U.S. Department
of Transportation
**Federal Highway
Administration**

Statistical Models of At-Grade Intersection Accidents

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

Publication No. FHWA-RD-96-125
November 1996

FOREWORD

This report is a preliminary step toward the accomplishment of the Interactive Highway Safety Design Model (IHSDM), which is a Federal Highway Administration (FHWA) objective to develop a highway safety evaluation tool. In particular, this research is a preliminary effort to relate multi-vehicle accidents of at-grade intersections to highway design elements. The results are of interest to those concerned with highway safety research for use by planners and designers. Furthermore, the results are useful to researchers who will eventually improve, validate, and finalize these intersection accident models.

Based on retrospective analysis, several statistical modeling techniques were tried. Besides using statistical techniques (such as regression models, discriminant analysis, cluster analysis, etc.), hard-copy accident reports were reviewed to determine the impact of design elements on accidents. Finally, five preliminary accident models were developed for at-grade intersections: (1) Rural, four-leg, stop-controlled; (2) Rural, three-leg, stop-controlled; (3) Urban, four-leg, stop-controlled; (4) Urban, three-leg, stop-controlled; and (5) Urban, four-leg, signalized.

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

NOTICE

Technical Report Documentation Page

1. Report No. FHWA-RD-96-125	2. Government Accession No.	3. Recipient's Catalog No.

4. Title and Subtitle Statistical Models of At-grade Intersection Accidents	5. Report Date November 1996
	6. Performing Organization Report No. 3195-0102
7. Author(s) K. M. Bauer and D. W. Harwood	
9. Performing Organization Name and Address Midwest Research Institute 425 Volker Boulevard Kansas City, Missouri 64110-2299	10. Work Unit No. (TRAIS) 3A5A
	11. Contract or Grant No. DTFH61-92-C-00031
12. Sponsoring Agency Name and Address Federal Highway Administration Office of Safety and Traffic Operations Research and Development 6300 Georgetown Pike McLean, Virginia 22101-2296	13. Type of Report and Period Covered Final Technical Report February 1993-June 1996
	14. Sponsoring Agency Code

15. Supplementary Notes

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

16. Abstract

The objective of this research was to develop statistical models of the relationship between traffic accidents and highway geometric elements for at-grade intersections. These models also incorporated the effect of traffic control features and traffic volumes on intersection accidents. The data base used to develop the models was obtained from the California Department of Transportation. Field data were also collected for a sample of urban, four-leg, signalized intersections to provide data on additional geometric design variables and turning-movement counts that were not available from existing highway agency files.

The statistical modeling approaches used in the research included Poisson, lognormal, negative binomial, and logistic regression, as well as discriminant and cluster analysis. Regression models of the relationships between accidents and intersection geometric design, traffic control, and traffic volume variables were found to explain between 16 and 38 percent of the variability in the accident data. However, most of that variability was explained by the traffic volume variables considered; geometric design variables accounted for only a very small additional portion of the variability.

An evaluation of hard-copy police accident reports by three independent reviewers for a sample of eight urban, four-leg, signalized intersections found that only 5 to 14 percent of the accidents had causes that appeared to be related to geometric design features of the intersections.

17. Key Words Accident Modeling Traffic Accidents Geometric Design At-grade Intersections	Poisson Regression Negative Binomial Regression Lognormal Regression	18. Distribution Statement No restrictions. This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161.

19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 157	22. Price

Form DOT F 1700.7 (8-72)

Reproduction of completed page authorized

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		
in²	square inches	645.2	square millimeters	mm²
ft²	square feet	0.093	square meters	m²
yd²	square yards	0.836	square meters	m²
ac	acres	0.405	hectares	ha
mi²	square miles	2.59	square kilometers	km²
		VOLUME		
fl oz	fluid ounces	29.57	milliliters	mL
gal	gallons	3.785	liters	L
ft³	cubic feet	0.028	cubic meters	m³
yd³	cubic yards	0.765	cubic meters	m³
NOTE: Volumes greater than 1000 l shall be shown in m³.				
		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/m²	cd/m²
		FORCE and PRESSURE or STRESS		
lbf	poundforce	4.45	newtons	N
lbf/in²	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		
mm²	square millimeters	0.0016	square inches	in²
m²	square meters	10.764	square feet	ft²
m²	square meters	1.195	square yards	yd²
ha	hectares	2.47	acres	ac
km²	square kilometers	0.386	square miles	mi²
		VOLUME		
mL	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m³	cubic meters	35.71	cubic feet	ft³
m³	cubic meters	1.307	cubic yards	yd³
		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/m²	candela/m²	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/in²

* 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

1. INTRODUCTION 1
   Research Overview 1
   Scope and Organization of This Report 2

2. LITERATURE OVERVIEW 5

3. DATA BASE SELECTION 9

4. DATA BASE DESCRIPTION 11
   Variables Available in the Caltrans Data Base 11
   Additional Variables for Which Data Were Obtained in Field Studies . . . 11

5. STATISTICAL MODELING 17
   Data Preparation 17
   Accident Frequency Distributions 22
   Lognormal and Loglinear Regression Models 31
   Accident Modeling Results 37

6. ALTERNATIVE STATISTICAL MODELING APPROACHES 95
   Introduction 95
   Logistic Regression Analysis 96
   Accident Analyses for Specific ADT Classes 107
   Discriminant Analysis 110
   Cluster Analysis 113

7. REVIEW OF HARD COPY POLICE ACCIDENT REPORTS 117
   Hard Copy Police Accident Reports 117
   Objectives of Hard Copy Accident Report Review 117
   Hard Copy Accident Report Review Approach 118
   Hard Copy Police Accident Report Review Results 121
   Findings and Recommendations 128

8. CONCLUSIONS 131

Appendixes
   A DEFINITIONS OF GEOMETRIC DESIGN, TRAFFIC CONTROL, AND TRAFFIC VOLUME VARIABLES 135
   B TOTAL MULTIPLE-VEHICLE ACCIDENT DATA DISTRIBUTIONS . . 141

REFERENCES 149

List of Figures

1. Accident Frequency Distributions at Rural, Four-leg, STOP-controlled Intersections 24
2. Accident Frequency Distributions at Rural, Three-leg, STOP-controlled Intersections 25
3. Accident Frequency Distributions at Urban, Four-leg, STOP-controlled Intersections 26
4. Accident Frequency Distributions at Urban, Three-leg, STOP-controlled Intersections 27
5. Accident Frequency Distributions at Urban, Four-leg, Signalized Intersections . 28
6. Accident Frequency Distributions at a Sample of 198 Urban, Four-leg, Signalized Intersections 30
7. Number of Multiple-Vehicle Accidents per Year as a Function of Traffic Volumes for Typical Rural, Four-leg, STOP-Controlled Intersections 48
8. Number of Multiple-Vehicle Accidents per Year as a Function of Traffic Volumes for Typical Rural, Three-leg, STOP-Controlled Intersections 54
9. Number of Multiple-Vehicle Accidents per Year as a Function of Traffic Volumes for Typical Urban, Four-leg, STOP-Controlled Intersections 61
10. Number of Multiple-Vehicle Accidents per Year as a Function of Traffic Volumes for Typical Urban, Three-leg, STOP-Controlled Intersections 67
11. Number of Multiple-Vehicle Accidents per Year as a Function of Traffic Volumes for Typical Urban, Four-leg, Signalized Intersections 75
12. Cluster Analysis Results for a Sample of 130 Urban, Four-leg, Signalized Intersections 115

List of Tables

1. Variables Available in the Existing Caltrans Data Base 12
2. Accident History Variables Derived from Caltrans Accident File 13
3. Variables Collected in Pilot Field Study of Urban, Four-leg, Signalized Intersections 15
4. Intersection Distribution in Caltrans Data Base by Type of Intersection 18
5. Accident Data Distributions in Caltrans Data Base 20
6. Annual Accident Statistics, 1990-1992 23
7. Annual Accident Statistics for Sample of 198 Urban, Four-leg, Signalized Intersections 29
8. Descriptive Statistics for Rural, Four-leg, STOP-controlled Intersections 38
9. Model Diagnostics for Total and Fatal and Injury Multiple-vehicle Accidents at Rural, Four-leg, STOP-controlled Intersections 41
10. Negative Binomial Regression Results for Total Multiple-vehicle Accidents at Rural, Four-leg, STOP-controlled Intersections 44
11. Negative Binomial Regression Results for Fatal and Injury Multiple-vehicle Accidents at Rural, Four-leg, STOP-controlled Intersections 45
12. Descriptive Statistics for Rural, Three-leg, STOP-controlled Intersections 49
13. Model Diagnostics for Total and Fatal and Injury Multiple-vehicle Accidents at Rural, Three-leg, STOP-controlled Intersections 50
14. Negative Binomial Regression Results for Total Multiple-vehicle Accidents at Rural, Three-leg, STOP-controlled Intersections 52
15. Negative Binomial Regression Results for Fatal and Injury Multiple-vehicle Accidents at Rural, Three-leg, STOP-controlled Intersections 53
16. Descriptive Statistics for Urban, Four-Leg, STOP-controlled Intersections 56
17. Model Diagnostics for Total and Fatal and Injury Multiple-vehicle Accidents at Urban, Four-leg, STOP-controlled Intersections 57
18. Lognormal Regression Results for Total Multiple-vehicle Accidents at Urban, Four-leg, STOP-controlled Intersections 59
19. Lognormal Regression Results for Fatal and Injury Multiple-vehicle Accidents at Urban, Four-leg, STOP-controlled Intersections 60
20. Descriptive Statistics for Urban, Three-Leg, STOP-controlled Intersections 62
21. Model Diagnostics for Total and Fatal and Injury Multiple-vehicle Accidents at Urban, Three-leg, STOP-controlled Intersections 63
22. Negative Binomial Regression Results for Total Multiple-vehicle Accidents at Urban, Three-leg, STOP-controlled Intersections 65
23. Negative Binomial Regression Results for Fatal and Injury Multiple-vehicle Accidents at Urban, Three-leg, STOP-controlled Intersections 66
24. Descriptive Statistics for Urban, Four-Leg, Signalized Intersections 69
25. Model Diagnostics for Total and Fatal and Injury Multiple-vehicle Accidents at Urban, Four-leg, Signalized Intersections 72
26. Lognormal Regression Results for Total Multiple-vehicle Accidents at Urban, Four-leg, Signalized Intersections 73

List of Tables (Continued)

27. Lognormal Regression Results for Fatal and Injury Multiple-vehicle Accidents at Urban, Four-leg, Signalized Intersections 74
28. Descriptive Statistics for Sample of 198 Urban, Four-Leg, Signalized Intersections 78
29. Descriptive Statistics of New Field Variables for Sample of 198 Urban, Four-leg, Signalized Intersections 81
30. Model Diagnostics for Total and Fatal and Injury Multiple-vehicle Accidents at a Sample of 198 Urban, Four-leg, Signalized Intersections 82
31. Negative Binomial Regression Results for Total Multiple-vehicle Accidents at Sample of 198 Urban, Four-leg, Signalized Intersections Using Caltrans Geometric Variables 84
32. Negative Binomial Regression Results for Total Multiple-vehicle Accidents at Sample of 198 Urban, Four-leg, Signalized Intersections Using Updated Values of Original Geometric Variables 85
33. Negative Binomial Regression Results for Total Multiple-vehicle Accidents at Sample of 198 Urban, Four-leg, Signalized Intersections Using New Field Geometric Variables 86
34. Negative Binomial Regression Results for Fatal and Injury Multiple-vehicle Accidents at Sample of 198 Urban, Four-leg, Signalized Intersections Using Caltrans Geometric Variables 87
35. Negative Binomial Regression Results for Fatal and Injury Multiple-vehicle Accidents at Sample of 198 Urban, Four-leg, Signalized Intersections Using Updated Values of Original Geometric Variables 88
36. Negative Binomial Regression Results for Fatal and Injury Multiple-vehicle Accidents at Sample of 198 Urban, Four-leg, Signalized Intersections Using New Field Geometric Variables 89
37. Left-turn Accident Frequencies at a Sample of 198 Urban, Four-leg, Signalized Intersections 91
38. Logistic Regression Results for Total Multiple-vehicle Accidents at Rural, Four-leg, STOP-controlled Intersections 99
39. Logistic Regression Results for Fatal and Injury Multiple-vehicle Accidents at Rural, Four-leg, STOP-controlled Intersections 100
40. Logistic Regression Results for Total Multiple-vehicle Accidents at Rural, Three-leg, STOP-controlled Intersections 101
41. Logistic Regression Results for Fatal and Injury Multiple-vehicle Accidents at Rural, Three-leg, STOP-controlled Intersections 102
42. Logistic Regression Results for Total Multiple-vehicle Accidents at Urban, Four-leg, STOP-controlled Intersections 103
43. Logistic Regression Results for Fatal and Injury Multiple-vehicle Accidents at Urban, Four-leg, STOP-controlled Intersections 104
44. Logistic Regression Results for Total Multiple-vehicle Accidents at Urban, Three-leg, STOP-controlled Intersections 105

List of Tables (Continued)

45. Logistic Regression Results for Fatal and Injury Multiple-vehicle Accidents at Urban, Three-leg, STOP-controlled Intersections 106
46. Major-road and Crossroad ADT Classes Used in Modeling Intersection Accidents 109
47. Cross-validation Summary Results from Discriminant Analyses 112
48. Stepwise Discriminant Analysis Results (0 or low vs high numbers of accidents) 113
49. Characteristics of Intersections Selected for Hard Copy Police Accident Report Review 119
50. Reviewers' Ratings of Number of Accidents Related to Each Intersection 121
51. Reviewers' Ratings of Number of Accidents Related to the Operation of the Traffic Signal 123
52. Reviewers' Ratings of Number of Accidents Related to Geometric Features of the Intersection 124
53. Reviewers' Ratings of Number of Accidents for Which Geometric Features Had a Role in Increasing Accident Severity 126
54. Reviewers' Ratings of Number of Accidents in Which Driver, Vehicle, and Roadway and Environmental Factors Had a Role 127
55. Definitions of Variables Available in the Existing Caltrans Data Base 136
56. Definitions of Additional Variables Obtained in Pilot Field Study 138
57. Total Multiple-vehicle Accident Data Distribution at Rural, Four-leg, STOP-controlled Intersections 141
58. Total Multiple-vehicle Accident Data Distribution at Rural, Three-leg, STOP-controlled Intersections 142
59. Total Multiple-vehicle Accident Data Distribution at Urban, Four-leg, STOP-controlled Intersections 143
60. Total Multiple-vehicle Accident Data Distribution at Urban, Three-leg, STOP-controlled Intersections 144
61. Total Multiple-vehicle Accident Data Distribution at Urban, Four-leg, Signalized Intersections 145
62. Total Multiple-vehicle Accident Data Distribution at a Sample of 198 Urban, Four-leg, Signalized Intersections 147

1. INTRODUCTION

This technical report presents the results of statistical analyses of multiple-vehicle accident experience for at-grade intersections. The objectives and scope of this research and the organization of this report are discussed below.

Research Overview

The objective of this research study was to develop statistical models for defining the relationships between traffic accidents and highway geometric elements for at-grade intersections. These models also incorporated the effects of traffic control features and traffic volumes on intersection accidents. It was hoped that these models could be used in predicting the effects on accidents of specific geometric design decisions at intersections.

Several major technical tasks were performed during the research, including:

- A review of previously published and unpublished literature and ongoing studies concerning the relationship between traffic accidents and intersection geometrics, as well as between traffic accidents and highway geometric design features in general.

- A review of existing policies, guidelines, standards, and practices for design of at-grade intersections.

- A review of existing highway agency files containing geometric design, traffic control, traffic volume, and accident data, including the data bases in the FHWA Highway Safety Information System (HSIS). As a result of these efforts, the data base of the California Department of Transportation (Caltrans) was found to be best suited for the investigation of relationships between intersection geometrics and accidents and was used for developing statistical models and testing statistical approaches in this research.

- Statistical models for relationships between traffic accidents and geometrics were developed. Alternative modeling approaches were investigated based on various assumptions about the distribution of accidents, including the Poisson, lognormal, negative binomial, and logistic distributions. The goodness of fit of these various alternative models and the role of geometric design variables in those models were assessed. Statistical models were developed for five specific types of intersections:

  — Rural, four-leg, STOP-controlled intersections
  — Rural, three-leg, STOP-controlled intersections
  — Urban, four-leg, STOP-controlled intersections

  - Urban, three-leg, STOP-controlled intersections
  - Urban, four-leg, signalized intersections

- A pilot field study to collect data on additional geometric design variables and turning movement volumes was conducted at a sample of the urban, four-leg, signalized intersections in California. Additional statistical analyses incorporating these field data were conducted.

- A review of hard copy policy accident reports was conducted to further investigate the role of geometric design features in the causation of intersection accidents.

Scope and Organization of This Report

This report is organized into eight main sections and two appendixes, in addition to this introduction. Each section is briefly described below.

Section 2 provides a brief overview of the literature related to modeling traffic accidents. Advantages and disadvantages of the various statistical modeling approaches are discussed.

Section 3 provides a review of available accident and roadway files of State highway agencies, including the States in the HSIS. The section documents the reasons for selecting the Caltrans data base for this work.

Section 4 provides details on the geometric design, traffic control, traffic volume, and accident variables for intersections in the Caltrans data base. A description is provided of the additional variables for which data were collected in the field for a sample of intersections.

Section 5 presents and compares the results from various statistical models that were developed with Poisson, negative binomial, and lognormal regression. These results were derived both from the Caltrans data base and from the new data collected in the field.

Section 6 presents the results of alternative statistical approaches to accident analysis for at-grade intersections. These alternative statistical approaches were investigated to determine whether the ability of models for predicting the safety effects of geometric design features of intersection could be improved.

Section 7 presents the results of a review that was conducted of hard copy police accident reports for eight selected urban, four-leg, signalized intersections. The objective of the hard copy police accident report review was to learn more about the role of geometric design features in accident causation at intersections.

Section 8 presents the conclusions of the study.

Appendix A presents definitions of geometric design, traffic control, and traffic volume variables from the Caltrans data base.

Appendix B presents the total multiple-vehicle accident data frequencies for the five types of intersections considered.

2. LITERATURE OVERVIEW

This section of the report presents a brief overview of the status of previous modeling of relationships between traffic accidents and highway geometric design variables including previous studies of intersection accidents. The purpose of this overview is to document those studies that have used nontraditional statistical approaches to accident modeling to illustrate how those nontraditional approaches have been applied to models for at-grade intersection accidents.

In past research, accident predictive models have often been developed with accident rates (i.e., accident frequencies per unit of exposure) as the dependent variable using simple multiple linear regression. In this traditional approach, the dependent variable (accident rate) was modeled as a linear combination of highway-related parameters, with or without interactions, under the assumption that the dependent variable follow a normal distribution. The results obtained from this approach have generally been disappointing both in terms of the proportion of the variation in accident rates explained by the models and the generally weak role of geometric design variables as accident predictors. Part of the reason for the disappointing results of past research may be that multiple regression is an inappropriate approach for developing such relationships.

There are several reasons for this concern. First, accident rates often do not follow a normal distribution. Traffic accidents are random, discrete events that are sporadic in nature. Normalizing accident frequencies with exposure estimates, such as million vehicle-miles of travel or million vehicles entering an intersection, to make accident rate appear to be a continuous random variable does not change the fundamentally discrete nature of accident data.

Second, accident frequencies for particular intersections or relatively small roadway sections are typically very small integers, even if several years of accident data are obtained for those intersections or roadway sections. In fact, it is not uncommon for a substantial proportion of the sites in an accident study to have experienced no accidents at all during the study period. Small integer counts, often zero or close to zero, do not typically follow a normal distribution. In fact, the Poisson and negative binomial distributions are often more appropriate for discrete counts of events during a given time period that are likely to be zero or small integer.

Finally, accident frequencies and accident rates are necessarily nonnegative. However, there is nothing to constrain traditional multiple regression models from predicting negative accident frequencies or accident rates, which confronts the accident analyst trying to use the predictive model with a meaningless result.

Research to develop accident predictive models published in recent literature has moved away from approaches based on multiple regression and has begun to use underlying distributional assumptions other than normal. As stated above, the Poisson distribution is appropriate for rare events like traffic accident counts where the number of events in a given time period is likely to be zero or a small integer. Bonneson and McCoy applied a regression model based on Poisson distribution in their effort to relate traffic volumes at unsignalized intersections to accident frequencies.(1) The authors found that their model based on traffic volume explained a large proportion of the variability in accidents among intersections of similar geometry and traffic control. Joshua et al. used multiple linear and Poisson regression models to describe the relationship between accident involvements of large trucks and associated traffic and geometric variables.(2) The authors concluded that the multiple linear regression model did not adequately describe that relationship, but that the Poisson model appeared to be adequate for this purpose. In another study of heavy truck-tractor accident rates, a Poisson regression model also was used by Blower et al.(3)

Miaou and Lum investigated four types of regression models to evaluate the relationship between truck accidents and highway geometric design elements.(4) The four models considered by the authors were two conventional linear regression models (one was normal or additive; the other was lognormal or multiplicative) and two multiplicative Poisson regression models (one using an exponential rate function; the other, a nonexponential rate function). Miaou and Lum concluded that of the four models tested, the Poisson model with the exponential rate function provided the best form of the relationship between truck accidents and highway geometric design elements in their study. The authors also identified the inherent limitations in using a Poisson model, which are discussed below.

One of the basic assumptions when choosing a Poisson model is that the mean and the variance of the error distribution are equal. However, in many applications, including the work that will be presented in this report, the data exhibit extra variation (i.e., the variance is greater than the mean of the estimated Poisson model). This situation is referred to as overdispersion. An alternative statistical model for addressing error structures with overdispersion like that often found in accident data is the negative binomial distribution. This approach has been used recently by several researchers, including Hauer et al., Knuiman et al., and Miaou et al., Shankar et al., and Hadi, et al.(5,6,7,8,9)

The performance of Poisson and negative binomial regression models was recently compared by Miaou.(10) The author applied these models to define a relationship between truck accidents and geometric design of road sections. The author concluded that, under moderate to high overdispersion in the data, the negative binomial model provides a sensible approach to modeling accidents in that particular application. However, under certain modeling estimation procedures, the regression coefficients are quite consistent between the Poisson and the negative binomial approach. In any case, Miaou suggests the use of Poisson regression as an initial step in the modeling effort, with the negative binomial model then being applied where appropriate. A 1987 paper by Lawless also examined the efficiency and robustness properties of the negative binomial and mixed Poisson regression models when applied to count data that exhibit extra variation.(11)

Most recently, Poch presented results of modeling four specific types of accidents as a function of geometric design and traffic variables for intersection approaches.(12) The author concluded that the negative binomial model provided an appropriate choice in identifying significant traffic and geometric elements affecting the four types of accidents studied: total, rear-end, angle, and turning accidents.

The results of these studies reported in the literature indicate the appropriateness of Poisson and negative binomial regression for development of predictive models for at-grade intersection accidents. This research was performed at about the same time as many of the studies discussed above were being published. During the research, various modeling approaches, including Poisson regression, negative binomial regression, and others, were tried as the research progressed and as the statistical software became more readily available and user-friendly. The results obtained are presented in the remainder of this report.

3. DATA BASE SELECTION

The first major activity in the research was to identify one or more existing data bases of geometric design, traffic control, traffic volume, and accident data for at-grade intersections that were suitable for testing the development of statistical models for accident prediction. In order to be useful in the planned statistical analyses, the various data files of geometric, traffic, and accident data needed to be linked together by a common location identification system, such as the milepost systems used by many State highway agencies.

The candidate data bases that were considered included the data bases available in the FHWA Highway Safety Information System (HSIS). The five original States included in the HSIS were Illinois, Maine, Michigan, Minnesota, and Utah. Of these five States, only the Maine, Michigan, and Minnesota data bases included files of intersection geometric design data. The Minnesota files appeared to be the most complete, so the files of the other HSIS States were eliminated from further consideration.

A review of data files from selected State highway agencies not included in HSIS was conducted to determine which States had existing data files of intersection geometric data. Candidate files from two States were identified: California and Washington.

Thus, data files for three candidate States—California, Minnesota, and Washington—were identified by the initial screening. The files available from these three States were reviewed in more detail to make a final selection. This review concluded that the files available from California were the most complete in terms of the range of geometric, traffic, and accident data available.

A primary disadvantage of the Minnesota accident data was that no data were included in the accident file on the directions of travel of the vehicles involved in particular accidents. This would limit the ability of the planned analyses to relate accident experience to the geometric features of particular intersection approaches or to any available turning movement data. A primary disadvantage of the data from Washington was that at intersections where the minor road (or crossroad) at an intersection was a roadway under local rather than State jurisdiction very few data were available for accidents that had occurred on the minor road. In general, the Washington data base includes accidents on a non-state-maintained crossroad only if they occur within 6 m (20 ft) of the intersection. The California data base had neither of these disadvantages; data were available on the directions of travel of accident-involved vehicles and accident data on non-state-maintained crossroads were generally included for distances up to 76 m (250 ft) from the intersection.

Two other disadvantages of the California data base were a concern, however. First, some the available data on the traffic volumes on the minor-road leg of intersections that were not maintained by the State were known to be based on estimates rather than actual counts, and there was concern that some of these estimates could be out of data. However, unlike some other States, the traffic volume data for California intersections were not incomplete. Average daily traffic (ADT) volumes were included in the file for the major- and minor-road legs of each intersection, even though in some cases these data were just estimates. Second, the California accident file did not include any specific variable identifying whether specific accidents were or were not intersection-related. Accidents that occurred within the curbline limits of the intersection could be identified explicitly, but for accidents outside the curbline limits of the intersection, there was no explicit identification of whether the cause of the accident was related to the operation of the intersection. The geometric file did, however, include a variable indicating the influence area of the intersection as a distance from the intersection in either direction of travel along the major road [typically 76 m (250 ft), but shorter in some instances such as between closely spaced intersections]. Thus, this influence distance, together with the milepost system used to identify intersection and accident locations, can be used to determine which accidents occurred within the influence area of each intersection.

Based on the factors discussed above, a decision was made that the California data base was most suited to the planned analyses and that its disadvantages were less serious than those of the other candidates considered. Copies of the data files were then obtained from the California Department of Transportation (Caltrans) and processed to conduct the analyses that are described in the remainder of the report. Data were available in the California files for a total of 19,940 intersections. The total accident frequencies experienced by these intersections for the 3-year period from 1990 to 1992 were as follows: 38,260 accidents in 1990; 34,820 accidents in 1991; and 33,203 accidents in 1992. Since these analyses began, California has been selected, along with North Carolina and Washington, for inclusion in the second stage of development of the HSIS. Thus, California data comparable to the data used for this report will be available to future analysts from the HSIS.

4. DATA BASE DESCRIPTION

This section describes the geometric design, traffic control, traffic volume, and accident history variables that were available in or were derived from the existing California data base and, thus, were available for the statistical analyses. This section also identifies the additional geometric design, traffic control, and traffic volume variables that were derived from field studies for selected California intersections.

Variables Available in the Caltrans Data Base

Table 1 presents a list of all relevant geometric design, traffic control, traffic volume, and other related variables from the Caltrans data base. Some of these variables were directly available in the data base, while others were derived or calculated from the available data (e.g., average lane width was calculated as the total traveled way width divided by the number of lanes). The variables in table 1 were selected for potential consideration in statistical modeling because it was postulated that each of these variables could possibly have an effect on intersection accidents. The results of statistical modeling to determine the effects of each of these variables on accidents are presented in section 5 of this report. Appendix A identifies each of these variables as continuous or categorical and defines the units for each continuous variable and the levels for each categorical variable.

Table 2 identifies the accident descriptors that were derived from each intersection from the existing computerized accident file provided by Caltrans. The descriptors were based on three calendar years of accident data—1990 through 1992, inclusively. For each intersection studied, the accidents of interest included those which occurred within the curbline limits of the intersection and those which occur within the influence area of the intersection, as defined by Caltrans [typically including 76 m (250 ft) along each leg of the intersection].

Additional Variables for Which Data Were Obtained in Field Studies

A number of geometric design, traffic control, and traffic volume variables of potential interest were not available in the existing Caltrans data base and, therefore, are not listed in table 1. A set of additional variables whose effects on accidents it would be desirable to examine was identified. A pilot study was conducted in which, for a selected subset of intersections, these data were collected in the field. The pilot study also provided an opportunity to verify, and where necessary update, the geometric and traffic control data in the existing file and, for some variables, to provide greater detail than was included in the existing file. For example, the existing Caltrans data, as shown in table 1, appears to imply that both major road and both crossroad approaches

Table 1. Variables Available in the Existing Caltrans Data Base

Geometric Design Features

- Intersection configuration (i.e., three-leg, four-leg, multileg, etc.)
- Number of lanes on major road
- Number of lanes on crossroad
- Presence of median on major road (i.e., divided/undivided)
- Median width on major road
- Average lane width on major road
- Shoulder width on major road
- Design speed of major road
- Functional classification of major road
- Presence of left-turn channelization on major road (i.e., separate left-turn lane)
- Presence of left-turn channelization on crossroad (i.e., separate left-turn lane)
- Presence of right-turn channelization on major road (i.e., separate roadway for free right turns)
- Presence of right-turn channelization on crossroad (i.e., separate roadway for free right turns)
- Presence of access control on major road (none/partial)

Traffic Control Features

- Type of intersection traffic control (STOP sign, traffic signal, etc.)
- One-way vs. two-way operation on major road
- Left-turn prohibition from major road
- Left-turn prohibition from crossroad
- Presence of mast arm signals on major road (signalized intersections only)
- Presence of mast arm signals on crossroad (signalized intersections only)
- Signal timing (i.e., pretimed/semiactuated/fully actuated)
- Signal phasing (i.e., two-phase/multiphase)

Traffic Volume Data

- Average daily traffic (ADT) of major road (veh/day)
- Average daily traffic (ADT) of crossroad (veh/day)

Other Related Data

- Rural/urban
- Terrain
- Presence of intersection lighting

Table 2. Accident History Variables Derived from Caltrans Accident File

- Total accidents for all 3 years combined

- Total accidents in calendar year 1990
- Total accidents in calendar year 1991
- Total accidents in calendar year 1992

- Total accidents for each calendar year by severity level:

  — fatal accidents
  — injury accidents
  — property-damage-only accidents

- Total accidents for each calendar year by location with respect to intersection:

  — within curbline limits of intersection
  — not within curbline limits, but on the major road within the influence area of the intersection
  — not within curbline limits, but on the crossroad within the influence area of the intersection

- Total accidents by calendar year and by accident type:

  Single-vehicle non-collision accidents:
  — ran-off-road
  — overturned in road
  — other single-vehicle non-collision accident

  Single-vehicle collision accidents:
  — collision with parked vehicle
  — collision with train
  — collision with pedestrian
  — collision with bicycle
  — collision with animal
  — collision with fixed object
  — other single-vehicle collision

  Multiple-vehicle collision accidents:
  — Head-on collision
  — Sideswipe collision
  — Rear-end collision
  — Angle collision
  — Right-turn collision
  — Left-turn collision (or U-turn)
  — Other multiple-vehicle collision

always have the same left-turn and right-turn channelization. In fact, the two major-road and crossroad approaches may differ in geometrics; therefore, data in the pilot field study were collected separately for each of the four intersection legs.

A decision was reached to focus the pilot study on just one of the five types of at-grade intersections that were addressed in the study—urban, four-leg, signalized intersections. A target sample size of 200 intersections was selected for the pilot study. In fact, the pilot field data were collected for a randomly selected sample of 198 of the 1,306 urban, four-leg, signalized intersections in the available sample.

Table 3 identifies the geometric design, traffic control, traffic volume, and related variables that were collected in the pilot field studies. As for the data from the existing Caltrans file, Appendix A identifies each of these variables as continuous or categorical and defines the units for each continuous variable and the levels for each categorical variable.

Table 3. Variables Collected in Pilot Field Study of Urban, Four-leg, Signalized Intersections

Geometric Design Features

- Number of through lanes on each approach
- Number of exclusive left-turn lanes on each approach
- Number of exclusive right-turn lanes on each approach
- Type of left-turn treatment on each approach
- Type of right-turn treatment on each approach
- Horizontal alignment of each approach
- Approach grades on each approach [within 76 m (250 ft) of the intersection]
- Presence of crest/sag vertical curve on each approach
- Total through lane width on each approach (ft)
- Total left-turn lane width on each approach (ft)
- Presence of median on each approach (i.e., divided/undivided)
- Type of median (if any) on each approach
- Median width on each approach (ft)
- Number of driveways within 76 m (250 ft) of the intersection on each approach
- Type of driveways on each approach
- Angle between intersecting approaches
- Curb return radius (ft) in intersection quadrant to the right of each approach (for selected intersections with high pedestrian activity)

Traffic Control Features

- One-way vs. two-way operation on each approach
- Presence of left-turn prohibition on each approach
- Curb parking within 76 m (250 ft) of the intersection on each approach
- Number of signal faces for each approach
- Signal head mounting for each approach (i.e., post-mounted/mast-arm)
- Left-turn phasing for each approach (i.e., presence of left-turn arrow phase)
- Presence of pedestrian signals for crossing each approach
- Presence of painted crosswalk for crossing each approach
- Presence of advance warning signs (e.g., SIGNAL AHEAD) for each approach
- Posted speed limit for each approach (mi/h)

Traffic Volume Data

- Turning movement volumes for all approaches by 15-min periods for a 2-h morning peak period (typically 7 to 9 a.m.) and a 2-h evening peak period (typically 4 to 6 p.m.)
- Level of pedestrian activity

Other Related Data

- Presence of intersection lighting
- Character of surrounding development

5. STATISTICAL MODELING

This section describes the statistical modeling of at-grade intersection accidents for five selected types of intersections that was conducted during the research based on the Caltrans data base. The discussion includes both the data preparation steps prior to the analysis and the analyses that were conducted for the five intersection types. The statistical models for urban, four-leg, signalized intersections are further investigated using additional data collected during the pilot field studies.

Data Preparation

The Caltrans data base, whose selection was described in chapter 3 of this report, contains information on geometric design features, traffic control features, and traffic volumes at over 19,000 intersections located on State highways in California. Total accident frequencies at these intersections for the 3-year period from 1990 to 1992 were: 38,260 accidents in 1990; 34,820 accidents in 1991; and 33,203 accidents in 1992. A preliminary assessment was made of the types of intersections that were present in sufficient numbers and had sufficient data available for statistical modeling of accidents to be conducted. The selection of intersections, accident types, geometric and traffic parameters, and volumes is discussed in the following sections.

Intersection Types

The Caltrans data base included data for a total of 19,398 intersections. These included rural and urban intersections, various intersection configurations (three-leg T intersections, three-leg Y intersections, four-leg intersections, four-leg offset intersections, and multileg intersections), and intersections with various types of traffic control (no control, two-way STOP control, four-way STOP control, yield control, and signal control). After reviewing the number of intersections of each type that were included in the Caltrans data base, a decision was reached to focus the analyses on three-leg and four-leg intersections with both two-way STOP-control and signal control. Only three-leg T intersections were considered because the available sample of three-leg Y intersections was quite small. A total of 15,369 intersections was found within the eight cells defined by these three factors (which is equivalent to approximately 79 percent of all of the available intersections). Next, intersections with major road average daily traffic (ADT) below 400 veh/day or crossroad ADT below 100 veh/day were deleted from the data base, resulting in a total of 11,165 intersections for consideration. This was done both to eliminate extremely low-volume sites, which typically have very few accidents and would be difficult to model reliably, and to eliminate intersections with low ADTs that appeared in error. The intersection breakdown in the Caltrans data base before and after eliminating the low-volume locations is illustrated in table 4. The total number of accidents for the intersections in each of these cells was reviewed, and a decision was reached to conduct statistical modeling for data from five of the eight cells.

1. Rural, four-leg, STOP-controlled intersections
2. Rural, three-leg, STOP-controlled intersections
3. Urban, four-leg, STOP-controlled intersections
4. Urban, three-leg, STOP-controlled intersections
5. Urban, four-leg, signalized intersections

These five intersection types of interest are identified by the shaded portions in the lower portion of table 4. Each of these five intersection types was considered separately in the lognormal and loglinear modeling activities that are described in this section.

Table 4. Intersection Distribution in Caltrans Data Base by Type of Intersection

	Four-leg intersections		Three-leg intersections		
	STOP controlled	Signal controlled	STOP controlled	Signalized	Total
Number of intersections in Caltrans data base					
Rural	2,281	103	5,512	32	7,928
Urban	1,726	1,514	3,773	428	7,441
Total	4,007	1,617	9,285	460	15,369
Number of intersections in ADT ranges considered[a]					
Rural	1,581	100	2,907	2	4,590
Urban	1,475	1,433	3,256	411	6,575
Total	3,056	1,533	6,163	413	11,165

[a] Major road ADT above 400 veh/day; crossroad ADT above 100 veh/day.
The shaded cells are those for which statistical analyses were performed.

Safety Measures of Effectiveness (Dependent Variables)

The accidents analyzed for each intersection included:

- All accidents within the curbline limits of the intersection
- All accidents that occurred on the major road within a specified influence area defined by Caltrans [typically 76 m (250 ft) but shorter or longer in some cases]
- All accidents that occurred on the crossroad within 76 m (250 ft) of the intersection

Table 5 presents the distribution of accidents by number of vehicles involved (single-vehicle vs. multiple-vehicle) for the five selected intersection types. The table includes all intersections in each of the five categories, as well as for the intersections in each group that were selected for inclusion in the analyses (see below).

The analyses performed in this research focused on multiple-vehicle accidents, because single-vehicle accidents generally occur less frequently and are, therefore, more difficult to model. The conceptual plan developed for the FHWA Interactive Highway Safety Design Model (IHSDM) recommends that the frequency of single-vehicle run-off-road accidents be predicted using an encroachment-based technique rather than a statistical model.

All of the modeling efforts in the research addressed total multiple-vehicle accidents (for all accident severity levels combined) and fatal and injury multiple-vehicle accidents. Property-damage-only (PDO) accidents were not analyzed separately because of concerns about incompleteness of accident reporting. It might have been desirable to evaluate PDO towaway accidents (accidents in which one or more of the involved vehicles was towed from the scene) or fatal-plus-injury-plus-towaway accidents, but unfortunately the California accident data do not explicitly identify towaway accidents. In summary, the two dependent variables most extensively used in the modeling effort were:

- Total multiple-vehicle accidents of all severity levels that occurred during the 3-year study period

- Fatal and injury multiple-vehicle accidents in the 3-year period

Selection of Geometric and Traffic Parameters of Interest (Independent Variables)

For each of the five categories of intersections mentioned above, a preliminary selection of geometric and traffic variables as candidate independent variables for the statistical modeling activities was made from among the variables included in the existing Caltrans data base (see discussion in section 4) based on engineering knowledge and statistical criteria. A few of the candidate independent variables were quantitative variables measured on a continuous scale (e.g., lane width or shoulder width); however, most of the candidate independent variables were categorical (i.e., having a finite number of discrete levels). Appendix A identifies whether each candidate variable was continuous or categorical in nature and also identifies the levels for each categorical variable.

Table 5. Accident Data Distributions in Caltrans Data Base

	All intersections					Selected intersections[a]		
	Number of accidents			Percent of total accidents		Multiple-vehicle accidents		
	Single vehicle	Multiple vehicle	Total	Single vehicle	Multiple vehicle	Total	Fatal and injury	Ratio F&I/Total
Rural, four-leg, STOP-controlled (2,262 intersections)						(1,434 Intersections)		
1990	387	2,229	2,616	14.8	85.2	1,735	883	0.51
1991	359	2,003	2,362	15.2	84.8	1,543	795	0.52
1992	337	2,010	2,347	14.4	85.6	1,580	814	0.52
3 years	1,083	6,242	7,325	14.8	85.2	4,858	2,492	0.51
Rural, three-leg, STOP-controlled (5,491 intersections)						(2,692 Intersections)		
1990	931	2,496	3,427	27.2	72.8	1,714	778	0.45
1991	837	2,382	3,219	26.0	74.0	1,577	700	0.44
1992	858	2,362	3,220	26.6	73.4	1,578	712	0.45
3 years	2,626	7,240	9,866	26.6	73.4	4,869	2,190	0.45
Urban, four-leg, STOP-controlled (1,551 intersections)						(1,342 Intersections)		
1990	483	3,542	4,025	12.0	88.0	3,193	1,392	0.44
1991	430	3,070	3,500	12.3	87.7	2,782	1,332	0.48
1992	399	3,081	3,480	11.5	88.5	2,819	1,327	0.47
3 years	1,312	9,693	11,005	11.9	88.1	8,794	4,051	0.46
Urban, three-leg, STOP-controlled (3,680 intersections)						(3,057 Intersections)		
1990	854	4,682	5,536	15.4	84.6	4,199	1,762	0.42
1991	703	4,392	5,095	13.8	86.2	3,972	1,755	0.44
1992	664	4,155	4,819	13.8	86.2	3,781	1,697	0.45
3 years	2,221	13,229	15,450	14.4	85.6	11,952	5,214	0.44
Urban, four-leg, signalized (1,448 intersections)						(1,306 Intersections)		
1990	870	11,120	11,990	7.3	92.7	10,291	4,116	0.40
1991	788	10,159	10,947	7.2	92.8	9,431	3,980	0.42
1992	732	9,322	10,054	7.3	92.7	8,684	3,819	0.44
3 years	2,390	30,601	32,991	7.2	92.8	28,406	11,915	0.42
All Intersections (14,432)						(9,831 Intersections)		
3 years	9,632	67,005	76,637	12.6	87.4	58,879	25,862	0.44

[a] Intersections selected for analysis

To determine which of the candidate independent variables were suitable for use in the statistical modeling activities, frequency tables were generated for each candidate variable. When the available sample size for any given level of any particular variable was too small, one of the following courses of action was taken: (a) the intersections in that level were pooled with an adjacent level (where this made engineering sense) or (b) the intersections in that level were deleted. After reviewing all levels of all categorical variables, the process was repeated to ensure that all the sample sizes were now sufficient for data analysis. Any further minor changes found to be necessary were then made. If, for a particular independent categorical variable, all but a small number of the intersections fall in a single category, then that variable had to be excluded from the modeling effort because no effect can be determined unless a substantial number of the intersections fall in each level.

Best Form for ADT Variables

ADT data were available for both major road and crossroad at each intersection. Three alternative forms for incorporating ADT variables in the models were considered:

- Separate independent variables representing the major-road and crossroad ADT's
- One combined variable representing the sum of the major-road and crossroad ADT's (equivalent to the total daily traffic volume entering the intersection)
- One combined variable representing the product of the major-road and crossroad ADT's (representing the potential number of vehicle-vehicle interactions that may occur at the intersection)

Each of these approaches has been postulated in previous research as representing the most appropriate treatment of the ADT in statistical modeling.

Preliminary modeling efforts showed that the best results were obtained when the major-road ADT and the crossroad ADT were treated as separate independent variables; therefore, this approach was used throughout the statistical modeling activities in the research.

Accident Frequency Distributions

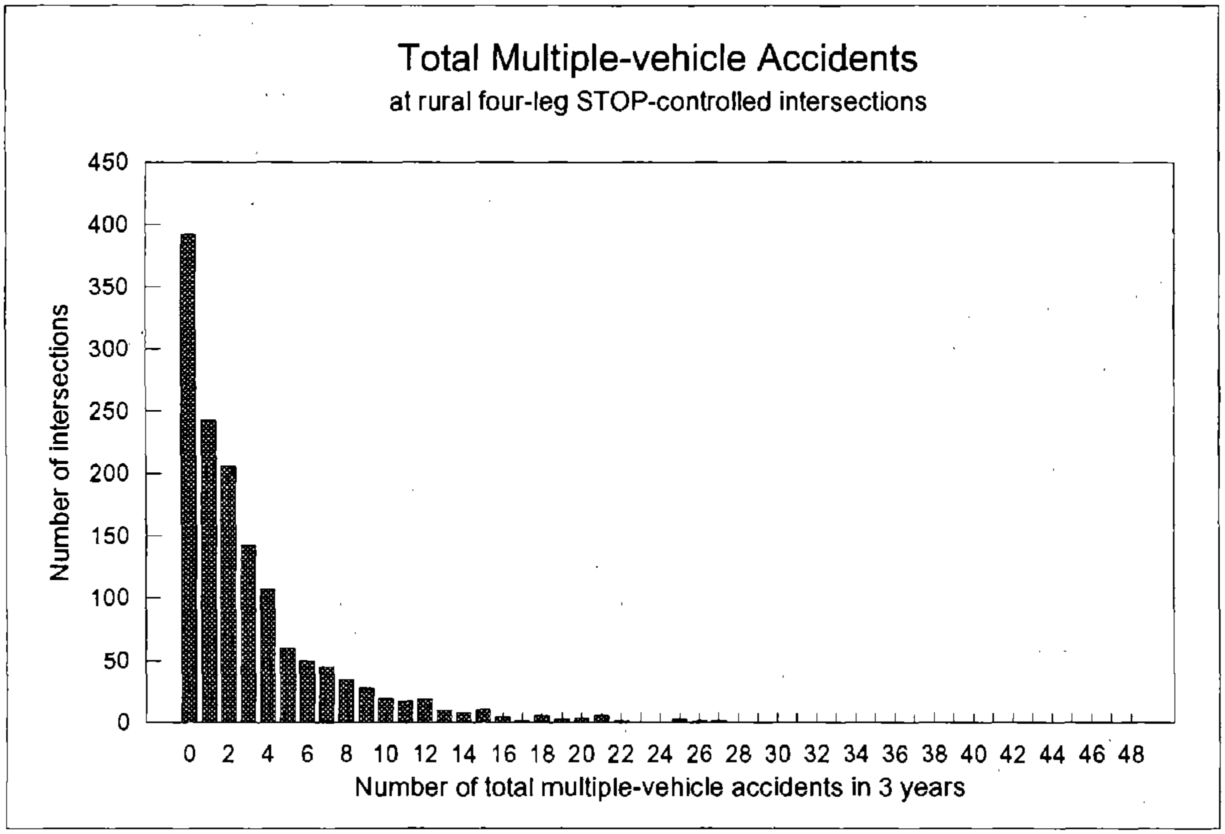
Prior to beginning the statistical modeling activities, the general shape of each accident frequency distribution was assessed for each of the five intersection types of interest. This was done visually by plotting the data for the 3-year totals and by calculating basic statistics. For each of the five intersection categories, table 6 shows yearly and 3-year total accident statistics (minimum, median, mean, maximum) for multiple-vehicle accidents. Total accident counts, as well as fatal and injury accident counts, are shown separately. Table 6 also shows the total number of accidents of each type in each given year. Next, the frequency data are plotted separately for each type of accident and each type of intersections in figures 1 through 5.

Similar statistics and frequency plots are shown in table 7 and figure 6, respectively, for the sample of 198 urban, four-leg signalized intersections investigated during the pilot field study.

The plots shown in figures 1 through 6 highlight the different shapes of accident frequencies. With large numbers of intersections with no or low accident experience, the distribution tends to follow the shape of a Poisson distribution. This observation clearly applies to rural, four- and three-leg, STOP-controlled intersections and to urban, three-leg, STOP-controlled intersections. When the number of intersections with no or low accident experience is relatively small, the distribution tends to follow the shape of a lognormal distribution. This is clearly seen in the case of urban, four-leg, signalized intersections (including the sample of 198 intersections) and also in the case of urban, four-leg, STOP-controlled intersections.

Table 6. Annual Accident Statistics, 1990-1992

	Total multiple-vehicle accidents					Fatal and injury multiple-vehicle accidents				
Year	Minimum	Median	Mean	Maximum	Total	Minimum	Median	Mean	Maximum	Total
Rural, Four-leg, STOP-controlled—1,434 intersections										
1990	0	1	1.21	14	1,735	0	0	0.62	9	883
1991	0	0	1.08	15	1,543	0	0	0.55	10	795
1992	0	0	1.10	19	1,580	0	0	0.57	8	814
1990-92	0	2	3.39	48	4,858	0	1	1.74	27	2,492
Rural, Three-leg, STOP-controlled—2,692 intersections										
1990	0	0	0.64	15	1,714	0	0	0.29	10	778
1991	0	0	0.59	12	1,577	0	0	0.26	8	700
1992	0	0	0.59	20	1,578	0	0	0.26	9	712
1990-92	0	1	1.81	45	4,869	0	0	0.81	21	2,190
Urban, Four-leg, STOP-controlled—1,342 intersections										
1990	0	1	2.38	26	3,193	0	1	1.04	12	1,392
1991	0	1	2.07	18	2,782	0	1	0.99	9	1,332
1992	0	1	2.10	21	2,819	0	0	0.99	12	1,327
1990-92	0	4	6.55	53	8,794	0	2	3.02	23	4,051
Urban, Three-leg, STOP-controlled—3,057 intersections										
1990	0	1	1.37	33	4,199	0	0	0.58	12	1,762
1991	0	1	1.30	31	3,972	0	0	0.57	10	1,755
1992	0	1	1.24	31	3,781	0	0	0.56	17	1,697
1990-92	0	2	3.91	80	11,952	0	1	1.71	28	5,214
Urban, Four-leg, Signalized—1,306 intersections										
1990	0	6	7.88	49	10,291	0	2	3.15	22	4,116
1991	0	6	7.22	53	9,431	0	2	3.05	20	3,980
1992	0	5	6.65	57	8,684	0	2	2.92	20	3,819
1990-92	0	18	21.75	147	28,406	0	8	9.12	50	11,915



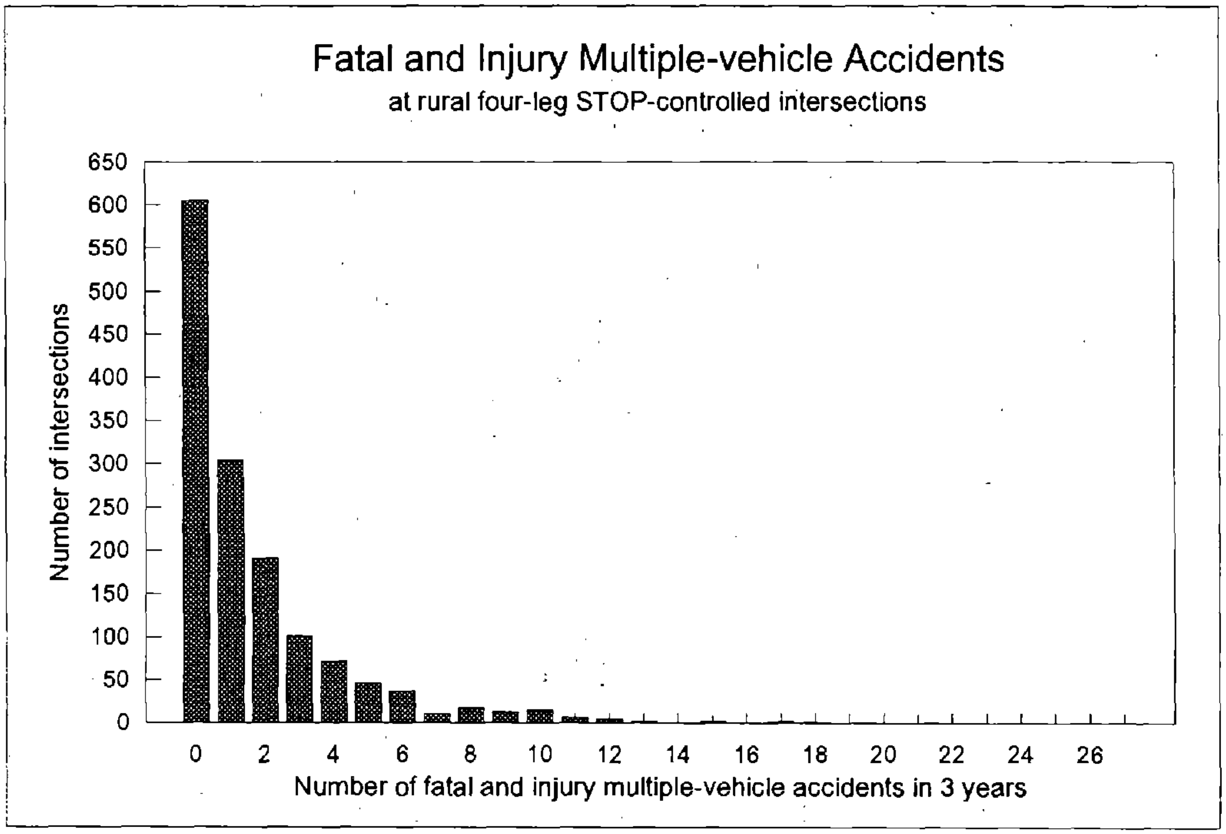


Figure 1. Accident Frequency Distributions at Rural, Four-leg, STOP-controlled Intersections

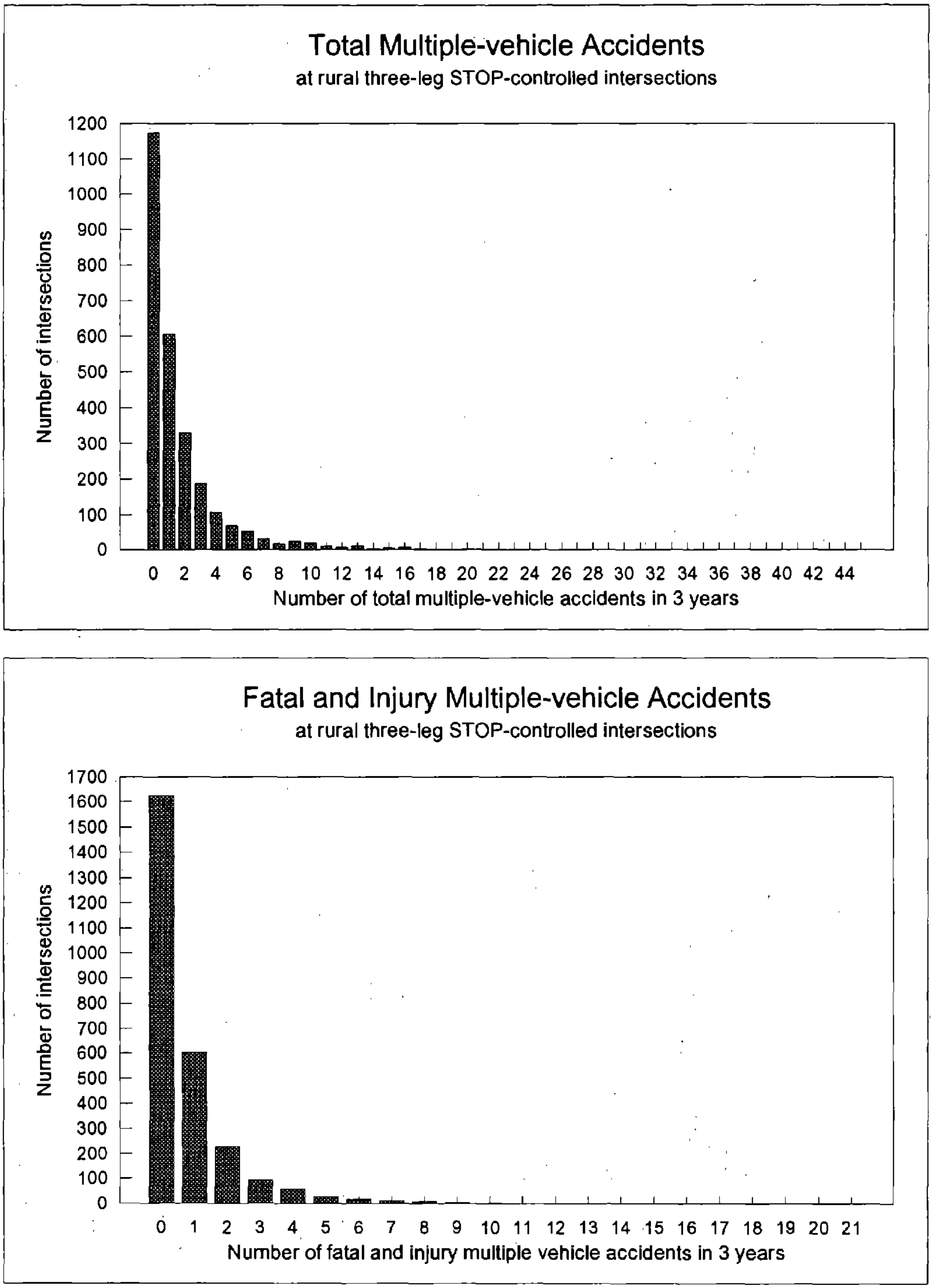
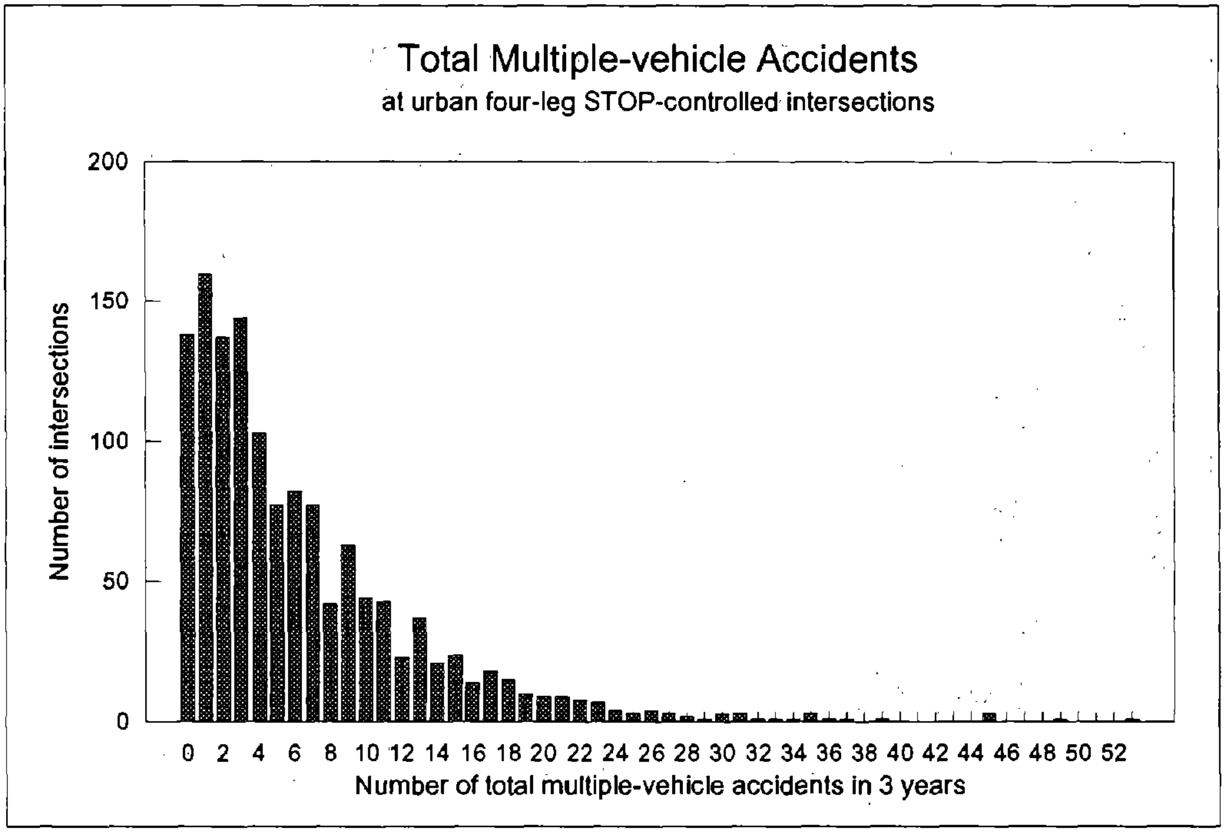


Figure 2. Accident Frequency Distributions at Rural, Three-leg, STOP-controlled Intersections



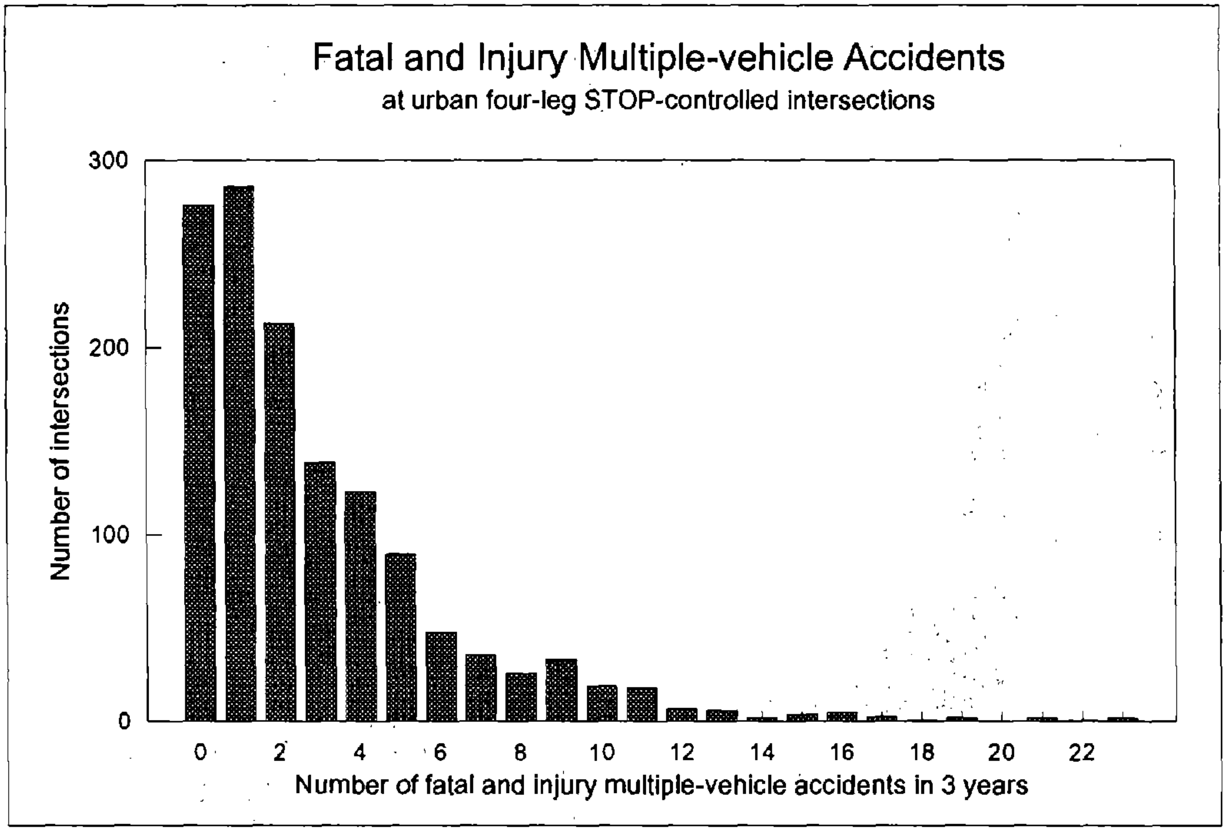
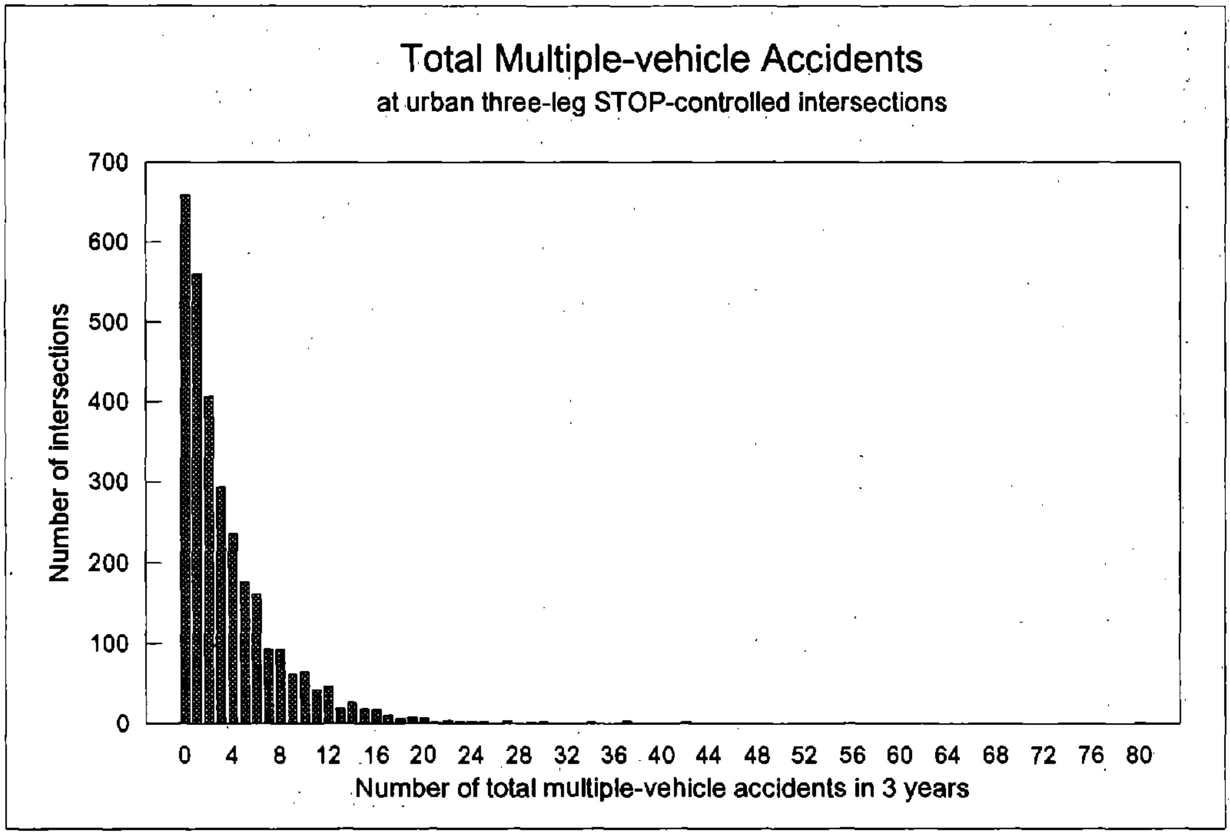


Figure 3. Accident Frequency Distributions at Urban, Four-leg, STOP-controlled Intersections



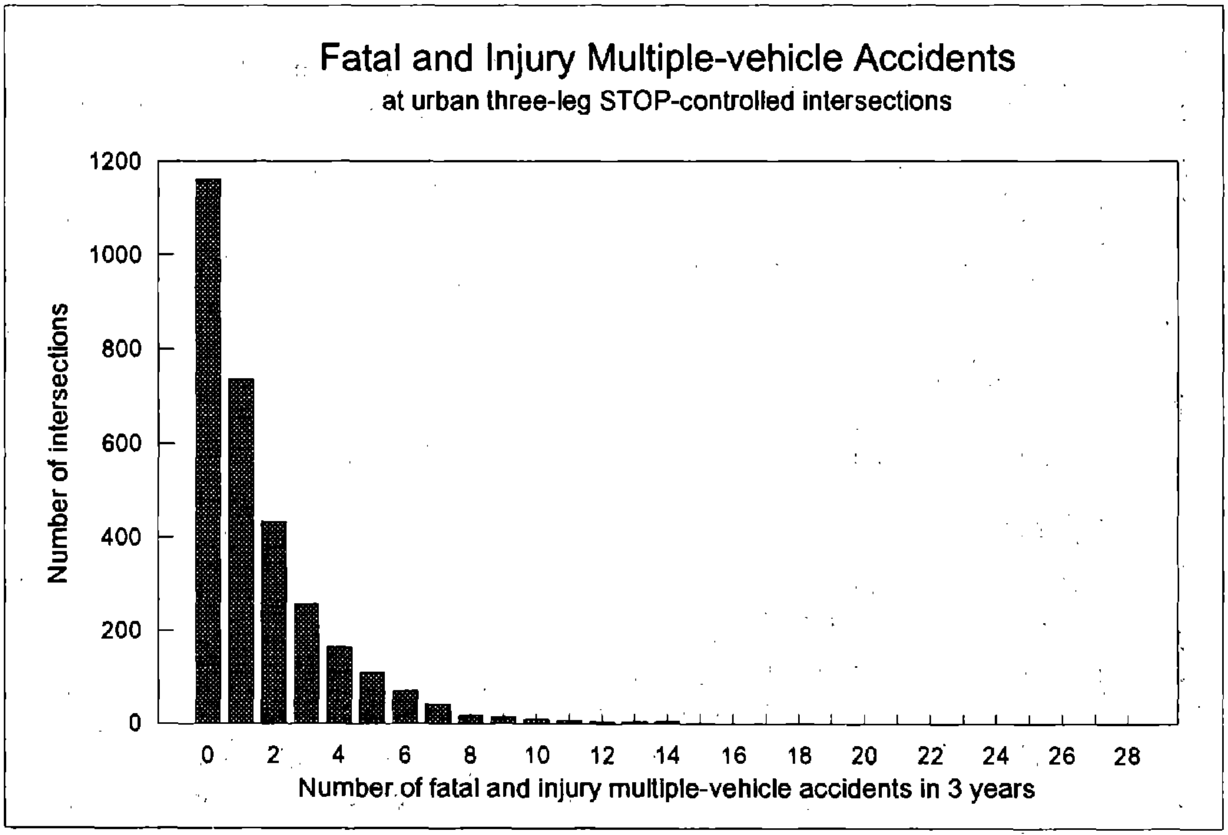
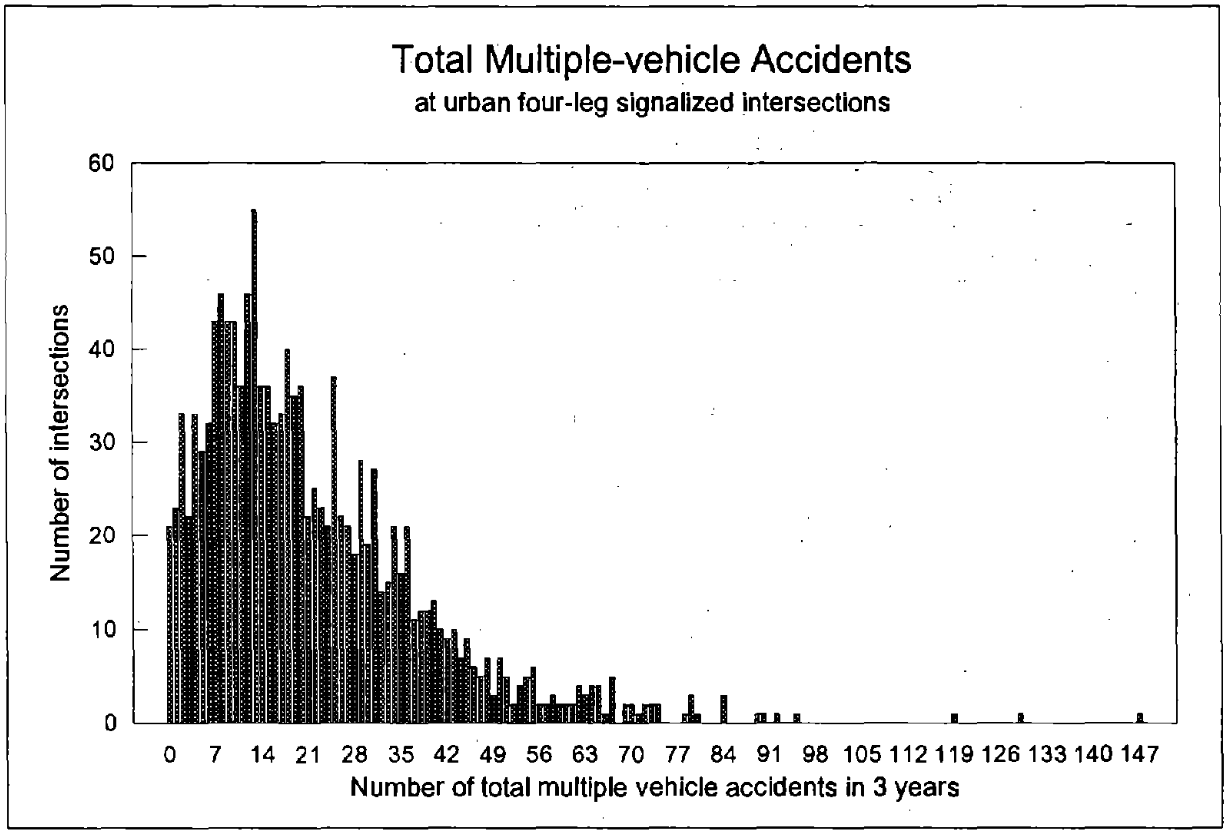


Figure 4. Accident Frequency Distributions at Urban, Three-leg, STOP-controlled Intersections



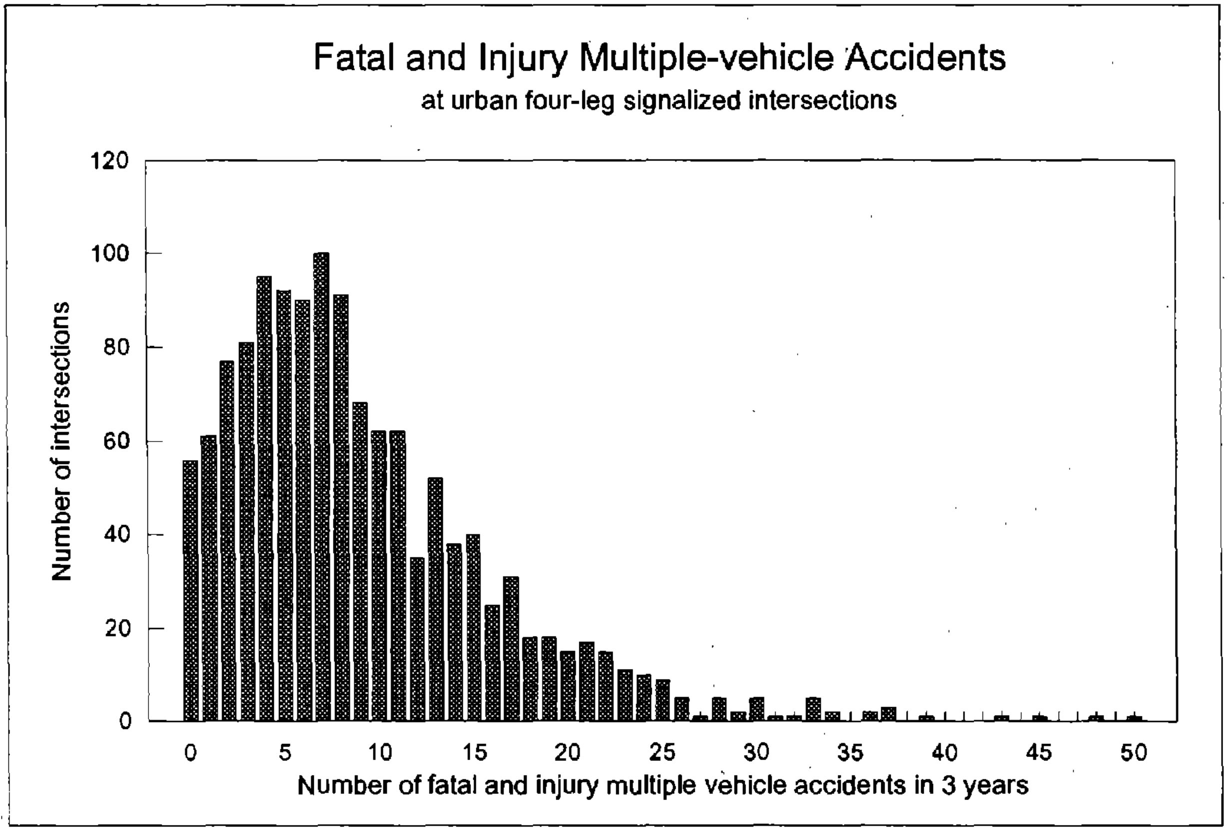
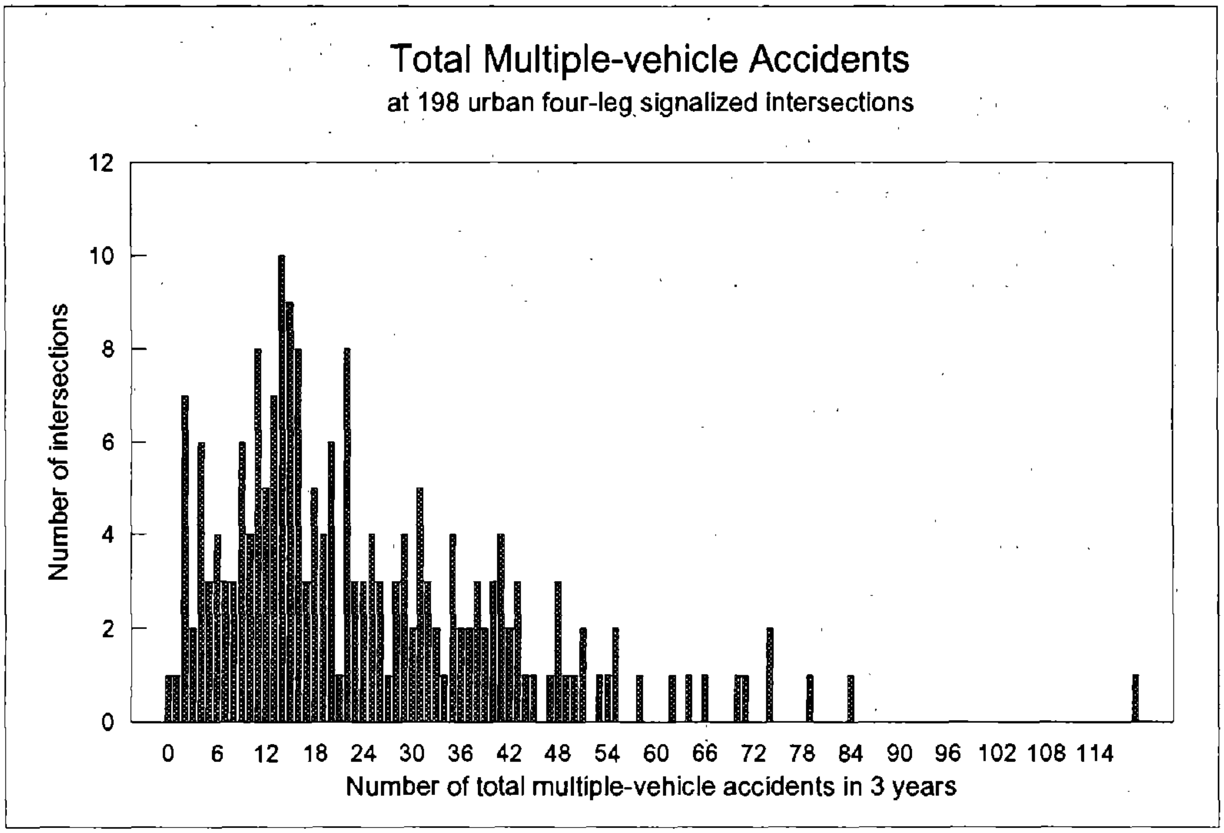


Figure 5. Accident Frequency Distributions at Urban, Four-leg, Signalized Intersections

Table 7. Annual Accident Statistics for Sample of 198 Urban, Four-leg, Signalized Intersections

Year	Minimum	Median	Mean	Maximum	Total
Total multiple-vehicle accidents					
1990	0	7	8.61	49	1,704
1991	0	7	8.15	36	1,613
1992	0	6	7.74	36	1,533
1990-92	0	20	24.49	119	4,850
Fatal and injury multiple-vehicle accidents					
1990	0	3	3.29	11	651
1991	0	3	3.39	16	671
1992	0	3	3.40	12	674
1990-92	0	9	10.08	30	1,996



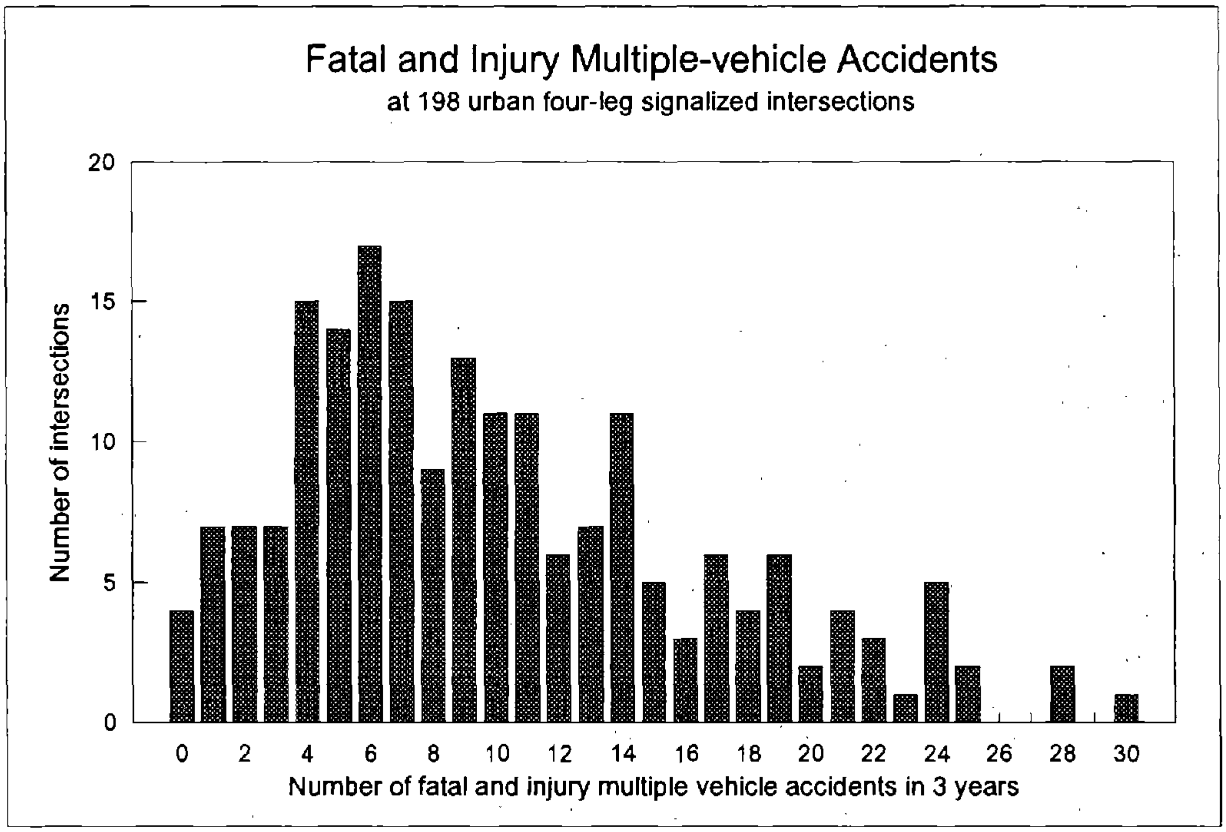


Figure 6. Accident Frequency Distributions at a Sample of 198 Urban, Four-leg, Signalized Intersections

Lognormal and Loglinear Regression Models

Several candidate analysis methods were investigated for application to the accident frequencies in the five types of at-grade intersections in this study. The analysis approach was driven by both the actual distribution of the accident frequencies and by recommendations and practices in the field of accident data analysis (see chapter 2). The frequency distributions of total multiple-vehicle accidents in the 3-year study period are shown in figures 1 through 6 above. The percentages of intersections with zero or one multiple-vehicle accident in the 3-year period are:

Intersection type	Percent of intersections with 0 or 1 accident in 3-year period
Rural, four-leg, STOP-controlled	44.3
Rural, three-leg, STOP-controlled	66.1
Urban, four-leg, STOP-controlled	22.2
Urban, three-leg, STOP-controlled	39.9
Urban, four-leg, signalized	3.4
Urban, four-leg, signalized (sample of 198 intersections)	1.0

For most types of intersections, a large proportion of the intersections experienced at most one accident over the 3-year period. This observation is not true for urban, four-leg, signalized intersections, where about 50 percent of the intersections experienced 19 or more multiple-vehicle accidents over the 3-year period, and only 10 percent of the intersections had five or fewer accidents. Also, the pattern for urban, four-leg, STOP-controlled intersections differs somewhat from the two extreme situations in that only 10 percent of the intersections experienced no accidents and about half of the intersections experienced at least four accidents in the 3-year period.

Two general types of statistical models were applied to the accident data in this study: (1) a lognormal regression model for all urban, four-leg intersections (both STOP-controlled and signalized); and (2) a loglinear regression model—either Poisson regression or negative binomial regression model—for all rural STOP-controlled (three- and four-leg) and urban, three-leg, STOP-controlled intersections. These models are presented next.

Consider a set of n intersections of a given class (e.g., rural, four-leg, STOP-controlled intersections). Associated with each intersection i, is a set of q parameters (X_{i1}, X_{i2}, ..., X_{iq}), describing the geometric design, traffic control, traffic volume, and other related characteristics of that intersection. Let the number of accidents occurring at the ith intersection during a 3-year period be denoted by Y_i, where i=1, ..., n. Next, denote by y_i the actual observation of Y_i during the 3-year period, that is, $y_i = 0, 1, 2, \ldots$ and i=1, ..., n.

The objective of a statistical model is to provide a relationship between a function of the expected number of accidents, $E(Y_i)=\mu_i$, at the ith intersection and the q intersection parameters, $X_{i1}, X_{i2}, \ldots, X_{iq}$. This relationship can be formulated through a general linear model of the form:

$$\text{Function}(\mu_i) = \beta_0 + \beta_1 X_{i1} + \ldots + \beta_q X_{iq} \tag{1}$$

where the regression coefficients, $\beta_0, \beta_1, \beta_2, \ldots \beta_q$, are to be estimated from the data. The estimation procedure used to obtain the regression coefficients is dependent on the assumption made about the distribution of the Y_i.

Note: Throughout this report, all logarithms are natural logarithms and are denoted by *log* in all equations.

Lognormal Regression Models

Lognormal regression models are based on the assumption that the natural logarithm of Y_i follows a normal distribution with mean μ_i and variance σ^2. In other words, it is assumed that Y_i follows a lognormal distribution, a reasonable choice whenever the data are inherently non-negative, suggesting that a model with positive skewness is needed and the mean is relatively large. This model also ensures that μ_i, the expected number of accidents, remains positive.

In this case, the relationship between the expected number of accidents at the ith intersection and the q predictor variables, $X_1, \ldots, X_q$, can be written as:

$$\log(\mu_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_q X_{iq} \tag{2}$$

or alternatively, in the multiplicative form, as

$$\mu_i = \exp(\beta_0)\exp(\beta_1 X_{i1})\exp(\beta_2 X_{i2}) \ldots \exp(\beta_q X_{iq}) \tag{3}$$

where the log(number of accidents) is assumed to follow a normal distribution with mean μ_i and variance σ^2. The coefficients, $\beta_0, \beta_1, \beta_2, \ldots \beta_q$, are the linear regression coefficients to be estimated by ordinary least-squares method. This is the classical case of a multiple linear relationship between the logarithm of the dependent variable and q independent predictor variables. For the lognormal regressions, the usual normal-theory tests of significance of the parameters and goodness-of-fit of the model measures apply.

Loglinear Regression Models

In the present study, two loglinear models were considered for application to at-grade accident frequencies: the Poisson and the negative binomial models. Their general forms are described below.

Poisson Model. When the average number of accidents at an intersection becomes small, the assumption of a lognormal distribution is no longer valid. The Poisson model then becomes a natural choice as it models the occurrence of rare discrete events well. The relationship between the expected number of accidents occurring at the ith intersection and the q intersection parameters, $X_{i1}, X_{i2}, ..., X_{iq}$, is assumed to be of the form:

$$\log(\mu_i) = \beta_0 + \sum_{j=1}^{q} \beta_j X_{ij} \tag{4}$$

However, the assumption is now made that the number of accidents, Y_i, follows a <u>Poisson</u> distribution with mean μ_i. That is, the probability that an intersection defined by a known set of predictor variables, $X_{i1}, X_{i2}, ..., X_{iq}$, experiences y_i accidents can be expressed as:

$$P(Y_i = y_i; \mu_i) = \mu_i^{y_i} e^{-\mu_i} / y_i! \tag{5}$$

where $y_i!$ denotes the factorial of y_i.

Note that the Poisson distribution has only one parameter, namely its mean, μ_i, with the limitation that the variance, σ^2, equals the mean of the distribution. Under the assumption of a Poisson distribution, the regression coefficients, $\beta_0, \beta_1, \beta_2, ... \beta_q$, are estimated by the maximum likelihood method. The assymptotic normality of maximum likelihood estimates is used to obtain tests of significance of the parameters and goodness-of-fit measures for the models.

In the case of a Poisson distribution, the model coefficients are estimated by maximum likelihood method. The likelihood function is the product of the terms in equation (2) over all n intersections in the class of intersections of interest. This function is viewed as a function of the parameters, μ_i, and through them, the parameters β_i. The parameters are estimated by maximizing the likelihood, or more usually, by maximizing the logarithm of the likelihood (denoted by log likelihood). Equivalently, the estimation can be done by minimizing the negative of the log likelihood. The log likelihood is given by the equation:

$$\log(L) = \sum_{i=1}^{n} [y_i \log(\mu_i) - \mu_i - \log(y_i!)] \tag{6}$$

The maximum value possible for the likelihood for a given data set occurs if the model fits the data exactly. This occurs if the μ_i are replaced by y_i in (3). The difference between the log-likelihood functions for two models is a measure of how much one model improves the fit over the other. A special case of this was defined as the **deviance** by Nelder and Wedderburn.(13) Specifically, they defined the deviance as minus twice the log of the ratio of the likelihood for a model to the maximum likelihood. For the Poisson, the deviance takes the form given in Equation (7):

$$D = 2[\sum_{i=1}^{n} y_i \ln(y_i/\mu_i) - \sum_{i=1}^{n} (y_i - \mu_i)] \tag{7}$$

where the second term is identically zero in the usual case that the model includes a constant or intercept term. The deviance so defined is measured from that of the saturated model and so terms involving constants, the data alone, or a scale factor alone are omitted. For a sample of n independent observations, the deviance for a model with p degrees of freedom (that is, p parameters estimated including the mean or constant) has residual (n – p) degrees of freedom. When the residual degrees of freedom of the current model are approximately equal to the deviance, it is unlikely that further fitting of systematic components is worthwhile.

Since the deviance is effectively –2 times the log of the likelihood ratio, it has an asymptotic distribution that is chi-squared with degrees of freedom equal to n – p, where n is the number of intersections and p is the number of parameters estimated. This result can be used to construct a goodness-of-fit test for the model. In addition, by forming the ratio of the deviance to its residual degrees of freedom, an estimate of the scale constant can be found. For the Poisson, this should theoretically be equal to one. Values substantially in excess of one reflect overdispersion of the data.

Negative Binomial Model. As mentioned above, a limitation of the Poisson distribution is that the mean equals the variance of the distribution. Previous work in the field of accident research has shown that this is not always the case. Suppose a Poisson model is used for modeling accidents and the variance, or dispersion, of the data exceeds the estimated mean of the accident data distribution. The data are then said to be overdispersed, and the underlying assumption of the variance being equal to the mean for the Poisson distribution is violated. The negative binomial, which is a discrete distribution, provides an alternative model to deal with overdispersion in count data such as accident frequencies.

Unlike the Poisson distribution, the negative binomial distribution has two parameters. As for the Poisson model above, the relationship between the expected number of accidents occurring at the ith intersection and the q intersection parameters, X_{i1}, X_{i2}, ..., X_{iq}, is still taken to be:

$$\text{Function}(\mu_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_q X_{iq} \tag{8}$$

However, the assumption is now made that the number of accidents, Y_i, follows a negative binomial distribution with parameters α and k (with $0 \leq \alpha \leq 1$ and $k \geq 0$). That is, the probability that an intersection defined by a known set of predictor variables, X_{i1}, X_{i2}, ..., X_{iq}, experiences $Y_i = y_i$ accidents can be expressed as:

$$\Pr\left(Y_i = y_i; \alpha, k\right) = \frac{\left(y_i + k - 1\right)!}{y_i!\,(k-1)!} \quad \frac{\alpha^{y_i}}{(1+\alpha)^{y_i+k}} ; \qquad y_i = 0,1,2\ldots \tag{9}$$

where $y_i!$ denotes the factorial of y_i.

The mean and variance of the negative binomial distribution of accident counts can then be expressed in terms of the parameters α and k as follows:

$$\text{mean} = E(Y) = \mu_i = k\alpha, \text{ and} \tag{10}$$

$$\text{variance} = \text{Var}(Y) = k\alpha + k\alpha^2 = \mu_i + \mu_i^2/k \tag{11}$$

The term μ_i can be referred to as the Poisson variance function and μ_i^2/k as the extra component arising from combining the Poisson distribution with a gamma distribution for the mean to obtain the negative binomial distribution. The parameter k is not known a priori, but can be estimated so that the mean deviance becomes unity or the Pearson chi-square statistic equals its expectation (i.e., equals its degrees of freedom).(14)

As for the Poisson model, the model regression coefficients, β_0, β_1, β_2, ... β_q, are estimated by the method of maximum likelihood. The asymptotic normality of maximum likelihood estimates is used to obtain tests of significance of the parameters and goodness-of-fit measures for the models. The estimation of the model parameters can be done by minimizing the negative of the log likelihood. For the negative binomial distribution, the log likelihood is given by the equation:

$$\log(L) = \sum_{i=1}^{n} y_i \log\left[\alpha/(1+\alpha)\right] - nk\log(1+\alpha) + \left(\text{function of } y_i, k\right) \tag{12}$$

Substituting $\alpha = \mu_i/k$ into the term $\log[\alpha/(1+\alpha)]$ of Eq. (12), gives the function

$$\log\left(\frac{\mu_i}{\mu_i + k}\right) = \beta_o + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_q X_{iq} \tag{13}$$

The parameters α and k of the negative binomial distribution can thus be indirectly estimated using a generalized linear model and, by means of equations (8) and (13), the model regression coefficients β_0, β_1, β_2, ... β_q are obtained. The Statistical Analysis System, SAS, provides a procedure, PROC GENMOD (a generalized linear model procedure) that can be used to estimate the regression coefficients by implementing equations (11) and (13).$^{(15)}$

Treatment of ADT Variables in Lognormal and Loglinear Regression Models

In all models in this study, the natural logarithm of the major-road and crossroad ADT variables was used. This parallels the approach taken by other researchers where accident counts rather than accident rates are modeled. On the log-scale, the ratio of accident counts over ADT becomes the difference between log(accident counts) and log(ADT). The difference here is that it is assumed that the coefficient of log(ADT) is not equal to one, but rather is a coefficient to be estimated through analysis. Thus, in the lognormal and Poisson and negative binomial models described above, X_1 and X_2 generally represent $\log(ADT_{\text{major road}})$ and $\log(ADT_{\text{crossroad}})$, respectively. The multiplicative model relating the expected accident counts and independent variables can thus be rewritten as:

$$\text{function}(\mu_i) = \exp(\beta_0)\left(ADT_{\text{major road}}\right)^{\beta 1}\left(ADT_{\text{crossroad}}\right)^{\beta 2}\exp(\beta_3 X_{i3}) \bullet \ldots \bullet \exp(\beta_q X_{iq}) \tag{14}$$

Accident Modeling Results

The following sections present the modeling results separately for each of the five selected types of intersections:

- Rural, four-leg, STOP-controlled
- Rural, three-leg, STOP-controlled
- Urban, four-leg, STOP-controlled
- Urban, three-leg, STOP-controlled
- Urban, four-leg, signalized

Rural, Four-leg, STOP-controlled Intersections

The first step in the analysis of rural, four-leg, STOP-controlled intersections was to select candidate independent variables for that particular group of intersections. Both engineering judgment and sample size requirements for the levels of each candidate variable were involved in the decision as to whether a particular variable was included in the modeling effort. A small number of independent categorical variables were not included in the full model because either all or nearly all intersections fell into one level of that variable. The variables originally considered that were not included for these reasons were:

- Major-road left-turn prohibition (all intersections had permitted left turns).

- Crossroad left-turn channelization (none of the intersections had left-turn lanes).

- Crossroad left-turn prohibition (none of the intersections had left turns prohibited).

- Number of lanes on the crossroad (99.8 percent had two lanes).

Table 8 identifies the variables that were selected for modeling of rural, four-leg, STOP-controlled intersections. This table also provides descriptive statistics for three types of variables: (1) total and fatal and injury multiple-vehicle accident frequencies in the 3-year study period (i.e., the dependent variables for the modeling effort); (2) all independent continuous variables considered; and (3) all independent categorical variables considered. Minimum, mean, median, and maximum values are given for the first two types of variables. For categorical variables, the percent of intersections within each level is given.

Next, using all the continuous and categorical variables shown in table 8, a Poisson regression model was fit separately to the data for total multiple-vehicle and fatal and injury multiple-vehicle accidents. This model is referred to as the full model because all the candidate independent variables are included in the model. The Poisson

Table 8. Descriptive Statistics for Rural, Four-leg, STOP-controlled Intersections

Parameter	Level	Percent of intersections	Mininum	Mean	Median	Maximum
Total multiple-vehicle accidents; 1990 through 1992 combined			0	3.4	2	48
Fatal and injury multiple-vehicle accidents; 1990 through 1992 combined			0	1.7	1	27
Major-road ADT (veh/day)			400	8,262	6,646	72,000
Crossroad ADT (veh/day)			100	630	351	9,585
Design speed of major road (mi/h)			25	56	60	70
Outside shoulder width on major road (ft)			0	6.7	8	15
Average lane width on major road (ft)			8	12.0	12	15
Terrain	Flat	64				
	Rolling	26				
	Mountainous	11				
Functional class of major road	Principal arterial	27		**1,434 intersections**		
	Minor arterial	61				
	Major collector	12				
Lighting	No	62				
	Yes	39				
Major-road left-turn channelization	No left-turn lane	63				
	Painted left-turn lane	32				
	Curbed left-turn lane	4.5				
Major-road right-turn channelization	No free right turns	90				
	Provision for free right turns	10				
Number of lanes on major road	3 or less	83				
	4 or more	17				
Crossroad right-turn channelization	No free right turns	95				
	Provision for free right turns	4.8				
Presence of median on major road	Divided	19				
	Undivided	81				
Access control on major road	None	81				
	Partial	19				

Conversion: 1 km/h = 0.621 mi/h; 1 m = 3.28 ft

modeling was performed with the GENMOD procedure of the SAS statistical software package. This procedure fits generalized linear models as defined by Nelder and Wedderburn and uses the maximum likelihood procedure to estimate the values of the regression coefficients. The GENMOD procedure can be customized for a large number of distributions, including the Poisson distribution.

Generally, the analysis results for the full model found some independent variables to be statistically significant at the 10 percent significance level and other variables to be not statistically significant. To obtain the best estimates of the regression coefficients for the independent variables that are statistically significant and the best estimate of the goodness of fit of the model as a whole, the Poisson regression model was fit again, including only those independent variables that were found to be statistically significant in the full model (i.e., the independent variables that were found to be not statistically significant were dropped). This model is referred to as the reduced model.

The first three columns in table 9 show various model diagnostics for the full and reduced Poisson regression models. The model diagnostics, which are shown separately in table 9 for each type of accident considered, include both basic statistics and goodness-of-fit criteria. The following model statistics are shown:

<u>Model statistic</u>	<u>Explanation</u>
Basic Statistics	
Number of intersections, n	Total sample size in that category of intersections
Number of parameters in model	Total number of independent variables, both categorical and continuous
Parameters degrees of freedom, p	Each continuous independent variable has 1 degree of freedom; the number of degrees of freedom associated with each categorical variable equals the number of levels minus 1. The intercept has one degree of freedom. The sum of these degrees of freedom is denoted as p.
k factor	Only applicable to the negative binomial distribution. The use of this factor results in a ratio of the deviance to its degrees of freedom of approximately 1 (see section 5).

Model statistic	Explanation
Criteria for Assessing Goodness of Fit	
Deviance/(n − p)	The deviance of the model containing all the parameters (including the intercept) divided by its degrees of freedom, n − p. This statistic (mean deviance) provides a test for overdispersion and a measure of fit of the model. Asymptotically, this value tends toward 1.
Pearson chi-square/(n − p)	The Pearson chi-square statistic divided by its degrees of freedom, n − p. This statistic provides another measure of fit of the model. Asymptotically, this value tends toward 1. This statistic is referred to as the Pearson chi-square ratio in subsequent sections.
R^2	A goodness-of-fit parameter based on the ordinary multiple correlation coefficient.
R^2_{FT}	A goodness-of-fit parameter based on the Freeman-Tukey variance stabilizing transformation of variables discussed in Fridstrøm et al.(16)

Two goodness-of-fit measures, the mean deviance and the Pearson chi-square ratio (the Pearson chi-square value divided by its degrees of freedom), were used to assess the fit of the model. Generally, if the Pearson chi-square ratio is between 0.8 and 1.2, this is an indication that the model can be assumed to be appropriate in modeling the data. Table 9 shows that the Pearson chi-square ratio is approximately 3.0 for total multiple-vehicle accidents and approximately 2.2 for fatal and injury multiple-vehicle accidents, an indication that the model does not fit the data well. In addition, the mean deviance is approximately 2.7 for total multiple-vehicle accidents and 2.0 for fatal and injury multiple-vehicle accidents, an indication of overdispersion. Assuming that the Poisson model is appropriate, then the mean deviance should be close to one. If the mean deviance exceeds one, then the data are said to display extra variation or overdispersion relative to a Poisson model. That is, the variance in the data is in fact greater than the Poisson model, in which the mean equals the variance, indicates. If the mean deviance is less than one, the data are said to display underdispersion relative to a Poisson model.

Two additional goodness-of-fit criteria are provided by R^2 and R^2_{FT}. These values are each approximately 41 percent for the reduced and the full model, for total multiple-vehicle accidents. The two R^2-values are slightly lower at 34 percent and 32 percent, respectively, for fatal and injury multiple-vehicle accidents.

Table 9. Model Diagnostics for Total and Fatal and Injury Multiple-vehicle Accidents at Rural, Four-leg, STOP-controlled Intersections

	Poisson regression		Negative binomial regression (reduced model)
	Full model	Reduced model	
Total Multiple-vehicle Accidents (3-year period)			
Number of intersections (n)	1,434	1,434	1,434
Number of parameters in model	14	10	8
Parameters degrees of freedom[a] (p)	18	14	12
k factor	na	na	0.71
Deviance/(n - p)	2.74	2.73	1.01
Pearson chi-square/(n - p)	3.01	3.00	1.01
R^2 (%)	40.67	40.79	38.16
R^2_{FT} (%)	41.35	41.29	40.51
Fatal and Injury Multiple-vehicle Accidents (3-year period)			
Number of intersections (n)	1,434	1,434	1,434
Number of parameters in model	14	10	9
Parameters degrees of freedom[a] (p)	18	14	13
k factor	na	na	0.71
Deviance/(n - p)	1.97	1.96	1.00
Pearson chi-square/(n - p)	2.15	2.14	1.04
R^2 (%)	34.20	34.07	32.17
R^2_{FT} (%)	32.02	32.03	31.35

[a] Includes one degree of freedom for the intercept.

In situations in which the use of a Poisson regression model appears to be inappropriate because of overdispersion, negative binomial regression usually provides an appropriate alternative approach. In addition to its mean, μ, the negative binomial distribution includes another parameter, α, generally referred to as a dispersion parameter, which allows the variance to exceed the mean of the distribution. Therefore, the data were modeled using the negative binomial distribution. The analyses were performed using the SAS GENMOD procedure with the negative binomial distribution and the appropriate deviance functions and variance adjustment factor, k.(14) Only those independent variables that were used in the reduced Poisson model were included in the negative binomial model. The significance of each regression coefficient was examined. If a coefficient was not significant at the 10 percent level, the corresponding variable was deleted from the model, and the negative binomial regression was rerun.

The choice of a 10 percent significance level or 90 percent confidence level reflects a moderately restrictive approach in the selection of independent variables that might significantly contribute to the variability in accidents. Many previous accident research efforts have used the more restrictive 5 percent significance level, which would generally include fewer independent variables in the predictive models. Thus, the choice of a 10 percent level retained some variables that would not have been significant at the 5 percent level. Since this step in the effort of identifying significant variables serves primarily as a screening step, this approach was considered appropriate. While it was not considered appropriate to include independent variables with significance levels above 10 percent in the models presented in this report, the text of the report identifies those independent variables that were found to have significance levels between 10 percent and 20 percent. This significance level, α, is indicated for each such variable.

The final model statistics for the negative binomial regression model are shown in the last column of table 9. Of the 14 original independent variables considered, only 8 remain statistically significant in the final negative binomial model for total multiple-vehicle accidents. A variance stabilizing factor, k, of 0.71 was needed to achieve a mean deviance of approximately one, an indication that the data are neither overdispersed nor underdispersed relative to the model. The Pearson chi-square ratio now equals approximately one, a value within the acceptable range of 0.8 to 1.2. These two goodness-of-fit results provide an indication that the choice of the negative binomial model appears appropriate.

The two additional measures of goodness of fit, R^2 and R^2_{FT}, are approximately 38 percent and 41 percent, respectively, for total multiple-vehicle accidents. It should be noted that despite the marked improvement of the negative binomial model over the Poisson model, as shown by the reduced mean deviance and Pearson chi-square ratio, the R^2 statistics vary only slightly between the full Poisson, the reduced Poisson, and the negative binomial models; this illustrates that the use of R^2 statistics alone is inappropriate in assessing the goodness of fit of a model.

Of the six independent variables considered in the full Poisson model but not in the final reduced negative binomial model, only one variable—lighting (α=0.12)—was not significant at the 10 percent but would have been at the 20 percent level.

The use of the negative binomial model had a similar impact on the results for fatal and injury multiple-vehicle accidents. Of the 14 original independent variables considered, only 9 remain statistically significant in the final negative binomial model for fatal and injury multiple-vehicle accidents. A variance stabilizing factor, k, of 0.71 was needed to achieve a mean deviance of approximately one. The Pearson chi-square ratio now equals approximately one, a value within the acceptable range of 0.8 to 1.2. These two goodness-of-fit results provide an indication that the choice of the negative binomial model appears appropriate.

Of the five independent variables considered in the full Poisson model but not in the final reduced negative binomial model, none that was not significant at the 10 percent would have been significant at the 20 percent level.

The two additional measures of goodness of fit, R^2 and R^2_{FT}, are approximately 32 percent and 31 percent, respectively, for fatal and injury multiple-vehicle accidents. Again, it should be noted that despite the marked improvement of the negative binomial model over the Poisson model, as shown by the reduced mean deviance and Pearson chi-square ratio, the R^2 statistics vary only slightly between the full Poisson, the reduced Poisson, and the negative binomial models.

Tables 10 and 11 summarize the regression results for the final negative binomial model for total multiple-vehicle accidents and fatal and injury multiple-vehicle accidents, respectively. Each table identifies the:

- Statistically significant variables remaining in the final model.
- Chi-square statistic for each remaining variable; all of these chi-square statistics are statistically significant at the 10 percent significance level or better.
- Levels of each statistically significant categorical variable.
- Direction of the effect if the effect was inverse to the expected direction.
- Value of the regression coefficient for each continuous variable or each level of each categorical variable in the model.
- Relative effect of a unit change in each variable on the expected accident frequency in a 3-year period (this is simply e^{β}, where β is the coefficient given in the table).
- Lower and upper 90 percent confidence limits of the regression coefficient.

Table 10. Negative Binomial Regression Results for Total Multiple-vehicle Accidents at Rural, Four-leg, STOP-controlled Intersections

Independent variable[a]	Chi square statistic[b]	Variable level	Direction of effect[c]	Coefficient	Relative effect[d]	90% confidence limits[e]	
						Lower	Upper
Intercept				−11.246		−12.185	−10.317
Crossroad ADT (log)	386.95		–	0.586	1.80	0.535	0.637
Major-road ADT (log)	306.24		–	0.797	2.22	0.720	0.875
Number of lanes on major road	24.54	3 or less	–	0.463	1.59	0.311	0.615
		4 or more	–	0			
Design speed on major road	22.53			0.013	1.01	0.009	0.018
Functional class of major road	11.46	Principal arterial	–	0			
		Minor arterial	–	0.244	1.28	0.125	0.362
		Major collector	–	0.241	1.27	0.055	0.427
Access control on major road	8.98	None	–	0.268	1.31	0.121	0.414
		Partial	–	0			
Terrain	8.87	Flat	I	0.155	1.17	0.039	0.270
		Rolling	I	0			
		Mountainous	I	−0.101	0.90	−0.284	0.083
Major road left-turn channelization	5.37	No left-turn lane	–	0.091	1.10	−0.021	0.203
		Painted left-turn lane	–	0			
		Curbed left-turn lane	I	0.313	1.37	0.063	0.567

NOTE: This analysis is based on the set of 1,434 intersections for which summary statistics are shown in table 8.

[a] All variables significant at the 90% confidence level or higher.

[b] Chi-square likelihood ratio statistic for testing the significance of the effect of the variable; with 1 degree of freedom for continuous variables; with (p-1) degrees of freedom for categorical variables with p levels.

[c] Direction of effect: I = Inverse of expected direction.

[d] Relative effect of unit change in the variable on the expected number of accidents, equals exp(coefficient).

[e] 90% lower and upper confidence limits of the estimated coefficient.

Table 11. Negative Binomial Regression Results for Fatal and Injury Multiple-vehicle Accidents at Rural, Four-leg, STOP-controlled Intersections

Independent variable[a]	Chi square statistic[b]	Variable level	Direction of effect[c]	Coefficient	Relative effect[d]	90% confidence limits[e] Lower	90% confidence limits[e] Upper
Intercept				−11.116		−12.201	−10.044
Crossroad ADT (log)	299.05		–	0.602	1.82	0.542	0.661
Major-road ADT (log)	165.23		–	0.674	1.96	0.586	0.763
Number of lanes on major road	22.61	3 or less	–	0.509	1.66	0.333	0.685
		4 or more	–	0			
Design speed on major road	20.00		–	0.016	1.02	0.010	0.021
Terrain	18.74	Flat	I	0.254	1.29	0.122	0.386
		Rolling	I	0			
		Mountainous	I	−0.185	0.83	−0.405	0.033
Functional class of major road	9.57	Principal arterial	–	0			
		Minor arterial	–	0.250	1.28	0.115	0.385
		Major collector	–	0.154	1.17	-0.060	0.368
Major road left-turn channelization	7.36	No left-turn lane	I	−0.045	0.96	−0.171	0.081
		Painted left-turn lane	I	0			
		Curbed left-turn lane	I	0.424	1.53	0.148	0.703
Lighting	6.36	No	–	0.191	1.21	0.066	0.315
		Yes	–	0			
Access control on major road	3.63	None	–	0.190	1.21	0.026	0.354
		Partial	–	0			

NOTE: This analysis is based on the set of 1,434 intersections for which summary statistics are shown in table 8.

[a] All variables significant at the 90% confidence level or higher.

[b] Chi-square likelihood ratio statistic for testing the significance of the effect of the variable; with 1 degree of freedom for continuous variables; with (p-1) degrees of freedom for categorical variables with p levels.

[c] Direction of effect: I = Inverse of expected direction.

[d] Relative effect of unit change in the variable on the expected number of accidents, equals exp(coefficient).

[e] 90% lower and upper confidence limits of the estimated coefficient.

In each table, the independent variables are listed in decreasing order of their ability to explain the variations in intersection accident frequencies as indicated by the chi-square values, which represents the strength of the relationship of each variable to accident frequency, taking into account all other variables in the model.

To predict the average accident frequency at rural, four-leg, STOP-controlled intersections, one replaces the regression coefficients, β_0, β_1, β_2, ... β_q, with the estimated values found in the table, and the variables X_1, X_2, ..., X_q, with their appropriate values or levels. For example, the expected 3-year total multiple-vehicle accident frequency can be estimated using the model presented in table 10 as:

$$Y = e^{-11.246} (X_1)^{0.586} (X_2)^{0.797} \exp(0.463X_3) \exp(0.013X_4) \exp(0.244X_5) \exp(0.241X_6) \exp(0.268X_7) \exp(0.155X_8) \exp(-0.101X_9) \exp(0.091X_{10}) \exp(0.313X_{11}) \quad (15)$$

where

Y = expected number of total multiple-vehicle accidents in a 3-year period
X_1 = ADT of the crossroad (veh/day)
X_2 = ADT of the major road (veh/day)
X_3 = 1 if the major road has 3 or fewer lanes in both direction of travel combined; 0 if 4 or more
X_4 = design speed on major road (mi/h)
X_5 = 1 if the major road is a minor arterial; 0 otherwise
X_6 = 1 if the major road is a major collector; 0 otherwise
X_7 = 1 if the major road has no access control; 0 if access control is partial
X_8 = 1 if terrain is flat; 0 otherwise
X_9 = 1 if terrain is mountainous; 0 otherwise
X_{10} = 1 if no left-turn lane is present on the major road; 0 otherwise
X_{11} = 1 if curbed left-turn lane is present on the major road; 0 otherwise

Note that when the level of a categorical variable is 0, the multiplicative term in Eq. (15) becomes $e^0 = 1$, and is therefore omitted from the model.

The relative effect of each variable, all other variables being held constant, can be calculated by simply taking the exponent of the corresponding coefficient. For example, the relative effect of having a major road with three lanes or fewer as opposed to four lanes or more is exp(0.463) = 1.59. In other words, decreasing the number of lanes on the major road from 4 or more to 3 or less, with all other factors being held constant, would increase the expected number of accidents by a factor of 1.59 or by 59 percent. Similarly, intersections on major roads without access control were found to have 31 percent more accidents than intersections on major roads with partial access control.

The results of the negative binomial regression modeling shown in tables 10 and 11 show that most of the variables of interest have effects in the direction expected. However, the observed effect of major-road left-turn lanes is not in the expected direction since the results imply that intersections with left-turn lanes in a curbed median have more accidents than intersections without left turn lanes. In addition, the results suggest that intersections in flat terrain have more accidents that intersections in rolling terrain which, in turn, have more accidents than intersections in mountainous terrain. Such effects that are opposite to the direction expected can represent situations in which a variable for which data are available is correlated with and serves as a surrogate for another variable for which data are not available.

Figure 7 illustrates the influence of the major-road and crossroad ADT on the annual number of accidents at rural, four-leg, STOP-controlled intersections with the typical geometrics identified in the box at the upper right corner of the figure. Each curve in the figure represents combinations of major-road and crossroad ADT that would be expected to result in a specific annual number of multiple-vehicle accidents, ranging from 0.5 to 5 accidents per year.

Rural, Three-leg, STOP-controlled Intersections

The statistical analysis approach used for rural, three-leg, STOP-controlled intersections was identical to that used for rural, four-leg, STOP-controlled intersections (see section 5). The median number of total multiple-vehicle accidents at any one intersection was one accident in the 3-year study period with a maximum of 45 accidents in the 3-year period. As shown in figure 2 and in table 58 in appendix B, approximately 66 percent of all 2,692 rural, three-leg, STOP-controlled intersections in the study experienced either zero accident or one accident in the 3-year period. Thus, the Poisson model appeared to be a logical choice for analysis of this data set.

The selection of independent variables was done in a similar fashion to that described earlier in section 5. Table 12 identifies the variables that were selected for modeling accidents at rural, three-leg, STOP-controlled intersections. As before, a small number of independent categorical variables were not included in the full model because either all or nearly all intersections fell into one level of that variable. The variables originally considered that were not included for this reason were:

- Major-road left-turn prohibition (no intersections had left turns prohibited).
- Crossroad left-turn channelization (none of the intersections had left-turn lanes).
- Crossroad left-turn prohibition (none of the intersections had left turns prohibited).
- Number of lanes on the crossroad (99.7 percent had two lanes).

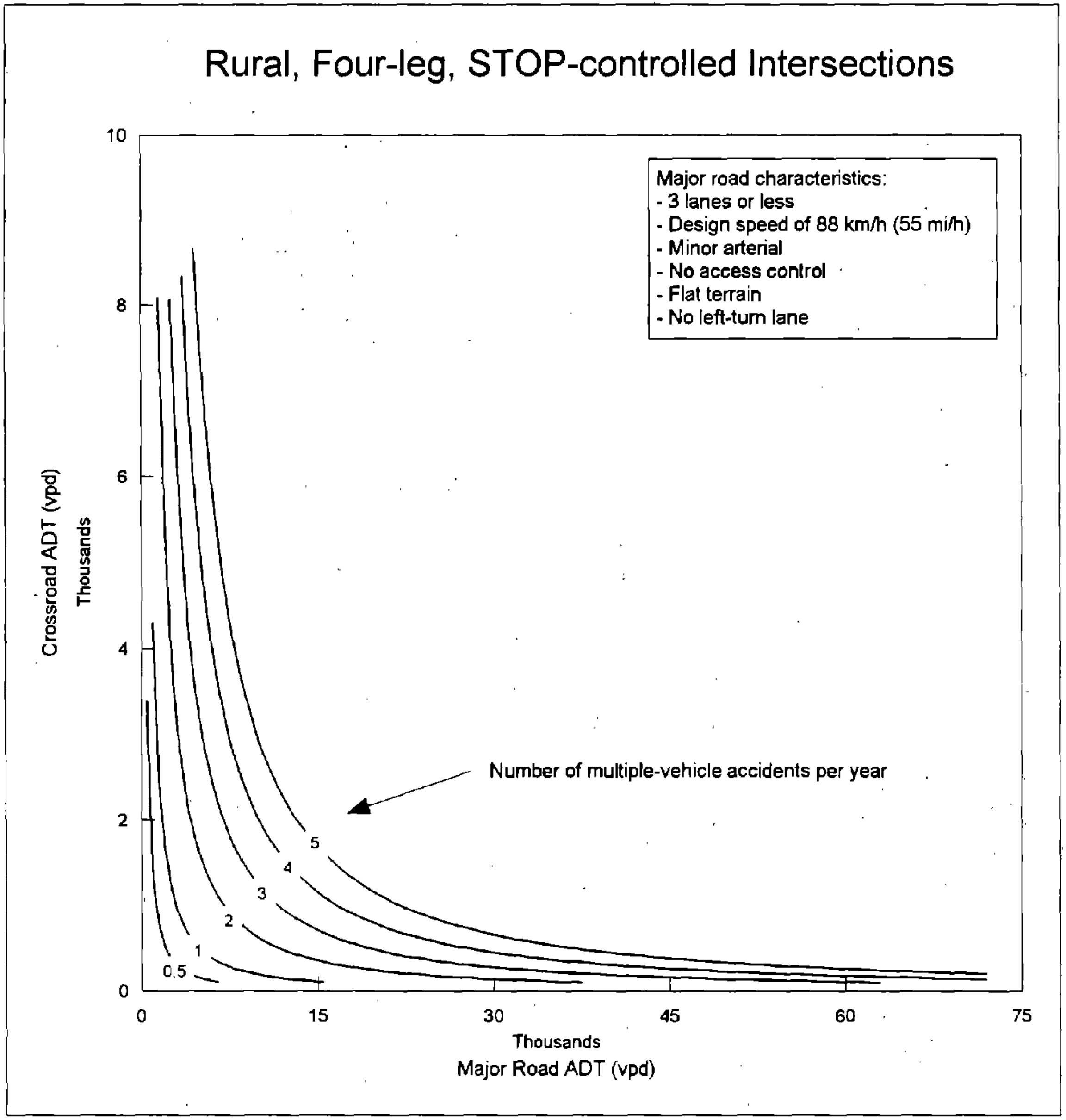


Figure 7. Number of Multiple-Vehicle Accidents per Year as a Function of Traffic Volumes for Typical Rural, Four-leg, STOP-Controlled Intersections

Table 12. Descriptive Statistics for Rural, Three-leg, STOP-controlled Intersections

Parameter	Level	Percent of intersections	Minimum	Mean	Median	Maximum
Total multiple-vehicle accidents; 1990 through 1992 combined			0	1.8	1	45
Fatal and injury multiple-vehicle accidents; 1990 through 1992 combined			0	0.8	0	21
Major-road ADT (veh/day)			400	8,288	6,138	72,000
Crossroad ADT (veh/day)			100	487	210	10,001
Design speed of major road (mi/h)			25	53	55	70
Outside shoulder width on major road (ft)			0	5.7	6	15
Average lane width on major road (ft)			8	11.8	12	15
Terrain	Flat	40				
	Rolling	35				
	Mountainous	25				
Functional class of major road	Principal arterial	20		**2,692 intersections**		
	Minor arterial	68				
	Major collector	12				
Lighting	No	70				
	Yes	30				
Major-road left-turn channelization	No left-turn lane	68				
	Painted left-turn lane	30				
	Curbed left-turn lane	2.1				
Major-road right-turn channelization	No free right turns	92				
	Provision for free right turns	7.7				
Number of lanes on major road	3 or less	88				
	4 or more	12				
Crossroad right-turn channelization	No free right turns	97				
	Provision for free right turns	2.6				
Presence of median on major road	Divided	16				
	Undivided	84				
Access control on major road	None	91				
	Partial	9.4				

Conversion: 1 km/h = 0.621 mi/h; 1 m = 3.28 ft

As shown in the table, 14 independent variables, both continuous and categorical, were considered in the full Poisson regression model. Of these 14 variables, 8 were found to have a statistically significant effect on accidents (both total and fatal and injury multiple-vehicle) at the 10 percent significance level. A reduced Poisson model was then rerun using only the eight statistically significant variables.

The first three columns in table 13 show the Poisson model statistics for both the full and the reduced model. For either model, the mean deviance is relatively large (approximately 2.1) for total multiple-vehicle accidents, indicating the presence of overdispersion in the data. Thus the Poisson model, which ideally would yield a ratio of one, appears to be inappropriate. This is further supported by the relatively large Pearson chi-square ratio (the chi-square value divided by its degrees of freedom) of approximately 2.6. The same conclusion can be drawn from the results for fatal and injury multiple-vehicle accidents, although the fatal and injury accident data have a smaller mean deviance and Pearson chi-square ratio (1.4 and 1.7, respectively) for both the full and reduced models.

Table 13. Model Diagnostics for Total and Fatal and Injury Multiple-vehicle Accidents at Rural, Three-leg, STOP-controlled Intersections

	Poisson regression		Negative binomial regression (reduced model)
	Full model	Reduced model	
Total Multiple-vehicle Accidents (3-year period)			
Number of intersections (n)	2,692	2,692	2,692
Number of parameters in model	14	8	5
Parameters degrees of freedom[a] (p)	18	12	8
k factor	na	na	0.70
Deviance/(n - p)	2.13	2.12	1.01
Pearson chi-square/(n - p)	2.59	2.58	1.17
R^2 (%)	36.76	36.65	35.16
R^2_{FT} (%)	36.71	36.68	36.26
Fatal and Injury Multiple-vehicle Accidents (3-year period)			
Number of intersections (n)	2,692	2,692	2,692
Number of parameters in model	14	8	6
Parameters degrees of freedom[a] (p)	18	11	9
k factor	na	na	0.38
Deviance/(n - p)	1.36	1.36	1.00
Pearson chi-square/(n - p)	1.73	1.73	1.30
R^2 (%)	29.28	28.99	27.82
R^2_{FT} (%)	26.41	26.26	25.92

[a] Includes one degree of freedom for the intercept.

Because of the overdispersion described above, a negative binomial regression model was then used with a variance adjustment factor, k, of 0.70 for total and 0.38 for fatal and injury multiple-vehicle accidents, respectively. In each case, this approach resulted in a mean deviance of approximately one. Also, the Pearson chi-square ratio was considerably reduced from 2.12 to 1.17 for total multiple-vehicle accidents and slightly less reduced from 1.73 to 1.30 for fatal and injury multiple-vehicle accidents. Note that the changes in both R^2 and R^2_{FT} were negligible for either type of accidents. Of the original 14 independent variables considered for modeling, only 5 remained statistically significant at the 90 percent confidence level for total multiple-vehicle accidents. A slightly different set of six variables remained statistically significant at the 90 percent confidence level for fatal and injury multiple-vehicle accidents.

Of the nine independent variables considered in the full Poisson model but not in the final reduced negative binomial model for total multiple-vehicle accidents, only one variable—terrain (α=0.13)—was not significant at the 10 percent level but would have been at the 20 percent level. Of the eight independent variables considered in the full Poisson model but not in the final reduced negative binomial model for fatal and injury multiple-vehicle accidents, only one variable—crossroad right-turn channelization—(α=0.13)—was not significant at the 10 percent level but would have been at the 20 percent level.

Tables 14 and 15 summarize the regression results for the final negative binomial model for total multiple-vehicle accidents and fatal and injury multiple-vehicle accidents, respectively. The tables show that for rural, three-leg, STOP-controlled intersections all of the independent variables evaluated had effects in the direction expected. For example, these data show that intersections with no separate left-turn lanes have 28 percent more accidents than intersections with left-turn lanes provided by painted channelization. Intersections with left-turn lanes in curbed medians had 7 percent fewer accidents than intersections with left-turn lanes provided by painted channelization.

Figure 8 illustrates the variations of the annual number of multiple-vehicle accidents with major-road and crossroad ADT for rural, three-leg, STOP-controlled intersections of the type specified at the upper right of the figure.

Table 14. Negative Binomial Regression Results for Total Multiple-vehicle Accidents at Rural, Three-leg, STOP-controlled Intersections

Independent variable[a]	Chi square statistic[b]	Variable level	Direction of effect[c]	Coefficient	Relative effect[d]	90% confidence limits[e]	
						Lower	Upper
Intercept				−11.364		−12.000	−10.736
Major-road ADT (log)	943.66		–	0.987	2.68	0.930	1.045
Crossroad road ADT (log)	333.82		–	0.429	1.54	0.389	0.469
Major-road left-turn channelization	19.86	No left-turn lane	–	0.249	1.28	0.154	0.344
		Painted left-turn lane	–	0			
		Curbed left-turn lane	–	−0.071	0.93	−0.334	0.197
Functional class of major road	10.30	Principal arterial	–	0			
		Minor arterial	–	0.201	1.22	0.098	0.304
		Major collector	–	0.196	1.22	0.030	0.363
Access control on major road	7.44	None	–	0.242	1.27	0.096	0.387
		Partial	–	0			

NOTE: This analysis is based on the set of 2,692 intersections for which summary statistics are shown in table 12.

[a] All variables significant at the 90% confidence level or higher.

[b] Chi-square likelihood ratio statistic for testing the significance of the effect of the variable; with 1 degree of freedom for continuous variables; with (p-1) degrees of freedom for categorical variables with p levels.

[c] Direction of effect: I = Inverse of expected direction.

[d] Relative effect of unit change in the variable on the expected number of accidents, equals exp(coefficient).

[e] 90% lower and upper confidence limits of the estimated coefficient.

Table 15. Negative Binomial Regression Results for Fatal and Injury Multiple-vehicle Accidents at Rural, Three-leg, STOP-controlled Intersections

Independent variable[a]	Chi square statistic[b]	Variable level	Direction of effect[c]	Coefficient	Relative effect[d]	90% confidence limits[e]	
						Lower	Upper
Intercept				−11.592		−12.315	−10.878
Major-road ADT (log)	650.23		–	0.953	2.59	0.888	1.020
Crossroad ADT (log)	282.17		–	0.439	1.55	0.396	0.483
Major-road left-turn channelization	10.34	No left-turn lane	–	0.195	1.22	0.093	0.298
		Painted left-turn lane	–	0			
		Curbed left-turn lane	–	−0.034	0.97	−0.296	0.223
Functional class of major road	8.60	Principal arterial	–	0			
		Minor arterial	–	0.178	1.19	0.069	0.287
		Major collector	–	0.258	1.29	0.077	0.438
Outside shoulder width on major road	5.68		–	−0.022	0.98	−0.037	−0.007
Lighting	4.07	No	–	0.115	1.12	0.021	0.209
		Yes	–	0			

NOTE: This analysis is based on the set of 2,692 intersections for which summary statistics are shown in table 12.

[a] All variables significant at the 90% confidence level or higher.

[b] Chi-square likelihood ratio statistic for testing the significance of the effect of the variable; with 1 degree of freedom for continuous variables; with (p-1) degrees of freedom for categorical variables with p levels.

[c] Direction of effect: I = Inverse of expected direction.

[d] Relative effect of unit change in the variable on the expected number of accidents, equals exp(coefficient).

[e] 90% lower and upper confidence limits of the estimated coefficient.

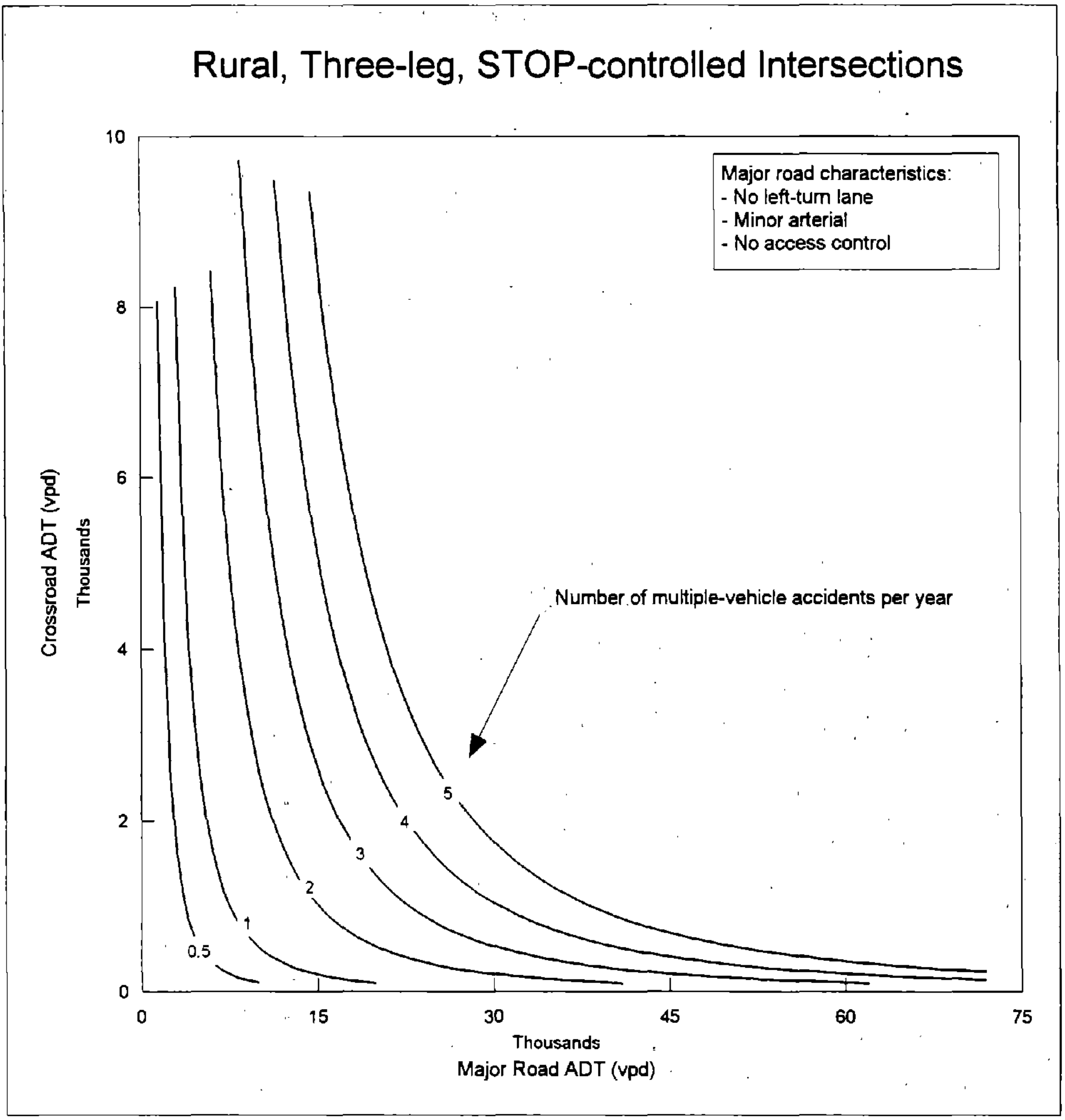


Figure 8. Number of Multiple-Vehicle Accidents per Year as a Function of Traffic Volumes for Typical Rural, Three-leg, STOP-Controlled Intersections

Urban, Four-leg, STOP-controlled Intersections

The statistical analysis approach used for urban, four-leg, STOP-controlled intersections was at first identical to that used for rural, four- and three-leg, STOP-controlled intersections. The median number of total multiple-vehicle accidents at any one intersection was four accidents in the 3-year study period with a maximum of 53 accidents in the 3-year period. As shown in figure 3 and in table 59 in appendix B, only approximately 10 percent of all 1,342 urban, four-leg, STOP-controlled intersections in the study experienced zero accidents in the 3-year period. Thus, the Poisson model was only a first attempt at modeling accidents in this data set.

The selection of independent variables was done in a similar fashion to that described earlier. Table 16 identifies the variables that were selected for modeling accidents at urban, four-leg, STOP-controlled intersections. As before, a small number of independent categorical variables were not included in the full model because either all or nearly all intersections fell into one level of that variable. The variables originally considered that were not included for this reason were:

- Crossroad left-turn prohibition (only 4 percent of the intersections had left turns prohibited; in addition, this variable showed a high negative correlation of -0.73 with the equivalent variable on the major road).

- Number of lanes on the crossroad (99.9 percent had two lanes).

As shown in the table, 16 independent variables, both continuous and categorical, were considered in the full Poisson regression model.

The first two columns in table 17 show the Poisson model statistics for the full model. The mean deviance is large (approximately 5) for total multiple-vehicle accidents, indicating the presence of considerable overdispersion in the data. Thus the Poisson model, which ideally would yield a ratio of one, appears to be inappropriate. This is further supported by the large Pearson chi-square ratio of approximately 5.7. The same conclusion can be drawn from the results for fatal and injury multiple-vehicle accidents, although the fatal and injury accident data have a smaller mean deviance and Pearson chi-square ratio (2.73 and 3.06, respectively) for both the full and reduced models. Also, the two R^2-values are relatively low at approximately 14.6 percent and 16.6 percent, respectively, for total multiple-vehicle accidents. These statistics are also low (14.4 percent and 15.7 percent) for fatal and injury multiple-vehicle accidents.

Based on these poor Poisson regression results and the shape of the accident data distributions (see figure 3), a lognormal regression model was used next to model accidents at this type of intersections. The natural logarithm of the accident counts was modeled using the full set of 16 independent variables. All modeling was performed using the SAS stepwise regression procedure.

Table 16. Descriptive Statistics for Urban, Four-Leg, STOP-controlled Intersections

Parameter	Level	Percent of intersections	Mininum	Mean	Median	Maximum
Total multiple-vehicle accidents; 1990 through 1992 combined			0	6.6	4	53
Fatal and injury multiple-vehicle accidents; 1990 through 1992 combined			0	3.0	2	23
Major-road ADT (veh/day)			1,100	23,240	21,217	79,000
Crossroad ADT (veh/day)			100	1,255	900	16,940
Design speed of major road (mi/h)			25	50	50	70
Outside shoulder width on major road (ft)			0	7.2	8	15
Average lane width on major road (ft)			8	12.1	12	15
Terrain	Flat	75				
	Rolling or mountainous	25				
Functional class of major road	Principal arterial	91				
	Minor arterial	7.7				
	Major collector	1.6				
Lighting	No	15				
	Yes	85				
Major-road left-turn channelization	No left-turn lane	44				
	Painted left-turn lane	40				
	Curbed left-turn lane	16				
Major-road right-turn channelization	No free right turns	96				
	Provision for free right turns	4.2				
Major-road left-turn prohibition	Left turns permitted	97				
	Left turns prohibited	3.1				
Number of lanes on major road	3 or less	31				
	4 or 5	61				
	6 or more	7.9				
Crossroad left-turn channelization	No left-turn lane	98				
	Painted left-turn lane	2.0				
Crossroad right-turn channelization	No free right turns	97				
	Provision for free right turns	3.3				
Presence of median on major road	Divided	55				
	Undivided	45				
Access control on major road	None	96				
	Partial	4.0				

1,342 Intersections

Table 17. Model Diagnostics for Total and Fatal and Injury Multiple-vehicle Accidents at Urban, Four-leg, STOP-controlled Intersections

	Poisson regression (full model)	Lognormal regression	
		Full model	Reduced model
Total Multiple-vehicle Accidents (3-year period)			
Number of intersections (n)	1,342	1,342	1,342
Number of parameters in model	16	16	8
Parameters degrees of freedom[a] (p)	20	20	10
Deviance/(n - p)	5.02	1.00	1.00
Pearson chi-square/(n - p)	5.74	1.00	1.00
R^2 (%)	14.57%	20.54%	20.58%
R^2_{FT} (%)	16.62%	na	na
Root mean square error	na	1.00	1.00
Fatal and Injury Multiple-vehicle Accidents (3-year period)			
Number of intersections (n)	1,342	1,342	1,342
Number of parameters in model	16	16	8
Parameters degrees of freedom[a] (p)	20	20	10
Deviance/(n - p)	2.73	0.80	0.80
Pearson chi-square/(n - p)	3.06	0.80	0.80
R^2 (%)	14.44%	18.17%	18.06%
R^2_{FT} (%)	15.69%	na	na
Root mean square error	na	0.89	0.90

[a] Includes one degree of freedom for the intercept.

The last two columns of table 17 show the model statistics for the full and reduced lognormal regression models. In this case, the root mean squared error has been added as a measure of fit of the model to the data. This statistic provides an estimate of the standard deviation of the error term (on the log scale).

The mean deviance and the Pearson chi-square ratio have each considerably decreased, an indication that the lognormal model appears to provide a better fit than the Poisson model. The R^2-values have slightly increased to approximately 21 percent for total multiple-vehicle accidents and to approximately 18 percent for fatal and injury multiple-vehicle accidents. These measure of fit, however, are relatively poor compared to those obtained for the previous types of intersections.

Of the original 16 independent variables considered for modeling, only 8 remained statistically significant at the 90 percent confidence level for total multiple-vehicle accidents. A slightly different set of eight variables remained statistically significant at the 90 percent confidence level for fatal and injury multiple-vehicle accidents.

Of the eight independent variables considered in the full but not in the final reduced lognormal model for total multiple-vehicle accidents, only one variable—outside shoulder width (α=0.13)—was not significant at the 10 percent level but would have been at the 20 percent level. Of the eight independent variables considered in the full but not in the final reduced lognormal model for fatal and injury multiple-vehicle accidents, only one variable—crossroad right-turn channelization (α=0.15)—was not significant at the 10 percent level but would have been at the 20 percent level.

Tables 18 and 19 summarize the regression results for the final negative binomial model for total multiple-vehicle accidents and fatal and injury multiple-vehicle accidents, respectively. The tables indicate that no statistically significant effect on accidents was found for major-road left-turn channelization. The average lane width on the major road was found to have an effect on intersection accidents in the expected direction [i.e., for each decrease of 0.3 m (1 ft) in lane width on the major-road approaches, multiple-vehicle intersection-related accidents increased by 9.1 percent]. Three of the variables evaluated had effects that were inverse to the direction expected: access control on the major road, crossroad right-turn channelization, and intersection lighting.

Figure 9 illustrates the variation of the annual number of multiple-vehicle intersection accidents with major-road and crossroad ADT for urban, four-leg, STOP-controlled intersections with the typical conditions specified in the figure.

Urban, Three-leg, STOP-controlled Intersections

The statistical analysis approach used for urban, three-leg, STOP-controlled intersections was identical to that used for rural, four- and three-leg, STOP-controlled intersections. The median number of total multiple-vehicle accidents at any 1 intersection was 2 accidents in the 3-year study period with a maximum of 80 accidents in the 3-year period. As shown in figure 4 and in table 60 in Appendix B, approximately 53 percent of all 3,057 urban, three-leg, STOP-controlled intersections in the study experienced two or fewer accidents in the 3-year period. Thus, the Poisson model appeared to be a logical choice for analysis of this data set.

The selection of independent variables was done in a similar fashion to that described in earlier sections. Table 20 identifies the variables that were selected for modeling accidents at urban, three-leg, STOP-controlled intersections. None of the variables considered for modeling were deleted due to small sample sizes. Although the percentage of intersections in some levels are relatively small (e.g, 0.7 percent of intersections had four or more lanes), the large number of intersections (3,057) in this category justified the inclusion of these variables and their levels in the analysis.

Table 18. Lognormal Regression Results for Total Multiple-vehicle Accidents at Urban, Four-leg, STOP-controlled Intersections

Independent variable[a]	Chi square statistic[b]	Variable level	Direction of effect[c]	Coefficient	Relative effect[d]	90% confidence limits[e]	
						Lower	Upper
Intercept				-5.073		-6.185	-3.961
Major-road ADT (log)	139.86		–	0.635	1.89	0.546	0.723
Crossroad ADT (log)	97.44		–	0.294	1.34	0.245	0.343
Major-road left-turn prohibition	33.60	Left turns prohibited	–	–0.969	0.38	–1.245	–0.694
		Left turns permitted	–	0			
Access control on major road	12.00	None	I	–0.518	0.60	–0.764	–0.272
		Partial	I	0			
Average lane width on major road	14.27		–	–0.091	0.91	–0.130	–0.051
Number of lanes on major road	12.29	3 or less	–	0.340	1.40	0.119	0.560
		4 or 5	–	0.087	1.09	–0.097	0.271
		6 or more	–	0			
Crossroad right-turn channelization	4.48	No free right turns	I	–0.331	0.72	–0.589	–0.074
		Provision for free right turns	I	0			
Lighting	4.16	No	I	–0.175	0.84	–0.316	–0.034
		Yes	I	0			

NOTE: This analysis is based on the set of 1,342 intersections for which summary statistics are shown in table 16.

[a] All variables significant at the 90% confidence level or higher.

[b] Chi-square likelihood ratio statistic for testing the significance of the effect of the variable; with 1 degree of freedom for continuous variables; with (p-1) degrees of freedom for categorical variables with p levels.

[c] Direction of effect: I = Inverse of expected direction.

[d] Relative effect of unit change in the variable on the expected number of accidents, equals exp(coefficient).

[e] 90% lower and upper confidence limits of the estimated coefficient.

Table 19. Lognormal Regression Results for Fatal and Injury Multiple-vehicle Accidents at Urban, Four-leg, STOP-controlled Intersections

Independent variable[a]	Chi square statistic[b]	Variable level	Direction of effect[c]	Coefficient	Relative effect[d]	90% confidence limits[e] Lower	90% confidence limits[e] Upper
Intercept				–4.745		–5.835	–3.654
Major-road ADT (log)	114.17		–	0.573	1.77	0.485	0.661
Crossroad ADT (log)	53.61		–	0.216	1.24	0.167	0.264
Major-road left-turn prohibition	21.09	Left turns prohibited	–	–0.768	0.46	–1.043	–0.493
		Left turns permitted	–	0			
Access control on major road	7.97	None	I	–0.398	0.67	–0.629	–0.166
		Partial	I	0			
Average lane width on major road	11.29		–	–0.081	0.92	–0.120	–0.041
Number of lanes on major road	6.84	3 or less	–	0.234	1.26	0.013	0.454
		4 or 5	–	0.044	1.04	–0.142	0.229
		6 or more	–	0			
Outside shoulder width on major road	5.30		–	–0.019	0.98	–0.032	–0.005
Crossroad right-turn channelization	3.29	No free right turns	I	–0.284	0.75	–0.542	–0.027
		Provision for free right turns	I	0			

NOTE: This analysis is based on the set of 1,342 intersections for which summary statistics are shown in table 16.

[a] All variables significant at the 90% confidence level or higher.

[b] Chi-square likelihood ratio statistic for testing the significance of the effect of the variable; with 1 degree of freedom for continuous variables; with (p-1) degrees of freedom for categorical variables with p levels.

[c] Direction of effect: I = Inverse of expected direction.

[d] Relative effect of unit change in the variable on the expected number of accidents, equals exp(coefficient).

[e] 90% lower and upper confidence limits of the estimated coefficient.

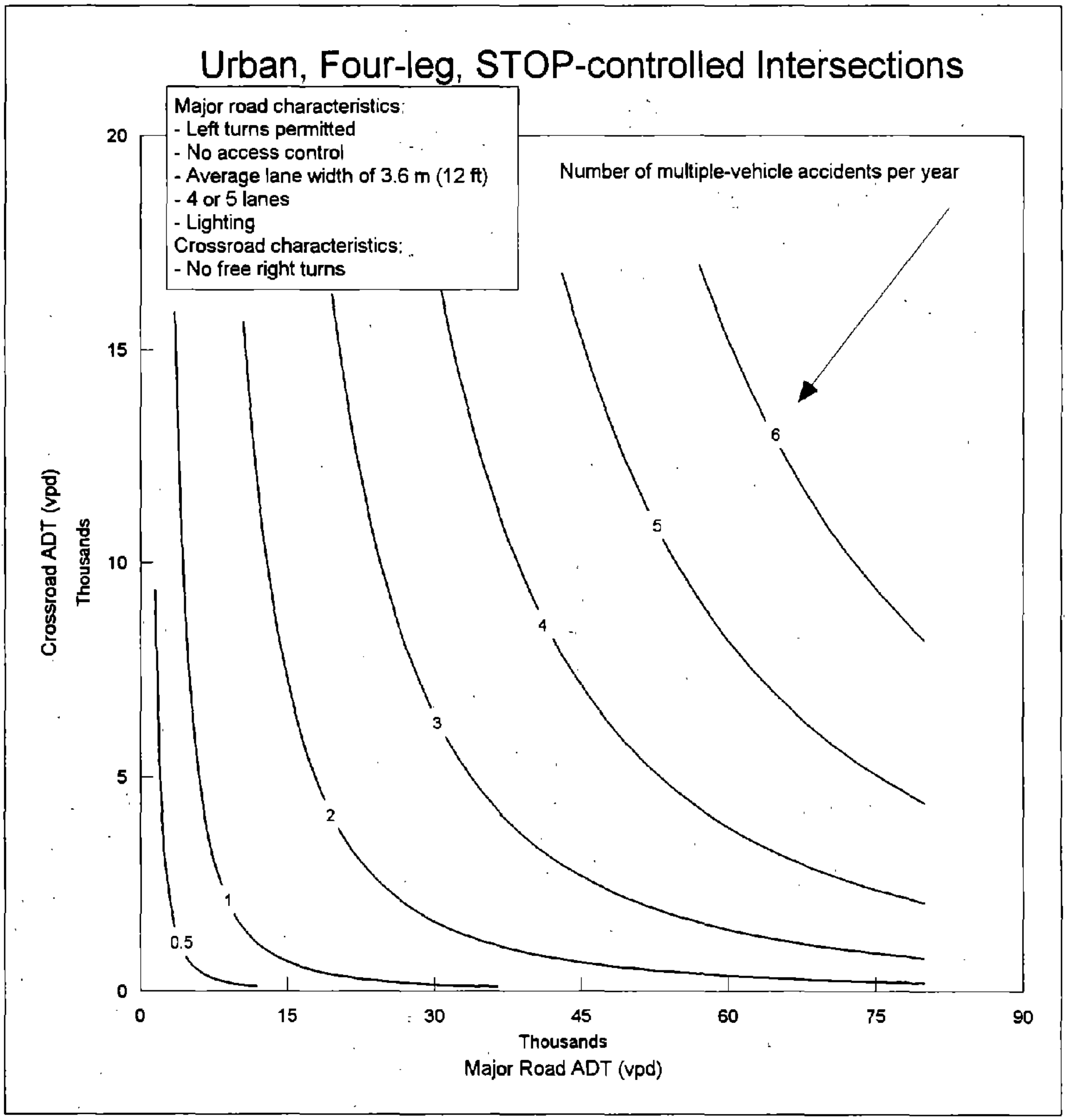


Figure 9. Number of Multiple-Vehicle Accidents per Year as a Function of Traffic Volumes for Typical Urban, Four-leg, STOP-Controlled Intersections

Table 20. Descriptive Statistics for Urban, Three-Leg, STOP-controlled Intersections

Parameter	Level	Percent of intersections	Minimum	Mean	Median	Maximum
Total multiple-vehicle accidents; 1990 through 1992 combined			0	3.9	2	80
Fatal and injury multiple-vehicle accidents; 1990 through 1992 combined			0	1.7	1	28
Major-road ADT (veh/day)			520	25,557	23,400	97,000
Crossroad ADT (veh/day)			100	808	501	21,800
Design speed of major road (mi/h)			25	50	50	70
Outside shoulder width on major road (ft)			0	7.0	8	15
Average lane width on major road (ft)			8	12.0	12	15
Terrain	Flat	72				
	Rolling or mountainous	29				
Functional class of major road	Principal arterial	90				
	Minor arterial	10				
Lighting	No	23				
	Yes	77				
Major-road left-turn channelization	No left-turn lane	50				
	Painted left-turn lane	36				
	Curbed left-turn lane	14				
Major-road right-turn channelization	No free right turns	97				
	Provision for free right turns	3.5				
Major-road left-turn prohibition	Left turns permitted	87				
	Left turns prohibited	13				
Number of lanes on major road	3 or less	32				
	4 or 5	56				
	6 or more	12				
Crossroad left-turn channelization	No left-turn lane	99				
	Painted or curbed left-turn lane	1.4				
Crossroad right-turn channelization	No free right turns	97				
	Provision for free right turns	3.0				
Crossroad left-turn prohibition	Left turns permitted	86				
	Left turns prohibited	13				
Number of lanes on crossroad	3 or less	99				
	4 or more	0.7				
Presence of median on major road	Divided	61				
	Undivided	39				
Access control on major road	None	98				
	Partial	2.4				

3,057 Intersections

As shown in the table, 18 independent variables, both continuous and categorical, were considered in the full Poisson regression model. Of these 18 variables, 13 were found to have a statistically significant effect on total multiple-vehicle accidents, and 11 variables were found to have a statistically significant effect on fatal and injury multiple-vehicle accidents at the 10 percent significance level. A reduced Poisson model was then rerun using only the 13 and 11 statistically significant variables, respectively.

The first three columns in table 21 show the Poisson model statistics for both the full and the reduced model. For either model, the mean deviance is relatively large (approximately 4) for total multiple-vehicle accidents, indicating the presence of overdispersion in the data. Thus the Poisson model, which ideally would yield a ratio of one, appears to be inappropriate. This is further supported by the relatively large Pearson chi-square ratio (the chi-square value divided by its degrees of freedom) of approximately 5.1. The same conclusion can be drawn from the results for fatal and injury multiple-vehicle accidents, although the fatal and injury accident data have a smaller mean deviance and Pearson chi-square ratio (2.1 and 2.5, respectively) for both the full and reduced models.

Table 21. Model Diagnostics for Total and Fatal and Injury Multiple-vehicle Accidents at Urban, Three-leg, STOP-controlled Intersections

	Poisson regression		Negative binomial regression (reduced model)
	Full model	Reduced model	
Total Multiple-vehicle Accidents (3-year period)			
Number of intersections (n)	3,057	3,057	3,057
Number of parameters in model	18	13	8
Parameters degrees of freedom[a] (p)	22	17	10
k factor	na	na	0.98
Deviance/(n - p)	3.99	3.98	1.00
Pearson chi-square/(n - p)	5.14	5.14	1.13
R^2 (%)	17.58	17.48	16.07
R^2_{FT} (%)	18.53	18.47	17.64
Fatal and Injury Multiple-vehicle Accidents (3-year period)			
Number of intersections (n)	3,057	3,057	3,057
Number of parameters in model	18	11	8
Parameters degrees of freedom[a] (p)	22	15	10
k factor	na	na	0.81
Deviance/(n - p)	2.12	2.12	1.01
Pearson chi-square/(n - p)	2.45	2.45	1.07
R^2 (%)	17.57	17.56	16.30
R^2_{FT} (%)	16.93	16.84	16.38

[a] Includes one degree of freedom for the intercept.

Because of the overdispersion described above, a negative binomial regression model was then used with a variance adjustment factor, k, of 0.98 for total and 0.81 for fatal and injury multiple-vehicle accidents, respectively. In each case, this approach resulted in a mean deviance of approximately one. Also, the Pearson chi-square ratio was considerably reduced from 5.14 to 1.13 for total multiple-vehicle accidents and from 2.45 to 1.07 for fatal and injury multiple-vehicle accidents. Note that the changes in both R^2 and R^2_{FT} were negligible for either type of accidents. For all regression models and both types of accidents, the models produced relatively poor results based on the two R^2-values, all in the range of 16 to 17.5 percent.

Of the original 18 independent variables considered for modeling, only 8 remained statistically significant at the 90 percent confidence level for total multiple-vehicle accidents. A slightly different set of eight variables remained statistically significant at the 90 percent confidence level for fatal and injury multiple-vehicle accidents.

Of the 10 independent variables considered in the full Poisson model but not in the final reduced negative binomial model for total multiple-vehicle accidents, five variables were not significant at the 10 percent level but would have been at the 20 percent level. These variables were outside shoulder width (α=0.16); number of lanes on major road (α=0.17); crossroad left-turn prohibition (α=0.13); number of lanes on crossroad (α=0.14); and access control on major road (α=0.15). Of the 10 independent variables considered in the full Poisson model but not in the final reduced negative binomial model for fatal and injury multiple-vehicle accidents, five variables were not significant at the 10 percent level but would have been at the 20 percent level. These variables were outside shoulder width (α=0.19); lighting (α=0.12); number of lanes on major road (α=0.11); crossroad left-turn prohibition (α=0.12); and number of lanes on crossroad (α=0.19).

Tables 22 and 23 summarize the regression results for the final negative binomial model for total multiple-vehicle accidents and fatal and injury multiple-vehicle accidents, respectively. The tables show an effect in the expected direction for the presence of a median on the major road; intersections on divided highways appear to have 15 percent fewer accidents than intersections on undivided highways. A concern with the models developed is that the effect on safety of major-road left-turn channelization is opposite to the direction expected. Other variables whose effects were found to be opposite to the direction expected were crossroad right-turn channelization and, in the models for fatal-and-injury accidents, access control on the major road.

Figure 10 shows the variation of the annual number of multiple-vehicle intersection accidents with major-road and crossroad ADT for urban, three-leg, STOP-controlled intersections with the typical conditions specified in the figure.

Table 22. Negative Binomial Regression Results for Total Multiple-vehicle Accidents at Urban, Three-leg, STOP-controlled Intersections

Independent variable[a]	Chi square statistic[b]	Variable level	Direction of effect[c]	Coefficient	Relative effect[d]	90% confidence limits[e]	
						Lower	Upper
Intercept				−6.808		−7.604	−6.014
Major-road ADT (log)	354.17		–	0.775	2.17	0.707	0.843
Crossroad ADT (log)	151.56		–	0.266	1.30	0.230	0.302
Major-road left-turn prohibition	33.61	Left turns prohibited	–	−0.478	0.62	−0.613	−0.343
		Left turns permitted	–	0			
Crossroad right-turn channelization	29.68	No free right turns	I	−0.601	0.55	−0.796	−0.412
		Provision for free right turns	I	0			
Major-road left-turn channelization	9.21	No left-turn lane	–	0.012	1.01	−0.090	0.113
		Painted left-turn lane	–	0			
		Curbed left-turn lane	I	0.192	1.21	0.086	0.298
Design speed of major road	7.00		–	−0.006	0.99	−0.009	−0.002
Presence of median on major road	6.27	Divided	–	−0.160	0.85	−0.266	−0.055
		Undivided	–	0			
Average lane width on major road	2.59		–	−0.030	0.97	−0.061	0.001

NOTE: This analysis is based on the set of 3,057 intersections for which summary statistics are shown in table 20.

[a] All variables significant at the 90% confidence level or higher.

[b] Chi-square likelihood ratio statistic for testing the significance of the effect of the variable; with 1 degree of freedom for continuous variables; with (p-1) degrees of freedom for categorical variables with p levels.

[c] Direction of effect: I = Inverse of expected direction.

[d] Relative effect of unit change in the variable on the expected number of accidents, equals exp(coefficient).

[e] 90% lower and upper confidence limits of the estimated coefficient.

Table 23. Negative Binomial Regression Results for Fatal and Injury Multiple-vehicle Accidents at Urban, Three-leg, STOP-controlled Intersections

Independent variable[a]	Chi square statistic[b]	Variable level	Direction of effect[c]	Coefficient	Relative effect[d]	90% confidence limits[e] Lower	90% confidence limits[e] Upper
Intercept				–7.358		–8.239	–6.483
Major-road ADT (log)	292.92		–	0.766	2.15	0.690	0.842
Crossroad ADT (log)	121.04		–	0.254	1.29	0.216	0.292
Major-road left-turn prohibition	26.76	Left turns prohibited	–	–0.458	0.63	–0.604	–0.312
		Left turns permitted	–	0			
Crossroad right-turn channelization	24.35	No free right turns	I	–0.575	0.56	–0.773	–0.380
		Provision for free right turns	I	0			
Major-road left-turn channelization	10.73	No left-turn lane	I	–0.055	0.95	–0.163	0.054
		Painted left-turn lane	I	0			
		Curbed left-turn lane	I	0.194	1.21	0.085	0.304
Presence of median on major road	7.37	Divided	–	–0.187	0.83	–0.301	–0.074
		Undivided	–	0			
Average lane width on major road	4.33		–	–0.042	0.96	–0.076	–0.009
Access control on major road	3.02	None	I	–0.234	0.79	–0.461	–0.013
		Partial	I	0			

NOTE: This analysis is based on the set of 3,057 intersections for which summary statistics are shown in table 20.

[a] All variables significant at the 90% confidence level or higher.

[b] Chi-square likelihood ratio statistic for testing the significance of the effect of the variable; with 1 degree of freedom for continuous variables; with (p-1) degrees of freedom for categorical variables with p levels.

[c] Direction of effect: I = Inverse of expected direction.

[d] Relative effect of unit change in the variable on the expected number of accidents, equals exp(coefficient).

[e] 90% lower and upper confidence limits of the estimated coefficient.

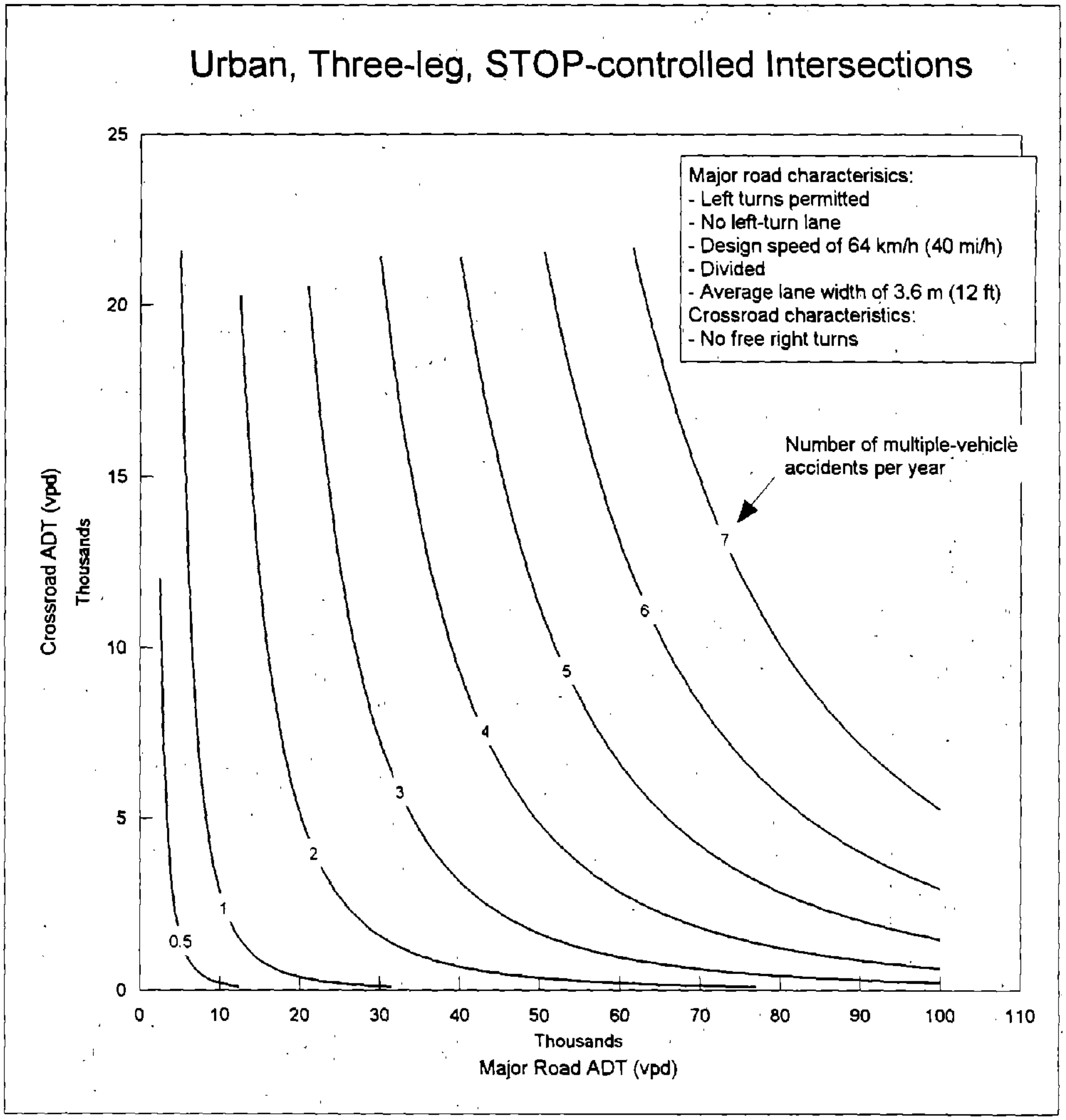


Figure 10. Number of Multiple-Vehicle Accidents per Year as a Function of Traffic Volumes for Typical Urban, Three-leg, STOP-Controlled Intersections

Urban, Four-leg, Signalized Intersections

Accident frequencies at urban, four-leg, signalized intersections are listed in table 61 in Appendix B and their distributions are shown in figure 5, for both total and fatal and injury multiple-vehicle accidents. A clear departure from the Poisson distribution is visible in the distribution plots. Only 21 out of 1,306 intersections (or 1.6 percent) experienced zero accident in the 3-year study period. Approximately half of all intersections in this category experienced 17 accidents or more in the 3-year period, with a maximum of 147 total multiple-vehicle accidents. Given these high accidents frequencies and the shape of the distribution for both types of accidents, a lognormal regression model presented a logical choice. Thus, the statistical analysis approach used for urban, four-leg, signalized intersections was identical to that ultimately used for urban, four-leg, STOP-controlled intersections.

The selection of independent variables was done in a similar fashion to that described earlier in section 5. Table 24 identifies the variables that were selected for modeling accidents at urban, four-leg, signalized intersections. As before, a small number of independent categorical variables was not included in the full model because either all or nearly all intersections fell into one level of that variable. The variables originally considered that were not included for this reason were:

- Lighting (all intersection were lighted).
- Presence of major-road signal mast arm (a mast arm was present on all intersections).
- Major-road left-turn prohibition (no intersections had left turns prohibited).

The natural logarithm of the accident counts was modeled using the full set of 19 independent variables listed in table 24. All modeling was performed using the SAS stepwise regression procedure.

Of the original 19 independent variables considered for modeling, only 8 remained statistically significant at the 90 percent confidence level for total multiple-vehicle accidents. A slightly different set of seven variables remained statistically significant at the 90 percent confidence level for fatal and injury multiple-vehicle accidents. The lognormal model was rerun using only the statistically significant variables to obtain the regression coefficients, their 90 percent confidence intervals, and other regression statistics.

Table 24. Descriptive Statistics for Urban, Four-Leg, Signalized Intersections

Parameter	Level	Percent of intersections	Minimum	Mean	Median	Maximum
Total multiple-vehicle accidents; 1990 through 1992 combined			0	21.8	18	147
Fatal and injury multiple-vehicle accidents; 1990 through 1992 combined			0	9.1	8	50
Major-road ADT (veh/day)			2,400	31,995	31,000	79,000
Crossroad ADT (veh/day)			101	8,061	5,501	48,000
Design speed of major road (mi/h)			25	51	50	70
Outside shoulder width on major road (ft)			0	7.0	8	15
Average lane width on major road (ft)			8	12.0	12	15
Terrain	Flat	80				
	Rolling or mountainous	20				
Functional class of major road	Principal arterial	96		**1,306 intersections**		
	Minor arterial	4.1				
Signal timing	Pretimed	2.5				
	Semiactuated	13				
	Fully actuated	85				
Signal phasing	Two-phase	21				
	Multiphase	79				
Major-road left-turn channelization	No left-turn lane	4.7				
	Painted left-turn lane	40				
	Curbed left-turn lane	56				
Major-road right-turn channelization	No free right turns	74				
	Provision for free right turns	26				
Number of lanes on major road	3 or less	7.9				
	4 or 5	72				
	6 or more	20				
Presence of crossroad signal mast arm	Mast arm not present	27				
	Mast arm present	73				

Table 24. Descriptive Statistics for Urban, Four-Leg, Signalized Intersections (Continued)

Parameter	Level	Percent of intersections	Minimum	Mean	Median	Maximum
Crossroad left-turn channelization	No left-turn lane	45				
	Painted left-turn lane	36				
	Curbed left-turn lane	19				
Crossroad right-turn channelization	No free right turns	70				
	Provision for free right turns	30				
Crossroad left-turn prohibition	Left turns prohibited	0.5				
	Left turns permitted	97				
Number of lanes on crossroad	3 or less	59				
	4 or 5	38				
	6 or more	3.0				
Presence of median on major road	Divided	83				
	Undivided	18				
Access control on major road	None	94				
	Partial	5.9				

Table 25 presents the model statistics for the full and reduced lognormal models. Although the lognormal models are significant at the 90 percent confidence level for both types of accidents, the percent variance explained by the model is relatively low, with an R^2-value of 25 percent for total multiple-vehicle accidents and approximately 24 percent for fatal and injury multiple-vehicle accidents. Also, the Pearson chi-square ratios (0.72 and 0.70, respectively, for both types of accidents) are below the 0.8 to 1.2 range, indicating that the lognormal model might not provide the best fit.

Of the 10 independent variables considered in the full but not in the final reduced lognormal model for total multiple-vehicle accidents, only one variable—number of lanes on crossroad (α=0.20)—was not significant at the 10 percent level but would have been at the 20 percent level. Of the 11 independent variables considered in the full but not in the final reduced lognormal model for fatal and injury multiple-vehicle accidents, only one variable—average lane width on major road (α=0.18)—was not significant at the 10 percent level but would have been at the 20 percent level.

Tables 26 and 27 summarize the regression results for the final lognormal model for total multiple-vehicle accidents and fatal and injury multiple-vehicle accidents, respectively. No statistically significant effect on accidents was found for either the major-road or crossroad left-turn channelization variable. However, it should be noted that major-road left-turn channelization could not be evaluated effectively because only 5 percent of the intersections had no left-turn lanes on the major road approaches. Only two variables appeared to have effects in the direction opposite to that expected: access control on the major road and major-road right-turn channelization.

Figure 11 shows the variation of the annual number of multiple-vehicle intersection accidents with major-road and crossroad ADT for urban, four-leg, signalized intersections with the typical conditions specified in the figure. As shown in the figure, typical accident experience at these intersections, in the range of data for which model predictions appear valid, extends up to 12 multiple-vehicle accidents per year.

Table 25. Model Diagnostics for Total and Fatal and Injury Multiple-vehicle Accidents at Urban, Four-leg, Signalized Intersections

	Lognormal regression	
	Full model	Reduced model
Total Multiple-vehicle Accidents (3-year period)		
Number of intersections (n)	1,306	1,306
Number of parameters in model	19	9
Parameters degrees of freedom[a] (p)	23	12
Deviance/(n - p)	0.72	0.72
Pearson chi-square/(n - p)	0.72	0.72
R^2 (%)	24.34	25.08
R^2_{FT} (%)	na	na
Root mean squared error	0.85	0.85
Fatal and Injury Multiple-vehicle Accidents (3-year period)		
Number of intersections (n)	1,306	1,306
Number of parameters in model	19	8
Parameters degrees of freedom[a] (p)	23	11
Deviance/(n - p)	0.70	0.70
Pearson chi-square/(n - p)	0.70	0.70
R^2 (%)	23.49	24.31
R^2_{FT} (%)	na	na
Root mean squared error	0.84	0.84

[a] Includes one degree of freedom for the intercept.

Table 26. Lognormal Regression Results for Total Multiple-vehicle Accidents at Urban, Four-leg, Signalized Intersections

Independent variable[a]	Chi square statistic[b]	Variable level	Direction of effect[c]	Coefficient	Relative effect[d]	90% confidence limits[e]	
						Lower	Upper
Intercept				−3.744		−5.187	−2.300
Crossroad ADT (log)	62.13		–	0.234	1.26	0.185	0.282
Major-road ADT (log)	53.04		–	0.517	1.68	0.400	0.633
Signal timing	59.70	Pretimed	–	0.032	1.03	−0.290	0.354
		Semiactuated	–	0			
		Fully actuated	–	0.636	1.89	0.489	0.784
Access control on major road	6.73	None	I	−0.312	0.73	−0.510	−0.114
		Partial	I	0			
Signal phasing	8.67	Two-phase	–	0			
		Multiphase	–	−0.221	0.80	−0.344	−0.097
Number of lanes on crossroad	4.14	3 or less	–	−0.134	0.87	−0.242	−0.026
		4 or more	–	0			
Average lane width on major road	3.21		–	−0.051	0.95	−0.098	−0.004
Number of lanes on major road	4.13	3 or less	–	−0.240	0.79	−0.471	−0.009
		4 or 5	–	−0.146	0.86	−0.272	−0.019
		6 or more	–	0			
Major road right-turn channelization	3.19	No free right turns	I	−0.119	0.89	−0.228	−0.009
		Provision for free right turns	I	0			

NOTE: This analysis is based on the set of 1,306 intersections for which summary statistics are shown in table 24.

[a] All variables significant at the 90% confidence level or higher.

[b] Chi-square likelihood ratio statistic for testing the significance of the effect of the variable; with 1 degree of freedom for continuous variables; with (p-1) degrees of freedom for categorical variables with p levels.

[c] Direction of effect: I = Inverse of expected direction.

[d] Relative effect of unit change in the variable on the expected number of accidents, equals exp(coefficient).

[e] 90% lower and upper confidence limits of the estimated coefficient.

Table 27. Lognormal Regression Results for Fatal and Injury Multiple-vehicle Accidents at Urban, Four-leg, Signalized Intersections

Independent variable[a]	Chi square statistic[b]	Variable level	Direction of effect[c]	Coefficient	Relative effect[d]	90% confidence limits[e]	
						Lower	Upper
Intercept				–5.845		–7.200	–4.490
Major-road ADT (log)	65.14		–	0.574	1.78	0.457	0.692
Crossroad ADT (log)	54.94		–	0.219	1.24	0.170	0.267
Signal timing	24.94	Pretimed	–	–0.073	0.93	–0.395	0.249
		Semiactuated	–	0			
		Fully actuated	–	0.389	1.48	0.242	0.536
Signal phasing	10.97	Two-phase	–	0			
		Multiphase	–	–0.247	0.78	–0.370	–0.124
Number of lanes on crossroad	5.41	3 or less	–	–0.153	0.86	–0.261	–0.045
		4 or more	–	0			
Access control on major road	4.84	None	I	–0.265	0.77	–0.463	–0.067
		Partial	I	0			
Number of lanes on major road	4.85	3 or less	–	–0.186	0.83	–0.416	0.043
		4 or 5	–	–0.168	0.85	–0.293	–0.042
		6 or more	–	0			
Design speed on major road	3.70		–	0.005	1.01	0.001	0.010

NOTE: This analysis is based on the set of 1,306 intersections for which summary statistics are shown in table 24.

[a] All variables significant at the 90% confidence level or higher.

[b] Chi-square likelihood ratio statistic for testing the significance of the effect of the variable; with 1 degree of freedom for continuous variables; with (p-1) degrees of freedom for categorical variables with p levels.

[c] Direction of effect: I = Inverse of expected direction.

[d] Relative effect of unit change in the variable on the expected number of accidents, equals exp(coefficient).

[e] 90% lower and upper confidence limits of the estimated coefficient.

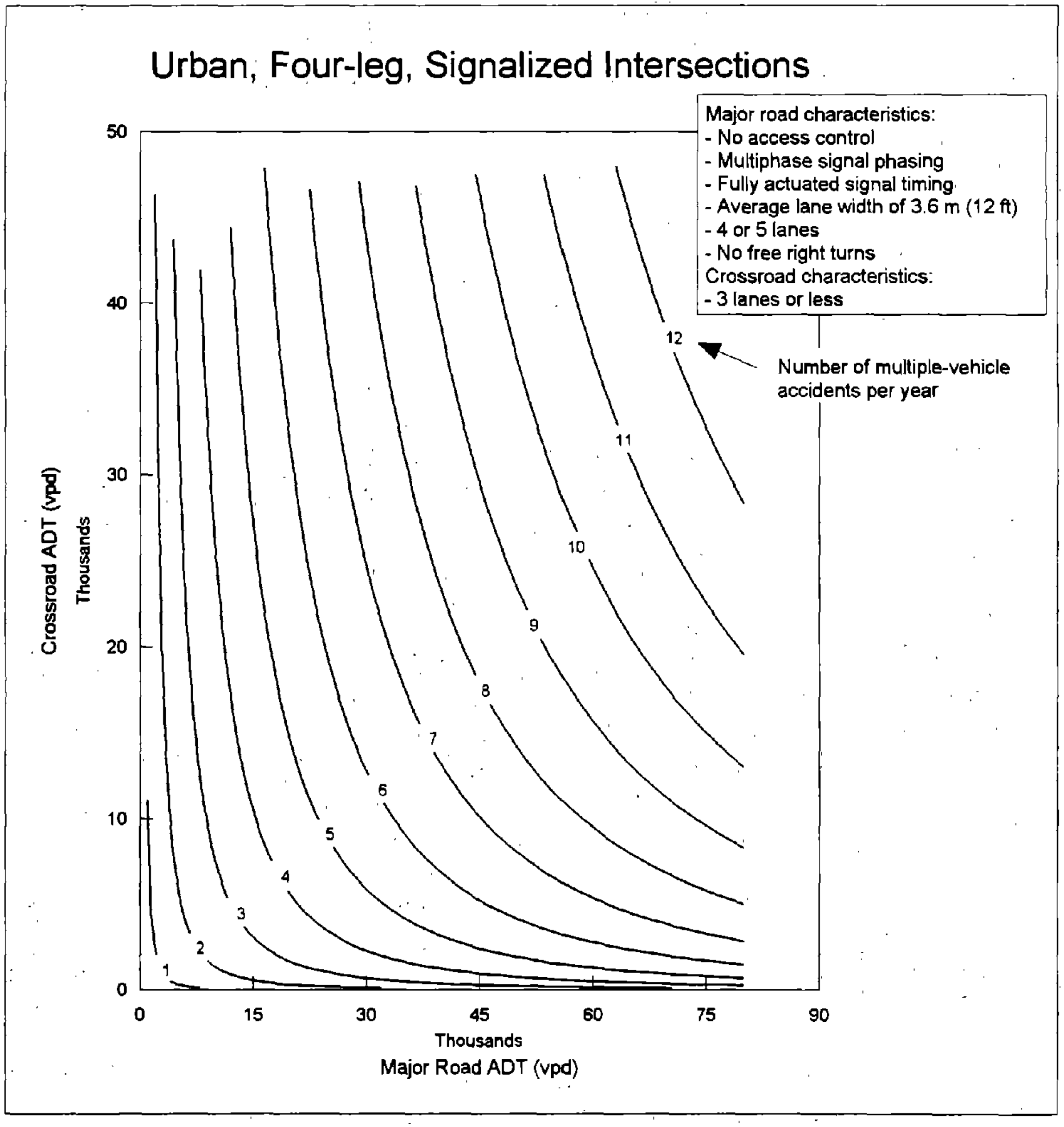


Figure 11. Number of Multiple-Vehicle Accidents per Year as a Function of Traffic Volumes for Typical Urban, Four-leg, Signalized Intersections

Urban, Four-leg, Signalized Intersections—Sample of 198 Intersections

A pilot field study was conducted at a randomly selected sample of 198 urban, four-leg, signalized intersections. The objective of this pilot study was to collect data for additional geometric, traffic control, and traffic volume variables (see section 4) and then to determine whether the availability of the additional data would improve the goodness of fit of the models. These intersections were randomly selected within four strata, of approximately equal size, defined by major-road ADT. These four strata of major-road ADT were:

- 20,000 veh/day or less.
- 20,000 to 30,000 veh/day.
- 30,000 to 40,000 veh/day.
- over 40,000 veh/day.

In addition, all intersections with no left-turn lanes on the major road were selected because there were only a limited number of such intersections. The sample was limited to a geographic area that included two major metropolitan areas, one smaller metropolitan area, and two adjacent smaller cities in California.

Table 3 in section 4 identifies the geometric design, traffic control, traffic volume, and related variables for which data were collected in the pilot field studies. Some data were collected to confirm the existing data obtained from the Caltrans data base. Other data were collected to expand the existing Caltrans data base. One advantage of the pilot field study was that geometric data were collected separately for each of the four intersection approaches; by contrast, the Caltrans data included combined data for both major-road and crossroad approaches, which does not allow for the possibility that the geometrics of these approaches might differ. In addition to collecting data for new geometric and traffic control variables, turning movement volumes for all approaches by 15-min periods for a 2-h morning peak period (typically 7 to 9 a.m.) and a 2-h evening peak period (typically 4 to 6 p.m.) were also recorded at all 198 intersections during the summer of 1994.

Reevaluation of Statistical Models for Sample of 198 Urban, Four-leg, Signalized Intersections: The purpose of the field data collection was to reevaluate the preliminary statistical model results obtained for the full sample of approximately 1,300 urban four-leg, signalized intersections (discussed in section 5 above). The following approach was taken to model total and fatal and injury multiple-vehicle accidents at the 198 intersections:

1. Estimate model coefficients using the same independent variables as those used in the model based on all 1,306 intersections (i.e., Caltrans data).

2. Estimate model coefficients using updated independent variables whenever available (i.e., a combination of data obtained from the Caltrans data base and data obtained from the field studies).

3. Estimate model coefficients using only additional variables for which data were obtained during the pilot field studies.

Accident frequencies at the 198 urban four-leg, signalized intersections are listed in table 62 in Appendix B and their distributions are shown in figure 6, for both total and fatal and injury multiple-vehicle accidents. A comparison of the distributions of the 1,306 and the sample of 198 intersections shown in figures 5 and 6, respectively, shows the similarities between the two sets. This was expected since the 198 intersections were randomly selected within major-road ADT strata. Only 1 of the 198 intersections experienced no accidents in the 3-year study period. Half the intersections experienced 19 accidents or more. Since accidents at the full set of approximately 1,300 urban, four-leg, signalized intersections were modeled using the lognormal distribution, accidents at the sample of 198 intersections were first modeled using a similar approach. However, due to the relatively small number of intersections, the negative binomial model approach appeared to be a logical choice in this case.

The selection of independent variables was done in a similar fashion to that described in section 5. Table 28 identifies the variables from the Caltrans data base that were selected for modeling accidents at the 198 urban four-leg, signalized intersections. The table shows the percent of intersections at each level of the categorical independent variables based on both the Caltrans data base and the updated data, when available. These percentages are in good agreement, indicating that no major changes in the geometric design, traffic control, and related variables have occurred between the two time periods. A small number of independent categorical variables was not included in the full model because either all or nearly all intersections fell into one level of that variable. The variables originally considered that were not included for this reason were:

- Lighting (all intersection were lighted).
- Presence of major-road signal mast arm (a mast arm was present on all intersections).
- Major-road left-turn prohibition (no intersections had left turns prohibited).
- Crossroad left-turn prohibition (no intersections had left turns prohibited).
- Access control on major road (191 of the 198 intersections had no access control on the major road).

Table 28. Descriptive Statistics for Sample of 198 Urban, Four-Leg, Signalized Intersections

Parameter	Mininum	Mean	Median	Maximum
Total multiple-vehicle accidents; 1990 through 1992 combined	0	24.5	20	119
Fatal and injury multiple-vehicle accidents; 1990 through 1992 combined	0	10.1	9	30
Major-road ADT (veh/day)	7,300	30,563	29,516	67,384
Crossroad ADT (veh/day)	101	9,276	6,601	48,000
Design speed of major road (mi/h)	25	51	50	65
Outside shoulder width on major road (ft)	0	7.1	8	15
Average lane width on major road (ft)	10	12.0	12	15

		Percent of intersections		
	Level	Caltrans data base	Updated data base	
Terrain	Flat	78		**Sample of 198 intersections**
	Rolling or mountainous	22		
Functional class of major road	Principal arterial	93		
	Minor arterial	7		
Signal timing	Semiactuated	13		
	Fully actuated	87		
Signal phasing	Two-phase	28		
	Multiphase	72		
Major-road left-turn channelization	No left-turn lane	15	13	
	Painted left-turn lane	33	31	
	Curbed left-turn lane	52	56	
Major-road right-turn channelization	No free right turns	75	64	
	Provision for free right turns	25	36	
Number of lanes on major road	3 or less (1)[a]	9	5	
	4 or 5 (2)	69	69	
	6 or more (3 or 4)	22	26	
Presence of crossroad signal mast arm	Mast arm not present	30	27	
	Mast arm present	70	73	

Conversion: 1 km/h = 0.621 mi/h; 1 m = 3.28 ft

Table 28. Descriptive Statistics for Sample of 198 Urban, Four-Leg, Signalized Intersections (Continued)

	Level	Percent of intersections Caltrans data base	 Updated data base
Crossroad left-turn channelization	No left-turn lane	43	36
	Painted left-turn lane	31	41
	Curbed left-turn lane	26	23
Crossroad right-turn channelization	No free right turns	70	59
	Provision for free right turns	30	41
Number of lanes on crossroad	3 or less (1)[a]	54	56
	4 or 5 (2 or more)	46	44
Presence of median on major road	Divided	79	65
	Undivided	21	35

[a] Number of lanes in Caltrans data base are both directions combined.
Number of lanes (indicated in parentheses) in updated data base are one direction only.

Table 29 identifies the new geometric variables for which data were collected in the field studies. The selection of these variables was based on both engineering judgement and the distribution of the 198 intersections across the levels of the full set of new geometric variables listed in table 3. Table 29 shows basic statistics (minimum, mean, median, maximum) for the continuous variables and the percent of intersections at each level of the categorical independent variables.

The 3-year accident counts were modeled using (1) the full set of 17 independent variables obtained from Caltrans, listed in table 28; (2) the full set of 17 independent variables obtained from Caltrans with values updated from the field studies when available, listed in table 28; and (3) the full set of 20 independent variables obtained in the field studies, listed in table 29. In all three cases, Caltrans traffic volume data (major-road and crossroad ADT's, on the log-scale) were used. All analyses were performed using the SAS GENMOD procedure with the negative binomial distribution and the appropriate deviance functions and variance adjustment factors, k. The significance of each regression coefficient in the full model was examined. If a coefficient was not significant at the 10 percent level, the corresponding independent variable was removed from the model and the negative binomial regression was rerun. In some cases, a second iteration of a reduced model had to be rerun to achieve significance of all the remaining variables.

Of the original 17 or 20 independent variables considered for modeling, only a small number were found to be statistically significant at the 10 percent level for total multiple-vehicle accidents for each of the three sets of data. Generally, slightly different sets of variables were statistically significant at the 10 percent level for fatal and injury multiple-vehicle accidents. In each case, the negative binomial model was rerun retaining only the statistically significant variables to obtain regression coefficients, their 90 percent confidence intervals, and other relevant regression statistics.

Table 30 presents the model statistics for the full and reduced negative binomial models using the three sets of independent variables, for total and fatal and injury multiple-vehicle accidents. As shown in table 30, the three sets of data lead to comparable model diagnostics for either type of accident. The fit of each model, however, is considerably improved over that obtained from the full set of approximately 1,300 intersections (see table 30), where R^2-values were in the mid-10 percent to low 20 percent. In the present case, R^2- and R^2_{FT}-values ranged consistently in the mid- to high 30 percent for total multiple-vehicle accidents, and in the mid- to high 20 percent for fatal and injury total-multiple accidents. For both types of accident, the Pearson chi-square ratios ranged from 0.84 to 0.89 for the reduced negative binomial models, an indication that these models provide an adequate fit to the data (values in the range of 0.8 to 1.2 are desirable).

Of the 14 independent variables considered in the full but not in the reduced negative binomial model (total multiple-vehicle accidents, Caltrans variables), 13 were not significant at the 10 percent level in the full model. Of these 13 variables, none

Table 29. Descriptive Statistics of New Field Variables for Sample of 198 Urban, Four-leg, Signalized Intersections

Parameter	Level	Percent of intersections	Minimum	Mean	Median	Maximum
Total multiple-vehicle accidents; 1990 through 1992 combined			0	24.5	20	119
Fatal and injury multiple-vehicle accidents; 1990 through 1992 combined			0	10	9	30
Major-road ADT (veh/day)			7,300	30,563	29,516	67,384
Crossroad ADT (veh/day)			101	9,276	6,601	48,000
Average lane width on major road (ft)			9	12.8	12.5	15
Median width on major road (ft)			0	15.6	16.5	68
Number of driveways on major road			0	2.8	3	10
Average lane width on crossroad (ft)			8.5	12.8	12.5	15
Median width on crossroad			0	5.1	0	35.5
Number of driveways on crossroad			0	3.0	3	9
Major-road left-turn channelization	No left-turn lane	13	**Sample of 198 intersections**			
	Painted or curbed left-turn lane	87				
Major-road right-turn channelization	No free right turns	64				
	Provision for free right turns	36				
Approach grade on major road	Grade	40				
	Level	60				
Curbed parking on major road	None	55				
	Parallel or angle	45				
Number of lanes on major road[a]	1 or 2	74				
	3 or 4	26				
Crossroad left-turn channelization	No left-turn lane	36				
	Painted or curbed left-turn lane	64				
Crossroad right-turn channelization	No free right turns	59				
	Provision for free right turns	41				
Approach grade on crossroad	Grade	39				
	Level	61				
Curbed parking on crossroad	None	49				
	Parallel or angle	52				
Number of lanes on crossroad[a]	1	56				
	2, 3, or 4	44				
Character development	B/C/I	59				
	M/R/X	41				
Skewness	Skewed	15				
	90°	85				

Conversion: 1 ft = 3.28 m

[a] Number of lanes are one direction only.

Table 30. Model Diagnostics for Total and Fatal and Injury Multiple-vehicle Accidents at a Sample of 198 Urban, Four-leg, Signalized Intersections

Negative Binomial Regression Models	Based on Caltrans geometric variables		Based on Caltrans geometric variables with updated values when available		Based on all new variables from field studies	
	Full model	Reduced model	Full model	Reduced model	Full model	Reduced model
Total Multiple-vehicle Accidents (3-year period)						
Number of intersections (n)	198	198	198	198	198	198
Number of parameters in model	17	3	17	6	20	5
Parameters degrees of freedom[a] (p)	21	5	20	9	21	6
k factor	0.34	0.34	0.32	0.32	0.32	0.32
Deviance/(n-p)	0.989	0.993	1.009	0.990	1.007	1.009
Pearson chi-square/(n-p)	0.868	0.888	0.869	0.838	0.863	0.868
R^2 (%)	41.31	31.88	39.75	37.56	40.11	36.28
R^2_{FT} (%)	44.26	35.09	44.05	40.66	43.99	38.79
Fatal and Injury Multiple-vehicle Accidents (3-year period)						
Number of intersections (n)	198	198	198	198	198	198
Number of parameters in model	17	4	17	2	20	4
Parameters degrees of freedom[a] (p)	21	6	20	3	21	5
k factor	0.27	0.26	0.27	0.27	0.26	0.26
Deviance/(n-p)	1.000	1.005	0.992	1.002	1.005	0.997
Pearson chi-square/(n-p)	0.899	0.886	0.872	0.865	0.870	0.865
R^2 (%)	32.43	26.58	33.41	24.51	34.57	28.44
R^2_{FT} (%)	34.76	28.65	34.81	25.52	35.42	29.28

[a] Includes one degree of freedom for the intercept

would have been significant at the 20 percent level. However, one variable—signal timing—was significant in the full model (α=0.03) but when considered in the reduced model, was no longer significant at the 10 percent level (α=0.13). The outcome of the models for fatal and injury multiple-vehicle accidents (using the original Caltrans geometric variables) was similar to that for total multiple-vehicle accidents. Of the 12 variables that were not significant at the 10 percent level in the full model, none would have been at the 20 percent. Again, signal timing was significant in the full model (α=0.05), but when considered in the reduced model, was no longer significant at the 10 percent level (α=0.14), and was therefore excluded from the final model.

Of the 11 independent variables considered in the full but not in the reduced negative binomial model (total multiple-vehicle accidents, updated Caltrans variables), two that were not significant at the 10 percent would have been at the 20 percent level. These variables were outside shoulder width (α=0.15) and crossroad right-turn channelization (α=0.17). In the full model for fatal and injury total multiple-vehicle accidents (using updated values for the original geometric variables), 13 were not significant at the 10 percent level, 2 of which would have been significant at the 20 percent level. These variables were major road left-turn channelization (α=0.15) and number of lanes on major road (α=0.18). Two additional variables—signal timing and design speed on major road—although significant in the full model (α=0.05 and 0.10, respectively), were subsequently excluded during a second iteration of the reduced model, with final α-values of 0.34 and 0.17, respectively.

Of the 15 independent variables considered in the full but not in the reduced negative binomial model (total multiple-vehicle accidents, new geometric field variables), four that were not significant at the 10 percent would have been at the 20 percent level. These variables were median width on major road (α=0.17), number of driveways on major road (α=0.18), crossroad right-turn channelization (α=0.12), and approach grade on crossroad (α=0.14). In the full model for fatal and injury total multiple-vehicle accidents (new geometric field variables), 14 were not significant at the 10 percent level. Of these, one—curbed parking on major road (α=0.19)—would have been significant at the 20 percent level. Two additional variables—average lane width on crossroad and crossroad left-turn channelization—although significant in the full model (α=0.05 and 0.07, respectively), were subsequently excluded during a second iteration of the reduced model, with final α-values of 0.12 and 0.21, respectively.

Tables 31 through 33 summarize the regression results for the final negative binomial models for total multiple-vehicle accidents, using the three sets of independent variables, respectively. Similarly, tables 34 through 36 summarize the final negative binomial regression results for fatal and injury multiple-vehicle accidents. While a considerable number of geometric and traffic control variables were found to contribute significantly to the variability in the 3-year accident counts for the full set of approximately 1,300 urban, four-leg, signalized intersections, only a small number of variables remained in the models for the sample of 198 intersections.

Table 31. Negative Binomial Regression Results for Total Multiple-vehicle Accidents at Sample of 198 Urban, Four-leg, Signalized Intersections Using Caltrans Geometric Variables

Independent variable[a]	Chi square statistic[b]	Variable level	Direction of effect[c]	Coefficient	Relative effect[d]	90% confidence limits[e]	
						Lower	Upper
Intercept				−5.775		−8.041	−3.511
Crossroad ADT (log)	32.15		—	0.258	1.29	0.185	0.330
Major-road ADT (log)	28.91		—	0.670	1.95	0.466	0.875
Number of lanes on major road	6.97	3 or less	—	−0.500	0.61	−0.854	−0.138
		4 or 5	—	−0.287	0.75	−0.487	−0.090
		6 or more	—	0			

NOTE: This analysis is based on the set of 198 intersections for which summary statistics are shown in table 28.

[a] All variables significant at the 90% confidence level or higher.

[b] Chi-square likelihood ratio statistic for testing the significance of the effect of the variable; with 1 degree of freedom for continuous variables; with (p-1) degrees of freedom for categorical variables with p levels.

[c] Direction of effect: I = Inverse of expected direction.

[d] Relative effect of unit change in the variable on the expected number of accidents, equals exp(coefficient).

[e] 90% lower and upper confidence limits of the estimated coefficient.

Table 32. Negative Binomial Regression Results for Total Multiple-vehicle Accidents at Sample of 198 Urban, Four-leg, Signalized Intersections Using Updated Values of Original Geometric Variables

Independent variable[a]	Chi square statistic[b]	Variable level	Direction of effect[c]	Coefficient	Relative effect[d]	90% confidence limits[e]	
						Lower	Upper
Intercept				–7.740		–9.762	–5.713
Major-road ADT (log)	66.07		–	0.909	2.48	0.727	1.092
Major road left-turn channelization	12.70	No left-turn lane	–	0.475	1.61	0.171	0.786
		Painted left-turn lane	–	0			
		Curbed left-turn lane	–	–0.176	0.84	–0.367	0.014
Crossroad ADT (log)	9.82			0.167	1.18	0.080	0.251
Crossroad left-turn channelization	8.06	No left-turn lane	I	–0.332	0.72	–0.528	–0.136
		Painted left-turn lane	I	0			
		Curbed left-turn lane	I	0.005	1.01	–0.193	0.205
Signal timing	4.44	Semiactuated	–	0			
		Fully actuated	–	0.368	1.45	0.082	0.651
Number of lanes on crossroad	3.62	1	–	–0.200	0.82	–0.368	–0.032
		2 or more	–	0			

NOTE: This analysis is based on the set of 198 intersections for which summary statistics are shown in table 28.

[a] All variables significant at the 90% confidence level or higher.

[b] Chi-square likelihood ratio statistic for testing the significance of the effect of the variable; with 1 degree of freedom for continuous variables; with (p-1) degrees of freedom for categorical variables with p levels.

[c] Direction of effect: I = Inverse of expected direction.

[d] Relative effect of unit change in the variable on the expected number of accidents, equals exp(coefficient).

[e] 90% lower and upper confidence limits of the estimated coefficient.

Table 33. Negative Binomial Regression Results for Total Multiple-vehicle Accidents at Sample of 198 Urban, Four-leg, Signalized Intersections Using New Field Geometric Variables

Independent variable[a]	Chi square statistic[b]	Variable level	Direction of effect[c]	Coefficient	Relative effect[d]	90% confidence limits[e]	
						Lower	Upper
Intercept				−7.206		−9.028	−5.379
Major-road ADT (log)	71.37		–	0.836	2.31	0.675	0.997
Crossroad ADT (log)	17.19		–	0.214	1.24	0.131	0.294
Crossroad left-turn channelization	11.64	No left-turn lane	I	−0.394	0.67	−0.583	−0.205
		Painted or curbed left-turn lane	I	0			
Major road left-turn channelization	6.11	No left-turn lane	–	0.346	1.41	0.114	0.584
		Painted or curbed left-turn lane	–	0			
Angle of intersection	3.25	Less than 90°	I	−0.234	0.79	−0.442	−0.021
		90°	I	0			

NOTE: This analysis is based on the set of 198 intersections for which summary statistics are shown in table 29.

[a] All variables significant at the 90% confidence level or higher.

[b] Chi-square likelihood ratio statistic for testing the significance of the effect of the variable; with 1 degree of freedom for continuous variables; with (p-1) degrees of freedom for categorical variables with p levels.

[c] Direction of effect: I = Inverse of expected direction.

[d] Relative effect of unit change in the variable on the expected number of accidents, equals exp(coefficient).

[e] 90% lower and upper confidence limits of the estimated coefficient.

Table 34. Negative Binomial Regression Results for Fatal and Injury Multiple-vehicle Accidents at Sample of 198 Urban, Four-leg, Signalized Intersections Using Caltrans Geometric Variables

Independent variable[a]	Chi square statistic[b]	Variable level	Direction of effect[c]	Coefficient	Relative effect[d]	90%n confidence limits[e]	
						Lower	Upper
Intercept				−4.406		−6.617	−2.201
Crossroad ADT (log)	18.10		—	0.189	1.21	0.116	0.261
Major-road ADT (log)	14.28		—	0.470	1.60	0.266	0.674
Number of lanes on major road	5.22	3 or less	—	−0.308	0.73	−0.659	0.045
		4 or 5	—	−0.262	0.77	−0.453	−0.073
		6 or more	—	0			
Design speed on major road	2.87		—	0.008	1.01	0.0002	0.016

NOTE: This analysis is based on the set of 198 intersections for which summary statistics are shown in table 28.

[a] All variables significant at the 90% confidence level or higher.

[b] Chi-square likelihood ratio statistic for testing the significance of the effect of the variable; with 1 degree of freedom for continuous variables; with (p-1) degrees of freedom for categorical variables with p levels.

[c] Direction of effect: I = Inverse of expected direction.

[d] Relative effect of unit change in the variable on the expected number of accidents, equals exp(coefficient).

[e] 90% lower and upper confidence limits of the estimated coefficient.

Table 35. Negative Binomial Regression Results for Fatal and Injury Multiple-vehicle Accidents at Sample of 198 Urban, Four-leg, Signalized Intersections Using Updated Values of Original Geometric Variables

Independent variable[a]	Chi square statistic[b]	Direction of effect[c]	Coefficient	Relative effect[d]	90% confidence limits[e]	
					Lower	Upper
Intercept			−5.977		−7.750	−4.211
Major-road ADT (log)	42.71	—	0.642	1.90	0.480	0.804
Crossroad ADT (log)	18.22	—	0.191	1.21	0.118	0.263

NOTE: This analysis is based on the set of 198 intersections for which summary statistics are shown in table 28.

[a] All variables significant at the 90% confidence level or higher.

[b] Chi-square likelihood ratio statistic for testing the significance of the effect of the variable; with 1 degree of freedom for continuous variables; with (p-1) degrees of freedom for categorical variables with p levels.

[c] Direction of effect: I = Inverse of expected direction.

[d] Relative effect of unit change in the variable on the expected number of accidents, equals exp(coefficient).

[e] 90% lower and upper confidence limits of the estimated coefficient.

Table 36. Negative Binomial Regression Results for Fatal and Injury Multiple-vehicle Accidents at Sample of 198 Urban, Four-leg, Signalized Intersections Using New Field Geometric Variables

Independent variable[a]	Chi square statistic[b]	Variable level	Direction of effect[c]	Coefficient	Relative effect[d]	90% confidence limits[e] Lower	90% confidence limits[e] Upper
Intercept				–5.838		–7.614	–4.068
Major-road ADT (log)	41.01		–	0.625	1.87	0.465	0.786
Crossroad ADT (log)	15.37		–	0.185	1.20	0.108	0.262
Curbed parking on major road	5.57	None	I	0.214	1.24	0.065	0.363
		Parallel or angle	I	0			
Angle of intersection	3.05	Less than 90°	I	–0.224	0.80	–0.431	–0.013
		90°	I	0			

NOTE: This analysis is based on the set of 198 intersections for which summary statistics are shown in table 29.

[a] All variables significant at the 90% confidence level or higher.

[b] Chi-square likelihood ratio statistic for testing the significance of the effect of the variable; with 1 degree of freedom for continuous variables; with (p-1) degrees of freedom for categorical variables with p levels.

[c] Direction of effect: I = Inverse of expected direction.

[d] Relative effect of unit change in the variable on the expected number of accidents, equals exp(coefficient).

[e] 90% lower and upper confidence limits of the estimated coefficient.

A comparison of table 31 with table 26 (total multiple-vehicle accidents, sample and full set of intersections) shows that only one of the original Caltrans geometric variables—number of lanes on major road—remained in the model. A possible explanation for the small number of significant variables in the models when applied to the sample of 198 intersections is the fact that the sample intersections may be fairly homogeneous in their geometrics and therefore show little variability above that explained by traffic volumes. A comparison of table 34 with table 27 (fatal and injury multiple-vehicle accidents, sample and full set of intersections) reveals similar patterns.

Substituting updated values for the original geometric variables, when available, slightly improved the fit of the model for total multiple-vehicle accidents but reduced it slightly for fatal and injury multiple-vehicle accident (compare columns 3 and 5 in table 30). A comparison of the model in table 26 with that in table 32 (total multiple-vehicle accidents) shows slight changes in the set of significant geometric design variables. Major-road left turn channelization was substituted for major-road right turn channelization, and crossroad left-turn channelization was added. Table 35 shows that, when modeling fatal and injury total-multiple accidents using the updated geometric variables, only traffic volumes on both roads remained significant.

The models including the new geometric and traffic control variables obtained in the field, shown in tables 33 and 36 (for total and fatal and injury multiple-vehicle accidents, respectively), are disappointing because several of the geometric design variables that were found to be statistically significant have effects on accidents in the opposite direction to that expected.

Left-turn Accident Analysis at Sample of 198 Urban, Four-leg, Signalized Intersections: Turning movement volumes were obtained for all approaches during the field studies. An additional analysis was conducted to apply the previous approach to specific types of accidents associated with specific turning movement combinations. This approach focussed specifically on multiple-vehicle left-turn accidents. This investigation was undertaken on the premise that models with better fit could be obtained if models were developed to relate accidents involving particular intersection turning movements to the traffic volumes and geometric elements associated with those turning movements. For example, Hauer et al. had reported success in relating particular accident types to the related turning volumes, although no goodness of fit measures for Hauer's models were reported.(5) The left-turn accident analysis reported here corresponds to Accident Pattern No. 6 in the Hauer et al. work, which includes only accidents involving a vehicle turning left which collides with an opposing through vehicle.

Dependent Variable: All multiple-vehicle left-turn accidents involving a left-turning vehicle on a particular approach colliding with a vehicle from the opposing approach were considered in the modeling approach. The following selection criteria apply to these accidents:

- As before, the total number of accidents from the 3 years 1990 through 1992 was considered (Caltrans data base).
- Left-turn accidents in the morning peak period were those between 6 and 10 a.m., which is centered on the period for which turning movement volumes were counted (7 to 9 a.m.).
- Left-turn accidents in the evening peak period were those between 3 and 7 p.m., which is centered on the period for which turning movement volumes were counted (4 to 6 p.m.).
- Each approach to each intersection was considered separately.
- Each peak period was considered separately.
- In total, left-turn accident data were available for 1,584 approach-by-peak-period combination sites (198 intersections x 4 approaches x 2 peak periods).

The 3-year left-turn accident experience at the 1,584 approach-by-peak-period observations is summarized in table 37.

Table 37. Left-turn Accident Frequencies at a Sample of 198 Urban, Four-leg, Signalized Intersections

Number of left-turn accidents[a]	Number of approaches[b]	Percent of approaches	Cumulative number of sites	Cumulative percent of sites (%)
0	1,335	84.28	1,335	84.28
1	210	13.26	1,545	97.54
2	28	1.77	1,573	99.31
3	5	0.32	1,578	99.62
4	3	0.19	1,581	99.81
5	1	0.06	1,582	99.87
8	1	0.06	1,583	99.94
16	1	0.06	1,584	100.00

[a] Accidents involving a left-turn vehicle and opposing through vehicle for a particular approach and peak period in the 3-year study period (1990 through 1992).

[b] Each approach is treated as a separate observation in the morning and evening peak periods.

From these data, the average number of left-turn accidents per observation is 0.20 (322 left-turn accidents in 1,584 observations). Based on the accident data distribution shown in the above table, a Poisson regression approach was a logical choice for modeling left-turn accidents at these intersection. In addition, a logistic regression was performed to assess which independent variables could discriminate between those observations without left-turn accidents (84.3 percent) and those with one or more accidents (15.7 percent). (More detail on logistic regression analysis is provided in section 6.)

Independent Variables: Based on the data collected in the field studies, the following independent variables were selected for inclusion in the regression models (Poisson or logistic):

- Left-turn and opposing through movement counts. These were determined for a 2-h morning (7 to 9 a.m.) and a 2-h evening (4 to 6 p.m.) peak period based on the turning movement counts (right, through, and left) collected in the field

- Other new and updated intersection geometric and traffic control variables from the field studies which included, depending on the model used:

Poisson Regression Model:

- Left-turn volume (log)
- Opposing left-turn volume (log)
- Opposing through volume (log)
- Type of left-turn treatment
- Angle of intersection (skewness)
- Average left-turn lane width
- Left-turn channelization
- Opposing direction median width
- Opposing direction number of through lanes
- Protected left-turn phase
- Same direction number of lanes
- Signal head mounting

Logistic Regression Model:

- Left-turn volume (log)
- Opposing through volume (log)
- Opposing left-turn volume (log)
- Type of left-turn treatment
- Angle of intersection (skewness)
- Average left-turn lane width
- Double left-turn channelization
- Left-turn channelization

- Opposing direction number of through lanes
- Opposing direction median width
- Protected left-turn phase
- Signal head mounting

Although both regression approaches (Poisson regression and logistic regression) produced models that were statistically significant at the 95 percent confidence level, the fit of the models was extremely poor, with R^2-values of approximately 4 to 5 percent. In both cases, left-turn and opposing through volumes (log) were significant contributors to the small amount of variability in the data that was explained by the models. These results were disappointing because the concept of relating turning movement data to specific related accident types appeared to hold promise. The poor results can probably be explained by the very small accident sample sizes that result from restricting the analysis to particular accident types, on particular intersection approaches, during particular times of the day.

6. ALTERNATIVE STATISTICAL MODELING APPROACHES

Introduction

This section describes several statistical approaches that were investigated as possible alternatives to the loglinear and lognormal regression analyses described in section 5. The purpose of this investigation was to determine whether any of these techniques might be more effective than loglinear or lognormal regression as a tool for developing relationships between accidents and geometric features of at-grade intersections or identifying particular geometric features that might have a role in such relationships.

The four alternative statistical approaches investigated were:

- Logistic regression analysis to relate geometric design features, traffic control features, and traffic volumes of intersections to the probability of accidents at those intersections.
- Repetition of the loglinear and lognormal regression analyses already performed, but restricting the analyses to specific ranges of ADT.
- Discriminant analysis of intersections based on pre-defined accident risk groupings (e.g., no-risk, low-risk, high-risk).
- Cluster analysis of intersections based on their geometric similarities and differences, followed by comparison of these intersection clusters based on geometrics with intersection clusters based on accidents.

Logistic regression was first considered during the research as a screening tool that might identify particular geometric design or traffic control features that should be considered in further analyses. Thus, logistic regression was initially applied to the California intersection data base prior to the loglinear and lognormal regression analyses reported in section 5. However, it was also recognized that, beyond being merely a screening tool, logistic regression could be effective in its own right as a modeling approach to establish predictive relationships for accident probabilities.

The three remaining statistical approaches were identified after initial results of the loglinear and lognormal modeling indicated that the geometric design, traffic control, and traffic volume variables that were considered generally explained only 18 to 37 percent of the variation in intersection accidents. These additional approaches were investigated as alternatives that might be effective in developing better predictive models.

Technical discussions of these four alternative statistical approaches and their results are presented below.

Preceding Page Blank

Logistic Regression Analysis

Logistic regression is a statistical technique for developing relationships that predict the probability that a particular event will occur rather than the number of events that will occur in a specified time period. As in any regression model, the dependent variable (probability of occurrence of the event of interest) is predicted by a linear combination of independent variables. In this application to the at-grade intersection data in the California intersection data base, the event for which probabilities were predicted was the occurrence of some particular number of intersection accidents and the independent variables of interest were geometric design features, traffic control features, and traffic volumes for the intersections in the data base.

Logistic regression was first envisioned in this research as a screening tool that could be applied to determine whether particular independent variables appeared to be useful as accident predictors. It was decided that logistic regression was likely to provide a better screening tool if it was applied to predict the probability that an intersection was experiencing the consistent occurrence of accidents from year to year, rather than just the probability of occurrence of a single accident. A dependent variable, Y, was therefore derived for each intersection as follows:

- Intersections with no accidents in the 3-year study period were assigned the value Y=0.

- Intersections for which the number of accidents exceeded a set minimum in each and every year of the 3-year study period were assigned the value Y=1. The minimum number of accidents considered depended on the class of intersections and was a function of the accident data distribution in that particular class.

- Intersections with accident frequencies ranging between these two groups (i.e., that experienced some accidents, but did not experience the specified minimum number of accidents in each and every year) were not considered in the analysis.

A logistic regression model then uses the selected independent variables of these two groups of intersections to predict the probability that the dependent variable (Y) takes on the value 0 or 1. This method provides a quantitative tool to assess the relative importance of intersection variables in their ability to predict group membership with a given degree of certainty.

Using the general intersection notation used in section 5, the logistic regression model can be written as follows:

$$\mathrm{logit}(p_i) = \log[p_i/(1-p_i)] = \beta_0 + \beta_1 X_{1i} + \ldots + \beta_q X_{qi} \tag{16}$$

where

p_i = Prob(Y_i = 1, given the set of q independent variables X_{1i}, ..., X_{qi} at intersection i)

and β_0, ..., β_q are the regression parameters, to be estimated by the method of maximum likelihood.

As in previous modeling (see section 5), X_1 and X_2 represent the $\log(ADT_{\text{major road}})$ and $\log(ADT_{\text{crossroad}})$, respectively.

Once the regression coefficients, β_0, ..., β_q, of Eq. (15) are estimated, one computes the logit of the probability that intersection i will experience a predetermined number of accidents by simply evaluating Eq. (16) at the values of the q parameters X_{1i}, ..., X_{qi}. Let L be the resulting value of Eq. (16). The predicted probability, p_i, is then computed from the logit as:

$$p_i = \exp(L)/[1+\exp(L)] \tag{17}$$

Logistic regression analysis was applied to four of the five classes of intersections. The probability of accidents occurring at urban, four-leg, signalized intersections was not investigated using this approach due to the very small number of intersections (only 21 of the 1,306) that experienced no accidents in the 3-year study period. Based on the observed distributions of total multiple-vehicle accidents (see appendix B), the following minimum numbers of accidents in each year were used to defined the Y = 1 group:

rural, four-leg, STOP-controlled:	2 multiple-vehicle accidents
rural, three-leg, STOP-controlled:	1 multiple-vehicle accident
urban, four-leg, STOP-controlled:	2 multiple-vehicle accidents
urban, three-leg, STOP-controlled:	2 multiple-vehicle accidents

A similar review of the distribution of fatal and injury multiple-vehicle accidents determined that one fatal or injury accident in each year was the appropriate cut-off point for all four classes of intersections.

For each intersection class and accident severity level, a stepwise logistic regression analysis was performed using PROC LOGISTIC in SAS to identify those parameters that significantly contribute to predicting the probability of one (or two) accidents occurring consistently each year at an intersection.(17,18) All confidence levels were set at 90 percent. The same sets of independent variables previously selected for use in the full Poisson, negative binomial or lognormal models (see section 5) were also used for these analyses. The logistic regression results and model diagnostics are shown in tables 38 through 45 for total and fatal and injury multiple-vehicle accidents at the four types of intersections.

The following information is provided in each table:

- The independent variable and its levels, if the variable is categorical.
- The direction of the effect of the variable on the predicted probability.
- The corresponding regression coefficient for use in Eq. (16).
- The Wald chi-square statistic (ratio of the square of the coefficient over its standard error estimate).
- The odds ratio, calculated as exp(coefficient). This value indicates the amount by which the probability that an accident will occur increases for each unit increase in the independent variable.
- The number (and percentage of total) intersections falling into each of the two categories defined by the selected cut-off point.
- The chi-square value, along with its degrees of freedom (df) and its significance level (p) for each of two criteria: the -2 log likelihood statistic and the score statistic. A significant p-value (e.g., less than 0.05) provides evidence that the regression coefficients are statistically nonzero.
- The residual chi-square value with its degrees of freedom (df) and probability level (p). The residual chi-square value is a measure of the variability in the data that is not explained by the reduced model but could be explained by the full model. A small residual chi-square value with a large p-value is desired. The degrees of freedom are those associated with the parameters not retained in the reduced model.
- Three estimated measures of model fit: the correlation-based R^2-value, the likelihood ratio R^2-value, and the adjusted likelihood ratio R^2-value. A discussion of the interpretation of these alternative R^2-values is provided by SAS.(18)

Table 38. Logistic Regression Results for Total Multiple-vehicle Accidents at Rural, Four-leg, STOP-controlled Intersections

Independent variable[a]	Variable level	Direction of effect[b]	Coefficient	Chi-square statistic[c]	Odds ratio[d]
Intercept			–40.82	104.47	
Crossroad ADT (log)		—	1.97	90.05	7.20
Major-road ADT (log)		—	2.88	71.57	17.76
Number of lanes on major road	1 if 3 or less; 0 if 4 or more	—	1.87	18.72	6.50
Terrain	1 if flat; 0 otherwise	I	0.67	3.82	1.96

Regression statistics

Number of intersections with 0 accidents (% of total)	392 (27%)
Number of intersections with at least 2 accidents (% of total)	147 (10%)
Chi-square value for	
a) the -2 log likelihood statistic	370.41 with 4 df ($p \leq 0.0001$)
b) the score statistic	278.40 with 4 df ($p \leq 0.0001$)
Residual chi-square	10.88 with 13 df ($p = 0.62$)
Estimated R^2 (correlation-based)	63.59%
Estimated R^2 (likelihood ratio)	49.70%
Adjusted R^2 (likelihood ratio)	72.01%

[a] All variables significant at the 90% confidence level or higher.

[b] Direction of effect: I = Inverse of expected direction.

[c] Wald chi-square statistic, with one degree of freedom, for testing the significance of the effect of the variable.

[d] Odds ratio equals exp(coefficient). It indicates the amount by which the odds of an accident occurring increase for each unit increase in the variable.

Conversion: 1 km/h = 0.621 mi/h

Table 39. Logistic Regression Results for Fatal and Injury Multiple-vehicle Accidents at Rural, Four-leg, STOP-controlled Intersections

Independent variable[a]	Variable level	Direction of effect[b]	Coefficient	Chi-square statistic[c]	Odds ratio[d]
Intercept			−32.11	135.31	
Crossroad ADT (log)		—	1.54	113.26	4.68
Major-road ADT (log)		—	1.82	66.07	6.16
Design speed of major road (mi/h)		—	0.06	27.39	1.07
Number of lanes on major road	1 if 3 or less; 0 if 4 or more	—	1.34	13.16	3.81
Terrain	1 if flat; 0 otherwise	I	0.94	12.06	2.56
Functional class of major road	1 if principal arterial; 0 otherwise	—	−0.93	10.50	0.39
Major-road left-turn channelization	1 if curbed left-turn lane; 0 otherwise	I	1.11	3.63	3.03

Regression statistics	
Number of intersections with 0 accidents (% of total)	605 (42%)
Number of intersections with at least 1 accident (% of total)	175 (12%)
Chi-square value for	
a) the -2 log likelihood statistic	347.78 with 7 df ($p \leq 0.0001$)
b) the score statistic	280.14 with 7 df ($p \leq 0.0001$)
Residual chi-square	9.53 with 10 df ($p = 0.48$)
Estimated R^2 (correlation-based)	64.45%
Estimated R^2 (likelihood ratio)	35.97%
Adjusted R^2 (likelihood ratio)	54.91%

Conversion: 1 km/h = 0.621 mi/h

[a] All variables significant at the 90% confidence level or higher.

[b] Direction of effect: I = Inverse of expected direction.

[c] Wald chi-square statistic, with one degree of freedom, for testing the significance of the effect of the variable.

[d] Odds ratio equals exp(coefficient). It indicates the amount by which the odds of an accident occurring increase for each unit increase in the variable.

Table 40. Logistic Regression Results for Total Multiple-vehicle Accidents at Rural, Three-leg, STOP-controlled Intersections

Independent variable[a]	Variable level	Direction of effect[b]	Coefficient	Chi-square statistic[c]	Odds ratio[d]
Intercept			−30.89	270.58	
Major-road ADT (log)		—	2.44	218.62	11.52
Crossroad ADT (log)		—	1.03	107.30	2.80
Access control on major road	1 if none; 0 if partial	I	1.00	10.10	2.71
Major-road left-turn channelization	1 if none; 0 otherwise	I	0.71	12.28	2.03
Functional class of major road	1 if minor arterial; 0 otherwise	—	0.51	6.37	1.66

Regression statistics	
Number of intersections with 0 accidents (% of total)	1,174 (44%)
Number of intersections with at least 1 accident (% of total)	290 (11%)
Chi-square value for	
a) the -2 log likelihood statistic	648.98 with 5 df ($p \leq 0.0001$)
b) the score statistic	517.12 with 5 df ($p \leq 0.0001$)
Residual chi-square	6.61 with 12 df ($p = 0.88$)
Estimated R^2 (correlation-based)	47.21%
Estimated R^2 (likelihood ratio)	35.81%
Adjusted R^2 (likelihood ratio)	56.80%

[a] All variables significant at the 90% confidence level or higher.

[b] Direction of effect: I = Inverse of expected direction.

[c] Wald chi-square statistic, with one degree of freedom, for testing the significance of the effect of the variable.

[d] Odds ratio equals exp(coefficient). It indicates the amount by which the odds of an accident occurring increase for each unit increase in the variable.

Table 41. Logistic Regression Results for Fatal and Injury Multiple-vehicle Accidents at Rural, Three-leg, STOP-controlled Intersections

Independent variable[a]	Variable level	Direction of effect[b]	Coefficient	Chi-square statistic[c]	Odds ratio[d]
Intercept			–29.07	159.56	
Major-road ADT (log)		—	2.30	103.36	9.95
Crossroad ADT (log)		—	0.96	64.78	2.60
Functional class of major road	1 if principal arterial; 0 otherwise	—	–0.60	3.99	0.55
Presence of median on major road	1 if divided; 0 if undivided	—	–0.52	2.83	0.60

Regression statistics	
Number of intersections with 0 accidents (% of total)	1,625 (60%)
Number of intersections with at least 1 accident (% of total)	103 (4%)
Chi-square value for	
a) the -2 log likelihood statistic	283.10 with 4 df ($p \leq 0.0001$)
b) the score statistic	252.26 with 4 df ($p \leq 0.0001$)
Residual chi-square	9.17 with 13 df ($p = 0.76$)
Estimated R^2 (correlation-based)	39.26%
Estimated R^2 (likelihood ratio)	15.11%
Adjusted R^2 (likelihood ratio)	41.57%

[a] All variables significant at the 90% confidence level or higher.

[b] Direction of effect: I = Inverse of expected direction.

[c] Wald chi-square statistic, with one degree of freedom, for testing the significance of the effect of the variable.

[d] Odds ratio equals exp(coefficient). It indicates the amount by which the odds of an accident occurring increase for each unit increase in the variable.

Table 42. Logistic Regression Results for Total Multiple-vehicle Accidents at Urban, Four-leg, STOP-controlled Intersections

Independent variable[a]	Variable level	Direction of effect[b]	Coefficient	Chi-square statistic[c]	Odds ratio[d]
Intercept			–19.50	35.05	
Crossroad ADT (log)		—	1.06	51.32	2.89
Major-road ADT (log)		—	1.89	47.72	6.64
Major-road left-turn prohibition	1 if left turns prohibited; 0 otherwise	—	–3.43	30.73	0.03
Average lane width on major road (ft)		—	–0.30	6.05	0.74
Number of lanes on major road	1 if 3 or less; 0 otherwise	—	0.73	4.54	2.08
Crossroad right-turn channelization	1 if no provision for free right turns; 0 otherwise	—	–1.90	3.65	0.15

Regression statistics	
Number of intersections with 0 accidents (% of total)	138 (10%)
Number of intersections with at least 2 accidents (% of total)	344 (26%)
Chi-square value for	
a) the -2 log likelihood statistic	176.68 with 6 df ($p \leq 0.0001$)
b) the score statistic	160.10 with 6 df ($p \leq 0.0001$)
Residual chi-square	9.50 with 13 df ($p = 0.73$)
Estimated R^2 (correlation-based)	38.52%
Estimated R^2 (likelihood ratio)	30.69%
Adjusted R^2 (likelihood ratio)	43.96%

[a] All variables significant at the 90% confidence level or higher.

[b] Direction of effect: I = Inverse of expected direction.

[c] Wald chi-square statistic, with one degree of freedom, for testing the significance of the effect of the variable.

[d] Odds ratio equals exp(coefficient). It indicates the amount by which the odds of an accident occurring increase for each unit increase in the variable.

Conversion: 1 m = 3.28 ft

Table 43. Logistic Regression Results for Fatal and Injury Multiple-vehicle Accidents at Urban, Four-leg, STOP-controlled Intersections

Independent variable[a]	Variable level	Direction of effect[b]	Coefficient	Chi-square statistic[c]	Odds ratio[d]
Intercept			–16.76	48.55	
Major-road ADT (log)		—	1.78	67.58	5.90
Major-road left-turn prohibition	1 if left turns prohibited; 0 otherwise	—	–2.52	21.26	0.08
Crossroad ADT (log)		—	0.52	24.81	1.68
Average lane width on major road (ft)		—	–0.25	9.49	0.78
Access control on major road	1 if none; 0 if partial	I	–1.22	5.77	0.30
Number of lanes on major road	1 if 3 or less; 0 otherwise	—	0.60	5.27	1.82

Regression statistics

Number of intersections with 0 accidents (% of total)	276 (21%)
Number of intersections with at least 1 accident (% of total)	315 (23%)
Chi-square value for	
a) the -2 log likelihood statistic	163.59 with 6 df ($p \leq 0.0001$)
b) the score statistic	143.39 with 6 df ($p \leq 0.0001$)
Residual chi-square	9.32 with 13 df ($p = 0.75$)
Estimated R^2 (correlation-based)	26.13%
Estimated R^2 (likelihood ratio)	24.18%
Adjusted R^2 (likelihood ratio)	32.29%

[a] All variables significant at the 90% confidence level or higher.

[b] Direction of effect: I = Inverse of expected direction.

[c] Wald chi-square statistic, with one degree of freedom, for testing the significance of the effect of the variable.

[d] Odds ratio equals exp(coefficient). It indicates the amount by which the odds of an accident occurring increase for each unit increase in the variable.

Conversion: 1 m = 3.28 ft

Table 44. Logistic Regression Results for Total Multiple-vehicle Accidents at Urban, Three-leg, STOP-controlled Intersections

Independent variable[a]	Variable level	Direction of effect[b]	Coefficient	Chi-square statistic[c]	Odds ratio[d]
Intercept			–23.82	120.93	
Major-road ADT (log)		—	2.43	152.65	11.41
Crossroad ADT (log)		—	0.69	57.00	1.99
Major-road left-turn prohibition	1 if left turns prohibited; 0 otherwise	—	–1.46	11.88	0.23
Crossroad right-turn channelization	1 if no provision for free right turns; 0 otherwise	I	–1.93	18.27	0.15
Average lane width on major road (ft)		—	–0.19	5.58	0.83
Access control on major road	1 if none; 0 if partial	I	–0.89	3.30	0.41
Outside shoulder width on major road (ft)		—	–0.05	3.16	0.96
Crossroad left-turn prohibition	1 if left turns prohibited; 0 otherwise	—	–0.69	2.71	0.50

Regression statistics	
Number of intersections with 0 accidents (% of total)	659 (22%)
Number of intersections with at least 2 accidents (% of total)	372 (12%)
Chi-square value for	
a) the -2 log likelihood statistic	461.88 with 8 df ($p \leq 0.0001$)
b) the score statistic	356.16 with 8 df ($p \leq 0.0001$)
Residual chi-square	13.83 with 13 df ($p = 0.39$)
Estimated R^2 (correlation-based)	39.23%
Estimated R^2 (likelihood ratio)	36.11%
Adjusted R^2 (likelihood ratio)	49.49%

[a] All variables significant at the 90% confidence level or higher.

[b] Direction of effect: I = Inverse of expected direction.

[c] Wald chi-square statistic, with one degree of freedom, for testing the significance of the effect of the variable.

[d] Odds ratio equals exp(coefficient). It indicates the amount by which the odds of an accident occurring increase for each unit increase in the variable.

Conversion: 1 m = 3.28 ft

Table 45. Logistic Regression Results for Fatal and Injury Multiple-vehicle Accidents at Urban, Three-leg, STOP-controlled Intersections

Independent variable[a]	Variable level	Direction of effect[b]	Coefficient	Chi-square statistic[c]	Odds ratio[d]
Intercept			−21.90	115.80	
Major-road ADT (log)		—	2.05	126.91	7.74
Crossroad ADT (log)		—	0.68	72.27	1.97
Major-road left-turn prohibition	1 if left turns prohibited; 0 otherwise	—	−1.30	31.10	0.27
Crossroad right-turn channelization	1 if no provision for free right turns; 0 otherwise	I	−1.42	16.11	0.24
Access control on major road	1 if none; 0 if partial	I	−1.28	8.93	0.28
Major-road left-turn channelization	1 if curbed left-turn lane; 0 otherwise	I	0.67	11.11	1.95
Presence of median on major road	1 if divided; 0 if undivided	—	−0.38	3.60	0.69
Average lane width on major road (ft)		—	−0.11	2.85	0.89

<u>Regression statistics</u>	
Number of intersections with 0 accidents (% of total)	1,162 (38%)
Number of intersections with at least 1 accident (% of total)	339 (11%)
Chi-square value for	
a) the -2 log likelihood statistic	434.99 with 8 df ($p \leq 0.0001$)
b) the score statistic	365.15 with 8 df ($p \leq 0.0001$)
Residual chi-square	8.36 with 13 df ($p = 0.82$)
Estimated R^2 (correlation-based)	38.42%
Estimated R^2 (likelihood ratio)	25.16%
Adjusted R^2 (likelihood ratio)	38.32%

[a] All variables significant at the 90% confidence level or higher.

[b] Direction of effect: I = Inverse of expected direction.

[c] Wald chi-square statistic, with one degree of freedom, for testing the significance of the effect of the variable.

[d] Odds ratio equals exp(coefficient). It indicates the amount by which the odds of an accident occurring increase for each unit increase in the variable.

Conversion: 1 m = 3.28 ft

The interpretation of tables 38 through 45 is described below, using table 35 as an example. The fitted model for total multiple-vehicle accidents at rural, four-leg, STOP-controlled intersections can be written as:

$$\text{logit}(p) = -40.82 + 1.97X_1 + 2.88X_2 + 1.87X_3 + 0.67X_4 \quad (18)$$

where:

X_1 = $\log(ADT_{crossroad})$
X_2 = $\log(ADT_{major\ road})$
X_3 = 1 if 3 or fewer lanes on the major road; 0 otherwise
X_4 = 1 if terrain is flat; 0 otherwise

From the logit(p) value obtained in Eq. (18), the probability, p, of at least two accidents occurring in a single year is then calculated by substituting that value into Eq. (17).

Thus, the values of the model coefficients given in tables 38 through 45 can be used to compute the probability that at least a specified number of accidents will occur. Tables 38 through 45 show that logistic models fit the data relatively well and, in general, better than the lognormal or loglinear models developed in section 5. The estimated likelihood R^2-values ranged from 31 to 50 percent for total multiple-vehicle accidents. The R^2-values are slightly lower for fatal and injury multiple-vehicle accidents, ranging from 15 to 36 percent. The slight improvement in model fit from the loglinear or lognormal to the logistic models can be explained in part by the manner in which the data for the modeling were selected. While all intersections were included in the lognormal and loglinear models, from 36 to 66 percent of the intersections were excluded from the logistic regressions depending on the selection criterion based on accident frequencies. These discarded intersections represent the middle of the accident frequency distributions. Restricting the analyses to the extremes of the accident frequency distributions, as was done in logistic modeling, may explain the better fit of these models.

The tables show that the major-road and crossroad ADT variables are statistically significant in each of the logistic regression models developed. In addition, a number of key geometric design variables are statistically significant including left-turn and right-turn channelization and major-road lane and shoulder widths. However, as was the case for the loglinear and lognormal models presented in section 5, several of the geometric design variables in these models have effects that are in the opposite direction to those expected. These could represent surrogate effects of variables for which data are not available.

Accident Analyses for Specific ADT Classes

One concern in all of the previous analyses of at-grade intersection accidents with loglinear, lognormal, and logistic regressions discussed above is that the traffic volume variables (major-road ADT and crossroad ADT) have much stronger relationships to accident frequency than the geometric variables of interest and, therefore, account for most of the variation in accident frequency explained by the statistical models. This raised a concern that the strong effects of the ADT variables could be masking less strong, but potentially useful, relationships between geometric features and accidents. An attempt was made therefore to examine the effect of geometric variables on accidents in an analysis without including major-road and crossroad ADT as independent variables. This, however, is appropriate only if the variation of ADT is restricted to a relatively narrow range, so as to keep the intersections homogenous with respect to ADT. If this could be achieved, then one could repeat the loglinear and lognormal analyses of at-grade intersection accidents that were performed and discussed in section 5, but without including the major-road and crossroad ADT's as independent variables.

The key to structuring this analysis approach was in defining appropriate ADT strata within which the loglinear and lognormal regression analyses could be performed. Since both major-road ADT and crossroad ADT showed correlations with accident occurrence, it was necessary to define cells of intersections with similar traffic volume levels based on stratifications of both major-road ADT and crossroad ADT.

The following approach was applied separately to each of the five types of intersections studied. Two-way contingency tables of intersections were constructed in which each cell was defined by a range of major-road ADT and crossroad ADT. The starting point for defining the ADT ranges for each cell was a review of the distributions of major-road and crossroad ADT. In general, increments of 1,000 veh/day were used for major-road ADT on rural highways and 5,000 veh/day on urban highways. Increments of 100 veh/day for crossroad ADT were appropriate for all types of intersections. Based on the number of intersections falling within these predetermined ADT cells, a decision was made to focus the analyses on those portions of the contingency tables that included a reasonably large number of intersections (e.g., over 200 intersections within relatively narrow ranges of the two ADT variables). Table 46 summarizes the combinations of major-road and crossroad ADT ranges for each of the five types of intersections that appear to include enough intersections to make a statistical analysis worthwhile.

Within each group of intersections defined in table 46, the distributions of the independent variables (i.e., geometric design and traffic control features) considered in the full models (see section 5) were reevaluated. As in previous analyses, some categorical independent variables originally considered for inclusion in the models had to be discarded because either all or nearly all intersections had the same level for that variable.

Full models analogous to those obtained in section 5 (Poisson or lognormal) were developed within each group of intersections shown in table 46 (for a total of 16 analyses). In all cases, the major-road and crossroad ADT (variables) were excluded from the models. These regression analyses did not provide an improvement over those

Table 46. Major-road and Crossroad ADT Classes Used in Modeling Intersection Accidents

Type of intersection	Major-road ADT range (veh/day)	Crossroad ADT range (veh/day)	Number of intersections
Rural, four-leg, STOP-controlled	21,000 to 5,000	100 to 500	334
	5,000 to 10,000	100 to 500	302
Rural, three-leg, STOP-controlled	400 to 5,000	100 to 200	523
	5,000 to 10,000	100 to 200	287
	1,000 to 5,000	200 to 400	304
	5,000 to 10,000	200 to 400	231
Urban, four-leg, STOP-controlled	5,00 to 20,000	100 to 400	159
	10,000 to 20,000	400 to 900	133
Urban, three-leg, STOP-controlled	5,000 to 15,000	100 to 400	421
	15,000 to 25,000	100 to 400	355
	15,000 to 25,00	300 to 600	218
	25,000 to 35,000	400 to 700	167
	25,000 to 40,000	700 to 1,100	202
Urban, four-leg, signalized	15,000 to 25,000	100 to 5,000	149
	25,000 to 35,000	2,000 to 6,000	152
	35,000 to 50,000	4,000 to 9,000	137

performed in section 5. The models generally explained less of the variation in accident experience (i.e., had lower R^2-values) than the models developed in section 5, although of course all of the variation explained was attributable to geometric design and traffic control variables. In summary, the fit of the models, expressed as estimated R^2-value, ranged from a low of 6.4 percent to a high of 15.7 percent for total multiple-vehicle accidents, and from a low of 2.8 percent to a high of 13.3 percent for fatal and injury multiple-vehicle accidents. One exception to the general finding was in the models for urban, three-leg, STOP-controlled intersections with a major-road ADT in the 5,000 to 15,000 veh/day range and a crossroad ADT in the 100 to 400 veh/day range. In this case, the R^2-value was estimated at 35 percent and at 31 percent for total multiple-vehicle accidents and fatal and injury multiple-vehicle accidents, respectively.

These generally poor results are not surprising given that the two traffic volume variables accounted for most of the variation in accident frequency explained by the statistical models in section 5. Once the variability in accident frequencies attributed to traffic volume is removed from the data, geometric variables can explain only a relatively small proportion of the remainder of the variation in accidents.

Discriminant Analysis

Discriminant analysis is a statistical technique for predicting whether a particular observation in a data set should be classified in one of two or more predefined groups of data. This classification is based on a linear combination of independent variables that determine a discriminant "score" which serves as the basis for assigning the observation to a particular group. Discriminant analysis was a third alternative statistical approach that was tried to determine whether improved relationships between geometric features and accidents could be discerned. As applied to the California intersection data base, the predefined groups consisted of two groups of intersections: those with no (or low) accident experience and those with high accident experience.

The goal of the discriminant analysis is somewhat similar to that of logistic regression. In the latter, the objective of the analysis is to develop a linear relationship (logistic function) between the probability of an accident occurring at a given intersection and intersection geometric, traffic control, and traffic volume variables. In discriminant analysis, the objective is to classify an intersection into one of several groups (e.g., accident risk categories) by means of a discriminant function derived from the independent variables measured at that intersection. In either case, the groups of intersections need to be defined and the individual intersections classified into those groups prior to the analysis.

As with logistic regression, the outcome of discriminant analysis is not a predictive model for accident counts at intersections. Rather, it is a mathematical rule to predict whether an intersection is likely to belong to one of two (or more) predefined groups. However, discriminant analysis could be a useful approach to intersection analyses because the results might indicate which geometric design variables are useful in classifying the intersections into the high- and low-risk categories. Another concern is that the loglinear and lognormal regression analyses performed in section 5 have shown that major-road and crossroad ADT's account for a large proportion of the variability of the data when predicting accidents. One could expect that by grouping the intersections into risk categories that the ADT variables will still be major contributors in the discriminant function. However, it might be possible that a clearer "line" could be drawn between the extreme groups of no-risk and high-risk intersections, thus providing results of value to highway designers.

Similar to the logistic regression approach, a function (called the discriminant function) was defined based on the accident frequencies of a specific type of intersection. For example, intersections with no multiple-vehicle accidents in any of the 3 years were grouped into the **low-risk category**; intersections with a minimum of 2 or 3 accidents in every single year were grouped into the **high-risk category**. All other intersections were excluded from the analysis. The cut-off points for accident frequencies were determined separately for each type of intersection using the same criteria that were used earlier in section 6.

Total multiple-vehicle accidents

- Rural, 4-leg, STOP-controlled: 0 vs at least 2 multiple-vehicle accidents per year.
- Rural, 3-leg, STOP-controlled: 0 vs at least 1 multiple-vehicle accident per year.
- Urban, 4-leg, STOP-controlled: 0 vs at least 2 multiple-vehicle accidents per year.
- Urban, 3-leg, STOP-controlled: 0 vs at least 2 multiple-vehicle accidents per year.

Exception (due to high accident frequencies):

- Urban, 4-leg, signalized-controlled: 0, 1 or 2 multiple-vehicle accidents per year vs at least 10 multiple-vehicle accidents per year.

Fatal and injury multiple-vehicle accidents

- Rural, 4-leg, STOP-controlled: 0 vs at least 1 fatal or injury multiple-vehicle accident per year.
- Rural, 3-leg, STOP-controlled: 0 vs at least 1 fatal or injury multiple-vehicle accident per year.
- Urban, 4-leg, STOP-controlled: 0 vs at least 1 fatal or injury multiple-vehicle accident per year.
- Urban, 3-leg, STOP-controlled: 0 vs at least 1 fatal or injury multiple-vehicle accident per year.

Exception (due to high accident frequencies):

- Urban, 4-leg, signalized-controlled: 0 or 1 fatal or injury multiple-vehicle accident per year vs at least 5 fatal or injury multiple-vehicle accidents per year.

The statistical approach is then to determine a rule—a discriminant function—that will allow the classification of each intersection into one of the two risk groups on the basis of the values of its geometric design, traffic control, and traffic volume variables. Discriminant analyses were performed separately for total and fatal and injury multiple-vehicle accidents for each of the five groups of intersections. The independent variables considered here were those used in the previous logistic (section 6) or loglinear or lognormal (section 5 regression analyses). All of the discriminant analyses were performed using the PROC DISCRIM in SAS.(19) A test of homogeneity of the covariance matrices within each of the low- and high-risk groups was performed prior to estimating the discriminant function. In all cases, this test was significant at the 90 percent level, and the within-group (rather than pooled) covariance matrices were then used to estimate a discriminant function.

The overall success or performance of the discriminant analysis is judged by estimating the error rates of the model, or the probabilities of misclassification if the discriminant function were used to classify intersections other than those used in its development. This is done by cross-validation where n-1 intersections are used to define the discriminant function and the results applied to the nth intersection left out of the analysis. This is done for all n intersections used in the analysis. An error is made whenever an intersection is classified incorrectly, and the misclassification rate for each of the two groups is the proportion of intersections in that group that are misclassified. An overall error rate can be calculated as the total proportion of misclassified intersections (i.e., low-risk into high-risk group and vice versa). Table 47 summarizes the cross-validation results from the 10 discriminant analyses performed, separately for total and multiple-vehicle accidents in each of the five intersection categories.

Table 47. Cross-validation Summary Results from Discriminant Analyses

Intersection type	Number of intersections: low- vs high-risk group	Cross-validation: % misclassification in low- and high-risk groups	Overall error rate (%)
Total Multiple-Vehicle Accidents			
Rural, 4-leg, STOP-controlled	392 vs 147	9% and 20%	12
Rural, 3-leg, STOP-controlled	1,176 vs 293	12% and 39%	17
Urban, 4-leg, STOP-controlled	121 vs 328	41% and 11%	19
Urban, 3-leg, STOP-controlled	652 vs 361	21% and 29%	24
Urban, 4-leg, signalized	129 vs 176	64% and 5%	30
Fatal and Injury Multiple-Vehicle Accidents			
Rural, 4-leg, STOP-controlled	605 vs 175	11% and 37%	17
Rural, 3-leg, STOP-controlled	1,629 vs 103	4% and 64%	8
Urban, 4-leg, STOP-controlled	251 vs 302	29% and 26%	28
Urban, 3-leg, STOP-controlled	1,154 vs 330	18% and 42%	24
Urban, 4-leg, signalized	176 vs 107	14% and 29%	20

For each analysis, the second column in table 47 presents the number of intersections in each of the two risk groups as defined by the criteria listed above; the third column shows the error rates in each of the two risk groups; and the final column presents the overall error rate across both risk groups.

As shown in table 47, the rate of misclassifying intersections from the low-risk group (i.e., no accidents) varies from 4 to 64 percent. Similar rates of misclassification were estimated for the high-risk intersections, with rates varying from 5 to 64 percent. The overall (i.e., combined) error rates range from 8 to 30 percent.

As mentioned earlier, one of the objectives of this approach was to determine whether geometric variables could significantly contribute in the separation of the low- and high-risk intersection groups beyond the contribution provided by traffic volumes.

In each discriminant analysis performed, major-road and crossroad ADT were the most significant factors in the discriminant function. Table 48 highlights the small contribution provided by the additional geometric design or traffic control variables that were found to be statistically significant in each discriminant analysis. The measure used is the average squared canonical correlation, which is a measure of separation of the low- and high-risk intersection groups. The second column in table 48 shows the average squared canonical correlation based on major-road and crossroad ADT variables only. The last column shows the squared correlation obtained using all statistically significant variables, including both the ADT variables and the additional geometric design and traffic control variables that were considered. The improvement in separation power provided by the additional geometric design and traffic control variables is minimal in most cases, which confirms again the difficulty in developing relationships between accidents and intersection geometrics.

Table 48. Stepwise Discriminant Analysis Results (0 or low vs high numbers of accidents)

	Average squared canonical correlation[a]	
Intersection type	ADT variables only	All variables significant at the 10% level
Total Multiple-Vehicle Accidents		
Rural, 4-leg, STOP-controlled	0.50	0.51
Rural, 3-leg, STOP-controlled	0.34	0.36
Urban, 4-leg, STOP-controlled	0.32	0.34
Urban, 3-leg, STOP-controlled	0.26	0.33
Urban, 4-leg, signalized	0.34	0.45
Fatal and Injury Multiple-Vehicle Accidents		
Rural, 4-leg, STOP-controlled	0.32	0.36
Rural, 3-leg, STOP-controlled	0.14	0.15
Urban, 4-leg, STOP-controlled	0.22	0.24
Urban, 3-leg, STOP-controlled	0.18	0.21
Urban, 4-leg, signalized	0.34	0.45

[a] measure of separation of intersection groups (i.e., low- and high-risk)

Cluster Analysis

A fourth alternative statistical approach that was investigated involved the classification of intersections by means of cluster analysis. This statistical method differs from the discriminant analysis discussed in section 6 in that no a priori group identification (e.g., low- or high-risk) of the intersections is necessary. The purpose of cluster analysis is to place intersections into groups or clusters based on similarities in their geometrics, traffic control, and/or traffic volumes, based on the values of these variables, so that intersections in a given cluster tend to be similar to each other in their geometric design and other features, and intersections in different clusters tend to be dissimilar. Thus, in the first step of this approach, accident data are not necessary and the classification into clusters is based on the characteristics of the intersections alone. In the event that the clustering procedure is successful, in the sense that some similarities and dissimilarities among intersections can be found, then one could superimpose the accident data (e.g., total multiple-vehicle accident frequency or fatal and injury multiple-vehicle accident frequency) onto these clusters. This approach might show that some of the intersections clusters (i.e., some combinations of geometric and traffic control variables) are associated with either low- or high-accident frequencies.

There are many clustering algorithms available, each with their own specific assumptions, strengths, and weaknesses discussed in the statistical literature and associated software manuals. The analyses shown here were performed using the SAS software (in particular, the PROC CLUSTER and PROC TREE procedures).(17,19)

An attempt was made to apply cluster analysis to the approximately 1,300 urban, four-leg, signalized intersections for which traffic volumes, traffic control and geometric variables are available in the Caltrans data base. The descriptive variables used were those selected for use in the loglinear and lognormal regression analyses (see section 5). This approach was unsuccessful in two areas. First, no distinct groups were identified by the analysis; second, the sample size of 1,300 intersections was too large to allow the classification tree to be printed out conveniently. A 10 percent random sample from among the 1,300 urban, four-leg, signalized intersections was then selected and the cluster analysis repeated. Again, no clear groups of intersections similar in their geometric and traffic volume and traffic control variables could be identified by this method.

The classification tree of the sample of 130 intersections is presented in figure 12. In this figure, each intersection is identified by its 3-year total multiple-vehicle accident frequency printed on top of the figure. Ideally, distinct clusters of intersections would be identified by groups of branches separated by large amounts of white space. As shown in this figure, the page presents practically no white space, indicating that each intersection represents its own cluster, showing no similarities between them. If one were to draw a line at the 0.9-value on the vertical axis (which represents the relative distance between cluster centroids), then three clusters emerge, defined by the groups of intersections separated by the two thin vertical white lines. However, the 3-year total multiple-vehicle accident frequencies associated with each intersection (top line of numbers on the figure) do not suggest that these three groups of intersections are distinct in their accident experience. This result supports the lack of association between intersection geometrics, traffic control, and traffic volumes and intersection accident experience, which has also been evident in the other statistical analysis approaches evaluated.

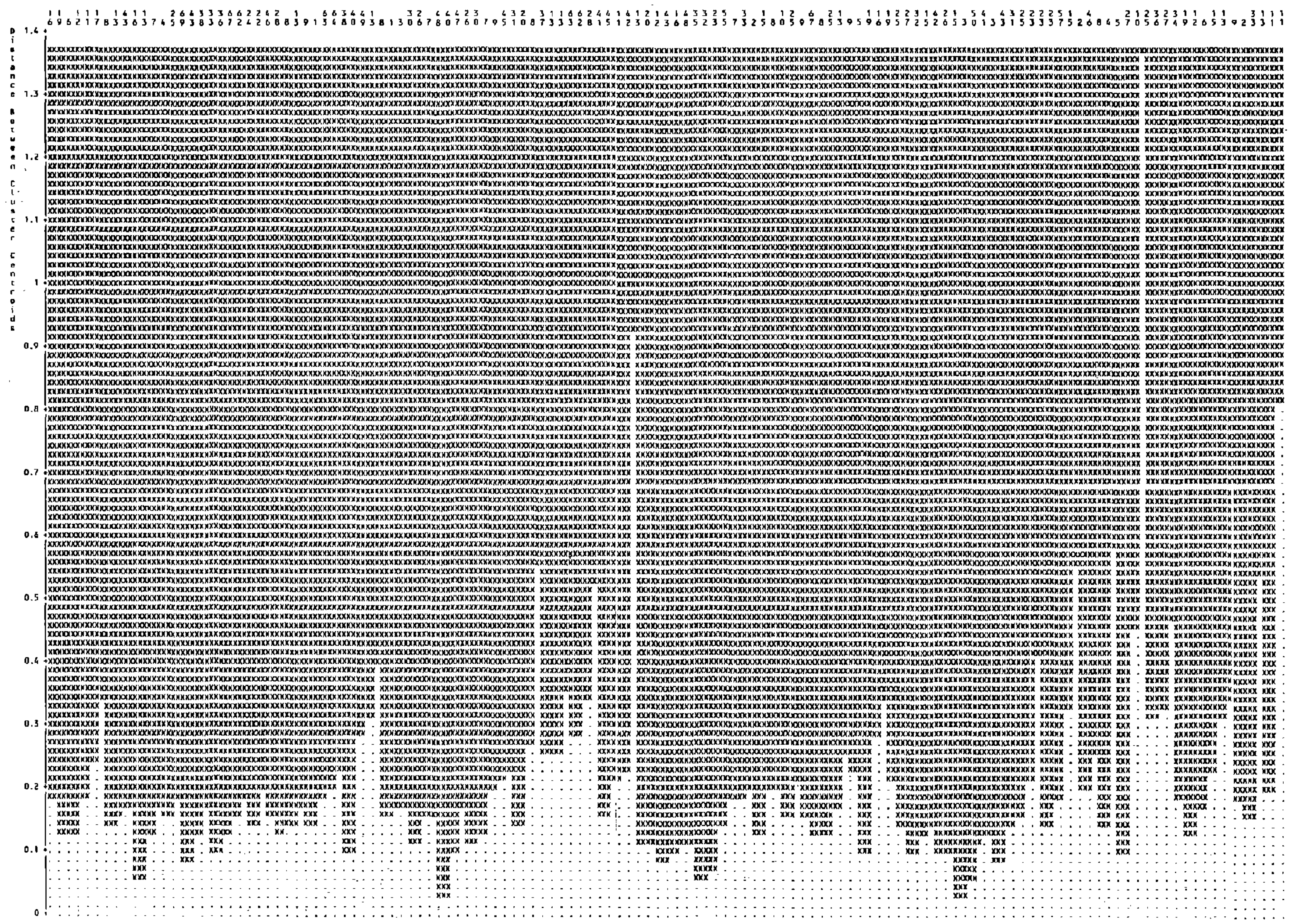


Figure 12. Cluster Analysis Results for a Sample of 130 Urban, Four-leg, Signalized Intersections

7. REVIEW OF HARD COPY POLICE ACCIDENT REPORTS

This section describes a review of hard copy police accident reports for selected intersections that was undertaken as part of the accident modeling research. This review was undertaken as part of the initial development of a technique by which accident report data could be used to assess the role of geometric elements as a causal or severity-increasing factor in traffic accidents. Intersection accidents were selected for this initial development work because of the availability of the data base for California intersections developed in this research.

Hard Copy Police Accident Reports

Hard copy police accident reports are the actual reports completed by police officers who conduct on-scene accident investigations. The computerized accident records systems maintained by highway and police agencies contain data that are extracted from the police accident report, but the police accident report also contains useful data that are not entered into the computer. For example, most police accident reports contain a diagram of the accident scene and the trajectories and/or final resting places of the involved vehicles. In addition, police accident reports include a narrative description prepared by the investigating officer that may include his assessment of the sequence of events in the accident, as well as the statements of the involved drivers and witnesses. The term "hard copy" is used because analysis of police accident reports is usually conducted using paper originals or photocopies of the reports, rather than microfilm copies or computerized data.

Objectives of Hard Copy Accident Report Review

The objective of the review was to provide a first step in the development of a technique for reviewing hard copy police accident reports to assess the extent that geometric elements caused, contributed to, or increased the severity of traffic accidents. The technique eventually developed from this approach might be a useful complement to statistical analyses of accident data and might even be effective in the development of hypotheses about accident causation that could be tested in formal statistical analyses. On the other hand, it was recognized that, in order to gain insights into accident causation, it might prove necessary to go beyond merely reviewing hard copy police accident reports and incorporate procedures such as making site visits to accident locations, performing accident reconstructions, or conducting multidisciplinary on-scene accident investigations. A diagnostic approach to accident investigation incorporating such procedures was recommended at a recent FHWA workshop on development of the IHSDM accident analysis module.(20) Thus, the review reported here is only the first step of a broader evaluation that could be performed to learn more about nontraditional methods for evaluating relationships between traffic accidents and geometric elements.

Preceding Page Blank

Hard Copy Accident Report Review Approach

Because of the availability of the geometric, traffic volume, traffic control, and accident data base for California intersections discussed earlier in this report, a decision was reached to use selected intersections from this data base for the hard copy accident report review. Eight intersections were selected from among the 198 urban, four-leg, signalized intersections that were included in the field study discussed in section 5 of this report. The eight intersections included four pairs of intersections with similar traffic volume levels. Each pair includes two intersections with very similar ADT levels on major road (i.e., the State highway) and the crossroad (usually not a State highway). However, one intersection of each pair had relatively high accident experience during the 3-year study period (1990-1992) and the other intersection had relatively low accident experience during that same period. The pairwise experimental design allowed us to consider whether geometric design elements can explain why one intersection in each pair had so many more accidents than the other, despite their similar ADT levels.

Table 49 summarizes the traffic volumes and accident experience of the selected intersections. As shown in the table, the eight study intersections experienced a total of 253 accidents during the 3-year study period. Hard copy police accident reports for 242 of these 253 accidents were obtained with the assistance of the California Department of Transportation. As in the previous analyses, the accident reports obtained included accidents that occurred within the curbline limits of the study intersections, as well as accidents that occurred up to 76 m (250 ft) from the intersection on each intersection leg.

Three reviewers were selected to independently review each of the 242 hard copy police accident reports. Each of the three reviewers was an experienced traffic safety researcher with a knowledge of past research results of geometric-safety relationships. The reviewers were asked to work independently so that FHWA would have the benefit of three different points of view concerning how such a review should be conducted. In other words, the three reviewers did not use a standardized procedure, but developed their own procedure. The questions that were asked of each reviewer, and the limited guidance that was provided to each reviewer concerning the interpretation of those questions, are summarized below. The reported results have been used to develop recommendations that will lead to a more standardized procedure for hard copy police accident report review.

Table 49. Characteristics of Intersections Selected for Hard Copy Police Accident Report Review

Intersection pair	Site number	ADT (veh/day)		Number of lanes						Number of multiple-vehicle accidents (1990-92)	Number of hard copy police accident reports obtained
				Major road			Minor road				
		Major road	Minor road	Right	Through	Left	Right	Through	Left		
1	2-40	21,200	6,200	0	2	1	1[a]	1	1[a]	9	9
	2-56	23,800	4,700	0	2	1	0	2	1	30	30
2	2-41	29,500	25,000	0	2	1	1[b]	2	1	4	4
	2-50	25,500	20,000	1	2	1	1[a]	2	1	45	45
3	4-39	58,300	3,500	0	4	1	0	1	0	15	15
	4-99	52,000	3,900	1	2	1	1[b]	2	1	43	40
4	4-04	42,000	9,000	0	2	1	1[c]	2	1	33	33
	4-01	45,000	9,500	0	4	1	0	2	1	74	67
								Total		253	242

[a] Right- and left-turn lanes on eastbound minor-road approach only.
[b] Right-turn lane on southbound minor-road approach only.
[c] Right-turn lane on westbound minor-road approach only.

Each reviewer was asked to use his own best judgement in answering the following questions concerning the accident described in each report:

- Was the cause of the accident related to the presence of the intersection (i.e., did this accident happen only because the intersection was there)? Generally, the answer should be YES if the accident occurred within the curbline limits of the intersection or if it occurred on an intersection approach and involved other traffic that was influenced by the presence of the intersection. The answer should generally be NO if the accident occurred outside the curbline limits of the intersection and the vehicles involved were headed AWAY from the intersection, or if a driveway was involved in the accident. The remaining questions were answered only for accidents that were classified as being related to the intersection.

- Was the cause of the accident DEFINITELY RELATED to the operation of the traffic signal at the intersection? POSSIBLY RELATED to the operation of the traffic signal at the intersection? or, NOT related to the operation of the traffic signal at the intersection? (select only one answer)

- Was the cause of the accident DEFINITELY RELATED to the geometric design of the intersection or its approaches? POSSIBLY RELATED to the geometric design of the intersection or its approaches? or, NOT RELATED to the geometric design of the intersection or its approaches? (select only one answer)

- Did geometric features DEFINITELY CONTRIBUTE to increasing the severity of the accident? POSSIBLY CONTRIBUTE to increasing the severity of the accident? NOT CONTRIBUTE to increasing the severity of the accident?

- Which geometric element(s) were most closely associated with the causation of the accident? (please describe)

- Were driver factors involved in the cause of the accident? Vehicle factors? Roadway and environment factors? (select all that apply)

  Examples of driver factors could include driver error or inattentiveness, disregard for traffic control devices, violation of the rules of the road, driver condition (e.g., fatigue, DWI), etc. Vehicle factors could include poor vehicle condition, faulty equipment, mechanical defects, loading or cargo problems, vehicle size and weight issues, etc. Roadway and environment factors could include pavement surface condition (wet surface, ice and snow, potholes, etc.), limited visibility (weather-related), influence of geometric design elements or traffic control devices, roadside design, etc.

In making these assessments, the reviewers had available a sketch of the layout of each intersection and a short videotape recorded on each intersection approach, but detailed data such as signal head placements, signal timing, and sight distance were not available.

Hard Copy Police Accident Report Review Results

This section summarizes the results of the review of hard copy police accident reports for urban, four-leg, signalized intersections.

Relationship of Accidents to the Study Intersections

Table 50 summarizes the results of the review concerning the relationship of each accident to the intersection being studied. For 212 of the 242 accidents (88 percent), all three reviewers agreed on whether the accident was, or was not, related to the intersection. The most common reason for finding that an accident was not related to the intersection were: (1) the accident occurred outside the curbline limits of the intersection and all of the involved vehicles were headed away from the intersection; and (2) the accident occurred outside the curbline limits of the intersection and the accident involved a vehicle entering or leaving an intersection approach at a driveway.

Table 50. Reviewers' Ratings of Number of Accidents Related to Each Intersection

	Reviewer 1		Reviewer 2		Reviewer 3	
Site	Yes	No	Yes	No	Yes	No
2-40	9	0	9	0	9	0
2-56	19	11	18	10	18	10
2-41	3	1	4	0	4	0
2-50	36	9	36	9	38	7
4-39	11	4	10	5	11	4
4-99	27	13	30	10	30	10
4-04	26	7	26	7	28	5
4-01	50	17	52	15	54	13
Total	181	62	185	56	192	49
Percentage	74.5	25.5	76.8	23.2	79.7	20.3

A total of 170 of the 242 accidents were found by all three reviewers to be related to the intersection. The evaluations presented below are all based on this set of 170 intersection-related accidents.

Relationship of Accidents to the Operation of the Traffic Signal at the Intersection

Table 51 summarizes the reviewers' ratings concerning a causal relationship between the accident and the operation of the traffic signal. As shown in the table, Reviewers 1 and 2 obtained very comparable results; they found that 90 and 85 percent of the accidents, respectively, had a causal relationship to the operation of the traffic signal. Reviewer 3, on the other hand, found that only 11 percent of the accidents had a causal relationship to the operation of the traffic signal. This difference arose from a difference in the definitions developed by each user. Reviewers 1 and 2 considered an accident to be related to the operation of the signal if the sequence of events in the accident was definitely or possibly influenced by the presence of the signal or if one or more of the involved vehicles disobeyed the signal. In contrast, Reviewer 3 identified the accident as related to the operation of the signal only if it appeared that the accident was definitely or possibly related to a correctable signal problem such as misplacement of signal heads, poor signal timing, or lack of a protected turn phase. The criterion used by Reviewers 1 and 2 is recommended for future review efforts, because the judgement made by Reviewer 3 concerning correctable signal problems seems better suited to a field visit than to an office review.

Relationship of Accidents to Geometric Features of the Intersection

Table 52 summarizes the reviewers' assessments of the relationship of the accidents to the geometric features of the intersections. Reviewers 1 through 3 found that 14, 5, and 7 percent of the accidents, respectively, were definitely or possible related to the geometric features of the intersections. Thus, there were only a very few accidents that appeared to involve the geometric features of the study intersections as a causal factor. For these few accidents, the geometric features that were noted as having some potential role in accident causation at the intersections were:

- Rear-end accidents resulting from left-turn lanes that were too short and from which traffic backed up into the through lanes.
- Left-turn accidents resulting from restricted sight distance due to the presence of one or more opposing left-turn vehicles.
- Sideswipe and turning accidents resulting from wide curb lanes on an intersection approach which created confusion about the proper path to follow and tempted some drivers to use the one wide lane as if it were two. The curb lane on an intersection approach is the farthest lane to the right, adjacent to the roadway curb.

Table 51. Reviewers' Ratings of Number of Accidents Related to the Operation of the Traffic Signal

Site	Reviewer 1			Reviewer 2			Reviewer 3		
	Definitely related	Possibly related	Not related	Definitely related	Possibly related	Not related	Definitely related	Possibly related	Not related
2-40	8	1	0	7	2	0	1	0	8
2-56	16	1	1	16	0	2	1	2	13
2-41	0	0	3	0	0	3	0	0	3
2-50	29	2	4	27	1	7	0	2	33
4-39	9	0	0		0	1	0	0	8
4-99	17	0	6	19	0	4	0	1	22
4-04	16	8	1	17	0	8	2	0	23
4-01	38	8	2	48	0	0	7	3	88
Total	133	20	17	142	3	25	11	8	148
Percentage	78.2	11.8	10.0	83.5	1.8	14.7	6.6	4.8	88.6

NOTE: Includes only accidents for which there was agreement that the accident was related to the intersection.

Table 52. Reviewers' Ratings of Number of Accidents Related to Geometric Features of the Intersection

Site	Reviewer 1			Reviewer 2			Reviewer 3		
	Definitely related	Possibly related	Not related	Definitely related	Possibly related	Not related	Definitely related	Possibly related	Not related
2-40	0	3	6	0	1	8	0	4	5
2-56	0	2	16	0	0	18	0	0	18
2-41	0	0	3	0	0	3	0	0	3
2-50	0	2	33	0	1	34	0	0	35
4-39	0	0	9	0	0	9	0	0	9
4-99	0	0	23	0	0	23	0	0	23
4-04	0	4	21	6	0	19	6	0	19
4-01	0	13	35	0	0	48	1	0	47
Total	0	24	146	6	2	162	7	4	159
Percentage	0.0	14.1	85.9	3.5	1.2	95.3	4.1	2.4	93.5

NOTE: Includes only accidents for which there was agreement that the accident was related to the intersection.

- Sideswipe and turning accidents resulting from short curb return radii which caused some heavy vehicles making right turns to swing wide and come into conflict with vehicles in adjacent lanes.
- Left-turning accidents that resulted from a horizontal curve on the opposing approach which limited the drivers view on oncoming vehicles.

All three reviewers considered an accident to have a causal relationship to a geometric feature if the accident occurred because the geometric feature was present or if the accident might have been prevented if that feature had been redesigned.

Role of Geometrics in Increasing Accident Severity

Table 53 summarizes the role of the geometric features of the intersections in increasing the severity of the accidents that occurred at the intersections. The three reviewers were in agreement that geometric features had no severity increasing effect in any of the accidents. The results appear to indicate quite definitely that geometrics had no role in increasing accident severity at intersections, but this finding should not be taken to indicate that geometrics might not have a role in increasing accident severity at other types of highway elements.

Factors Involved in Accident Causation

Table 54 presents the reviewers' ratings of the role of driver, vehicle, and roadway and environmental factors involved in accident causation at the study intersections. The three reviewers were in agreement that driver factors were involved in at least 97 percent of the accidents reviewed and that vehicle factors were involved in 6 to 9 percent of the accidents. There were some differences between the reviewers in the identification of roadway and environmental factors. Reviewer 1 identified a roadway or environmental factor related to 74 percent of the accidents because this reviewer noted a roadway factor in most cases in which the accident was related to the operation of the traffic signal. Reviewer 2 had used a definition of relationship to traffic signal similar to that used by Reviewer 1, but then identified roadway and environmental factors only for those accidents in which some roadway or environmental factor other than the traffic signal was noted. Reviewer 3 used a similar criterion to that used by Reviewer 2, but noted more accidents related to roadway and environmental factors (23 vs. 6 percent). The approach used by Reviewers 2 and 3 appears more appropriate because it eliminates a potential overlap between questions.

Table 53. Reviewers' Ratings of Number of Accidents for Which Geometric Features Had a Role in Increasing Accident Severity

Site	Reviewer 1			Reviewer 2			Reviewer 3		
	Definitely related	Possibly related	Not related	Definitely related	Possibly related	Not related	Definitely related	Possibly related	Not related
2-40	0	0	9	0	0	9	0	0	9
2-56	0	0	18	0	0	18	0	0	18
2-41	0	0	3	0	0	3	0	0	3
2-50	0	0	35	0	0	35	0	0	35
4-39	0	0	9	0	0	9	0	0	9
4-99	0	0	23	0	0	23	0	0	23
4-04	0	0	25	0	0	25	0	0	25
4-01	0	0	48	0	0	48	0	0	48
Total	0	0	170	0	0	170	0	0	170
Percentage	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0

NOTE: Includes only accidents for which there was agreement that the accident was related to the intersection.

Table 54. Reviewers' Ratings of Number of Accidents in Which Driver, Vehicle, and Roadway and Environmental Factors Had a Role

Site	Reviewer 1			Reviewer 2			Reviewer 3		
	Driver factors	Vehicle factors	Roadway and environment factors	Driver factors	Vehicle factors	Roadway and environment factors	Driver factors	Vehicle factors	Roadway and environment factors
2-40	8	1	9	9	1	1	8	2	4
2-56	18	0	4	18	0	1	18	0	4
2-41	3	0	3	3	0	3	3	0	3
2-50	34	6	23	35	5	3	34	5	7
4-39	9	0	8	9	0	0	9	0	0
4-99	23	0	19	23	0	0	23	0	0
4-04	25	7	16	23	6	0	23	3	8
4-01	48	2	44	48	3	3	48	2	14
Total	168	16	126	168	15	11	166	12	40
Percentage	98.2	9.4	73.7	98.2	8.8	6.4	97.1	7.0	23.4

NOTE: Includes only accidents for which there was agreement that the accident was related to the intersection. Column percentages for each reviewer add to more than 100% because many accidents involve more than one factor.

Findings and Recommendations

The hard copy police accident report review found that geometric features contributed to accident causation in only 5 to 14 percent of the accidents that occurred at the set of eight urban, four-leg, signalized intersections. This helps to explain the disappointing results of many of the statistical analyses presented earlier in the report. If 85 to 90 percent of the accidents that occur at this type of intersection have no obvious causal relationship to geometric features, it seems unlikely that statistical techniques would discern relationships that are not evident when the data are examined one accident at a time.

The hard copy accident report review did not lead to any specific insights concerning why some intersections had many more accidents than other intersections with similar traffic volume levels. This result seem inevitable given the low overall level of causal relationships between geometrics and accidents.

One reason for these finding may be that all of the intersections studied had relatively good geometrics, as they are located on the highway system operated by a State agency with a substantial construction budget and a large professional staff. More useful results might have been obtained with data from intersections operated by local agencies with more varied professional staff capabilities and funding levels. However, local agency data are much less accessible than state highway agency data and are not included in existing national data bases such as the FHWA Highway Safety Information System (HSIS). An equivalent data base of intersections maintained by local agencies would require a much greater effort to assemble.

Although the review of hard copy police accident reports found no indication of a relationship between intersection geometrics and accidents, a review of this could well provide useful results for some other highway features for which there is a stronger underlying relationship between geometrics and accidents. Furthermore, a hard copy accident report review could be a useful screening technique in setting directions for future safety research. For example, the hard copy police accident review technique used in this research may provide a useful tool for preliminary analyses in future research to develop statistical relationships between geometric features and accidents. This tool also may have appropriate applications to highway features other than intersections. For example, one could envision conducting a preliminary review of hard copy police accident reports for a small sample of locations before beginning a large statistical evaluation. If the preliminary review found very few accidents with an obvious relationship to the geometric feature(s) of interest, then the plans for the statistical research should perhaps be reconsidered. If, on the other hand, the review does find some role of geometric features in accident causation, a more detailed diagnostic accident investigation, including site visits to accident locations, accident reconstructions, or multidisciplinary on-scene accident investigations, might better define those causal relationships, prior to or in parallel with statistical research.

In order to serve this role, the hard copy accident report review technique will need to be further developed and standardized. It is recommended that this technique be considered in future research, either alone or in conjunction with more complete diagnostic studies, as a precursor and/or as a complement to more traditional statistical analyses of accident data.

8. CONCLUSIONS

The following conclusions were reached as a result of the statistical analysis of relationships between traffic accidents and geometrics of at-grade intersections conducted in this research.

1. Traditional multiple linear regression is generally not an appropriate statistical approach to modeling of accident relationships because accidents are discrete, non-negative events that often do not follow a normal distribution.

2. The Poisson, negative binomial, lognormal, and logistic distributions appear to be better suited to modeling of accident relationships than the normal distribution. In all cases, the form of the statistical distribution selected for any particular modeling should be chosen based on a review of the data to be modeled.

3. The choice between the Poisson and the negative binomial distributions should be based on the overdispersion observed in the accident data. Overdispersion results when the variance of the accident data exceeds the mean of the Poisson distribution. In the modeling of accidents for at-grade intersections, overdispersion was commonly observed and, therefore, the negative binomial distribution was preferred.

4. Regression models to determine relationships between accidents and intersection geometric design, traffic control, and traffic volume variables based on the negative binomial distribution explained between 16 and 38 percent of the variability in the accident data.

5. Models developed to predict total multiple-vehicle accidents generally performed slightly better than did models for fatal and injury multiple-vehicle accidents.

6. In general, the consideration of major-road ADT and crossroad ADT as separate independent variables provided better modeling results than consideration of a single variable representing either the sum or the product of the two ADT variables.

7. In negative binomial regression models for three of five specific intersection types, the major-road ADT and crossroad ADT variables accounted for most of the variability in accident data that was explained by the models. Geometric design variables accounted for a very small additional portion of the variability. The geometric design features of at-grade intersections whose effects on safety were statistically significant in negative binomial regression models included: presence of a separate left-turn lane; provision of channelization for free right turns; number of lanes on major road; average lane width on the major road; presence of a median on the major road; outside shoulder width on the major road; and access control on the major road. In some cases, however, the observed effects of these geometric design variables on accidents were in the opposite direction to that expected.

8. For urban, four-leg, signalized and STOP-controlled intersections, the lognormal distribution was found to be an appropriate choice for modeling of intersection accidents. These urban intersections experienced many more accidents than the other types of intersections evaluated and only a small number of them experienced no accidents in the 3-year period. The Poisson distribution was not an appropriate choice for modeling accidents at either of these two types of intersection. The negative binomial could have been used but, in all likelihood, would have had only a slight effect on the regression coefficient estimates. As in the models for the other three intersection types, the major-road ADT and the crossroad ADT variables accounted for most of the variability in accident data that was explained by the models. Geometric design features whose effects were found to be statistically significant included: major-road left-turn prohibition; access control on major road; average lane width on major road; number of lanes on major road; crossroad right-turn channelization; intersection lighting; design speed on major road; and outside shoulder width on major road. Traffic control features that were significant at signalized intersections were signal timing and signal phasing. In some cases, however, the observed effects of these geometric design and traffic control variables on accidents were in the opposite direction to that expected.

9. Field data were collected for a sample of urban, four-leg, signalized intersections, to provide data on geometric design variables and turning-movement counts that were not available from existing highway agency files. However, addition of these data to the existing data set did not increase the proportion in variation in accidents explained by the lognormal regression models.

10. An analysis was conducted to relate accidents involving collisions between left-turning vehicles and opposing through vehicles to the corresponding turning movement volumes and to geometric design and traffic control features related to left turns. However, these models explained only about 5 percent of the variability in the left-turn accident data. One reason that this modeling approach may not have been successful is that the limitation of the analysis to particular intersection approaches, and to particular times of the day for which turning counts were available, greatly reduced the sample size of accidents available for analysis.

11. Logistic regression was applied to develop models to predict the probability that an intersection would experience a specified number of accidents as a function of intersection geometric design, traffic control, and traffic volume variables. The models obtained explained slightly more of the accident variability than the lognormal or loglinear models. However, geometric design variables still accounted for only a small portion of the variability in accidents data explained by the models.

12. Several other statistical modeling approaches, including modeling within specific ADT classes, discriminant analysis, and cluster analysis, were investigated to see whether models in which geometric design variables explain more of the variation in accident data could be developed. However, none of these approaches provided

results that were preferable to those obtained from the negative binomial, lognormal, and logistic regressions.

13. An evaluation, by three independent reviewers, of hard-copy police accident reports for a sample of eight urban, four-leg, signalized intersections found that only 5 to 14 percent of the accidents had causes that appeared to be related to geometric design features of the intersections. This finding confirms the results of the statistical analyses, which indicated that geometric design features explain relatively little of the variability in intersection accident data for at-grade intersections.

14. While the models presented in this report are the best that can be developed from the available data, they do not appear to be of direct use to practioners. The models do not include effects for all geometric variables of potential interest to highway designers, and some of the effects they do include are in a direction opposite to that expected. Furthermore, the goodness of fit of the models is not as high as would be desired. Therefore, the models presented here are appropriate as a guide to future research but do not appear to be appropriate for direct application by practitioners.

APPENDIX A

DEFINITIONS OF GEOMETRIC DESIGN, TRAFFIC CONTROL, AND TRAFFIC VOLUME VARIABLES

This appendix presents definitions of the geometric design, traffic control, and traffic volume variables considered in the statistical modeling of at-grade intersection accidents. Table 55 presents the definitions of variables from the existing Caltrans data base, while table 56 presents the definitions of variables that were collected for a selected sample of 198 urban, four-leg, signalized intersections in the pilot field studies. Each table identifies the variables that were considered and identifies whether each variable was continuous or categorical in nature. Continuous variables are those with quantitative values on a continuous scale. Categorical variables are those with a finite number of discrete levels or categories. For each categorical variable, the tables in this appendix also identify the levels or categories that were available for that variable. Not all levels were considered in the statistical modeling; in some cases, because of sample size considerations, a particular category had to be excluded from the analyses or merged with adjacent categories. This process is described in section 5 of the main text of this report.

Table 55. Definitions of Variables Available in the Existing Caltrans Data Base

Variable	Variable type	Range of values/levels
Geometric Design Features		
Intersection configuration	Categorical	Three-leg T intersection Three-leg Y intersection Four-leg intersection Four-leg offset intersection Multi-leg intersection
Number of lanes on major road (both directions of travel combined)	Categorical	Range: 2 to 8 lanes
Number of lanes on crossroad	Categorical	Range: 2 to 8 lanes
Presence of median on major road	Categorical	Divided Undivided
Median width on major road (ft)	Continuous	Range: 0 to 99 ft or more
Average lane width on major road (ft) (computed as total traveled way width divided by total number of lanes)	Continuous	Range: 8 to 15 ft
Outside shoulder width on major road (i.e., shoulder width on right side of road)	Continuous	Range: 0 to 15 ft
Design speed of major road (mi/h)	Categorical	less than 30 mi/h 30 mi/h 35 mi/h 40 mi/h 45 mi/h 50 mi/h 55 mi/h 60 mi/h 65 mi/h greater than 65 mi/h
Functional classification of major road	Categorical	Principal arterial Minor arterial Major collector Minor collector Local
Presence of left-turn channelization on major road	Categorical	No left-turn lane Painted left-turn lane Curbed left-turn lane
Presence of left-turn channelization on crossroad	Categorical	No left-turn lane Painted left-turn lane Curbed left-turn lane
Presence of right-turn channelization on major road	Categorical	Provision for free right turns No provision for free right turns
Presence of right-turn channelization on crossroad	Categorical	Provision for free right turns No provision for free right turns
Presence of access control on major road	Categorical	None Partial

Conversion: 1 km/h = 0.621 mi/h; 1 m = 3.28 ft

Table 55. Definitions of Variables Available in the Existing Caltrans Data Base (Continued)

Variable	Variable type	Range of values/levels
Traffic Control Features		
Type of intersection control	Categorical	No control Two-way STOP control Four-way STOP control YIELD control Signal control
One-way vs. two-way operation on major road	Categorical	Two-way operation One-way operation
Left-turn prohibition from major road	Categorical	Left turn permitted Left turn prohibited
Left-turn prohibition from crossroad	Categorical	Left turn permitted Left turn prohibited
Presence of mast-arm signals on major road	Categorical	Mast-arm signal present No mast-arm signals
Presence of mast-arm signals on crossroad	Categorical	Mast-arm signal present No mast-arm signals
Signal timing	Categorical	Pretimed Semi-actuated Fully actuated
Signal phasing	Categorical	Two-phase Multiphase
Traffic Volume Data		
ADT of major road (veh/day)	Continuous	Range: 400 to 97,000 veh/day
ADT of crossroad (veh/day)	Continuous	Range: 100 to 48,000 veh/day
Other Related Data		
Rural/urban	Categorical	Rural Urban
Terrain	Categorical	Level Rolling Mountainous
Presence of intersection lighting	Categorical	Intersection lighted Intersection not lighted

Conversion: 1 km/h = 0.621 mi/h; 1 m = 3.28 ft

Table 56. Definitions of Additional Variables Obtained in Pilot Field Study

Variable	Variable type	Range of values/levels
Geometric Design Features		
Number of through lanes on each approach	Categorical	Range: 1 to 4 lanes
Number of exclusive left-turn lanes on each approach	Categorical	Range: 0 to 2 lanes
Number of exclusive right-turn lanes on each approach	Categorical	Range: 0 to 2 lanes
Type of left-turn treatment on each approach	Categorical	No left-turn lane Painted left-turn lane Curbed left-turn lane
Type of right-turn treatment on each approach	Categorical	No right-turn treatment Separate right-turn roadway without exclusive turn lane Separate right-turn roadway with exclusive turn lane Exclusive right-turn lane without separate right-turn roadway
Horizontal alignment of each approach (based on curves within 76 m or 250 ft of the intersection)	Categorical	Tangent Gentle curve (radius over 2,000 ft) Moderate curve (radius from 500 to 2,000 ft) Sharp curve (radius less than 500 ft)
Percent grade on each approach	Categorical	Level (less than 2% grade) Moderate grade (2% to 4% grade) Steep grade (over 4% grade)
Direction of grade on each approach (for each approach with a moderate or steep grade)	Categorical	Upgrade Downgrade
Crest/sag vertical curve on each approach	Categorical	None Crest vertical curve on approach Sag vertical curve on approach
Total through lane width on each approach (ft)	Continuous	Range: 8.5 to 59.5 ft
Total left-turn lane width on each approach (ft)	Continuous	Range: 8 to 25 ft
Presence of median on each approach	Categorical	Divided Undivided
Type of median on each approach	Categorical	No median Raised median (curbed) Depressed median Flush median

Conversion: 1 m = 3.28 ft

Table 56. Definitions of Additional Variables Obtained in Pilot Field Study (Continued)

Variable	Variable type	Range of values/levels
Median width on each approach (ft)	Continuous	Range: 0 to 99 ft or more
Number of driveways within 76 m (250 ft) of intersection on each approach	Continuous	Range: 0 to 10
Type of driveways on each approach	Categorical	No driveways Commercial Industrial Residential Combinations of above types
Angle between intersection approaches	Continuous	Range: 45° to 90°
Traffic Control Features		
One-way vs. two-way operation on each approach	Categorical	Two-way operation One-way operation
Presence of left-turn prohibition on each approach	Categorical	Left turns permitted Left turns prohibited
Curb parking on right side of roadway within 250 ft of intersection on each approach	Categorical	No curb parking Parallel parking Angle parking
Number of signal faces for each approach	Continuous	Range: 2 to 8
Signal head mounting for each approach	Categorical	Post-mounted signals Mast-arm signals
Left-turn phasing for each approach	Categorical	No separate left-turn phase Protected left-turn phase Protected left-turn phase with left turns permitted on green ball
Presence of pedestrian signals for crossing each approach	Categorical	Pedestrian signals present Pedestrian signals not present
Presence of painted crosswalk for crossing each approach	Categorical	Crosswalk marked Crosswalk not marked
Presence of advance warning signs (e.g., SIGNAL AHEAD) on each approach	Categorical	Advance warning signs present Advance warning signs not present
Posted speed limit for each approach (mi/h)	Categorical	Range: 15 to 55 mi/h

Conversion: 1 km/h = 0.621 mi/h; 1 m = 3.28 ft

Table 56. Definitions of Additional Variables Obtained in Pilot Field Study (Continued)

Variable	Variable type	Range of values/levels
Traffic Volume Data		
Turning movement volumes from each approach per 15-min period in morning peak hour (typically 7 to 9 a.m.) and evening peak hour (typically 4 to 6 p.m.)	Continuous	Through volume range: 0 to 807 vehicles Left-turn volume range: 0 to 215 vehicles Right-turn volume range: 0 to 326 vehicles
Level of pedestrian activity for each intersection as a whole	Categorical	Low (almost no pedestrian activity) Medium (pedestrian activity with some frequency) High (pedestrian activity during every signal cycle)
Other Related Data		
Presence of intersection lighting	Categorical	None Continuous street lighting Lighting at intersection only
Character of surrounding development	Categorical	Central business district Outlying business district Industrial district Mixed commercial and residential development Residential development Other

APPENDIX B

TOTAL MULTIPLE-VEHICLE ACCIDENT DATA DISTRIBUTIONS

Table 57. Total Multiple-vehicle Accident Data Distribution at Rural, Four-leg, STOP-controlled Intersections

Number of accidents in 3 years	Number of intersections	Percent of intersections	Cumulative number of intersections	Cumulative percent of intersections
0	392	27.3	392	27.3
1	243	16.9	635	44.3
2	206	14.4	841	58.6
3	142	9.9	983	68.5
4	107	7.5	1,090	76.0
5	60	4.2	1,150	80.2
6	50	3.5	1,200	83.7
7	45	3.1	1,245	86.8
8	35	2.4	1,280	89.3
9	28	2.0	1,308	91.2
10	20	1.4	1,328	92.6
11	18	1.3	1,346	93.9
12	19	1.3	1,365	95.2
13	10	0.7	1,375	95.9
14	8	0.6	1,383	96.4
15	11	0.8	1,394	97.2
16	5	0.3	1,399	97.6
17	2	0.1	1,401	97.7
18	6	0.4	1,407	98.1
19	3	0.2	1,410	98.3
20	4	0.3	1,414	98.6
21	6	0.4	1,420	99.0
22	2	0.1	1,422	99.2
23	1	0.1	1,423	99.2
24	1	0.1	1,424	99.3
25	3	0.2	1,427	99.5
26	2	0.1	1,429	99.7
27	2	0.1	1,431	99.8
30	1	0.1	1,432	99.9
34	1	0.1	1,433	99.9
48	1	0.1	1,434	100

Table 58. Total Multiple-vehicle Accident Data Distribution at Rural, Three-leg, STOP-controlled Intersections

Number of accidents in 3 years	Number of intersections	Percent of intersections	Cumulative number of intersections	Cumulative percent of intersections
0	1,174	43.6	1,174	43.6
1	605	22.5	1,779	66.1
2	331	12.3	2,110	78.4
3	186	6.9	2,296	85.3
4	107	4.0	2,403	89.3
5	69	2.6	2,472	91.8
6	54	2.0	2,526	93.8
7	33	1.2	2,559	95.1
8	18	0.7	2,577	95.7
9	25	0.9	2,602	96.7
10	20	0.7	2,622	97.4
11	11	0.4	2,633	97.8
12	7	0.3	2,640	98.1
13	12	0.4	2,652	98.5
14	5	0.2	2,657	98.7
15	6	0.2	2,663	98.9
16	7	0.3	2,670	99.2
17	4	0.1	2,674	99.3
18	2	0.1	2,676	99.4
19	1	0.04	2,677	99.4
20	2	0.1	2,679	99.5
21	3	0.1	2,682	99.6
22	2	0.1	2,684	99.7
23	1	0.04	2,685	99.7
24	1	0.04	2,686	99.8
25	1	0.04	2,687	99.8
28	1	0.04	2,688	99.9
29	1	0.04	2,689	99.9
33	1	0.04	2,690	99.9
37	1	0.04	2,691	100
45	1	0.04	2,692	100

Table 59. Total Multiple-vehicle Accident Data Distribution at Urban, Four-leg, STOP-controlled Intersections

Number of accidents in 3 years	Number of intersections	Percent of intersections	Cumulative number of intersections	Cumulative percent of intersections
0	138	10.3	138	10.3
1	160	11.9	298	22.2
2	137	10.2	435	32.4
3	144	10.7	579	43.1
4	103	7.7	682	50.8
5	77	5.7	759	56.6
6	82	6.1	841	62.7
7	77	5.7	918	68.4
8	42	3.1	960	71.5
9	63	4.7	1,023	76.2
10	44	3.3	1,067	79.5
11	43	3.2	1,110	82.7
12	23	1.7	1,133	84.4
13	37	2.8	1,170	87.2
14	21	1.6	1,191	88.7
15	24	1.8	1,215	90.5
16	14	1.0	1,229	91.6
17	18	1.3	1,247	92.9
18	15	1.1	1,262	94.0
19	10	0.7	1,272	94.8
20	9	0.7	1,281	95.5
21	9	0.7	1,290	96.1
22	8	0.6	1,298	96.7
23	7	0.5	1,305	97.2
24	4	0.3	1,309	97.5
25	3	0.2	1,312	97.8
26	4	0.3	1,316	98.1
27	3	0.2	1,319	98.3
28	2	0.1	1,321	98.4
29	1	0.1	1,322	98.5
30	3	0.2	1,325	98.7
31	3	0.2	1,328	99.0
32	1	0.1	1,329	99.0
33	1	0.1	1,330	99.1
34	1	0.1	1,331	99.2
35	3	0.2	1,334	99.4
36	1	0.1	1,335	99.5
37	1	0.1	1,336	99.5
39	1	0.1	1,337	99.6
45	3	0.2	1,340	99.8
49	1	0.1	1,341	99.9
53	1	0.1	1,342	100

Table 60. Total Multiple-vehicle Accident Data Distribution at Urban, Three-leg, STOP-controlled Intersections

Number of accidents in 3 years	Number of intersections	Percent of intersections	Cumulative number of intersections	Cumulative percent of intersections
0	659	21.6	659	21.6
1	560	18.3	1,219	39.9
2	407	13.3	1,626	53.2
3	294	9.6	1,920	62.8
4	236	7.7	2,156	70.5
5	177	5.8	2,333	76.3
6	161	5.3	2,494	81.6
7	93	3.0	2,587	84.6
8	93	3.0	2,680	87.7
9	62	2.0	2,742	89.7
10	65	2.1	2,807	91.8
11	42	1.4	2,849	93.2
12	47	1.5	2,896	94.7
13	20	0.7	2,916	95.4
14	26	0.9	2,942	96.2
15	19	0.6	2,961	96.9
16	18	0.6	2,979	97.5
17	11	0.4	2,990	97.8
18	6	0.2	2,996	98.0
19	9	0.3	3,005	98.3
20	7	0.2	3,012	98.5
21	2	0.1	3,014	98.6
22	5	0.2	3,019	98.8
23	3	0.1	3,022	98.9
24	3	0.1	3,025	99.0
25	2	0.1	3,027	99.0
26	1	0.03	3,028	99.1
27	4	0.1	3,032	99.2
28	1	0.03	3,033	99.2
29	2	0.1	3,035	99.3
30	3	0.1	3,038	99.4
31	1	0.03	3,039	99.4
33	1	0.03	3,040	99.4
34	2	0.1	3,042	99.5
35	1	0.03	3,043	99.5
36	1	0.03	3,044	99.6
37	4	0.1	3,048	99.7
38	1	0.03	3,049	99.7
40	1	0.03	3,050	99.8
42	3	0.1	3,053	99.9
44	1	0.03	3,054	99.9
57	1	0.03	3,055	99.9
80	2	0.07	3,057	100

Table 61. Total Multiple-vehicle Accident Data Distribution at Urban, Four-leg, Signalized Intersections

Number of accidents in 3 years	Number of intersections	Percent of intersections	Cumulative number of intersections	Cumulative percent of intersections
0	21	1.6	21	1.6
1	23	1.8	44	3.4
2	33	2.5	77	5.9
3	22	1.7	99	7.6
4	33	2.5	132	10.1
5	29	2.2	161	12.3
6	32	2.5	193	14.8
7	43	3.3	236	18.1
8	46	3.5	282	21.6
9	43	3.3	325	24.9
10	43	3.3	368	28.2
11	36	2.8	404	30.9
12	46	3.5	450	34.5
13	55	4.2	505	38.7
14	36	2.8	541	41.4
15	36	2.8	577	44.2
16	32	2.5	609	46.6
17	33	2.5	642	49.2
18	40	3.1	682	52.2
19	35	2.7	717	54.9
20	36	2.8	753	57.7
21	22	1.7	775	59.3
22	25	1.9	800	61.3
23	23	1.8	823	63.0
24	21	1.6	844	64.6
25	37	2.8	881	67.5
26	22	1.7	903	69.1
27	21	1.6	924	70.8
28	18	1.4	942	72.1
29	28	2.1	970	74.3
30	19	1.5	989	75.7
31	27	2.1	1,016	77.8
32	14	1.1	1,030	78.9
33	15	1.1	1,045	80.0
34	21	1.6	1,066	81.6
35	16	1.2	1,082	82.9
36	21	1.6	1,103	84.5
37	11	0.8	1,114	85.3
38	12	0.9	1,126	86.2
39	12	0.9	1,138	87.1
40	13	1.0	1,151	88.1
41	10	0.8	1,161	88.9
42	9	0.7	1,170	89.6
43	10	0.8	1,180	90.4

Table 61. Total Multiple-vehicle Accident Data Distribution at Urban, Four-leg, Signalized Intersections (Continued)

Number of accidents in 3 years	Number of intersections	Percent of intersections	Cumulative number of intersections	Cumulative percent of intersections
44	7	0.5	1,187	90.9
45	9	0.7	1,196	91.6
46	6	0.5	1,202	92.0
47	5	0.4	1,207	92.4
48	7	0.5	1,214	93.0
49	3	0.2	1,217	93.2
50	7	0.5	1,224	93.7
51	5	0.4	1,229	94.1
52	2	0.2	1,231	94.3
53	4	0.3	1,235	94.6
54	5	0.4	1,240	94.9
55	6	0.5	1,246	95.4
56	2	0.2	1,248	95.6
57	2	0.2	1,250	95.7
58	3	0.2	1,253	95.9
59	2	0.2	1,255	96.1
60	2	0.2	1,257	96.3
61	2	0.2	1,259	96.4
62	4	0.3	1,263	96.7
63	3	0.2	1,266	96.9
64	4	0.3	1,270	97.2
65	4	0.3	1,274	97.6
66	1	0.1	1,275	97.6
67	5	0.4	1,280	98.0
69	2	0.2	1,282	98.2
70	2	0.2	1,284	98.3
71	1	0.1	1,285	98.4
72	2	0.2	1,287	98.5
73	2	0.2	1,289	98.7
74	2	0.2	1,291	98.9
78	1	0.1	1,292	98.9
79	3	0.2	1,295	99.2
80	1	0.1	1,296	99.2
84	3	0.2	1,299	99.5
89	1	0.1	1,300	99.5
90	1	0.1	1,301	99.6
92	1	0.1	1,302	99.7
95	1	0.1	1,303	99.8
119	1	0.1	1,304	99.8
129	1	0.1	1,305	99.9
147	1	0.1	1,306	100

Table 62. Total Multiple-vehicle Accident Data Distribution at a Sample of 198 Urban, Four-leg, Signalized Intersections

Number of accidents in 3 years	Number of intersections	Percent of intersections	Cumulative number of intersections	Cumulative percent of intersections
0	1	0.5	1	0.5
1	1	0.5	2	1.0
2	7	3.5	9	4.6
3	2	1.0	11	5.6
4	6	3.0	17	8.6
5	3	1.5	20	10.1
6	4	2.0	24	12.1
7	3	1.5	27	13.6
8	3	1.5	30	15.2
9	6	3.0	36	18.2
10	4	2.0	40	20.2
11	8	4.0	48	24.2
12	5	2.5	53	26.8
13	7	3.5	60	30.3
14	10	5.1	70	35.4
15	9	4.5	79	39.9
16	8	4.0	87	43.9
17	3	1.5	90	45.5
18	5	2.5	95	48.0
19	4	2.0	99	50.0
20	6	3.0	105	53.0
21	1	0.5	106	53.5
22	8	4.0	114	57.6
23	3	1.5	117	59.1
24	3	1.5	120	60.6
25	4	2.0	124	62.6
26	3	1.5	127	64.1
27	1	0.5	128	64.7
28	3	1.5	131	66.2
29	4	2.0	135	68.2
30	2	1.0	137	69.2
31	5	2.5	142	71.7
32	3	1.5	145	73.2
33	2	1.0	147	74.2
34	1	0.5	148	74.8
35	4	2.0	152	76.8
36	2	1.0	154	77.8
37	2	1.0	156	78.8
38	3	1.5	159	80.3
39	2	1.0	161	81.3
40	3	1.5	164	82.8
41	4	2.0	168	84.9
42	2	1.0	170	85.9

Table 62. Total Multiple-vehicle Accident Data Distribution at a Sample of 198 Urban, Four-leg, Signalized Intersections (Continued)

Number of accidents in 3 years	Number of intersections	Percent of intersections	Cumulative number of intersections	Cumulative percent of intersections
43	3	1.5	173	87.4
44	1	0.5	174	87.9
45	1	0.5	175	88.4
47	1	0.5	176	88.9
48	3	1.5	179	90.4
49	1	0.5	180	90.9
50	1	0.5	181	91.4
51	2	1.0	183	92.4
53	1	0.5	184	92.9
54	1	0.5	185	93.4
55	2	1.0	187	94.4
58	1	0.5	188	95.0
62	1	0.5	189	95.5
64	1	0.5	190	96.0
66	1	0.5	191	96.5
70	1	0.5	192	97.0
71	1	0.5	193	97.5
74	2	1.0	195	98.5
79	1	0.5	196	99.0
84	1	0.5	197	99.5
119	1	0.5	198	100

REFERENCES

1. Bonneson, J. A. and McCoy, P. T. Estimation of Safety at Two-Way Stop-Controlled Intersections on Rural Highways, presented at the 72nd Annual Meeting of the Transportation Research Board, January 1993.

2. Joshua, S. C. and Garber N. J. Estimating Truck Accident Rate and Involvements Using Linear and Poisson Regression Models. *Transportation Planning and Technology*, Vol. 15. pp. 41-58. 1990.

3. Blower, D., Campbell, K. and Gree, P. E. Accident Rates for Heavy Truck-tractors in Michigan. *Accident Analysis and Prevention*, Vol. 25(3), pp. 307-321, 1993.

4. Miaou, S.-H. and Lum, H. Modeling Vehicle Accidents and Highway Geometric Design Relationships. *Accident Analysis and Prevention*, Vol. 25(6), pp. 689-709. 1993.

5. Hauer, E., Ng, J. C. N. , and Lovell, J. "Estimation of Safety at Signalized Intersections." *Transportation Research Record 1185*, 1988.

6. Knuiman, M. W., Council, F. M., and Reinfurt, D. W. The Effect of Median Width on Highway Accident Rates. Transportation Research Record 1401. 1993.

7. Miaou, S.-P. et al. Development of Relationship Between Truck Accidents and Geometric Design: Phase I. FHWA Publication No. FHWA-RD-91-124. Federal Highway Administration, McLean, VA. March 1993.

8. Shankar, V., F. L. Mannering, and W. Barfield, "Effect of Roadway Geometrics and Environmental Factors on Rural Accident Frequencies," presented at the 74th Annual Meeting of the Transportation Research Board, January 1995.

9. Hadi, M. A., J. Aruldhas, L. Chow, and J. A. Wattleworth, "Estimating the Safety Effects of Cross-Section Design for Various Highway Types Using Negative Binomial Regression," presented at the 74th Annual Meeting of the Transportation Research Board, January 1995.

10. Miaou, S.-P. The Relationship between Truck Accidents and Geometric Design of Road Sections: Poisson Versus Negative Binomial Regressions. *Accident Analysis and Prevention*, Vol. 26(4), pp. 471-482. 1994.

11. Lawless, J. F. Negative Binomial and Mixed Poisson Regression. *Canadian Journal of Statistics*, 15, pp. 209-225. 1987.

12. Poch, M. and F. L. Mannering, "Negative Binomial Analysis of Intersection Accident Frequencies," presented at the 75th Annual Meeting of the 75th Annual Meeting of the Transportation Research Board, January 1996.

13. Nelder, J. A., and R. W. M. Wedderburn, "Generalized Linear Models," *JRSS A* (1972), 135 Part 3, p. 370.

14. McCullagh, P., and Nelder, J. A., *Generalized Linear Models,* Second Edition (1989), New York: Chapman & Hall.

15. SAS Institute Inc., SAS® Technical Report P-243, SAS/STAT® Software: The GENMOD Procedure, Release 6.09, Cary, NC: SAS Institute Inc., 1993, 88 pp.

16. Fridstrøm, L., et al., "Measuring the Contribution of Randomness, Exposure, Weather, and Daylight to the Variation in Road Accident Counts," *Accident Analysis and Prevention* (1995), Vol. 27, pp. 1-20.

17. SAS Institute Inc., SAS® User's Guide, Version 6, Fourth Edition, Volume 2, Cary, NC: SAS Institute Inc., 1989, 446 pp.

18. SAS Institute Inc., Logistic Regression Examples Using the SAS® System, Version 6, First Edition, Volume 2, Cary, NC: SAS Institute Inc., 1995, 163 pp.

19. SAS Institute Inc., SAS/STAT® User's Guide, Version 6, Fourth Edition, Volume 1, Cary, NC: SAS Institute Inc., 1989, 943 pp.

20. Harwood, D. W., K. M. Bauer, J. M. Mason, and W. A. Stimpson, "Workshop on Development of the Interactive Highway Safety Design Model Accident Analysis Module," Report No. FHWA-RD-96-075, Federal Highway Administration, McLean, VA, 1996.