0.0107, 0.0391)	
ABSGRD1 major road				0.1553 (0.1123, 0.1666)	
PK%TURN					0.0351 (0.0238, 0.1404)
100×RSDR2 (100×1/ft)	1 ft = 0.305 m				2.284 (1.503, 0.1286)
N, p	72, 3	72, 5	72, 5	70, 7	72, 6
K	0.6144 (0.1562, 0.0001)	0.4578 (0.1307, 0.0005)	0.4820 (0.1425, 0.0007)	0.3682 (0.1124, 0.0011)	0.4183 (0.1147, 0.0003)
R_K^2	0.3801	0.5381	0.5136	0.5953	0.5780
R^2	0.2565	0.3109	0.2520	0.4797	0.4494
R_D^2, P_D^2	0.1080, 0.6011	0.1623, 0.5858	0.2557, 0.5874	0.1575, 0.5666	0.1635, 0.5792
R_{PD}^2	0.1796	0.2771	0.2557	0.2780	0.2822
$CAIC_{NB}$	385.168	374.948	377.165	369.829	374.870

TABLE 33. Negative Binomial Models for Crashes per Year (TOTACCI), Four-Legged Intersections

Estimated regression coefficients (estimated standard error and P-value in parentheses).

Variables	ADT only	Main Model	Variant 1	Variant 2	Variant 3
Intercept	-7.2501 (2.9094, 0.0130)	-11.1096 (3.3345, 0.0008)	-10.9008 (3.3257, 0.0010)	-11.8796 (3.2980, 0.0004)	-13.2806 (3.2833, 0.0001)
Log of ADT1	0.4582 (0.2844, 0.1071)	0.9299 (0.3433, 0.0067)	0.9325 (0.3452, 0.0069)	1.0161 (0.3382, 0.0027)	1.2160 (0.3434, 0.0004)
Log of ADT2	0.5311 (0.0996, 0.0001)	0.3536 (0.1163, 0.0024)	0.3498 (0.1300, 0.0071)	0.2866 (0.1336, 0.0319)	0.2195 (0.1279, 0.0862)
PK%LEFT1 major road		0.1491 (0.0586, 0.0110)	0.1427 (0.0583, 0.0144)	0.3854 (0.1674, 0.0213)	0.1396 (0.0540, 0.0097)
LTLN1S (0 or 1)			-0.2891 (0.2920, 0.3222)		-0.4890 (0.2970, 0.0998)
PK%LEFT1 squared				-0.0172 (0.0111, 0.1221)	
PK%THRU2 minor road					0.0284 (0.0145, 0.0511)
ABSGRD1 major road					0.1698 (0.1353, 0.2093)
N, p	72, 3	72, 4	72, 5	72, 5	70, 7
K R^2_K	0.8814 (0.2267, 0.0001) 0.3338	0.7096 (0.1906, 0.0002) 0.4637	0.6901 (0.1827, 0.0002) 0.4784	0.6548 (0.1779, 0.0002) 0.5051	0.5556 (0.1512, 0.0002) 0.5498
R^2	0.2323	0.1587	0.1952	0.1646	0.3675
R^2_D, P^2_D R^2_{PD}	0.0814, 0.5802 0.1403	0.1334, 0.5715 0.2334	0.1273, 0.5646 0.2255	0.1410, 0.5665 0.2488	0.1275, 0.5470 0.2332
$CAIC_{NB}$	350.887	341.404	342.541	340.135	337.889

TABLE 34. Negative Binomial Models for Crashes per Year (INJACC and INJACCI), Four-Legged Intersections

Estimated regression coefficients (estimated standard error and P-value in parentheses).

Variables	ADT-only INJACC	Variant 1 INJACC	ADT-only INJACCI	Variant 1 INJACCI
Intercept	-9.8454 (2.5675, 0.0001)	-12.5296 (2.9908, 0.0001)	-9.7977 (3.2819, 0.0028)	-13.5576 (3.9998, 0.0008)
Log of ADT1	0.7224 (0.2591, 0.0053)	0.9505 (0.3284, 0.0038)	0.6735 (0.3285, 0.0403)	0.9918 (0.4268, 0.0201)
Log of ADT2	0.4778 (0.1401, 0.0007)	0.3237 (0.1645, 0.0491)	0.5138 (0.1604, 0.0014)	0.3310 (0.1894, 0.0805)
PK%LEFT1 major road		0.0994 (0.0433, 0.0216)		0.1228 (0.0614, 0.0457)
SPD2 (in mph)		0.0339 (0.0179, 0.0577)		0.0429 (0.0240, 0.0740)
N, p	72, 3	72, 5	72, 3	72, 5
K R^2_K	0.5741 (0.1821, 0.0016) 0.4218	0.4308 (0.1824, 0.0182) 0.5662	0.9671 (0.2899, 0.0009) 0.3449	0.7178 (0.2716, 0.0082) 0.5138
R^2	0.2445	0.3565	0.1987	0.3237
R^2_D, P^2_D R^2_{PD}	0.1197, 0.4817 0.2485	0.1550, 0.4654 0.3331	0.0834, 0.4812 0.1734	0.1214, 0.4680 0.2593
$CAIC_{NB}$	294.271	289.919	275.196	270.302

1 mph = 1.61 km/h

STATE, and STATE ceases to correlate significantly with the residual of this new model. PK%LEFT1 and LTLN1S correlate with each other as well, having a correlation coefficient of -0.2288 and a P-value of 0.0532, but both of them still seem to contribute to the accident count.

The original major road left-turn lane variable LTLN1 takes values 0, 1, and 2, but only 4 out of 72 four-legged intersections (see table 6) have exactly one left-turn lane. One can model the left-turn lanes on the major road using two regression coefficients (dividing the intersections into three subclasses), but the quantity of data does not support this option. If one uses only the variable LTLN1, one is adopting the bias that two turning lanes have double the safety effect of one. Our data suggest that two left-turning lanes are less safe than one: a model with coefficients for each case gives a larger negative coefficient when there is one turning lane than when there are two. The variable LTLN1S takes the value 1 when there is at least one left-turning lane on the major road. This, in effect, divides the intersections into two classes and avoids any assumptions about the relative safety of intersections with one versus two left-turning lanes. (A similar approach is used in the next section with the signalized data where PROT_LT indicates at least one protected left turn on the major road.) Variant 1 is a model that indicates a quadratic dependence of crash count on PK%LEFT1. This model does not perform as well as the main model, but is present because residuals for the main model correlate negatively with PK%LEFT1. We will discuss this issue further in connection with Table 33. If LTLN1S is added to the Variant 1 model of Table 32, the quadratic term in PK%LEFT1 becomes insignificant, with a P-value going to 0.3962.

The next most significant variable after those in the main model is PK%THRU2. There are reasons to be wary of adding two turning percentages because of the strong correlations among ADT1, ADT2, PK%LEFT1, and PK%LEFT2 (see Tables 16 and 19). Also, two intersections must be removed from the sample for which the minor legs had no traffic approaching the intersection during the peak-hour visits. However, if PK%THRU2 is added, it is significant. The design variable ABSGRD1 is also included in Variant 2.

Variant 3 in Table 32 is obtained by using PK%TURN rather than PK%LEFT1 and proceeding to add significant variables. The average sight distance right from the minor road in feet, represented here by its reciprocal multiplied by 100, is known to be correlated with all types of crashes (see Table 12). Also, it has a strong correlation with the residuals from other models, but when it is added to models, Variant 3 is the only model where its regression coefficient achieves a relatively small P-value.

In Table 33, similar models are done for TOTACCI. The results are similar except that LTLN1S is less significant. A version of Table 32, Variant 3, is not shown because the P-values of LTLN1S and RSDR2 rise from 0.0252 and 0.1286 to 0.2594 and 0.3108, respectively.

We discern again a quadratic dependency on PK%LEFT1 (compare Variant 2 in Table 33 with Variant 1 in Table 32). A quadratic of the form $ax - bx^2$ with a and b positive has its maximum when $x = a/2b$. The two quadratic models have maximum contribution from PK%LEFT1 at the values $0.2976/(2\times0.0131) = 11.36$ and $0.3854/(2\times0.0172) = 11.20$, respectively. This suggests that when the left-turn percentage from the major road is less than 11%, crashes rise with increasing percentage, but that when it is greater, crashes fall with increasing percentage. Among the 72 intersections upon which the model is based, 5 of them have PK%LEFT1 in excess of 11%.

Variant 1 in Table 33 includes LTLN1S, and yields improvement in K and R^2, but not in R^2_{PD} or $CAIC_{NB}$ or P-values. Likewise, Variant 3 in Table 33 includes the design variables LTLN1S and ABSGRD1. Without them, but with PK%LEFT1 and PK%THRU2 retained, in a model for TOTACCI that we do not display, the overdispersion parameter K is larger (0.6261), R^2 is smaller (0.1260), and R^2_{PD} is smaller (0.2261), but $CAIC_{NB}$ is also smaller (337.641). This is a reminder that the various criteria do not always have consistent trends. In addition, the behavior of $CAIC_{NB}$ suggests the possibility that a regime is being entered where overfitting occurs. Overfitting occurs when random variation in a set of input variables is used to explain the random variation in a single output variable. When the number of input variables is a significant fraction of the sample size, some combination of the noises in the input variables may, by coincidence, approximate the variation in the output variable without having predictive value.

Models for INJACC and INJACCI are shown in Table 34. Turning percentages are significant for these models and so is posted minor road speed SPD2, but other design variables fail to be. PK%THRU2 is marginally significant with a P-value of about 0.19, but it has been omitted, in part, because the interpretation is unclear. Since SPD2 correlates negatively with STATE, one might suspect that its influence is due to that source, but STATE itself is not significant in the presence of the ADT variables and PK%LEFT1.

The general features of the models for the four-legged intersections are:

- Turning percentage, along with major and minor road ADT, are influential for all crash types.
- LTLN1S, which registers the presence of one or more left-turn lanes on the major road, is influential for TOTACC and marginally so for TOTACCI.
- Grade and poor sight distance right from the minor road are marginally significant for TOTACC and TOTACCI.
- High minor road posted speed appears to contribute to serious crashes.

MODELS FOR THE SIGNALIZED INTERSECTIONS

Negative Binomial Models

The signalized intersections present special difficulties As shown in Table 26 and Figure 9, at first appearances, the dependence of crashes on major road ADT is negligible. Likewise, the correlation coefficient between crashes and ADT1 is insignificant in Table 12. An ADT-only model for TOTACC in terms of the logs of ADT1 and ADT2 actually assigns a negative (but insignificant) regression coefficient to the log of ADT1.

Part of the insignificance perhaps stems from the small sample size — only 49 signalized intersections. However, at signalized intersections, minor and major roads tend to have more equal standing. If their standing is equal, their ADT's should enter into any model symmetrically. For example, the coefficient of ADT1 would be the same as that of ADT2 except for noise. We have attempted to address that possibility by using the log of the product, log(ADT1×ADT2), as a variable in some of the signalized models below. At the same time, ADT by itself becomes less important. Signalized intersections, one may argue, are less stereotypical than other rural intersections. On the latter, the division between major road and minor road is more pronounced and the turning percentages on each fall into a narrower range. More important on the signalized intersections, one would judge, are the movements of the vehicles through the intersection. Turning percentages, left and right, from all approaches and flows along each approach are likely to be more determinative of crashes.

There is also the issue of how to define the major road. Usually, and in this study, it is taken to be the road with the larger ADT. But if there is significant turning along certain legs, legs of the same road may have drastically different ADT. Most of the ADT may be on two adjacent legs, say legs 1 and 4, and very little on the other two adjacent legs, legs 2 and 3. See Figure 1. Usually, the major road has a lower percentage of turning traffic than the minor road, but it is possible that a road with less traffic would have virtually no turning traffic (all of it through), while the crossroad has much more traffic and a significant amount of it is turning. In the data, an asymmetry can occur between minor road turning traffic and major road turning traffic. This can be caused by failure of the morning and evening peak hours to match up, by unusual travel hours to and from locations, or even by alternative routes.

Despite these considerations, the models exhibited here take ADT to be primary, in part because of its familiarity and acceptability to the engineering community and in part to permit comparisons with other models that use ADT. Yet, it should be recognized that rural signalized intersections are a transitional class where variables other than ADT may prove to be more appropriate. This is addressed in the subsection after this one, where models of single and multiple-vehicle crashes in terms of traffic flows are briefly investigated.

In Tables 35, 36, and 37 are shown negative binomial models for crashes on the signalized two-lane by two-lane intersections. ADT-only models are omitted since when ADT1 and ADT2 are separated, ADT1 is insignificant and has a coefficient of negative sign, and when they are united in the form log(ADT1×ADT2), the model coefficients are somewhat unstable (SAS and LIMDEP give rather different values for the regression coefficients, but the same log likelihood, indicating that the maximum occurs at a hard-to-find set of values on a large relatively flat plateau). When variables that correlate well with the residuals to these models are added, the models settle down and the ADT variables share in the significance.

Table 35 shows models for TOTACC. The existence of one or more protected left turns on the major road at the signal is an influential variable. It correlates strongly with STATE, as noted earlier, and it is possible that there is a combined effect here. A total of 17 out of 18 California signalized intersections had one or more protected left turns on the major road, while only 4 out of 31 Michigan signalized intersections did. Nonetheless, when PROT_LT is added to the model versus STATE, the former improves the model more than the latter, and the correlation between the residual of a PROT_LT model and the STATE variable is negligible.

TABLE 35. Negative Binomial Models for Crashes per Year (TOTACC), Signalized Intersections

Estimated regression coefficients (estimated standard error and P-value in parentheses).

Variables	Main Model	Variant 1	Variant 2	Variant 3
Intercept	**-6.9536** (2.7911, 0.0132)	**-6.1236** (2.5973, 0.0184)	**-6.3658** (3.3207, 0.0552)	**-5.4091** (3.0054, 0.0718)
Log of ADT1	**0.6199** (0.2504, 0.0133)		**0.6475** (0.3156, 0.0402)	
Log of ADT2	**0.3948** (0.1737, 0.0230)		**0.2104** (0.2232, 0.3459)	
Log of ADT1×ADT2		**0.4643** (0.1483, 0.0017)		**0.3914** (0.1732, 0.0238)
PROT_LT 0 = no, 1 = yes	**-0.6754** (0.1824, 0.0002)	**-0.6110** (0.1507, 0.0001)	**-0.7181** (0.1973, 0.0003)	**-0.5980** (0.1690, 0.0004)
PK%LEFT2 minor road	**-0.0142** (0.0047, 0.0023)	**-0.0134** (0.0048, 0.0052)		
PK%LEFT1 major road			**0.0220** (0.0142, 0.1207)	**0.0137** (0.0116, 0.2388)
VEICOM vertical, all legs	**0.1299** (0.0450, 0.0039)	**0.1243** (0.0507, 0.0142)	**0.1001** (0.0508, 0.0486)	**0.1044** (0.0618, 0.0914)
PK%TRUCK truck %	**0.0315** (0.0143, 0.0275)	**0.0300** (0.0141, 0.0331)	**0.0353** (0.0175, 0.0441)	**0.0317** (0.0167, 0.0573)
N, p	49, 7	49, 6	49, 7	49, 6
K **R^2_K**	**0.1161** (0.0323, 0.0003) **0.6490**	**0.1186** (0.0317, 0.0002) **0.6414**	**0.1353** (0.0341, 0.0001) **0.5910**	**0.1422** (0.0375, 0.0002) **0.5701**
R^2	**0.5053**	**0.5208**	**0.5134**	**0.5172**
R^2_D, P^2_D **R^2_{PD}**	**0.1479, 0.6262** **0.2362**	**0.1619, 0.6349** **0.2550**	**0.1059, 0.6263** **0.1691**	**0.1123, 0.6351** **0.1768**
$CAIC_{NB}$	358.508	356.471	363.937	363.044

TABLE 36. Negative Binomial Models for Crashes per Year (TOTACCI), Signalized Intersections

Estimated regression coefficients (estimated standard error and P-value in parentheses).

Variables	Main Model	Variant 1	Variant 2	Variant 3
Intercept	-6.0841 (3.3865, 0.0724)	-4.9564 (3.0779, 0.1074)	-4.1075 (2.9461, 0.1633)	-5.4581 (3.1937, 0.0874)
Log of ADT1	0.5951 (0.2847, 0.0366)			0.5995 (0.2795, 0.0319)
Log of ADT2	0.2935 (0.1972, 0.1366)			0.2015 (0.1917, 0.2932)
Log of ADT1×ADT2		0.3857 (0.1788, 0.0309)	0.3320 (0.1719, 0.0534)	
PROT_LT 0 = no, 1 = yes	-0.4708 (0.2000, 0.0186)	-0.3822 (0.1668, 0.0220)	-0.3025 (0.1745, 0.0830)	-0.4041 (0.1883, 0.0319)
PK%LEFT2 minor road	-0.0165 (0.0057, 0.0036)	-0.0153 (0.0060, 0.0101)	-0.0160 (0.0055, 0.0038)	-0.0177 (0.0050, 0.0005)
VEICOM vertical, all legs	0.1126 (0.0365, 0.0020)	0.1033 (0.0416, 0.0130)	0.0996 (0.0382, 0.0091)	0.1114 (0.0326, 0.0006)
PK%TRUCK truck %	0.0289 (0.0131, 0.0276)	0.0268 (0.0131, 0.0398)	0.0234 (0.0122, 0.0547)	0.0256 (0.0117, 0.0287)
NODRWY1 major road			0.0347 (0.0270, 0.1986)	0.0407 (0.0246, 0.0983)
N, p	49, 7	49, 6	49, 7	49, 8
K R_K^2	0.1313 (0.0392, 0.0008) 0.5521	0.1354 (0.0390, 0.0005) 0.5382	0.1222 (0.0374, 0.0011) 0.5834	0.1145 (0.0401, 0.0043) 0.6094
R^2	0.3650	0.3913	0.4563	0.4327
R_D^2, P_D^2 R_{PD}^2	0.0944, 0.5854 0.1612	0.1053, 0.5951 0.1770	0.1067, 0.5853 0.1822	0.1044, 0.5751 0.1816
$CAIC_{NB}$	342.266	340.672	340.831	341.551

TABLE 37. Negative Binomial Models for Crashes per Year (INJACC, INJACCI), Signalized Intersections

Estimated regression coefficients (estimated standard error and P-value in parentheses).

Variables	INJACC	INJACCI
Intercept	-3.2562 (2.9932, 0.2767)	-1.5475 (3.0298, 0.6095)
Log of ADT1×ADT2	0.2358 (0.1722, 0.1707)	0.1290 (0.1757, 0.4627)
PROT_LT 0 = no, 1 = yes	-0.2943 (0.1864, 0.1144)	
PK%LEFT2 minor road	-0.0113 (0.0062, 0.0678)	-0.0149 (0.0066, 0.0250)
VEICOM vertical, all legs	0.0822 (0.0551, 0.1358)	0.0686 (0.0692, 0.1858)
PK%TRUCK truck %	0.0323 (0.0146, 0.0267)	0.0282 (0.0152, 0.0628)
N, p	49, 6	49, 5
K R_K^2	0.1630 (0.0662, 0.0138) 0.4474	0.1433 (0.0692, 0.0385) 0.4829
R^2	0.3275	0.3488
R_D^2, P_D^2 R_{PD}^2	0.0420, 0.4926 0.0853	0.0665, 0.4565 0.1458
$CAIC_{NB}$	285.287	265.687

Other significant variables shown in the main model of Table 35 include PK%TRUCK, VEICOM, and PK%LEFT2.

Crashes rise at signalized intersections with a higher percentage of truck traffic and with more vertical curvature out to 800 feet (244 meters) on any or all approaches. Trucks at a signal, as well as having greater destructive capacity than passenger vehicles, take a long time to engage in turning maneuvers and block visibility during this time. In Table 13, almost all vertical variables correlate positively with crashes, although few have significant P-values. The combination that is most significant in the modeling is VEICOM. VEICOM is an average change of grade per 100 feet (30.5 meters) along both major and minor roads for vertical curves at least partly within 800 feet (244 meters) of the intersection center. One may wonder why VEICOM is more significant than VICOM, the comparable measure out to 250 feet (76 meters). Signalized intersections are rarely placed immediately beside vertical curves, but are often found to be displaced from them by hundreds of yards or meters. The mean, median, and standard deviation of VICOM are 1.79, 1.2, and 0.28, respectively, while those for VEICOM are 1.88, 1.43, and 0.27. The difference in medians, in particular, shows that vertical curves partly within 800 feet (244 meters) of the intersection, but not within 250 feet (76 meters), increase the average.

Crashes fall with increasing PK%LEFT2, the left-turn percentage on the minor road. PK%LEFT2 is the most significant of the turning percentage variables, but the others are also significant. PK%LEFT2 is, of course, equivalent to (100 - PK%THRU2 - PK%RIGHT2), i.e., to the sum of PK%THRU2 and PK%RIGHT2, and each of the latter two variables correlates positively with crashes (see Table 23). PK%LEFT2 correlates negatively with PK%LEFT1 and thus the latter should increase crashes. Variants 2 and 3 in Table 35 show that this is indeed the case, but that the P-value rises. Note also that the P-value for the log of ADT2 becomes rather large in Variant 2, presumably due to the strong positive correlation between PK%LEFT1 and ADT2 (Table 17).

The difference between the Main Model in Table 35 and Variant 1 is in the use of Log(ADT1×ADT2) rather than the individual logs. In fact, Variant 1 gives a smaller value of $CAIC_{NB}$ and a larger value of R^2. The decrease in $CAIC_{NB}$ suggests that Variant 1 may be the superior model: it has about the same explanatory value, but with fewer variables. In the Main Model, we have elected to exhibit coefficients for ADT1 and ADT2 separately, partly to allow comparison with other models. When they are combined in Variant 1, the new coefficient is intermediate between the separate coefficients. The estimated difference in the two coefficients in the Main Model is, of course, 0.6199 - 0.3948 = 0.2251. Using the estimated covariance matrix for the model, we find that the estimated standard error of the estimated difference is $[(0.2504)^2 + (0.1737)^2 - 2\times 0.0039226]^{1/2} \approx 0.2916$. This gives a P-value of 0.4401 for testing the hypothesis that the coefficients are different. In other words, the Main Model does not allow us to reject the hypothesis that the regression coefficients of the logs of ADT1 and ADT2 are the same.

The models for TOTACCI in Table 36 are similar to those in Table 35, except that the P-value of ADT2 increases and the variable NODRWY1 is marginally significant in Variant 2 and significant in Variant 3. Variant 3 has an unacceptably high P-value for ADT2. NODRWY2 and the combined variable NODRWYCOM, although positively correlated with crashes, do not achieve as good a P-value as NODRWY1. NODRWY1 also correlates positively with TOTACC, and in a TOTACC model with the same variables as Variant 2 of Table 36, gives a P-value of 0.2270. Surprisingly, its P-value in Variant 2 of Table 36, 0.1986, is lower. This is a surprise because TOTACCI attempts to eliminate driveway crashes with no intersection involvement. We have omitted variant models in which PK%LEFT1 is used instead of PK%LEFT2. In one such model, the P-value of ADT2 jumps to 0.6481, although other variables behave well; in another model with LOG(ADT1×ADT2), ADT behaves well, but VEICOM and PK%LEFT1 have P-values of 0.2402 and 0.3263, respectively.

For both TOTACC and TOTACCI, variables such as LIGHT and LTLN1 correlate well with the residuals of the models shown, LIGHT positively and LTLN1 negatively. When these variables are added to the models, they are also marginally significant. However, the values of $CAIC_{NB}$ do not decrease, and concern about overfitting leads us to omit them.

In Table 37, we present one model each for INJACC and INJACCI. The coefficient of the log of ADT2 is quite insignificant because of large standard error. So we only exhibit models using LOG(ADT1×ADT2). Even with these, the P-value deteriorates substantially. In addition, VEICOM becomes less significant and PROT_LT attains, in the case of INJACCI, a P-value of 0.5666 (not shown).

The main features of the signalized intersection models are:

- ADT1 is insignificant for all crash types when ADT2 is present but without other variables.
- PK%TRUCK and the turning percentages, especially PK%LEFT2, are significant for all crash types.
- The existence of one or more protected left turns on the major road, as well as major and minor road vertical curves, is significant for TOTACC and TOTACCI, becoming less significant for INJACC and insignificant for INJACCI.
- NODRWY1 is marginally significant for TOTACC and TOTACCI, but not for serious accidents.
- For TOTACC and TOTACCI models, in general, ADT1 becomes more significant as variables are added, while ADT2 gets less significant, sharing its influence with turning percentage.

Flow Models

The signalized intersections, as noted, behave somewhat peculiarly with respect to ADT. This suggests a more detailed analysis, making use of flows and crash types. Here we examine a few models based on the decomposition of TOTACC into single-vehicle crashes and multiple-vehicle crashes by the variables TOTACCS and TOTACCM. Although many single-vehicle crashes may in fact be multiple-vehicle crashes in which other vehicles escape unscathed, we proceed as if this decomposition is valid.

For single-vehicle crashes, one approach is to regard them as functions of incoming flows, with minor and major legs treated on an equal footing and without interaction terms. An underlying rationale is that single-vehicle crashes depend on same-direction traffic as well as intersection features other than traffic, such features including perhaps pedestrian traffic, intersection geometry, roadside hazards, obstructions that limit sight distances, signal timing, etc. Then one might expect that the number of such crashes is proportional to some power of the flow. Although such a view is not particularly consistent with the evidence in Table 21, we pursue the approach and indicate the outcome.

A negative binomial model with mean number of single-vehicle crashes per unit time of the form

$$\mu = C \times [(F_1)^a + (F_2)^a + (F_3)^a + (F_4)^a]$$

is sought with variance equal to $\mu + K\mu^2$. The unknowns are the multiplicative constant C, the power a, and the overdispersion parameter K. For a given power a, LIMDEP or SAS will choose the pair (C, K) to maximize the probability (or log-likelihood) of the observed numbers y of single-vehicle crashes given the observed values of F_1, F_2, F_3, and F_4. The crude strategy we follow, suggested by the measure R_K^2, is to vary a and choose the triple (C, a, K) that yields the smallest value for K (and hence the largest for R_K^2).

When this is done, the resulting model is as follows:

SINGLE-VEHICLE CRASH MODEL, SIGNALIZED INTERSECTIONS

$$\mu = (\exp -1.9218) \times [(F_1)^{0.01} + (F_2)^{0.01} + (F_3)^{0.01} + (F_4)^{0.01}]$$

where

μ is the mean number of single-vehicle crashes per year,

the intersection flows are F_1, F_2 F_3, and F_4 in thousands of vehicles per day, and

the overdispersion parameter K = 0.4670.

The constant term -1.9218 has an estimated standard error of 0.1419 and a P-value of 0.0001, and the overdispersion parameter 0.4670 has an estimated standard error of 0.1993 and a P-value of 0.0192. Because of the modeling technique, an estimated standard error for the power a = 0.01 is not available.

The power 0.01 is evidently quite small. Indeed, the so-called zero model is not substantially different from the one above. It is:

$$\mu = \exp -1.6727$$

with overdispersion parameter K = 0.4674. Here the intercept -1.6727 has an estimated standard error of 0.1419 and a P-value of 0.0001, and the overdispersion parameter has a standard error of 0.1998 and a P-value of 0.0193. The overdispersion parameter of the zero model is only slightly larger than that of the proposed single-vehicle crash model. Given the size of the standard errors involved, this suggests that single-vehicle crashes are not appropriately estimated by this model form.

Turning to the multiple-vehicle crashes, we look for a negative binomial model of the form

$$\mu = C \times [(F_1)^a(F_4)^b + (F_4)^a(F_2)^b + (F_2)^a(F_3)^b + (F_3)^a(F_1)^b + p(F_1F_2)^c + p(F_3F_4)^c]$$

for which there are six unknown parameters: C, a, b, p, c, and the overdispersion parameter K. The first four terms represent interactions of adjacent flows and the last two represent interactions of opposing flows. Minor and major roads are represented symmetrically in this model form, but left versus right distinctions are maintained since a need not be equal to b, and adjacent flow interactions are not assumed to be of the same magnitude as opposite flow interactions, i.e., p need not be equal to 1 nor are the powers a and b constrained in relation to the power c.

The modeling methodology employed here, similar to that for the single-vehicle crash model, is to fix the quadruple (a, b, c, p) and apply SAS or LIMDEP to yield a maximum likelihood model for the observed number y of multiple-vehicle crashes given the observed flows F_1, F_2, F_3, and F_4. This yields values for the pair (C, K). Then the values of the quadruple (a, b, c, p) are varied in such a way as to minimize K.

The resulting model is the following:

MULTIPLE-VEHICLE CRASH MODEL, SIGNALIZED INTERSECTIONS

$$\mu = (\exp -0.4420) \times [(F_1F_4)^{0.3} + (F_4F_2)^{0.3} + (F_2F_3)^{0.3} + (F_3F_1)^{0.3} + 0.95]$$

where

μ is the mean number of multiple-vehicle crashes per year,

the intersection flows are F_1, F_2, F_3, and F_4 in thousands of vehicles per day, and

the overdispersion parameter K = 0.2936.

The constant term -0.4420 has an estimated standard error of 0.1015 and a P-value of 0.0001, and the overdispersion parameter 0.2936 has an estimated standard error of 0.0696 and a P-value of 0.0001. Because of the modeling technique, estimated standard errors for the powers a = b = 0.3, c = 0, and p = 0.95 are not available.

These results indicate that the product of opposing flows, at least when summed over both approaches, does not significantly contribute to the crash rate. The sum of the 0.3 powers of adjacent flow products is the relevant variable, and a linear transformation is applied to it. If the flow on any two opposite legs is zero, the mean number of multiple-vehicle crashes per year is estimated to be $\exp(-0.4420)\times(0.95) \approx 0.61$. Perhaps the chief point of interest is that the powers a and b turn out to be at least roughly equal, with values close to those in the models of Tables 35 and 36 (Variant 1). Note also that the overdispersion parameter for TOTACCM, 0.2936, is significantly larger than those shown in the TOTACC models of Table 35.

Many additional ideas could be explored along the lines introduced here. In particular, in view of Table 21, model forms that stress minor leg flows could be considered. Other crash decompositions could be considered, including TOTACCI, INJACC, INJACCI, time of day, or crash type.

RESIDUAL ANALYSIS

For the three Main Models of TOTACCI, from Tables 29, 33, and 36, graphs of cumulative scaled residuals versus explanatory variables are plotted in Figures 11 through 22. For an explanatory variable x, a plot is made of J versus the quantity

$$\sum_{\{ i : x_i \leq J \}} \frac{y_i - \hat{y}_i}{\sqrt{\hat{y}_i + K(\hat{y}_i)^2}} \tag{5.9}$$

called the cumulative scaled residual. The variable J runs through the set of values that the explanatory variable x assumes on the data set. The terms in (5.9) are scaled residuals and should each be approximately unbiased with mean square equal to 1 if the model form and estimated $\hat{y}_i$ and K are essentially correct. However, if the sum depends in some regular way on the values of 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 $x \leq J$. For the three samples $\sqrt{84} \approx 9$, $\sqrt{72} \approx 8.5$, and $\sqrt{49} = 7$, and these numbers are indications of the permissible order of magnitude of the sum. 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 Main Models in Tables 29, 33, and 36, the overall sums of the scaled residuals are 5.7, -0.5, and -0.2, respectively. Thus, the corresponding graphs should wander from a height of 0 to these heights in a random manner.

Figures 11 through 14 refer to the Main Model for TOTACCI in Table 29 (three-legged intersections). The explanatory variables are ADT1, ADT2, MEDWIDTH1, and NODRWY1. The graphs of scaled residuals versus each of these four variables exhibit regions of systematic trends. This suggests that separate models might capture the crash counts better with variables restricted to smaller ranges.

Figures 15, 16, and 17 refer to the Main Model for TOTACCI in Table 33 (four-legged intersections). The explanatory variables are ADT1, ADT2, and PK%LEFT1. Another variable, LTLN1S, which indicates the presence of left-turn lanes on the major road, is marginally significant, but is categorical in nature and hence does not lend itself to detailed residual analysis. In any case, it is not included in the Main Model. Figures 16 and 17 indicate that there may be quadratic dependence on ADT2 (Log of ADT2) and/or PK%LEFT1. Table 33 does include a model (Variant 2) with quadratic dependence on PK%LEFT2, which appears to be an improvement over the Main Model according to the various R-squared measures. The horizontal outlier in Figure 15 is a four-legged intersection with a major road ADT of 73,000. When it was removed from the sample and modeling was done without it, there were small but insignificant changes to the estimated regression coefficients and the estimated overdispersion parameter. It was not found to be unduly influential.

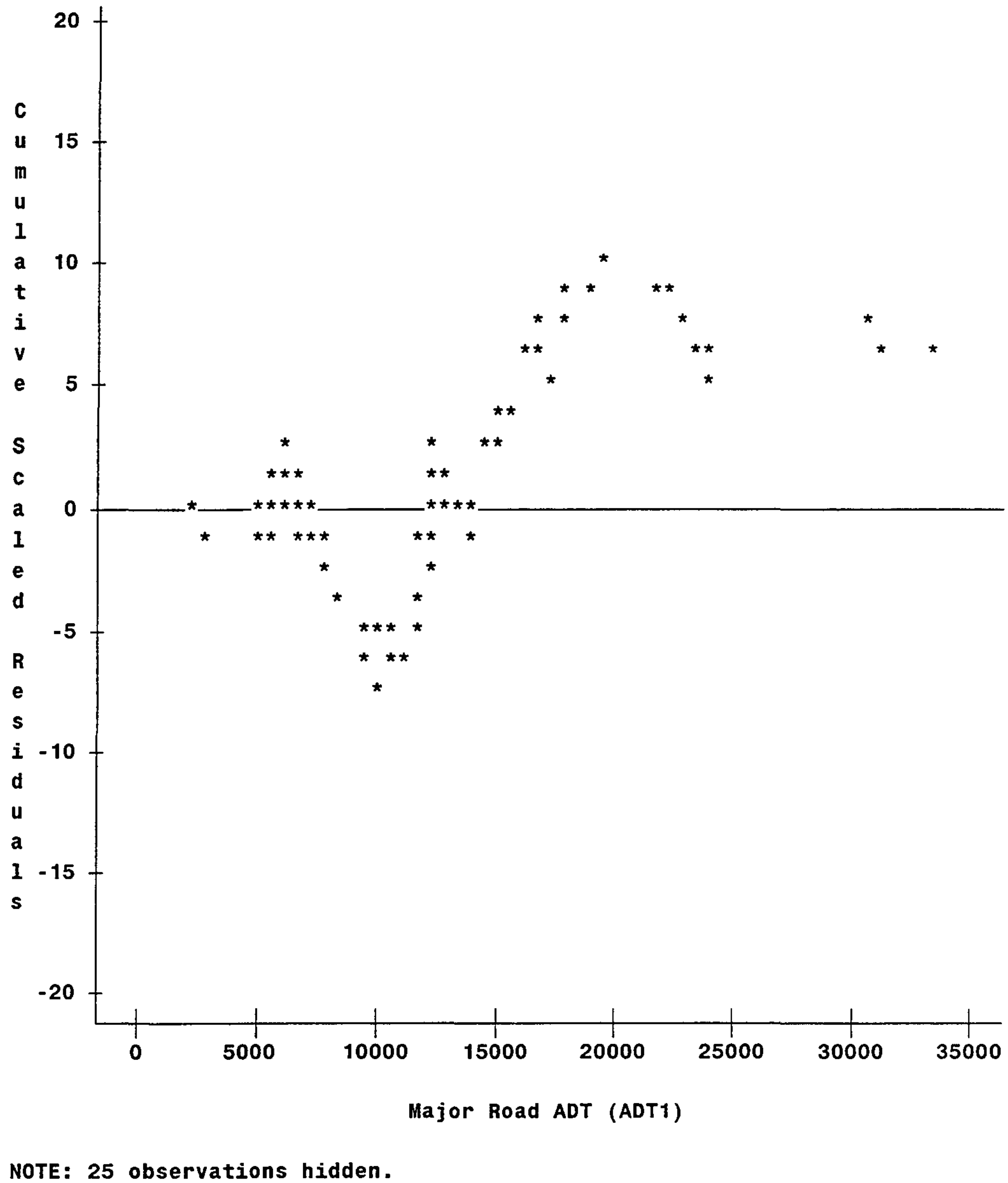


FIGURE 11. Cumulative Scaled Residual Versus ADT1 for Three-Legged Intersections, TOTACCI Main Model of Table 29

The cumulative scaled residual varies from -7.2 to 10.2, ending at 5.7. It is positive for 38 out of 84 intersections. In the middle range of ADT, the model at first overpredicts (negative slope) and then underpredicts (positive slope).

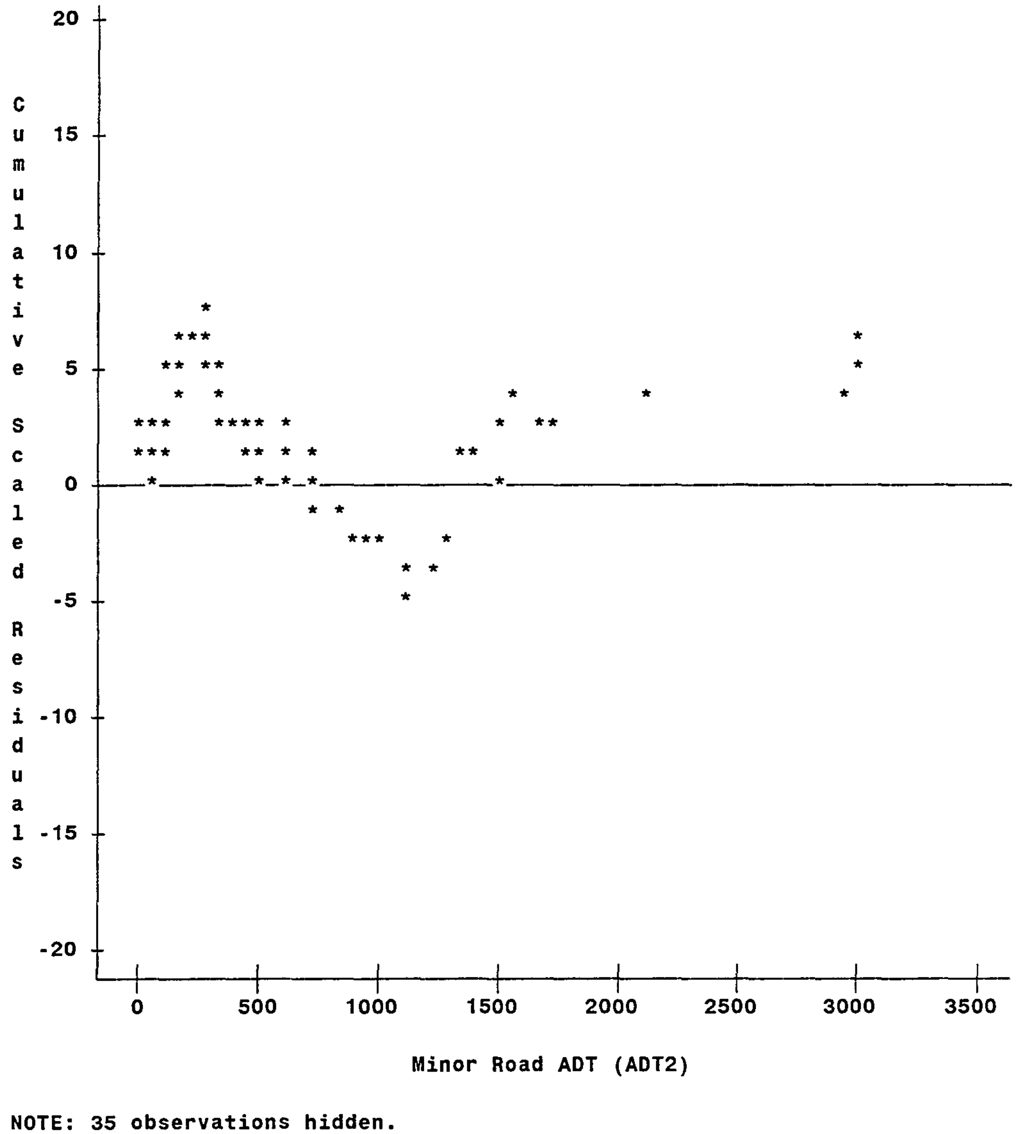
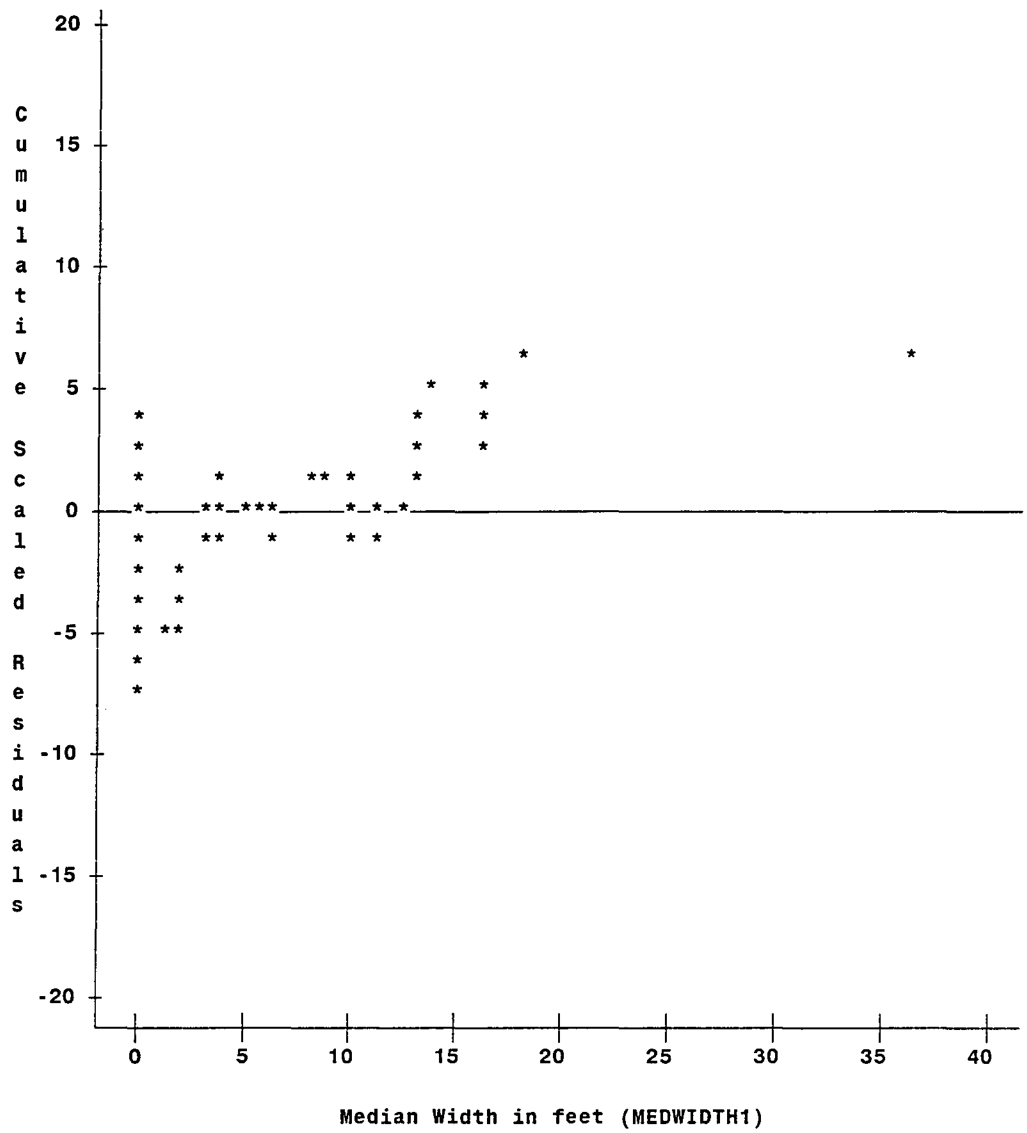


FIGURE 12. Cumulative Scaled Residual Versus ADT2 for Three-Legged Intersections, TOTACCI Main Model of Table 29

The cumulative scaled residual varies from -4.5 to 7.9, ending at 5.7. It is positive for 65 out of 84 intersections. For low values of ADT2, the model underpredicts.



NOTE: 44 observations hidden.
1 ft = 0.305 m

FIGURE 13. Cumulative Scaled Residual Versus MEDWIDTH1 for Three-Legged Intersections, TOTACCI Main Model of Table 29

The cumulative scaled residual varies from -7.2 to 6.2, ending at 5.7. It is positive for 32 out of 84 intersections. On the eight intersections with median widths from 12 to 16 feet (3.7 to 4.9 meters), the model underpredicts on average.

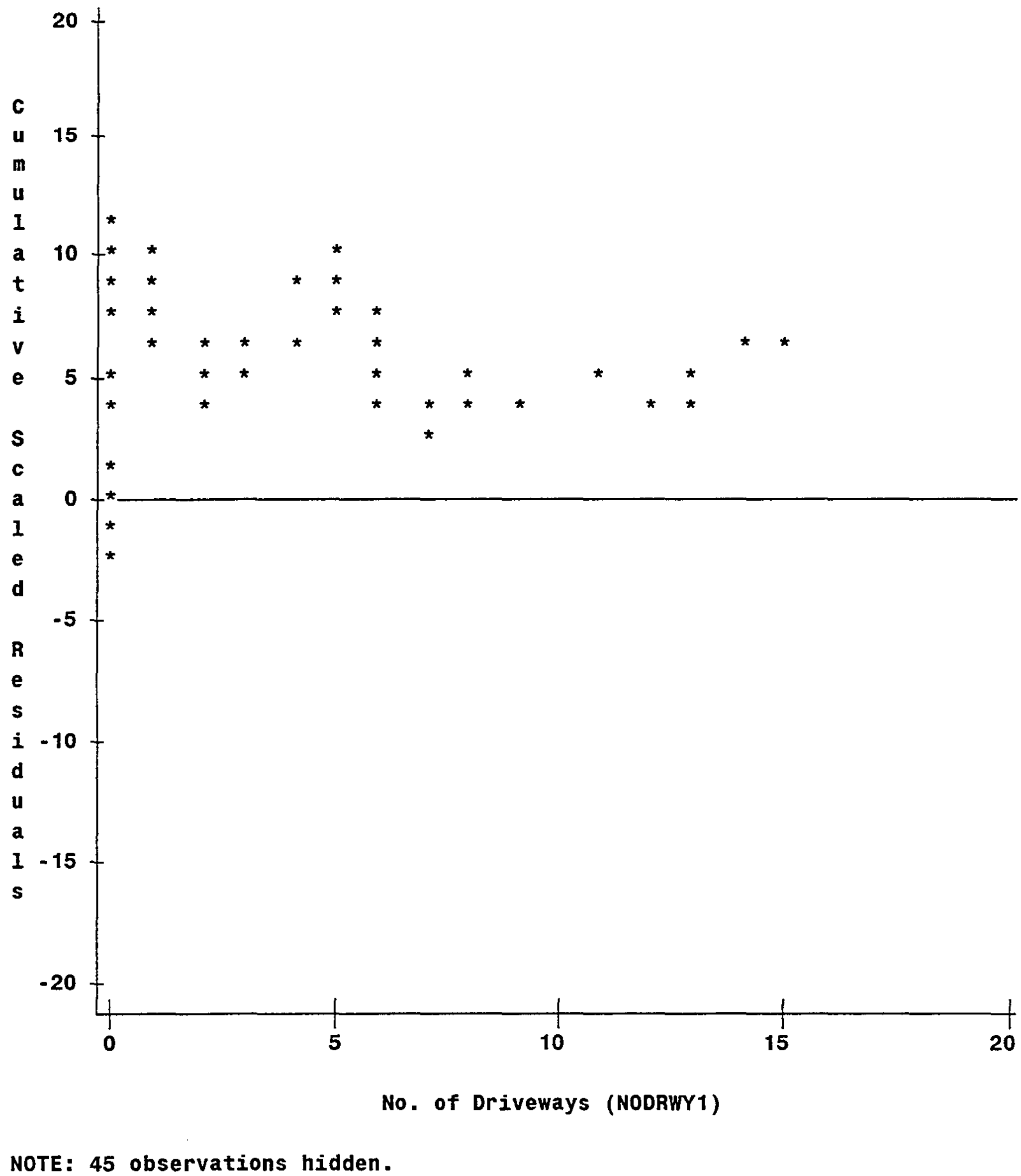


FIGURE 14. Cumulative Scaled Residual Versus NODRWY1 for Three-Legged Intersections, TOTACCI Main Model of Table 29

The cumulative scaled residual varies from -3.1 to 10.9, ending at 5.7. It is positive for 69 out of 84 intersections. When there are few driveways, the model tends to underpredict crashes.

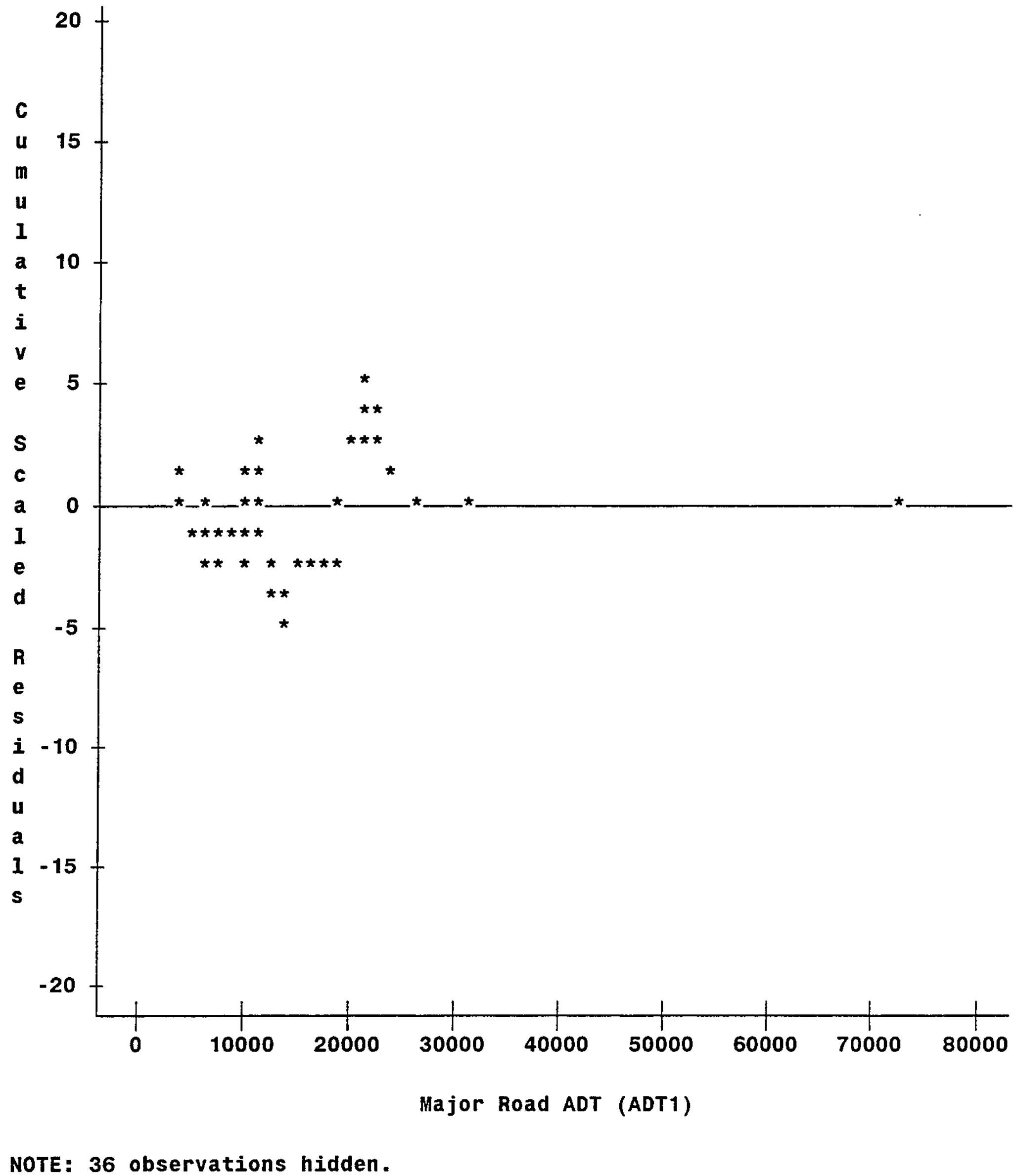


FIGURE 15. Cumulative Scaled Residual Versus ADT1 for Four-Legged Intersections, TOTACCI Main Model of Table 33

The cumulative scaled residual varies from -4.5 to 5.2, ending at -0.5. It is positive for 27 out of 72 intersections.

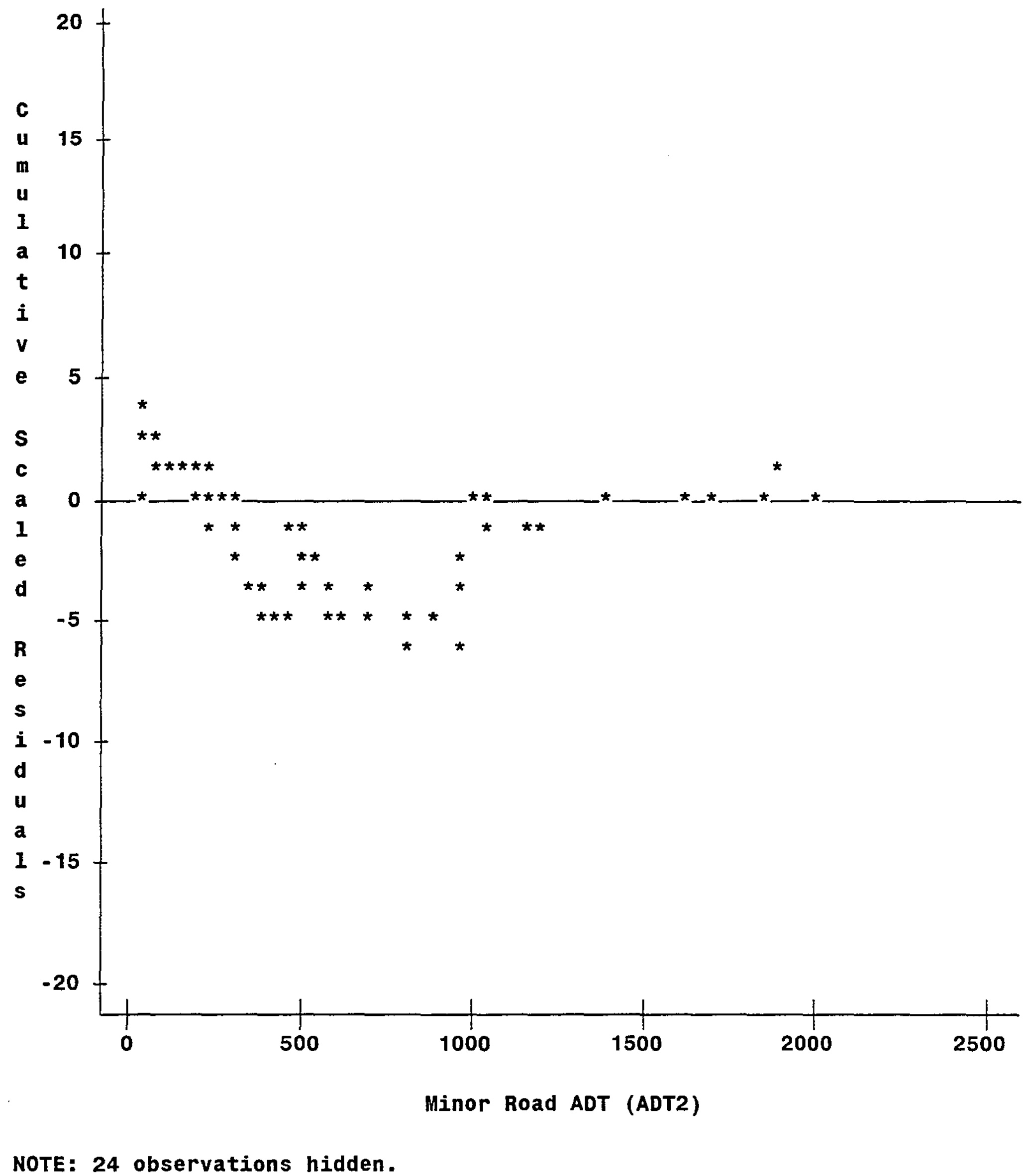


FIGURE 16. Cumulative Scaled Residual Versus ADT2 for Four-Legged Intersections, TOTACCI Main Model of Table 33

The cumulative scaled residual varies from -6.1 to 3.5, ending at -0.5. It is positive for 26 out of 72 intersections. There is some indication of quadratic dependence on ADT2 or Log of ADT2 to describe overprediction at low values of ADT2 and underprediction at higher values.

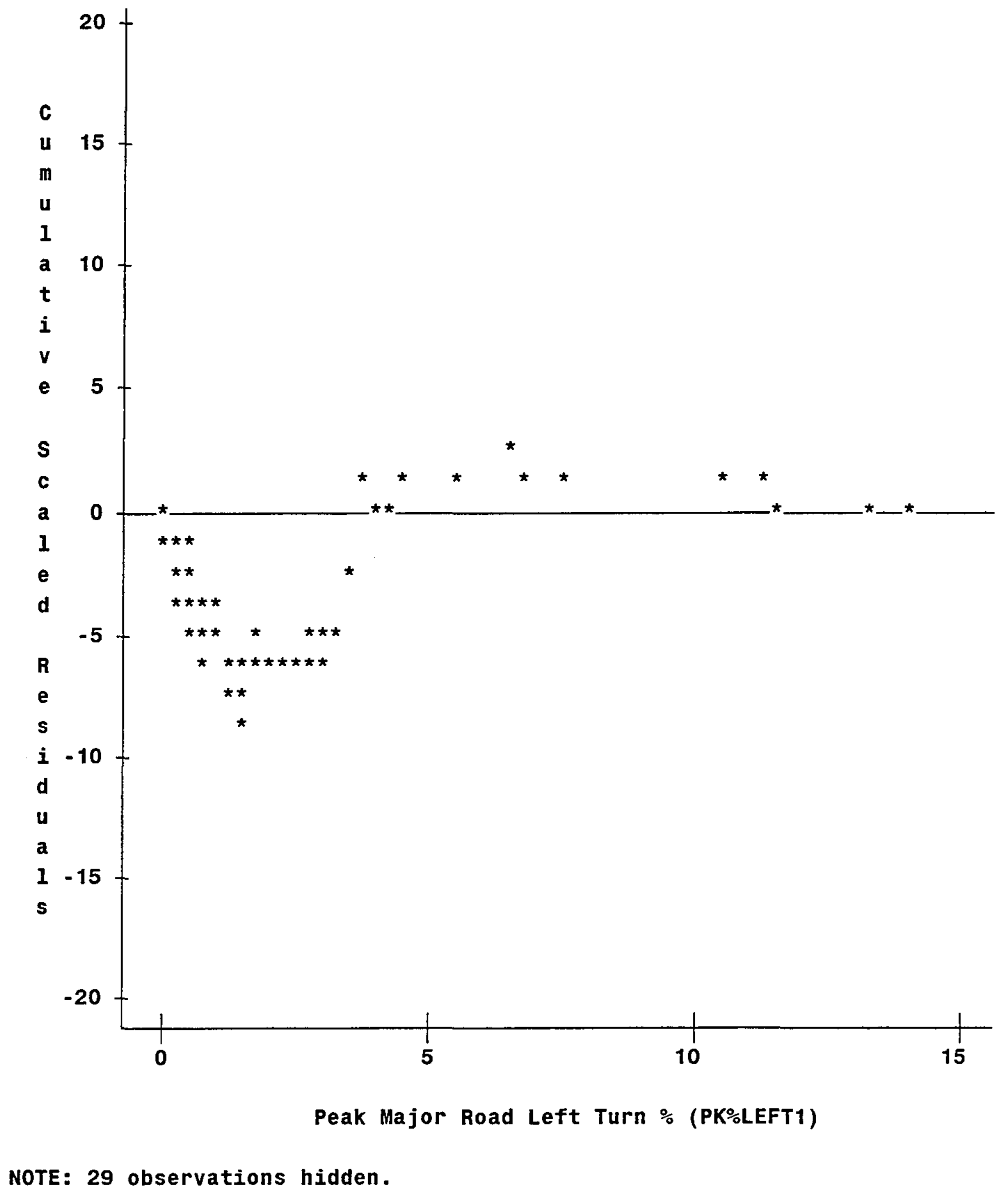


FIGURE 17. Cumulative Scaled Residual Versus PK%LEFT1 for Four-Legged Intersections, TOTACCI Main Model of Table 33

The cumulative scaled residual varies from -8.4 to 2.6, ending at -0.5. It is positive for 12 out of 72 intersections. There is some indication of overprediction at lower turning percentages, followed by underprediction at somewhat higher turning percentages. A quadratic model addresses this matter in Variant 2 of Table 33, starting out with a smaller intercept and steeper slope, but with the slope becoming smaller as PK%LEFT1 increases.

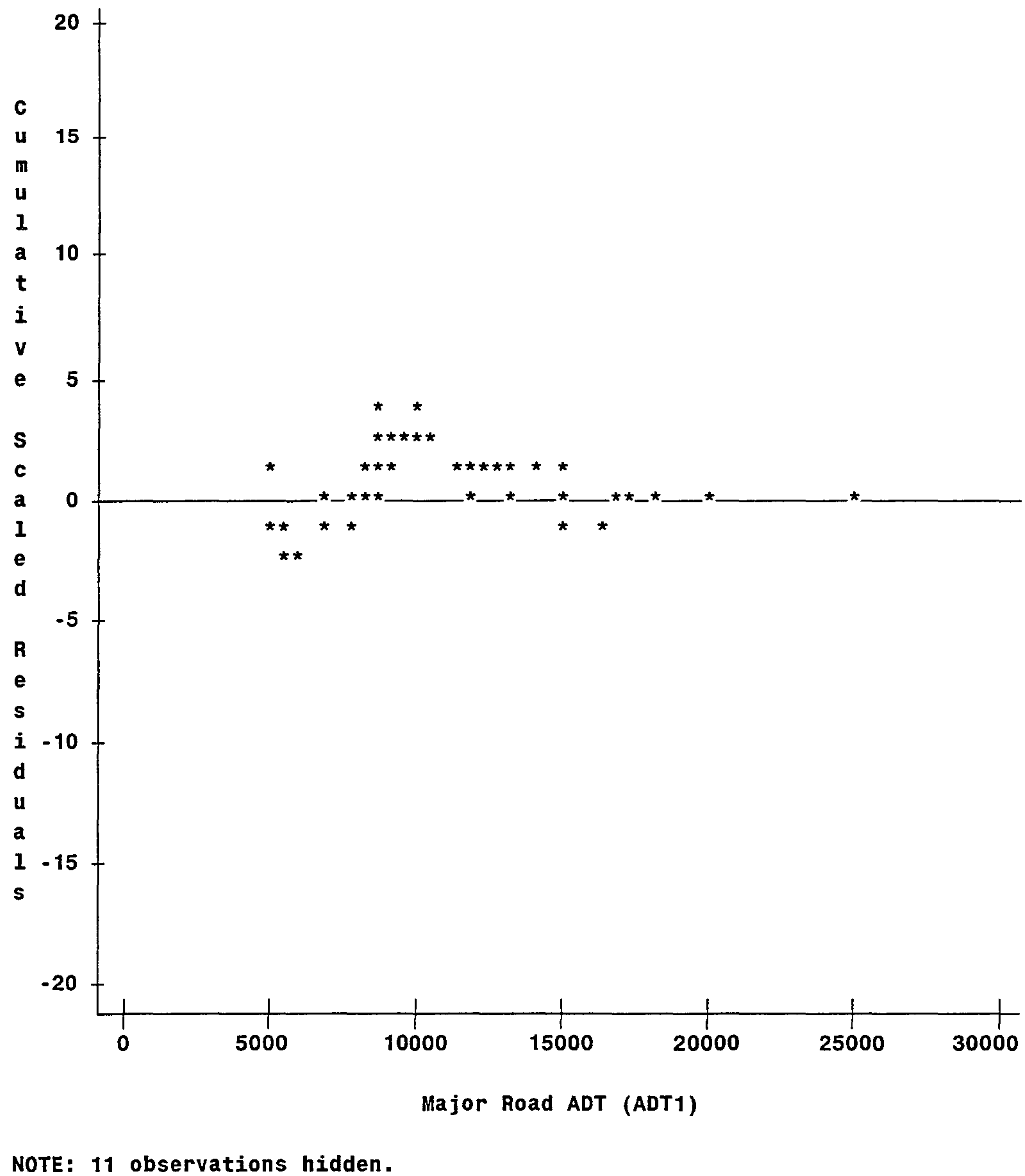
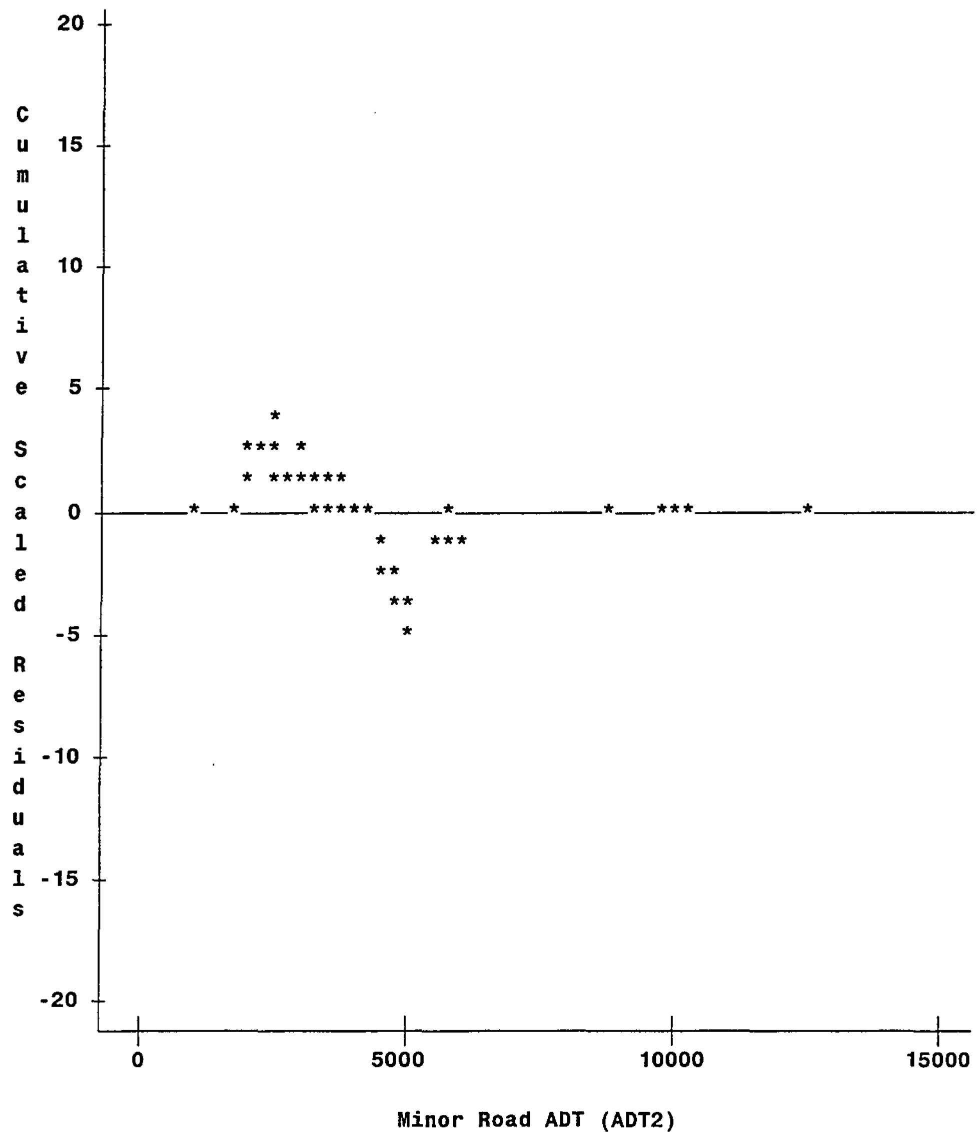


FIGURE 18. Cumulative Scaled Residual Versus ADT1 for Signalized Intersections, TOTACCI Main Model of Table 36

The cumulative scaled residual varies from -3.1 to 3.9, ending at -0.2. It is positive for 31 out of 49 intersections.



NOTE: 15 observations hidden.

FIGURE 19. Cumulative Scaled Residual Versus ADT2 for Signalized Intersections, TOTACCI Main Model of Table 36

The cumulative scaled residual varies from -5.4 to 4.2, ending at -0.2. It is positive for 29 out of 49 intersections.

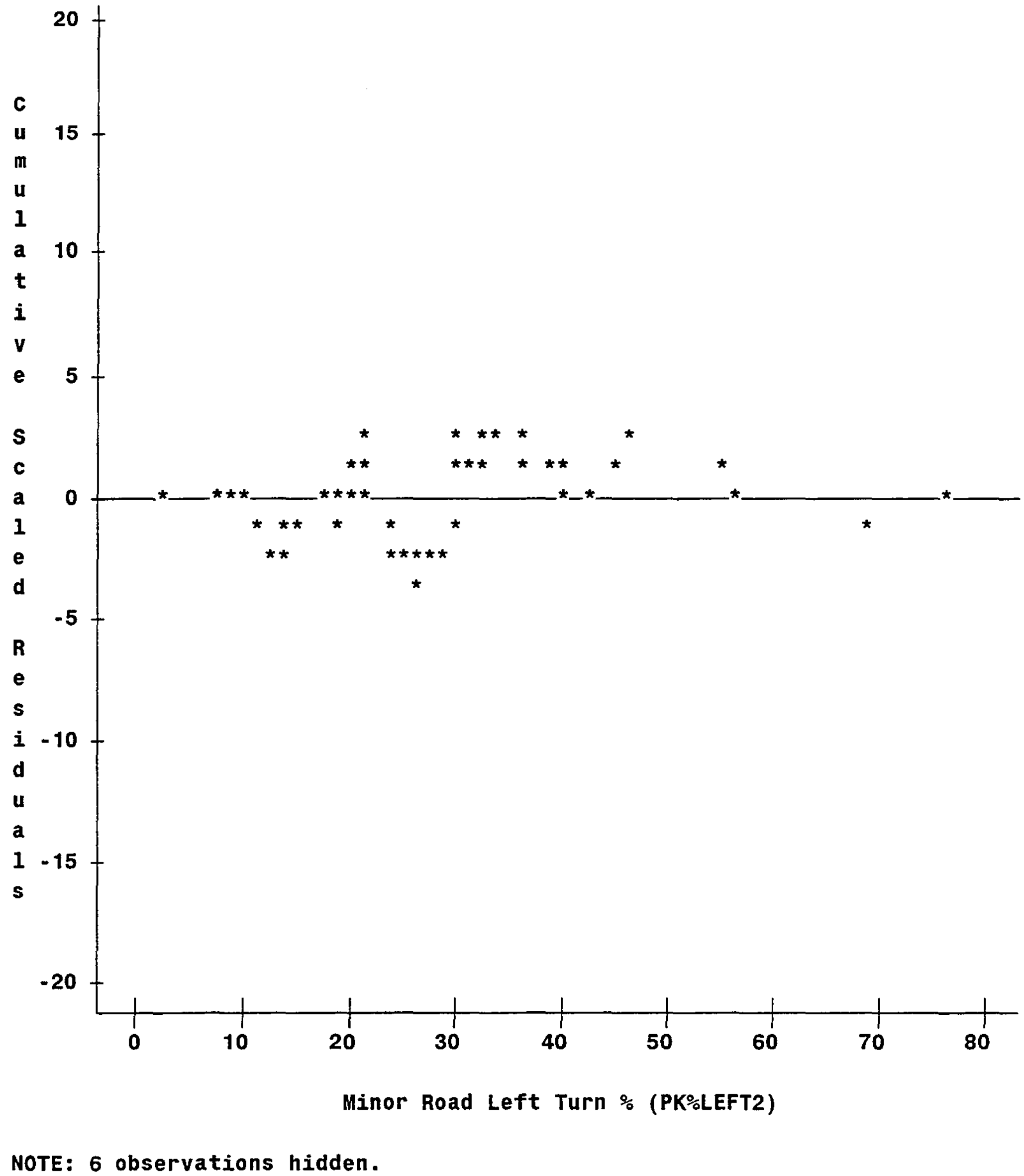
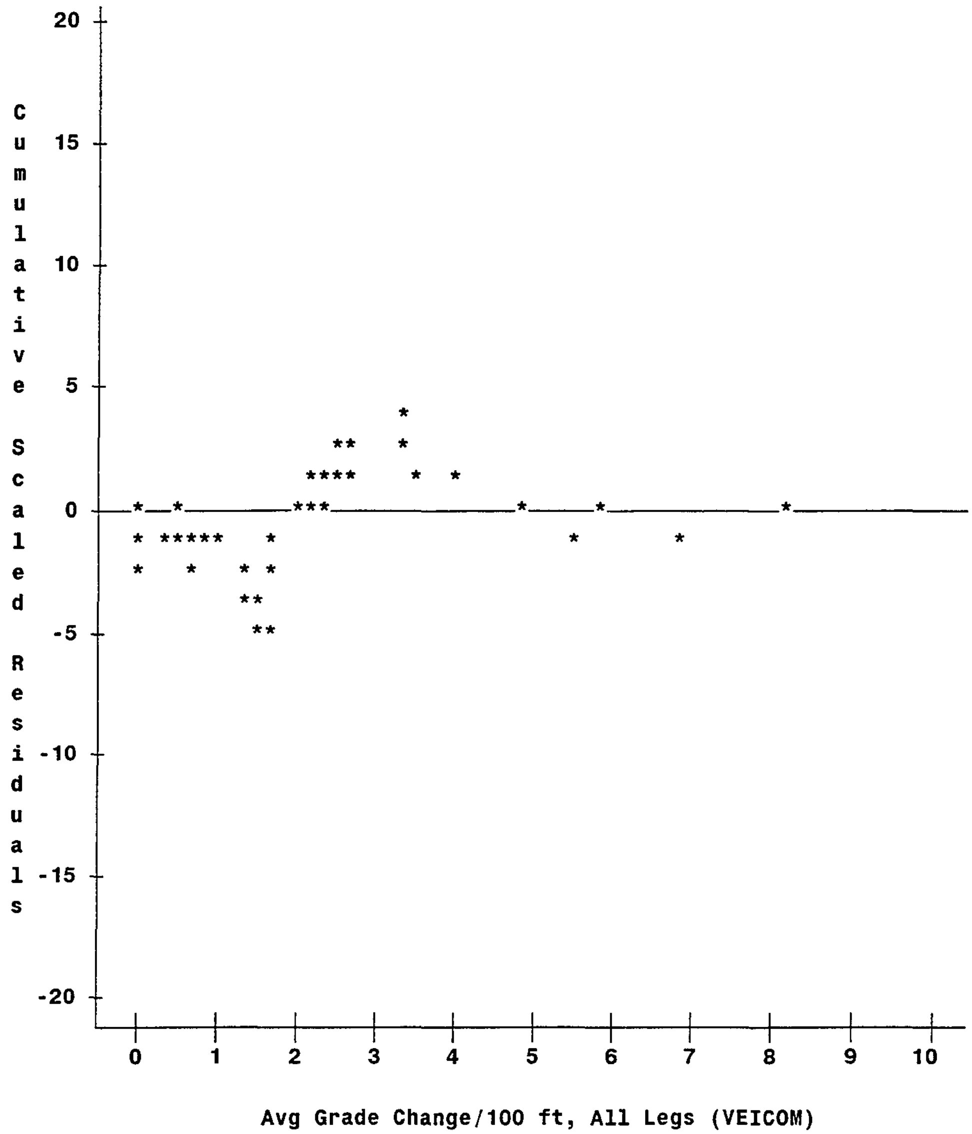


FIGURE 20. Cumulative Scaled Residual Versus PK%LEFT2 for Signalized Intersections, TOTACCI Main Model of Table 36

The cumulative scaled residual varies from -3.2 to 2.9, ending at -0.2. It is positive for 24 out of 49 intersections.



NOTE: 14 observations hidden.
1 ft = 0.305 m

FIGURE 21. Cumulative Scaled Residual Versus VEICOM for Signalized Intersections, TOTACCI Main Model of Table 36

The cumulative scaled residual varies from -4.6 to 3.2, ending at -0.2. It is positive for 19 out of 49 intersections.

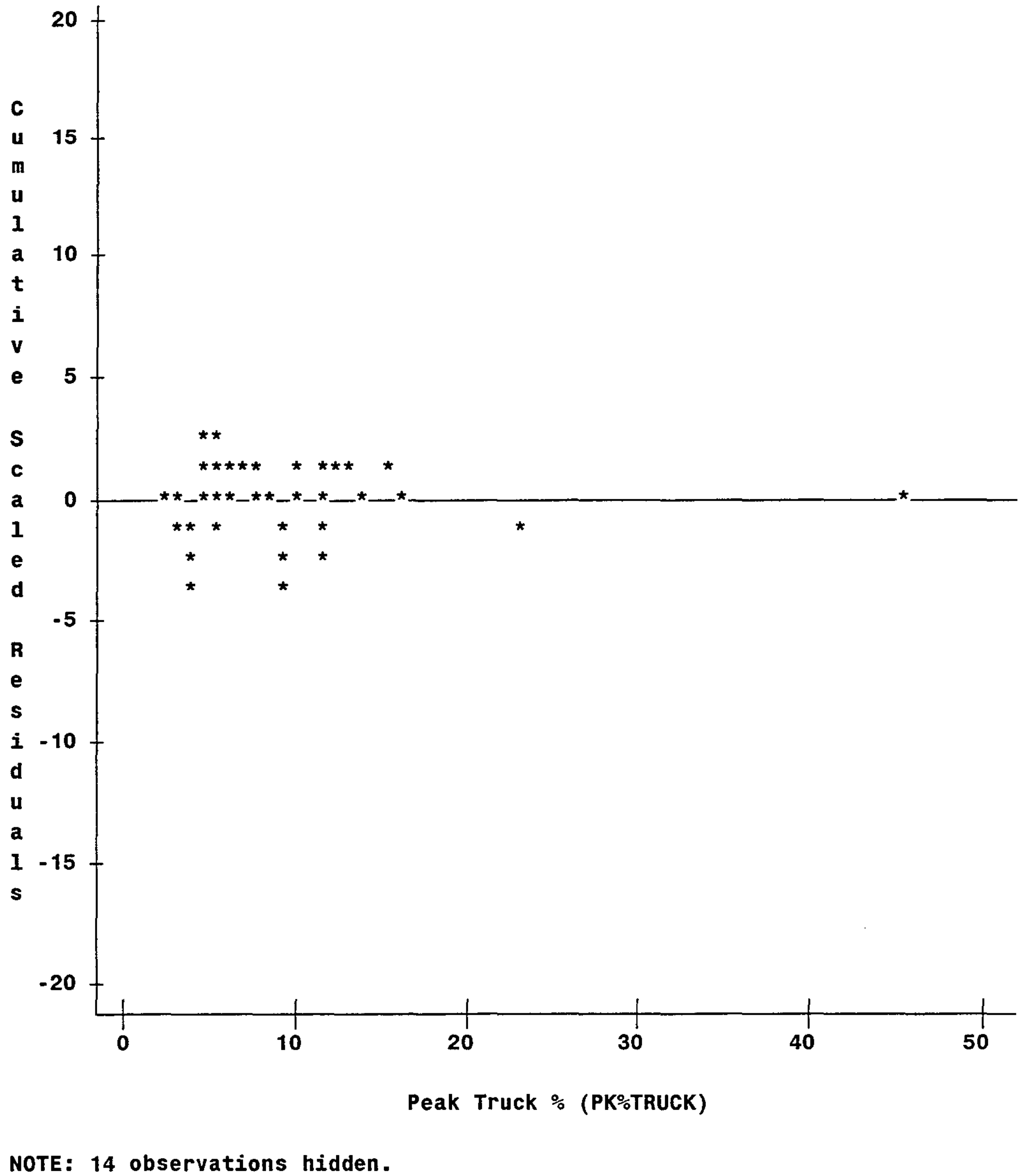


FIGURE 22. Cumulative Scaled Residual Versus PK%TRUCK for Signalized Intersections, TOTACCI Main Model of Table 36

The cumulative scaled residual varies from -3.4 to 2.0, ending at -0.2. It is positive for 27 out of 49 intersections.

Cumulative residuals for the TOTACCI Main Model of Table 36 (signalized intersections) are plotted in Figures 18 through 22. The explanatory variables are ADT1, ADT2, PROT_LT, PK%LEFT2, VEICOM, and PK%TRUCK. The variable PROT_LT is not used in the residual analysis since it is categorical. The figures show what appear to be random walks with no particular systematic effects. Indeed, the fact that they stay relatively close to zero suggests that possibly overfitting is occurring.

Table 38 shows the range of values for the cumulative scaled residuals of all variables in the TOTACCI Main Models. The range is quite consistent with the square roots of the sample sizes. For PROT_LT, the sum of the scaled residuals over all signalized intersections without a protected left turn is 0.45, so that the signalized models slightly underpredict crashes on intersections without major road protected left turns. Since there are 28 signalized intersections without protected left turns, the average scaled residual is $0.45/28 \approx 0.016$. The overall cumulative sum being -0.2, it fol-

TABLE 38. Cumulative Scaled Residuals Versus Increasing Value of Intersection Variables, TOTACCI Main Models

	Intersection Variable	Range of Cumulative Scaled Residual
84 three-legged intersections (Table 29 Main Model) $\sqrt{84} \approx 9$	ADT1	-7.2 to +10.2
	ADT2	-4.5 to +7.9
	MEDWIDTH1	-7.2 to +6.2
	NODRWY1	-3.1 to +10.9
72 four-legged intersections (Table 33 Main Model) $\sqrt{72} \approx 8.5$	ADT1	-4.5 to +5.2
	ADT2	-6.1 to +3.5
	PK%LEFT1	-8.4 to +2.6
49 signalized intersections (Table 35 Main Model) $\sqrt{49} = 7$	ADT1	-3.1 to +3.9
	ADT2	-5.4 to +4.2
	PK%LEFT2	-3.2 to +2.9
	VEICOM	-4.6 to +3.2
	PK%TRUCK	-3.4 to +2.0
	PROT_LT	-0.2, +0.45

lows that the sum of the scaled residuals on the intersections where major road protected left turns are present is -0.65, for an average on the latter intersections of -0.65/21 ≈ -0.031. Thus, the model slightly overpredicts on the intersections that have major road protected left turns.

In summary,

- The three-legged Main Model for TOTACCI might be improved by partitioning the intersection variables into smaller ranges and developing models for each range.
- The four-legged Main Model for TOTACCI might be improved by including quadratic dependence on ADT2 or the log of ADT2 and/or PK%LEFT1.
- The signalized Main Model for TOTACCI has well-behaved residuals, possibly an indication of overfitting.

In view of the relatively small sample sizes, the models all behave reasonably well.

A residual analysis was not done for the TOTACC models, although it is believed that it would yield similar results.

6. CONCLUSIONS

In this chapter, we exhibit the Main Models for TOTACC and TOTACCI again. Then, we use these models and the log-likelihood R^2 to decompose the variation in crashes into proportions due to different variables. We also develop Accident Reduction Factors for the models. Finally, we review and summarize ideas in this study.

THE MAIN MODELS

Three-Legged Intersections

I. Three-legged rural intersections of a four-lane major road with stop-controlled two-lane minor road, **TOTACC** Main Model (Table 28)

Negative Binomial Model with K = 0.389

$$\hat{y} = \boldsymbol{\mathit{NUMBER\ OF\ YEARS}} \times (\boldsymbol{\mathit{ADT1}})^{1.148} \times (\boldsymbol{\mathit{ADT2}})^{0.262} \times \mathbf{exp(-12.220)} \times \mathbf{exp(-0.0546} \times \boldsymbol{\mathit{MEDWIDTH1}} + \mathbf{0.0391} \times \boldsymbol{\mathit{NODRWY1}}\mathbf{)}$$

where the variables are:

$\hat{y}$ = predicted mean number of crashes within 250 feet (76 meters) of the intersection center

NUMBER OF YEARS

ADT1 = average two-way major road traffic in vehicles per day

ADT2 = average two-way minor road traffic in vehicles per day

MEDWIDTH1 = the major road median width in feet

NODRWY1 = the number of residential and commercial driveways on the major road within 250 feet (76 meters) of the intersection center.

NOTE: A metric version of this model is obtained by replacing - 0.0546×MEDWIDTH1 above with -0.179× MEDWIDTH1_m, where MEDWIDTH1_m = the major road median width in meters.

II. Three-legged rural intersections of a four-lane major road with stop-controlled two-lane minor road, **TOTACCI** Main Model (Table 29)

Negative Binomial Model with K = 0.512

$$\hat{y} = \textbf{\textit{NUMBER OF YEARS}} \times (\textbf{\textit{ADT1}})^{1.433} \times (\textbf{\textit{ADT2}})^{0.269} \times \mathbf{exp(-15.466)} \times \mathbf{exp(-0.0612}\times\textbf{\textit{MEDWIDTH1}} + \mathbf{0.0560}\times\textbf{\textit{NODRWY1}})$$

where the variables are:

$\hat{y}$ = predicted mean number of intersection-related crashes within 250 feet (76 meters) of the intersection center

NUMBER OF YEARS

ADT1 = average two-way major road traffic in vehicles per day

ADT2 = average two-way minor road traffic in vehicles per day

MEDWIDTH1 = the major road median width in feet

NODRWY1 = the number of residential and commercial driveways on the major road within 250 feet (76 meters) of the intersection center.

NOTE: A metric version of this model is obtained by replacing - 0.0612×MEDWIDTH1 above with -0.201× $MEDWIDTH1_m$, where $MEDWIDTH1_m$ = the major road median width in meters.

Four-Legged Intersections

I. Four-legged rural intersections of a four-lane major road with stop-controlled two-lane minor roads, **TOTACC** Main Model (Table 32)

Negative Binomial Model with K = 0.458

$$\hat{y} = \textit{NUMBER OF YEARS} \times (\textit{ADT1})^{0.850} \times (\textit{ADT2})^{0.329} \times \exp(-9.463) \times \exp(0.110 \times \textit{PK\%LEFT1} - 0.484 \times \textit{LTLN1S})$$

where the variables are:

$\hat{y}$ = predicted mean number of crashes within 250 feet (76 meters) of the intersection center

NUMBER OF YEARS

ADT1 = average two-way major road traffic in vehicles per day

ADT2 = average two-way minor road traffic in vehicles per day

PK%LEFT1 = the percentage of incoming major road traffic during peak hours that turns left

LTLN1S = 0 if the major road has no left-turn lane, 1 if the major road has at least one left-turn lane.

II. Four-legged rural intersections of a four-lane major road with stop-controlled two-lane minor roads, **TOTACCI** Main Model (Table 33)

Negative Binomial Model with K = 0.710

$$\hat{y} = \textit{NUMBER OF YEARS} \times (ADT1)^{0.930} \times (ADT2)^{0.354} \times \exp(-11.110) \times \exp(0.149 \times PK\%LEFT1)$$

where the variables are:

$\hat{y}$ = predicted mean number of intersection-related crashes within 250 feet (76 meters) of the intersection center

NUMBER OF YEARS

ADT1 = average two-way major road traffic in vehicles per day

ADT2 = average two-way minor road traffic in vehicles per day

PK%LEFT1 = the percentage of all incoming major road traffic during peak hours that turns left.

Signalized Intersections

I. Signalized four-legged rural intersections of two-lane roads, **TOTACC** Main Model (Table 35)

Negative Binomial Model with K = 0.116

$$\hat{y} = \textit{NUMBER OF YEARS} \times (ADT1)^{0.620} \times (ADT2)^{0.395} \times \exp(-6.954) \times \exp(-0.0142 \times PK\%LEFT2 + 0.0315 \times PK\%TRUCK) \times \exp(-0.675 \times PROT\_LT + 0.130 \times VEICOM)$$

where the variables are:

$\hat{y}$ = predicted mean number of crashes within 250 feet (76 meters) of the intersection center

NUMBER OF YEARS

ADT1 = average two-way major road traffic in vehicles per day

ADT2 = average two-way minor road traffic in vehicles per day

PK%LEFT2 = the percentage of all incoming minor road traffic during peak hours that turns left

PK%TRUCK = the percentage of all incoming traffic during peak hours that consists of trucks

PROT_LT = 0 if the major road has no protected left turn, 1 if the major road has at least one protected left turn

VEICOM = (1/2) (VEI-1 + VEI-2)

VEI-1 = the sum of absolute percent grade change per 100 feet (30.5 meters) for each vertical curve along the major road, any portion of which is within 800 feet (244 meters) of the intersection center, divided by the number of such curves

VEI-2 = the sum of absolute percent grade change per 100 feet (30.5 meters) for each vertical curve along the minor road, any portion of which is within 800 feet (244 meters) of the intersection center, divided by the number of such curves.

NOTE: A metric version of this model is obtained by replacing 0.130×VEICOM above with

$0.0396 \times VEICOM_m$, where $VEICOM_m = (1/2)(VEI\text{-}1_m + VEI\text{-}2_m)$ and $VEI\text{-}1_m$ and $VEI\text{-}2_m$ are as above, except that units of absolute grade change per 100 meters are used for each vertical curve, any portion of which is within 244 meters of the intersection center.

II. Signalized four-legged rural intersections of two-lane roads, **TOTACCI** Main Model (Table 36)

Negative Binomial Model with K = 0.131

$$\begin{aligned}\hat{y} = \; & \boldsymbol{NUMBER\ OF\ YEARS} \times (\boldsymbol{ADT1})^{0.595} \times (\boldsymbol{ADT2})^{0.294} \times \mathbf{exp(-6.084)} \\ & \times \mathbf{exp(-0.0165} \times \boldsymbol{PK\%LEFT2} + \mathbf{0.0289} \times \boldsymbol{PK\%TRUCK}) \\ & \times \mathbf{exp(-0.471} \times \boldsymbol{PROT\_LT} + \mathbf{0.113} \times \boldsymbol{VEICOM})\end{aligned}$$

where the variables are:

$\hat{y}$ = predicted mean number of intersection-related crashes within 250 feet (76 meters) of the intersection center

NUMBER OF YEARS

ADT1 = average two-way major road traffic in vehicles per day

ADT2 = average two-way minor road traffic in vehicles per day

PK%LEFT2 = the percentage of all incoming minor road traffic during peak hours that turns left

PK%TRUCK = the percentage of all incoming traffic during peak hours that consists of trucks

PROT_LT = 0 if the major road has no protected left turn, 1 if the major road has at least one protected left turn

VEICOM = (1/2) (VEI-1 + VEI-2)

VEI-1 = the sum of absolute percent grade change per 100 feet (30.5 meters) for each vertical curve along the major road, any portion of which is within 800 feet (244 meters) of the intersection center, divided by the number of such curves

VEI-2 = the sum of absolute percent grade change per 100 feet (30.5 meters) for each vertical curve along the minor road, any portion of which is within 800 feet (244 meters) of the intersection center, divided by the number of such curves.

NOTE: A <u>metric</u> version of this model is obtained by replacing 0.113×VEICOM above with 0.0344×$VEICOM_m$, where $VEICOM_m$ = (1/2)(VEI-1$_m$ + VEI-2$_m$) and VEI-1$_m$ and VEI-2$_m$ are as above, except that units of absolute grade change per 100 meters are used for each vertical curve, any portion of which is within 244 meters of the intersection center.

EXPLANATORY VALUE OF MAIN MODELS

A customary way to measure the explanatory value of variables in a model is to note the increment to a goodness-of-fit measure as each variable is added to the model. For Poisson and negative binomial models, as Fridstrøm et al. (1995) have observed, there is inherent randomness in the model that needs no explanation. With respect to the log-likelihood R-squared measure proposed by Fridstrøm et al., negative binomial randomness is represented by 1 - P_D^2 where P_D^2 is as in equation (5.5) of Chapter 5. The contribution of other factors is represented by: (i) R_D^2 for the first variable when a model with that variable present is used, and (ii) the increment in R_D^2 for each additional variable as it is successively added to the model. Recall the definition of R_D^2 in equation (5.4) of Chapter 5. 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.

Tables 39, 40, and 41 and Figures 23, 24, and 25 decompose the variation according to this method for each of the Main Models.

TABLE 39. Explanation of Variation in Total Crashes by Groups of Covariates, Main Three-Legged Intersection Models

3-Legged Intersection Main Models (Tables 28 and 29)	Log-Likelihood Coefficient of Determination (%)	
	TOTACC	TOTACCI
Randomness	44.11	46.26
Exposure (ADT1, ADT2)	18.21	17.31
Design (MEDWIDTH1, NODRWY1)	4.26	5.02
Unexplained	33.42	31.41
TOTAL	100.00	100.00

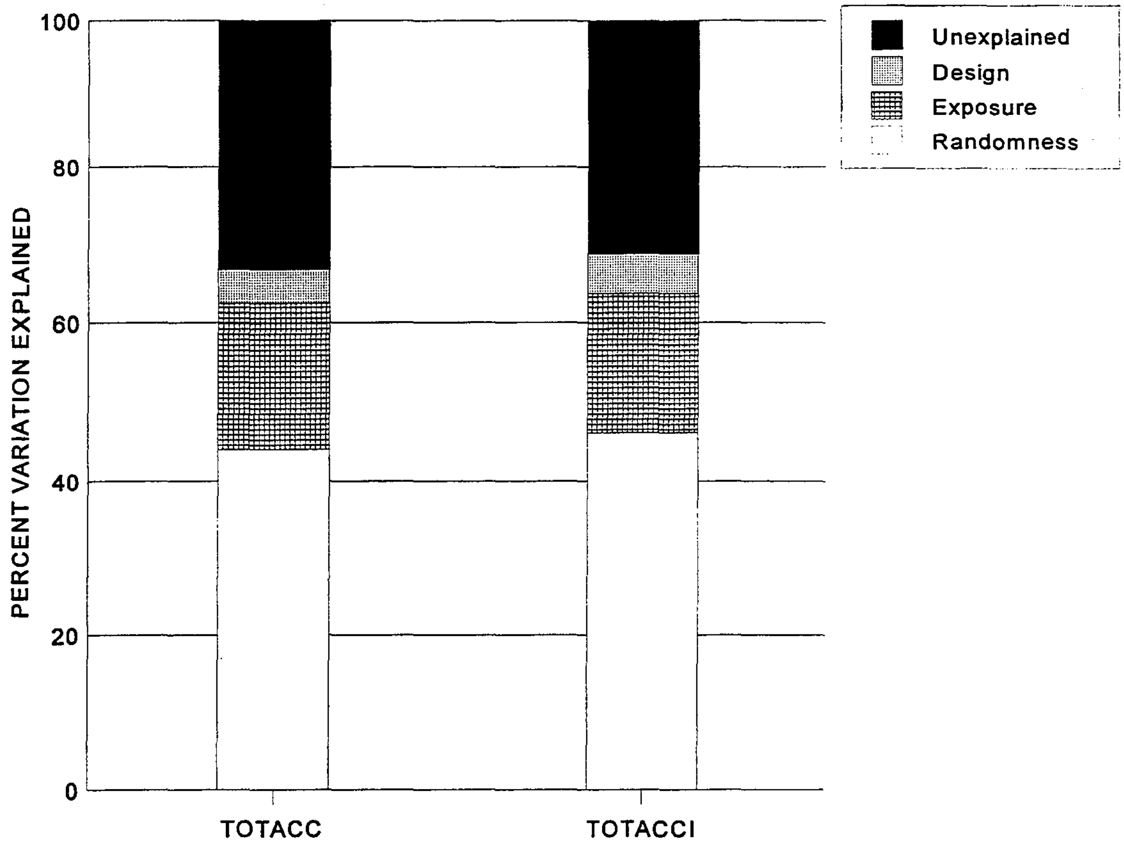


FIGURE 23. Explanation of Variation of TOTACC and TOTACCI by Groups of Covariates, Main Negative Binomial Models for Three-Legged Intersections, Log-Likelihood R-Squared

For the three-legged intersections, ADT explains 17 to 18% of the variation, while MEDWIDTH1 and NODRWY1 explain another 4 to 5%. For the four-legged intersections, ADT explains 8 to 10% of the variation, while major road left-turn percentage and/or the presence of a major road left-turn explains another 5%.

In sharp contrast, for the signalized intersections, ADT by itself explains a negligible percentage of crashes. Turning and truck percentages explain 1 to 3% and the design variables PROT_LT and VEICOM explain between 6 and 13%, depending on the model. As Fridstrøm et al. (1995, p. 11) point out, the explanatory value of a variable may well be affected by the order in which variables are added. This is amply demonstrated by Table 41 and Figure 25. A more cautious interpretation of Table 41 is that in the case of the TOTACC model, 0.34 + 1.46 + 12.99 = 14.79% of the variation is explained by the six intersection variables, and in the case of the TOTACCI model, 0.00 + 3.27

+ 6.16 = 9.43% of the variation is explained by the six intersection variables. Furthermore, the proportion of the explanatory power that is attributable to the individual variables is uncertain. ADT alone does not explain much.

TABLE 40. Explanation of Variation in Total Crashes by Groups of Covariates, Main Four-Legged Intersection Models

4-Legged Intersection Main Models (Tables 32 and 33)	Log-Likelihood Coefficient of Determination (%)	
	TOTACC	TOTACCI
Randomness	41.42	42.85
Exposure (ADT1, ADT2)	10.79	8.14
PK%LEFT1	2.55	5.20
LTLN1S	2.89	-
Unexplained	42.35	43.81
TOTAL	100.00	100.00

TABLE 41. Explanation of Variation in Total Crashes by Groups of Covariates, Main Signalized Intersection Models

Signalized Intersection Main Models (Tables 35 and 36)	Log-Likelihood Coefficient of Determination (%)	
	TOTACC	TOTACCI
Randomness	37.38	41.46
Exposure (ADT1, ADT2)	0.34	0.00
PK%LEFT2, PK%TRUCK1	1.46	3.27
VEICOM, PROT_LT	12.99	6.16
Unexplained	47.83	49.11
TOTAL	100.00	100.00

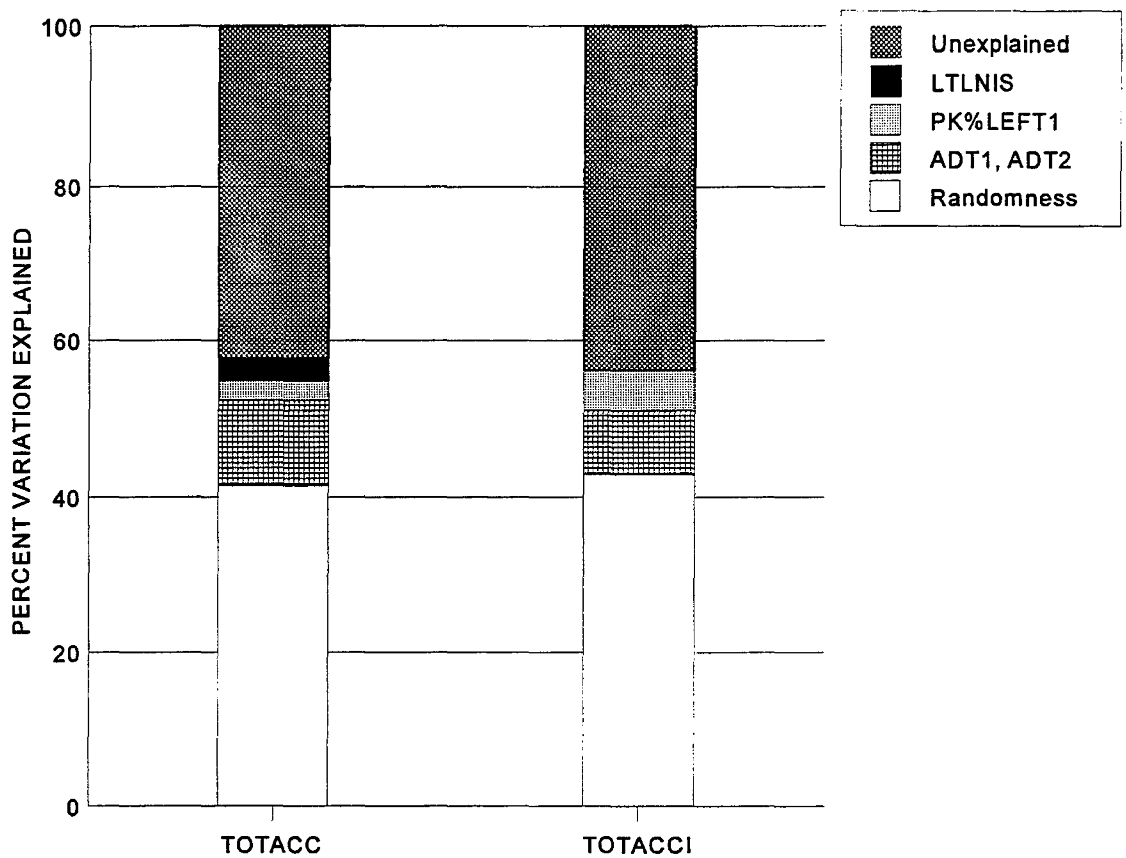


FIGURE 24. Explanation of Variation of TOTACC and TOTACCI by Groups of Covariates, Main Negative Binomial Models for Four-Legged Intersections, Log-Likelihood R-Squared

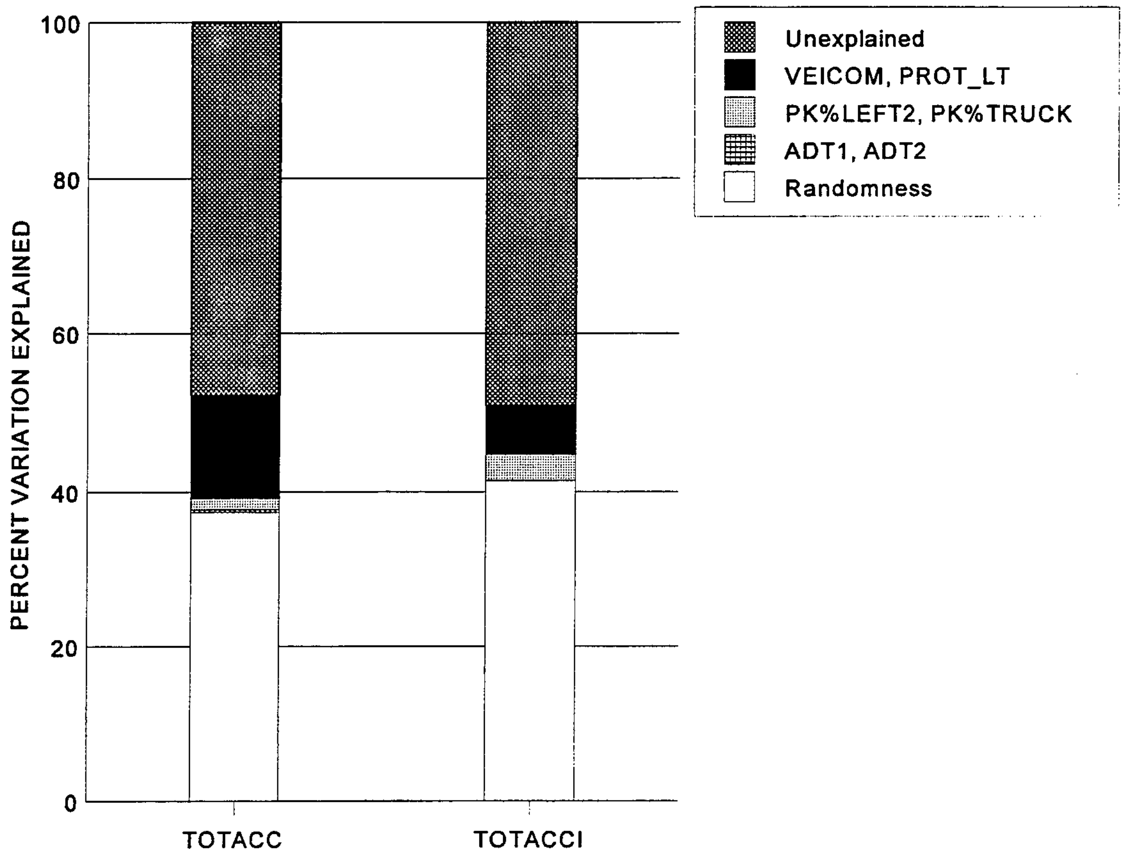


FIGURE 25. Explanation of Variation of TOTACC and TOTACCI by Groups of Covariates, Main Negative Binomial Models for Signalized Intersections, Log-Likelihood R-Squared

ACCIDENT REDUCTION FACTORS

The Main Models yield the Accident Reduction Factors shown in Table 42. Recall that the Accident Reduction Factor is the percentage decrease in mean predicted crash count when a variable is increased by one unit, all other variables being held fixed. A negative value signifies that crashes increase by that percentage when the variable is increased by one unit.

For the three-legged intersections, the TOTACC and TOTACCI models yield similar results. It is a curiosity that the number of driveways is more significant for intersection-related crashes than for all crashes, given that the former attempts to exclude driveway crashes and the latter does not.

TABLE 42. Accident Reduction Factors for the Main Models

3-Legged intersections		
	TOTACC Main Model Table 28	TOTACCI Main Model Table 29
MEDWIDTH1	5.3%	6.6%
NODRWY1	-4.0%	-5.7%
4-Legged intersections		
	TOTACC Main Model Table 32	TOTACCI Main Model Table 33
PK%LEFT1	-11.6%	-16.1%
LTLN1S	38.4%	-
Signalized intersections		
	TOTACC Main Model Table 35	TOTACCI Main Model Table 36
PK%LEFT2	1.4%	1.6%
PK%TRUCK	-3.2%	-2.9%
PROT_LT	49.1%	37.5%
VEICOM	-13.9%	-11.9%

Note: Negative Accident Reduction Factors signify an increase in accidents.

For the four-legged intersections, the TOTACC model declares that the presence of one or more left-turn lanes reduces crashes by 38.4%. LTLN1S had a high P-value (0.3222) when applied to TOTACCI and appears only in the Variant 1 and Variant 3 models of Table 33. In the Variant 1 model, its Accident Reduction Factor is 25.1%, while that of PK%LEFT1 is -15.3%. The number 25.1% is not as large as 38.4%, but is still quite substantial.

Variables in the signalized intersection models show similar Accident Reduction Factors as one passes from TOTACC to TOTACCI. Only PROT_LT shows a dramatic decreases in its effectiveness by going from 49.1% to 37.5%. The two regression coefficients for PROT_LT on which these estimates are based have overlapping confidence intervals so that the difference between 49.1% and 37.5% may be illusory.

The effect of PK%LEFT1 in different models is of some interest. Consider TOTACC models for all three classes of intersections containing this variable, namely, the Variant Model in Table 28, the Main Model in Table 32, and Variant 2 in Table 35. The respective Accident Reduction Factors are -5.6%, -11.6%, and -2.2%. For each 1-percent increase in left turns from the major road, crashes increase by 5.6%, 11.6%, and 2.2% at three-legged, four-legged, and signalized intersections, respectively. A superficial justification of the relative sizes of these numbers runs as follows: at the four-legged intersections, a driver turning left from the major road has to worry about traffic from both minor legs; at a three-legged intersection, the driver has to worry about traffic from only one minor leg; and at a signalized intersection, the driver has to worry about neither minor leg (as long as the signal is green). Even if minor road ADT is low, the presence of minor legs requires some division of attention.

SUMMARY

The Main Models presented at the beginning of this chapter are the primary product of this study. There are six such models, one for each of the three intersection classes and for each of the two crash types TOTACC and TOTACCI. Because our sample sizes were small, we judged it expedient to use all observations for model development and reserve none for prediction, so no efforts have been made to test the predictive powers of the models. The models are, however, reasonably stable: potentially influential observations were removed and the models retained similar coefficients and P-values.

With regard to the two crash types TOTACC and TOTACCI, we do not make a selection. The models for each are reasonably consistent with one another, the variables are mostly the same, and the regression coefficients are similar. For the three-legged and four-legged intersections, the exception is that as one passes from TOTACC to TOTACCI, the intercept gets smaller and the coefficient of the log of ADT1 gets larger. TOTACCI is more sensitive to major road ADT than TOTACC. On the signalized intersections, in the same transition, the intercept gets larger and the coefficients of the logs of both major and minor ADT get smaller. TOTACCI is less sensitive to ADT than TOTACC. These trends are systematic, but not too much weight should be put on them since the standard errors of the coefficients do not preclude the possibility that the true coefficients are equal (but there must be a net adjustment downward somewhere since TOTACCI < TOTACC).

Both the TOTACC models and the TOTACCI models are equally serviceable. A decision on which to use should be based on what they will be used for and how overlapping models will be assembled to represent all crashes. Of some importance is agreement among interested parties as to what an intersection-related crash is. Desirable properties include simplicity, i.e., an understandable definition, and practicality, i.e., one that can be used to extract data from existing or soon-to-exist data bases. The treatment of driveway crashes, run-off-road crashes, and minor road crashes that are not intersection-related should be addressed. Also, a decision is needed about whether the same criteria can be used to define intersection-related crashes for different kinds of intersections: ones with two-lane versus four-lane major roads, ones with or without signals, and ones in urban versus rural environments. The BMI criteria discussed at the beginning of Chapter 4 were used in this study, but they had a limited purpose and scope and their overall applicability should be reassessed.

The same considerations apply to INJACC versus INJACCI models. Their differences and similarities mirror those between the TOTACC and TOTACCI models.

A separate issue is whether injury crash models are needed. A reason not to develop them is that it may suffice to apply a percentage to TOTACC or to TOTACCI in order to estimate INJACC or INJACCI. Tables 9 and 10 in Chapter 4 suggest that injury crashes as a proportion of all accidents vary at least by State and by intersection class. However, the State variable in this study seemed to have no independent influence, and this is a significant finding of our study. Our evidence suggests that serious crashes at three-legged and four-legged intersections are not distributed in the same proportion relative to all crashes at different intersections. Although we do not identify Main Models for INJACC or INJACCI, we do develop INJACC and INJACCI models. It is worthwhile to compare such models with TOTACC/TOTACCI models. For the three-legged intersections, the angle variable HAU assumes prominence and median width loses importance. For the four-legged intersections, minor road posted speed gains influence and channelization loses influence. On the other hand, INJACC/INJACCI models for signalized intersections are similar to the TOTACC/TOTACCI models. Since injury crashes are of greater concern to society and are better reported, contrasts between models for injury crashes and all crashes deserve attention.

We also argue that the variant models shown in the tables of Chapter 5 are worthy of attention. When P-values are large, it is not possible to confirm that the true regression coefficient is non-zero. Thus, an estimated regression coefficient of 0.3 with an estimated standard error of 0.3 could well be a fluctuation for a variable whose true coefficient is zero, the variable thus having no bearing on crash experience. On the other hand, the fluctuation could run in the opposite direction and the true regression coefficient might be 0.6. The estimated coefficient 0.3 summarizes the sample at hand accurately (as does its standard error 0.3) and may be regarded as a point estimate for the true regression coefficient. It is the single best guess as to what that coefficient is. If its standard error is large, there is the possibility that this coefficient might be zero, but the true answer might also be twice as large. If engineering judgment supports the sign and rough magnitude of a regression coefficient, some latitude is in order.

Variant 1 in Table 33 is such a case. The variable LTLN1S, representing the existence of a left-turn lane, has an estimated coefficient of -0.2891 with an estimated standard error of 0.2920 (and a P-value of 0.3222) in a TOTACCI model. This variable is significant in the TOTACC model and is significant in another TOTACCI model, Variant 3 in Table 33.

All of the models are, of course, subject to caveats. The definitions of TOTACC and TOTACCI are imperfect. California and Michigan assign crashes to an intersection out to different distances along the minor road. TOTACCI in California, but not in Michigan, may contain some driveway crashes where a car is entering a driveway. Alignments, sight distances, grades, and median widths are subject to measurement errors, and any and all variables may have changed from the time period 1993-1995 to the time of the field work (1997-1998).

Of special concern, since they are so prominent in the models, are the peak-hour traffic data. We have referred to some of them as turning percentages or peak turning percentages. But they are in fact merely a sample of peak-hour turning percentages collected during a portion of peak hours on a particular day in 1997-1998 and are averaged between morning and evening. They can be regarded as crude estimates of the true average peak-hour turning percentages or truck percentages during 1993-1995. But even the variable one is trying to estimate is somewhat suspect. A peak hour can be defined by a clock definition or by actual experience along a highway. The latter seems more pertinent to crash experience, but the former is presumably closer to what we have.

Yet another issue, one that has not been addressed in this study, is how peak-hour turning percentages should relate to ADT. If one were dealing with true mean turning percentage, it would seem that a model form would be required that yields zero crashes when all turning percentages are zero. As a practical matter, if, for example, major road turning percentages are zero, then we can usually assume that there is zero minor road traffic. Relationships can be built into the model form to ensure that this happens. Since we are dealing with peak-hour turning percentages rather than true mean 24-hour turning percentages, it is possible in principle that the former could be zero without the latter being zero and that the latter could adjust itself to be compatible with almost any observed values of ADT2 or ADT1. Rather than address these thorny issues, we have taken an empirical point of view and allowed interrelated variables, such as the log of ADT1, the log of ADT2, PK%LEFT1, PK%LEFT2, and PK%THRU2, to appear in generalized linear expressions without regard to their hypothetical mutual constraints.

Indeed, especially in the case of the signalized intersections where ADT behaves somewhat peculiarly when other variables are missing, as confirmed in Table 26 and Figure 9 as well as Table 41 and Figure 25, new model forms should be explored that might better describe the data. The limited data in this study suggest that at signalized intersections, some measure of turning percentage (e.g., PROT_LT, PK%LEFT1, PK%LEFT2) should be adjoined to major and minor road ADT as the primary intersection variables. It would also be desirable if new model forms retained some affinity with existing forms that have been adequate for other classes.

One caveat for all of the models is that some variables have rather wide ranges, e.g., NODRWY1, PK%LEFT2, PK%TR. The coefficients assigned to these variables represent their behavior as linear. Over such wide ranges, piecewise linear or quadratic dependencies might be more appropriate. Ezra Hauer has suggested that model forms where the mean number of crashes depends on major road ADT through expressions of the form $(ADT1)^a \times \exp(-b \times ADT1)$ or $\exp(a \times (\text{Log of } ADT1) - b \times ADT1)$, with a and b positive, should be explored. Figures 5 and 9, for three-legged and signalized intersections, respectively, suggest such a possibility. A similar form could be applied to minor road ADT. More elaborate forms could also be considered that allow crash frequency to rise to a maximum as ADT1 increases, with the value of ADT1 at which the maximum occurs depending on ADT2.

We recapitulate the main points below:

- The data in this study have shortcomings. These include relatively small sample sizes, peak turning percentages and truck percentages measured by samples not contemporary with the crash data, and the difficulty of measuring and defining crash and intersection variables.

- In addition to the six Main Models, alternate models deserve consideration. These include variants given in the tables using other variables, the Flow Models in Chapter 5, models that restrict the range of certain inputs (piecewise linear) or allow quadratic dependencies, and model forms suggested by Hauer.

- Major road ADT plays a lesser role as one passes from three-legged to four-legged to signalized intersections, with turning percentage measures becoming more important, and unexplained crash frequency variation increasing (Figures 23, 24, and 25).

- The six Main Models adequately summarize the data in this study, with the choice of a crash variable TOTACC (all crashes within 250 feet (76 meters)) or TOTACCI (all intersection-related crashes within 250 feet (76 meters)) to be determined by other criteria.

APPENDIX. DATA FROM PILOT STUDY PHASE OF DATA COLLECTION

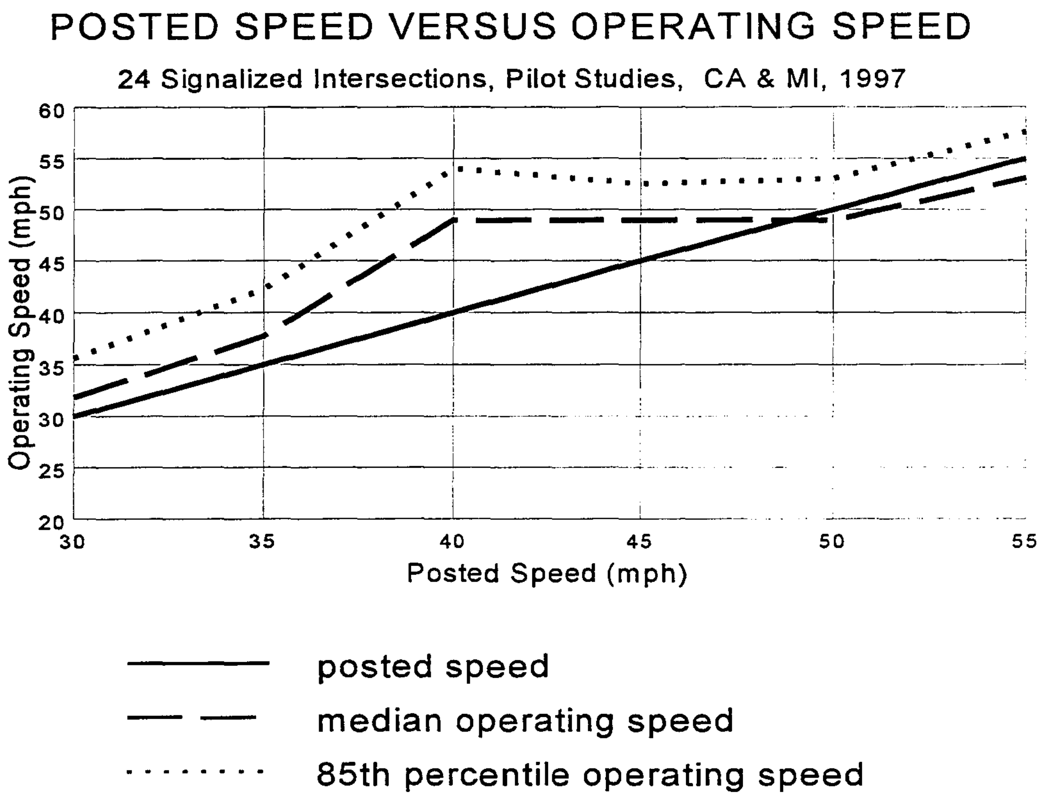
During the pilot study phase of data collection for this report, in March and May 1997, plates were used to collect traffic data on minor legs of signalized intersections in Michigan and radar guns were used to measure operating speeds on all legs of intersections in both California and Michigan. Figures A-1, A-2, and A-3 exhibit some relationships obtained from these data. In addition an area-of-influence study was done on a few selected intersections to judge whether crashes near the intersection were intersection-related. Figure A-4 shows the findings for one such intersection.

Figure A-1 is a graph of posted speed versus observed operating speed at signalized intersections. Operating speeds were determined by radar guns aimed along the road toward the intersection during daytime hours out of sight of the intersection or far enough away so that drivers typically had not begun to slow. The graph shows that many drivers exceed the posted speed limit, but that the excess tends to be less at low and high speeds and greater at intermediate speeds.

Figure A-2 is a graph of daytime speeds versus 24-hour speeds along minor legs approaching Michigan signalized intersections. Daytime speeds were measured by radar guns, and 24-hour speeds by HISTAR/NU-METRICS plate counts. It is apparent that the 24-hour speeds are lower, although some of the extreme cases may represent miscalibration of the radar guns and/or the plates.

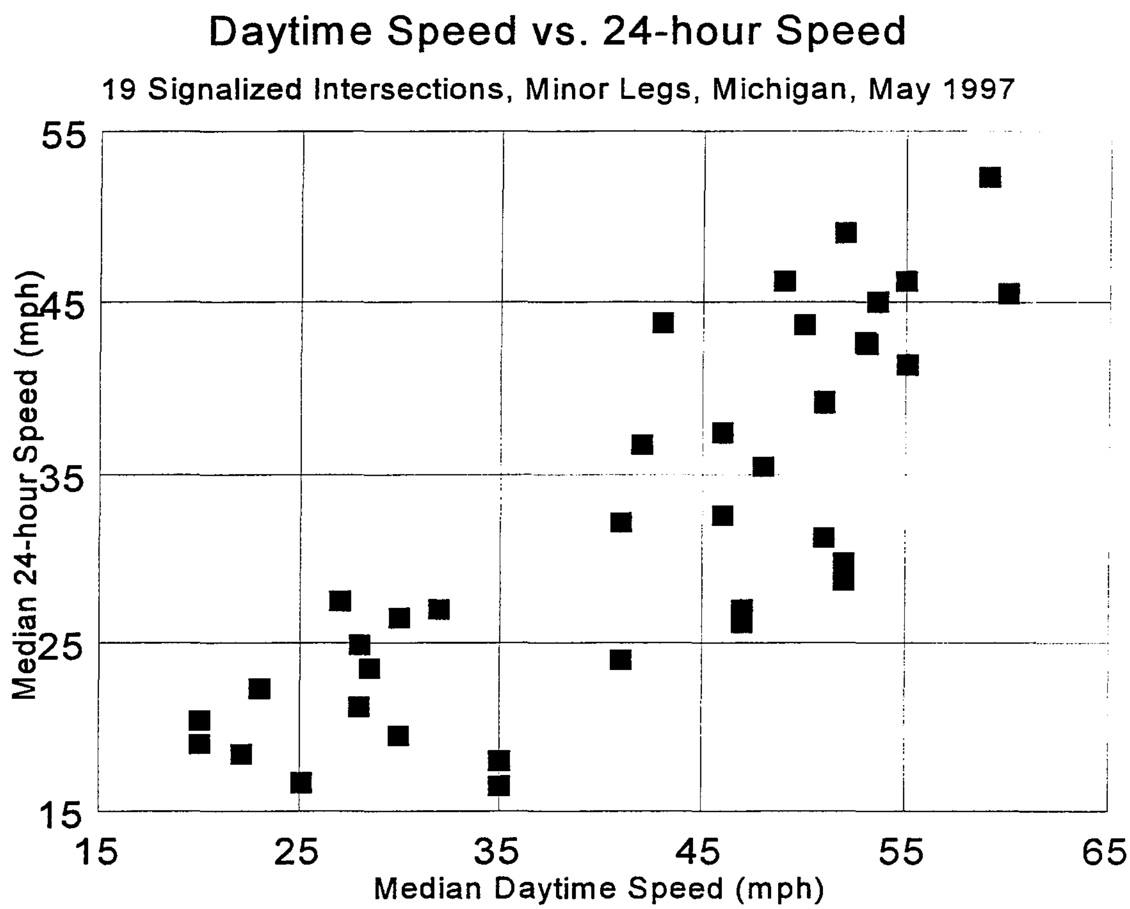
Figure A-3 shows that truck percentage in off-hours tends to be higher than in peak hours. At the end of Chapter 4, it is noted that a.m. truck percentage is higher than p.m. truck percentage as well, and that Miaou et al. (1988) have called for studies of truck percentage by time-of-day. Between a.m. and p.m., a rough reversal of movements was found for all traffic (e.g., southbound predominance in a.m., northbound in p.m.) although variances were large. Truck percentage is a small portion of the total during peak hours, and may be larger in off-hours, chiefly because noncommercial traffic lessens.

A few intersections in this study were examined in detail, in an effort to analyze the area of influence of an intersection, i.e., how far out from the intersection center intersection-related crashes are likely to be found. For this purpose, all crashes within 500 feet (152 meters) of the intersection center were examined. Figure A-4 shows crash locations for one such intersection. A distance of 250 feet (76 meters) from the intersection center includes most intersection-related crashes, misses a few, and picks up a few that are not intersection-related. Crashes that are not intersection-related are more likely to be found on the outward bound lanes from the intersection center. One State highway engineer reported intersection-related crashes that occurred on roads that did not pass through the intersection. During heavy traffic, a driver turning onto an intersection leg from a side road is sometimes involved in a crash related to the main intersection.



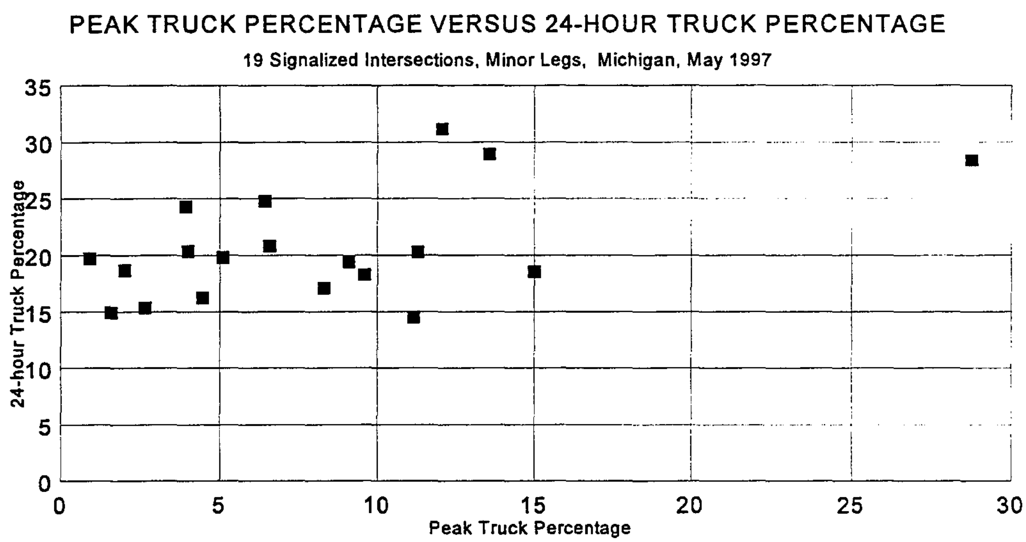
Operating speeds were measured during daytime hours
by radar guns well away from the intersection.
1 mph = 1.61 km/h

FIGURE A-1. Posted Speed Versus Operating Speed



Daytime speeds were measured by radar guns.
24-hour speeds were measured by HISTAR/NU-METRICS plate counts.
1 mph = 1.61 km/h

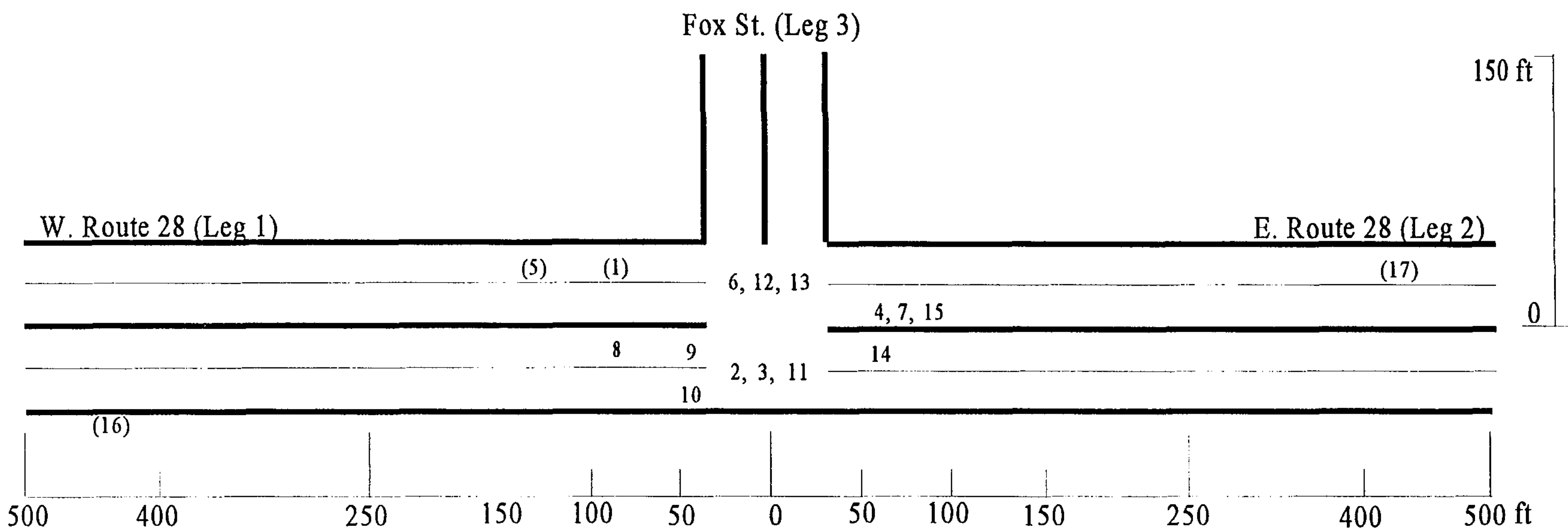
FIGURE A-2. Daytime Speed Versus 24-Hour Speed



Peak Truck Percentages were measured by daytime manual count.
24-hour Truck Percentages were measured by HISTAR/NU-METRICS plate counts.

FIGURE A-3. Peak Truck Percentage Versus 24-Hour Truck Percentage

CA State Route 28 intersects Fox Street in Placer County. This intersection (cnty_rte = "03028 31Z", milepost = 10.025) has minor leg stop control, is of the T type, four-lane by two-lane, with a right angle, and no medians on any leg. The intersection is in rolling terrain with a HAZRAT equal to 2. The longitudinal sight distance for leg 1 is 800 feet (244 meters). Although the intersection is defined as "rural," it is in the Lake Tahoe resort area with 12 commercial driveways along legs 1 and 2 within ±250 feet (76 meters) of the intersection center. This is a high-crash intersection with 17 crashes occurring within ±500 feet (152 meters) of the intersection center during the years 1993-1995. On the basis of review of HSIS files, the crashes with numbers in parentheses were deemed not to be intersection-related.



Crash locations are numbered; those with parentheses
are not thought to be intersection-related.
1 ft = 0.305 meters

FIGURE A-4. Crash Locations and Relationships at a Three-Legged Intersection

REFERENCES

American Association of State Highway and Transportation Officials, *A Policy on Geometric Design of Highways and Intersections*, Washington, D.C., 1994.

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

K.M. Bauer and D. Harwood, *Statistical Models of At-Grade Intersection Accidents*, Report No. FHWA-RD-96-125, Federal Highway Administration, McLean, Va., 1996.

D. Blower, K.L. Campbell, and P.E. Green, "Accident Rates for Heavy Truck-Tractors in Michigan," *Accident Analysis and Prevention* 25 (3): 307-321, 1993.

J.A. Bonneson and P.T. McCoy, "Estimation of Safety at Two-Way Stop-Controlled Intersections on Rural Highways," *Transportation Research Record* 1401: 83-89, 1993.

U. Brüde and J. Larsson, "Models for Predicting Accidents at Junctions Where Pedestrians and Cyclists Are Involved: How Well Do They Fit?" *Accident Analysis and Prevention* 25 (5): 499-509, 1993.

N.A. David and J.R. Norman, *Motor Vehicle Accidents in Relation to Geometric and Traffic Features of Highway Intersections, Volume II - Research Report*, Report No. FHWA-RD-76-129, Federal Highway Administration and National Highway Traffic Safety Administration , Washington, D.C., 1975.

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

R. Elvik, "Meta-Analysis of Evaluations of Public Lighting as Accident Countermeasures," *Transportation Research Record* 1485: 112-123, 1995.

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

N.J. Garber and R. Srinivasan, "Risk Assessment of Elderly Drivers at Intersections: Statistical Modeling," *Transportation Research Record* 1325: 17-22, 1991.

A.S. Hakkert and D. Mahalel, "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.

J.T. Hanna, T.E. Flynn, and W.L. Tyler, "Characteristics of Intersection Accidents in Rural Municipalities," *Transportation Research Record* 601: 79-82, 1976.

D.W. Harwood, M.T. Pietrucha, M.D. Woolridge, R.E. Brydia, and K. Fitzpatrick, *Median Intersection Design*, National Cooperative Highway Research Program Report 375, Transportation Research Board, National Research Council, National Academy Press, Washington, D.C., 1995.

E. Hauer, *Observational Before-After Studies in Road Safety*, Pergamon Press, Oxford, U.K., 1997.

E. Hauer and A.S. Hakkert, "Extent and Some Implications of Incomplete Accident Reporting," *Transportation Research Record* 1185: 1-10, 1988.

E. Hauer, J.C.N. Ng, and J. Lovell, "Estimation of Safety at Signalized Intersections," *Transportation Research Record* 1185: 48-61, 1988.

E. Hauer, E. Pagitsas, and B.T. Shin, "Estimation of Turning Flows From Automatic Counts," *Transportation Research Record* 795: 1-7, 1981.

E. Hauer and B. Persaud, *Safety Analysis of Roadway Geometry and Ancillary Features*, Transportation Association of Canada, Ottawa, 1996.

E. Hauer, D. Terry, and M.S. Griffith, "Effect of Resurfacing on Safety of Two-Lane Rural Roads in New York State," *Transportation Research Record* 1467: 30-37, 1994.

H.C. Joksch and L.P. Kostyniuk, *Modeling Intersection Crash Counts and Traffic Volume*, Report No. FHWA-RD-98-096, Federal Highway Administration, McLean,Va., 1998.

G.F. King and R.B. Goldblatt, "Relationship of Accident Patterns to Type of Intersection Control," *Transportation Research Record* 540: 1-12, 1975.

M.W. Knuiman, F.M. Council, and D.W. Reinfurt, "Association of Median Width and Highway Accident Rates," *Transportation Research Record* 1401: 70-82, 1993.

S.R. Kuciemba and J.A. Cirillo, *Safety Effectiveness of Highway Design Features, Volume V - Intersections*, Report No. FHWA-RD-91-048, Federal Highway Administration, Washington, D.C., 1992.

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

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

H. Lum and J. Reagan, *Interactive Highway Safety Design Model: Accident Predictive Module*, FHWA Draft 8-22-94.

G. Maycock and R.D. Hall, *Accidents at 4-Arm Roundabouts*, Laboratory Report 1120, Transport and Road Research Laboratory, Department of Transport, Crowthorne, Berkshire, United Kingdom, 1984.

P.T. McCoy, W.J. Hoppe, and D.V. Dvorak, "Benefit-Cost Evaluation of Left-Turn Lanes on Uncontrolled Approaches of Rural Intersections (Abridgement)," *Transportation Research Record* 1025: 40-43, 1985.

P.T. McCoy, E.J. Tripi, and J.A. Bonneson, *Guidelines for Realignment of Skewed Intersections*, Nebraska Department of Roads Research Project Number RES1 (0099) P471, 1994.

J.W. McDonald, "Relation Between Number of Accidents and Traffic Volume at Divided- Highway Intersections," *Highway Research Board Bulletin 74, Traffic-Accident Studies*, pp. 7-17, National Academy of Sciences, National Research Council, Washington, D.C., 1953.

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

S.-P. Miaou, "The Relationship Between Truck Accidents and Geometric Design of Road Sections: Poisson Versus Negative Binomial Regressions," *Accident Analysis and Prevention* 26 (4): 471-482, 1994.

S.-P. Miaou, *Measuring the Goodness-of-Fit of Accident Prediction Models*, Report No. FHWA-RD-96-040, McLean, Va., 1996.

S.-P. Miaou, P.S. Hu, T. Wright, S.C. Davis, and A.K. Rathi, *Development of Relationship Between Truck Accidents and Geometric Design: Phase I*, Report No. FHWA-RD-91-124, Federal Highway Administration, McLean, Va., 1993.

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

Midwest Research Institute, *Critical Reviews of Intersection Safety Studies Task K Resource Paper*, MRI Report, MRI Project No. 4584-09, Kansas City, Mo., 1997.

J.A. Nelder and R.W. Wedderburn, "Generalized Linear Models," *Journal of the Royal Statistical Society*, Series A, 135 (3): 370-384, 1972.

O. Pendleton, *Evaluation of Accident Methodology*, Report No. FHWA-RD-96-039, McLean, Va., 1996.

D. Pickering, R.D. Hall, and M. Grimmer, *Accidents at Rural T-Junctions*, Research Report 65, Transport and Road Research Laboratory, Department of Transport, Crowthorne, Berkshire, United Kingdom, 1986.

M. Poch and F. Mannering, "Negative Binomial Analysis of Intersection-Accident Frequencies," *Journal of Transportation Engineering* 122 (2): 105-113, 1996.

J.A. Reagan, "The Interactive Highway Safety Design Model: Designing for Safety by Analyzing Road Geometrics," *Public Roads* 37-43, Summer 1994.

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

J.M. Tessmer, *Rural and Urban Crashes: A Comparative Analysis*, Report No. DOT-HS-808-450, U.S. Department of Transportation, National Highway Traffic Safety Administration Technical Report, Washington, D.C., 1996.

A. Vogt and J.G. Bared, *Accident Models for Two-Lane Rural Roads: Segments and Intersections*, Report No. FHWA-RD-98-133, Federal Highway Administration, McLean, Va., 1998.

A. Vogt and J.G. Bared, "Accident Models for Two-Lane Rural Segments and Intersections," *Transportation Research Record* 1635: 18-29, 1998.

G.M. Webb, "The Relation Between Accidents and Traffic Volumes at Signalized Intersections," *Institute of Transportation Engineers Proceedings*, Technical Session No. 3B, pp. 149-167, 1955.

C.V. Zegeer, J. Hummer, L. Herf, D. Reinfurt, and W. Hunter, *Safety Cost-Effectiveness of Incremental Changes in Cross-Section Design — Informational Guide*, Report No. FHWA-RD-87-094, McLean, Va., 1987.