Title

An Ordinal Multivariate Analysis of Accident Counts as Functions of Traffic Approach Volumes at Intersections

Authors
Golob, Thomas F.
Ruhl, Bolie
Meurs, Henk
et al.

Publication Date
1987-10-01

An Ordinal Multivariate Analysis of Accident Counts as Functions of Traffic Approach Volumes at Intersections

UCI-ITS-WP-87-6

Thomas F. Golob [1]
Bolie Ruhl [2]
Henk Meurs [2]
Leo van Wissen [2]

[1] Institute of Transportation Studies
University of California, Irvine

[2] Bureau Goudappel Coffeng, Deventer, The Netherlands

October 1987

Institute of Transportation Studies
University of California, Irvine
Irvine, CA 92697-3600, U.S.A.
http://www.its.uci.edu

This research was sponsored in part by the Division of Traffic Engineering of the Ministry of Transportation and Public Works, The Netherlands.

Published in: *Accident Analysis and Prevention*, 20(5), 1988, 335-355.

ABSTRACT

This research is concerned with the statistical analysis of accident counts at non-signalized intersections. The objective is to develop a method for determining general (non-linear) relationships between approach volumes and accident counts. The method must accommodate the testing of whether intersections of differing physical designs have higher or lower rates of accidents than predicted by traffic levels. It is assumed that only aggregate data are available: (1) counts of total accidents by type (e.g., injury versus property damage) without details concerning the locational position(s) of the vehicle(s) involved; and (2) aggregate traffic intensity on each intersection entry without details concerning turning volumes. The method involves the application of non-linear multivariate methods to variables treated as ordinal scales. A case study application involving four-leg and three-leg ("T") non-signalized major arterial intersections in the Netherlands is described. The effect of bicycle traffic on accident rates is included in the case study analysis. The results indicate that there are three groups of each of the two types of intersections based on traffic flow patterns. For each group, a different functional form was found to relate accident rates and specific variables measuring traffic volumes. There were no significant differences among the physical design categories of the intersections in each group that were not accounted for by differences in traffic intensities.

1. OBJECTIVE AND SCOPE

The objective of the reported research is to develop and test a methodology for characterizing relationships between accident counts and traffic volume measures for major road intersections. The methodology allows tests to be conducted regarding whether the traffic-accident count relationships vary according to (categorical) variables such as the physical design of the intersection. The methodology is intended for use in situations where there are no data available on the specific traffic movements through an intersection, the only available data being traffic volume estimates for each leg of an intersection. Moreover, it is assumed that the only available data on accidents are the total count of accidents over a given period of time (possibly broken down by injury versus property damage) that occurred within each intersection. The method does not require strong assumptions about the scale properties of the data, and it is designed to accommodate general non-linear relationships.

A case study application of the methodology was conducted using data for approximately five hundred non-signalized intersections in the Netherlands. Because bicycle traffic is heavy at many Dutch intersections, and because bicycle safety is a major issue in traffic engineering in that country, traffic intensity variables were included for both motorized and bicycle modes. In the case study, separate analyses were conducted for three-leg and four-leg intersections. Finally, there were three dependent accident count variables: total accidents, accidents involving injury or death, and accidents involving bicycles and injury or death.

2. BACKGROUND

Safety comparisons among road sections are typically conducted in terms of accidents per vehicle mile of travel per time period. With regard to intersections, several authors have pointed out that accident risk is partially dependent on the numbers of vehicles whose paths cross one another, so that simple vehicle counts are inappropriate measures of exposure (Tanner, 1953; Grossman, 1954; Mathewson and Brenner, 1957; Breunning and Bone, 1959).

As reviewed in Chapman (1973) and Hakkert and Mahalel (1978), the search for appropriate exposure measures for intersections focused on variables representing the product of vehicle flows. The simplest of these variables is the product of the total traffic on the two intersecting roads, usually taken to the 0.5 power (Tanner, 1953). Variations of this variable can be formed by flow products with the traffic count on each road raised to a different power (McDonald, 1953), and comparisons of the simple form and its variations are provided by Leong (1973).

If, instead of total road volumes, traffic volumes on all approaches to intersections are available, more complicated expressions of traffic flow conflicts can be used as exposure measures. Some researchers have used the sum of the products of crossing flows (e.g., Breunning and Bone, 1959; and Surti, 1965, 1969). Finally, if additional traffic volume information is available on turning maneuvers, flow product sums can be estimated for up to 24 conflict points for four-leg intersections; and Hakkert and Mahalel (1978) computed an exposure measure based on the sum of all 24 products.

Regarding the functional form between accident counts and any of these exposure measures, it has been shown in many studies that a power function is effective in explaining accident counts (Jorgensen, 1969; Chapman, 1973;

Hakkert and Mahalel, 1978; and Ceder and Livneh, 1978). The parameters of such power functions are typically estimated using log-log regressions. Weighted least squares (e.g., Hakkert and Mahalel, 1978) and maximum likelihood techniques (e.g., Zalinger, et al., 1977) can be used to determine unbiased and minimum variance parameter estimates for Poisson distributions of accident counts (Erlander, et al., 1969).

The proposed method differs from previous approaches in three respects. First, the traffic intensity data (in terms of intersection approach volumes) are treated as approximations of true intensities. Second, no assumptions are made regarding the exact functional forms of the relationships between accident counts and traffic approach volumes. Third and finally, the method is aimed at determining the best measures of exposure from the set of all traffic approach volumes and the interactions among these volumes.

3. OUTLINE OF THE METHOD

The methodology is divided into four steps. In the first step, the interrelationships among the traffic volume variables are explored using non-linear principal components analysis. The goal of this step is to eliminate redundancies among the intensity variables by developing weighted averages which summarize the information contained in the redundant variables. Using a non-linear version of principal components analysis allows the variables measuring traffic volumes to be treated as ordinal scales. This is appropriate in view of the approximate nature of the data; results might show that slightly different values cannot be distinguished, and there is less possibility of undue influences of outlying observations. Most importantly, no assumption needs to be made regarding shapes of the curves best depicting relationships between accidents and traffic intensities. These curves might be linear, or they might be convex or concave (representing diminishing or increasing marginal effects), sigmoid (representing relative insensitivity at the extremes, but higher sensitivity in the mid-range), binomial functions (two distinct levels), or any other monotonic form. The specific method used in this step is discussed in more detail in the next section.

The goal of the second step is to find homogeneous groups of intersections on the basis of traffic intensity characteristics as summarized by the non-linear principal components developed in the first step. This is accomplished using the method of cluster analysis in which specific groupings are found that have minimum pooled within-groups variance. By using the scores of the intersections on the principal components, the clustering is performed on dimensions which are statistically independent so that redundant variables do not dominate the results. The mean values of the clusters on the original traffic volume variables can be computed and analysis-of-variance tests conducted to identify significant differences. This facilitates direct interpretations of the clusters in terms of a typology of intersections based on traffic intensity characteristics.

The third step of the methodology involves explorations of the relationships between the traffic intensity variables and the accident rate variables separately for each cluster of intersections. This is accomplished using non-linear canonical correlation analysis, which can capture non-linear relationships between multiple accident count variables and multiple traffic volume variables. Results can show which traffic volume variables, when scaled according to the most effective ordinal relationship, are most highly correlated with the different accident count variables. Comparisons of results among the clusters can lead to better understanding of how the accident and traffic intensity relationships vary according to the traffic characteristics typology represented by the clustering.

The fourth and final step is to estimate the functional forms of the relationships between accidents and the selected intensity variables for each cluster of intersections and to identify differences among intersection design types. This is approached by regressing accident count variables on those traffic volume measures found to be most strongly related to accident counts in the previous step of the analysis. In general, non-linear functional forms can be expected. The non-linear canonical correlation results guide in the selection of functional forms for testing. The residuals, calculated as the difference between the actual number of accidents and the number predicted by the fitted regression equations, are then analyzed with respect to variables that further distinguish the intersections. In the case study, simple F-tests are conducted to determine whether or not the mean differences between actual and predicted numbers of accidents are a function of the physical design, as measured by a nine-category variable.

4. THE CASE STUDY DATA

The case study application of the method involved a total of 257 three-leg and 174 four-leg intersections in the Netherlands. Accident counts for these non-signalized intersections were available for the years 1975 and 1976, and the traffic volumes were counted or estimated variously during the time period 1971 through 1976. The traffic volumes represented the total daily traffic passing through a given leg of an intersection.

The traffic volume data were standardized by specifying variable subscripts so that the same relationship among priority and non-priority legs applied for all intersections. In the case of three-leg intersections, the variables applying to the non-priority leg (the vertical leg of the "T") were always assigned the subscript 3. The variables applying to the priority leg clockwise from the non-priority were subscripted 1, and the remaining priority leg was designated as leg 2. In the case of four-leg intersections, the non-priority leg with the lowest car traffic intensity was designated as leg 4. The priority leg clockwise from leg 4 was designated as leg 1; the next leg clockwise from leg 1 (the busier non-priority leg) was designated as leg 2; and the remaining (priority) leg as leg 3. These schemes are depicted in Figure 1 (three-leg intersections) and Figure 2 (four-leg intersections), where C represents car traffic intensity and B represents bicycle traffic intensity.

In order to estimate potential traffic conflicts, a number of interaction variables were computed. For example, a variable designated as C_{13} represents the product of variables C_1 and C_3, the car traffic volumes on legs 1 and 3 of a three-leg intersection. If both C_1 and C_3 are high, C_{13} will take on a high value; if only one of the two variables is high while the other is low, C_{13} may, in general, take on a mid-range value; and if both C_1 and C_3 are low, C_{13} will take on a low value. Several variables represent interactions between car and bicycle traffic. These are in response to the need to develop better exposure measures for cyclists (Howarth, 1982). The complete set of computed interaction variables are listed in Table 1 (three-leg intersections) and Table 2 (four-leg intersections).

The sets of explanatory variables listed in Tables 1 and 2 encompass most types of exposure measures used in previous studies of intersection accidents related to traffic volumes. In terms of the present variables, a

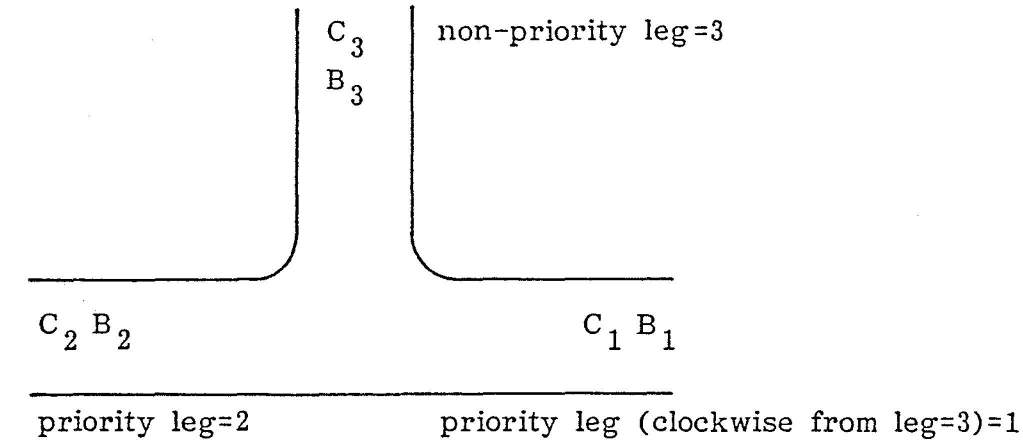

FIGURE 1

ORIENTATION OF VARIABLE SUBSCRIPTS: THREE-LEG INTERSECTIONS

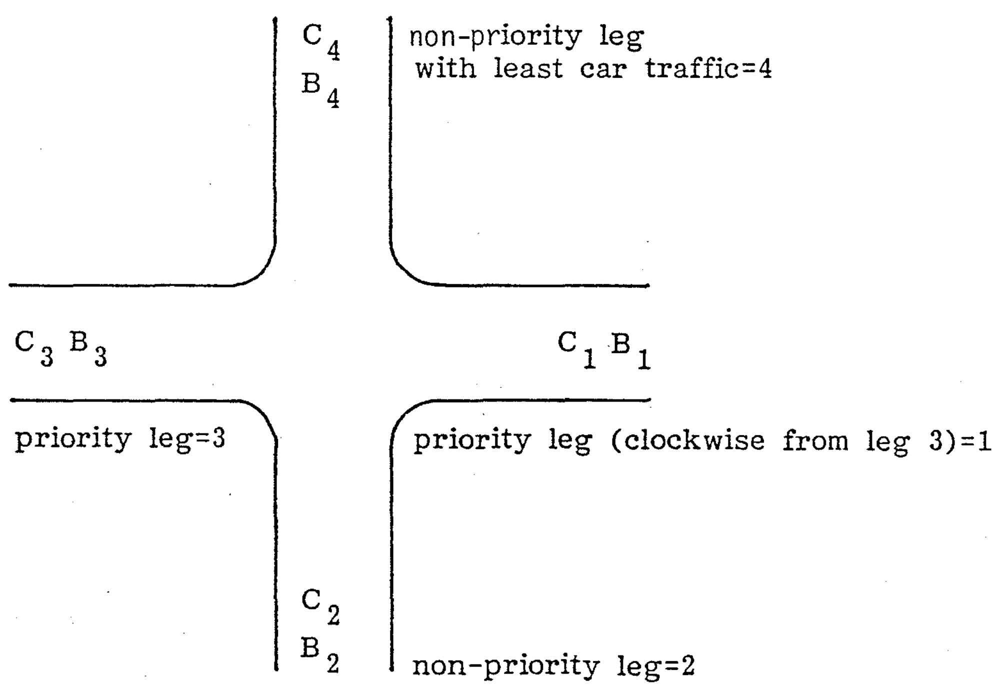

FIGURE 2

ORIENTATION OF VARIABLE SUBSCRIPTS: FOUR-LEG INTERSECTIONS

TABLE 1

VARIABLE DEFINITIONS FOR THREE-LEG INTERSECTIONS

VARIABLE	COMPUTATION	DESCRIPTION
C_1	-	Car intensity: priority leg clockwise from non-priority leg
C_2	-	Car intensity: priority leg counter-clockwise from non-priority leg
C_3	-	Car intensity: non-priority leg
C_{13}	C_1*C_3	Car interaction: C_1 by C_3
C_{23}	C_2*C_3	Car interaction: C_2 by C_3
C_{12+3}	$C_1*(C_2+C_3)$	Car interaction: C_1 by (C_2+C_3)
B_1	-	Bike intensity: priority leg clockwise from non-priority leg
B_2	-	Bike intensity: priority leg counter-clockwise from non-priority leg
B_3	-	Bike intensity: non-priority leg
B_{13}	B_1*B_3	Bike interaction: B_1 by B_3
B_{23}	B_2*B_3	Bike interaction: B_2 by B_3
B_{12+3}	$B_1*(B_2+B_3)$	Bike interaction: B_1 by (C_2+C_3)
$B_{23}C_T$	$B_2*B_3*(C_1+C_2+C_3)$	Bike-car interaction: B_{23} by $(C_1+C_2+C_3)$
$B_{13}C_T$	$B_1*B_3*(C_1+C_2+C_3)$	Bike-car interaction: B_{13} by $(C_1+C_2+C_3)$
B_1C_{13}	$B_1*C_1*C_3$	Bike-car interaction: B_1 by C_{13}
B_2C_{23}	$B_2*C_2*C_3$	Bike-car interaction: B_2 by C_{23}
$B_{12}C_2$	$B_1*B_2*C_2$	Bike-car interaction: B_{12} by C_2
ACC_T	-	Total accidents
ACC_I	-	Injury accidents
ACC_B	-	Bicycle-injury accidents
DESIGN	-	Physical design-categories described in Table 3

TABLE 2

VARIABLE DEFINITIONS FOR FOUR-LEG INTERSECTIONS

VARIABLE	COMPUTATION	DESCRIPTION
C_1	-	Car intensity: priority leg clockwise from non-priority leg with lowest car intensity
C_2	-	Car intensity: non-priority leg with highest car intensity
C_3	-	Car intensity: priority leg clockwise from non-priority leg with highest car intensity
C_4	-	Car intensity: non-priority leg
C_{12}	C_1*C_2	Car interaction: C_1 by C_2
C_{23}	C_2*C_3	Car interaction: C_2 by C_3
C_{34}	C_3*C_4	Car interaction: C_3 by C_4
C_{41}	C_4*C_1	Car interaction: C_4 by C_1
B_1	-	Bike intensity: priority leg clockwise from non-priority leg with lowest car intensity
B_2	-	Bike intensity: non-priority leg with highest car intensity
B_3	-	Bike intensity: priority leg clockwise fron non-priority leg with highest car intensity
B_4	-	Bike intensity: non-priority leg
B_{12}	B_1*B_2	Bike interaction: C_1 by C_2
B_{23}	B_2*B_3	Bike interaction: C_2 by C_3
B_{34}	B_3*B_4	Bike interaction: C_3 by C_4
B_{41}	B_4*B_1	Bike interaction: C_4 by C_1
$B_{12}C_T$	$B_1*B_2*(C_1+C_2+C_3+C_4)$	Bike-car interaction: B_{12} by $(C_1+C_2+C_3+C_4)$
$B_{23}C_T$	$B_2*B_3*(C_1+C_2+C_3+C_4)$	Bike-car interaction: B_{23} by $(C_1+C_2+C_3+C_4)$
$B_{34}C_T$	$B_3*B_4*(C_1+C_2+C_3+C_4)$	Bike-car interaction: B_{34} by $(C_1+C_2+C_3+C_4)$
$B_{41}C_T$	$B_4*B_1*(C_1+C_2+C_3+C_4)$	Bike-car interaction: B_{41} by $(C_1+C_2+C_3+C_4)$
$B_{24+}C_{1+3}$	$B_2*B_4*(C_1+C_3)$	Bike interaction: B_{24} by (C_1+C_3)
ACC_T	-	Total accidents
ACC_I	-	Injury accidents
ACC_B	-	Bicycle-injury accidents
DESIGN	-	Physical design-categories described in Table 3

power function explaining accident counts in terms of exposure would be expected to take one of two forms:

$$A = c\, X_i^a\, X_j^b \qquad (1)$$

or

$$A = c\, X_{ij}^a \qquad (2)$$

where A = accident count, X_i and X_j represent single-subscripted variables in Tables 1 or 2; X_{ij} represents a double-subscripted or compound variable; and a, b, and c are parameters to be estimated. Form (1) is of the product flow form with exponents for each intersecting flow (e.g., McDonald, 1953; Leong, 1973), and form (2) is the (root) product flow form with a single exponent (e.g., Tanner, 1953; Bennett, 1966; Leong, 1973). The choice between the forms in the present approach is generalized to select the most effective explanatory variable, and this is accomplished using the ordinal multivariate methods outlined in Section 3.

In addition to the traffic volume and accident count variables, there is a categorical variable (called DESIGN) which records the major design characteristics of the intersection. The categories are similar for the three-leg and four-leg intersections and are specified in Table 3. The methodology is designed so as to determine how a variable such as physical design category is interrelated with traffic volumes and accident counts.

5. INTERRELATIONSHIPS AMONG TRAFFIC VOLUME VARIABLES

The first step in developing a typology of intersections based on traffic intensity characteristics is to analyze the interrelationships among the intensity variables. When all variables are measured numerically, such

TABLE 3

CATEGORIES OF THE DESIGN VARIABLE FOR BOTH THREE- AND FOUR-LEG INTERSECTIONS

Category	Description		Frequencies	
			3-leg inter-sections	4-leg inter-sections
1.1	Lane separation (non-priority leg(s)): Lane separation (priority legs): Left-turn lane: Number of through-lanes:	yes (any type) barrier yes 1	26	38
1.2	Lane separation (non-priority leg(s)): Lane separation (priority legs): Left-turn lane: Number of through-lanes:	yes (any type) barrier yes 2	26	25
2	Lane separation (non-priority leg(s)): Lane separation (priority legs): Left-turn lane:	yes (any type) painted yes	53	30
3	Lane separation (non-priority leg(s)): Lane separation (priority legs): Left-turn lane:	yes (any type) barrier no	6	4
4	Lane separation (non-priority leg(s)): Lane separation (priority legs): Left-turn lane:	yes (any type) none no	82	40
5	Lane separation (non-priority leg(s)): Lane separation (priority legs): Left-turn lane:	none barrier yes	5	7
6	Lane separation (non-priority leg(s)): Lane separation (priority legs): Left-turn lane:	none painted yes	2	5
7	Lane separation (non-priority leg(s)): Lane separation (priority legs): Left-turn lane:	none barrier no	2	2
8	Lane separation (non-priority leg(s)): Lane separation (priority legs): Left-turn lane:	none none no	55	23

interrelationships are typically studied using principal components analysis or other techniques (e.g., factor analysis, confirmatory factor analysis, or canonical factor analysis) that are based on eigenvalue extractions of the variable variance-covariance or correlation matrices. In the present situation, the traffic intensity variables are ordinally scaled so conventional principal components analysis is inappropriate. However, a non-linear version of principal components analysis has been developed and is one of a series of computer programs developed by the Department of Data Theory of the University of Leiden (Gifi, 1981). This program, called PRINCALS, was applied to both the three-leg intersections and four-leg intersections data sets in the present study. The use of the program is outlined in Gifi (1983).

The objective of the PRINCALS algorithm is similar to that of conventional principal components analysis: namely, to reduce the rank of a correlation matrix by determining the linear combinations (weighted averages) of the original variables that account for the maximum variance under the constraint that these linear combinations are mutually orthogonal. In conventional principal components analysis, these linear combinations (components) are determined through a closed-form solution involving the latent roots and vectors (eigenvalues and eigenvectors) of the correlation matrix. In non-linear principal components analysis, an iterative solution is required: the correlation matrix itself must be determined by finding the scores for the categories of the ordinal or nominal variables in such a way that the latent roots of the matrix are maximized. In the PRINCALS program, this is accomplished using the technique of alternating least squares (Young, et al., 1976).

For three-leg intersections, the overall fits of PRINCALS solutions with from one to four dimensions are shown in Table 4. The three-dimensional solution is preferred on the basis of the latent roots. This solution accounts for almost 92 percent of the variance among the seventeen optimally-scaled traffic intensity variables, and the addition of a fourth component adds only about 4 percent to the explained sum, which is less than the expected random contribution of one variable.

Table 5 shows the correlations between each optimally scaled traffic intensity variable and each of the three principal components. The last column in Table 5 lists indices of overall fit for each variable, calculated as the square root of the sums of squared correlations between the variable and the three components. Such an index can range from zero to one and can be interpreted as the proportion of variance in the optimally scaled variable which is explained cumulatively by the principal components. The lowest levels of fit are for the three simple car traffic intensity variables, but

TABLE 4

NON-LINEAR PRINCIPAL COMPONENTS ANALYSIS (PRINCALS) SOLUTIONS FOR THREE-LEG INTERSECTIONS

Number of dimensions	Eigenvalues 1	2	3	4	Cumulative percentage variance accounted for
1	0.54	-	-	-	54.2
2	0.44	0.36	-	-	80.0
3	0.43	0.35	0.13	-	91.9
4	0.42	0.36	0.13	0.03	95.2

TABLE 5

CORRELATIONS AMONG THE PRINCALS COMPONENTS AND THE TRAFFIC INTENSITY VARIABLES FOR THREE-LEG INTERSECTIONS
(ONLY CORRELATIONS SHOWN WITH ABSOLUTE VALUE GREATER THAN 0.30)

Variable	Correlations with components 1	2	3	Proportion of variance accounted for
C_1	0.78	0.35		0.79
C_2	0.79			0.77
C_3	0.77		-0.36	0.76
C_{13}	0.87	0.31	-0.34	0.97
C_{23}	0.88		-0.34	0.97
C_{123}	0.88	0.31	-0.34	0.99
B_1		0.68	0.63	0.87
B_2		0.62	0.68	0.89
B_3	-0.61	0.73		0.98
B_{13}	-0.61	0.75		0.98
B_{23}	-0.60	0.75		0.97
B_{123}	-0.61	0.75		0.99
$B_{23}C_T$	-0.60	0.75		0.97
$B_{13}C_T$	-0.61	0.75		0.98
B_1C_{13}	0.67	0.63		0.88
B_2C_{23}	0.67	0.62		0.90
$B_{12}C_2$	0.35	0.69	0.59	0.96

even these three indices are above 0.75. Thus, the seventeen variables can be collapsed to three statistically-independent weighted averages without loss of much information.

The correlation coefficients shown in Table 5 can be used to interpret the three principal components. (To aid in interpretation, only correlations with absolute value greater than 0.30 are shown). The first component reflects high levels of car traffic on all legs of an intersection, high levels of interactions B_1C_{13} and B_2C_{23} between bicycle and car traffic, but relatively low levels of bicycle traffic on the non-priority leg. The second component primarily reflects high bicycle traffic on all legs, and high bicycle-car interaction terms. Finally, the third component reflects high bicycle traffic on the priority legs only, and relatively low car traffic on the non-priority leg (and hence also low levels of the interaction terms involving C_3).

For four-leg intersections, the latent roots for PRINCALS solutions of four dimensionalities are shown in Table 6. The overall fits for these solutions are lower than for the corresponding solutions for the three-leg intersections. The four-dimension solution was chosen because the fourth latent root was greater than the expected contribution of a single random variable (1 : 21 = 0.04) and because the three-dimension solution had low fits for some variables. As shown in Table 7, the total fits (proportions of variance accounted for) in the four-dimension solution are all 0.86 or greater, except for C_2, car traffic intensity on the busier of the two non-priority legs (0.73) and B_1, bicycle traffic intensity on one priority leg

TABLE 6

NON-LINEAR PRINCIPAL COMPONENTS ANALYSIS (PRINCALS) SOLUTIONS FOR FOUR-LEG INTERSECTIONS

Number of dimensions	Eigenvalues				Cumulative percentage variance accounted for
	1	2	3	4	
1	0.56	-	-	-	56.3
2	0.54	0.20	-	-	74.3
3	0.53	0.21	0.10	-	83.1
4	0.51	0.22	0.10	0.05	88.3

TABLE 7

CORRELATIONS AMONG THE PRINCALS COMPONENTS AND THE TRAFFIC INTENSITY VARIABLES FOR FOUR-LEG INTERSECTIONS
(ONLY CORRELATIONS SHOWN WITH ABSOLUTE VALUE GREATER THAN 0.30)

Variable	Correlations with components				Proportion of variance accounted for
	1	2	3	4	
C_1	0.54	0.63		-0.42	0.93
C_2	0.48	0.60		0.35	0.73
C_3	0.58	0.63		-0.34	0.89
C_4	0.30	0.49	0.37	0.66	0.89
C_{12}	0.61	0.73			0.92
C_{23}	0.62	0.73			0.92
C_{34}	0.57	0.75			0.90
C_{41}	0.55	0.76			0.89
B_1	0.63		-0.57		0.80
B_2	0.68		0.57		0.91
B_3	0.60		-0.60		0.87
B_4	0.64		0.64		0.86
B_{12}	0.84	-0.42			0.89
B_{23}	0.84	-0.43			0.89
B_{34}	0.83	-0.36			0.87
B_{41}	0.86	-0.35			0.87
$B_{12}C_T$	0.88				0.89
$B_{23}C_T$	0.92				0.89
$B_{34}C_T$	0.89				0.88
$B_{41}C_T$	0.91				0.88
$B_{24}C_{1+3}$	0.75		0.57		0.96

(0.80). Thus, while the traffic intensity variables for four-leg intersections are more heterogeneous than the variables for three-leg intersections, a significant reduction in dimensionality (from twenty-one to four) is still possible.

Interpretations of the four principal components can be made using the correlation coefficients listed in Table 7: the first component reflects high overall levels of traffic intensity on all legs, but it is particularly associated with high bicycle-car and bicycle-bicycle intensity interactions. The second component is associated with high car traffic intensity on all legs and relatively low bicycle intensity interaction terms. The third component captures a pattern of high bicycle traffic intensity on the non-priority legs but low bicycle traffic intensity on the priority legs. The fourth component similarly captures a pattern of high and low car traffic intensities on the non-priority and priority legs, respectively.

6. TRAFFIC-VOLUME TYPOLOGIES OF INTERSECTIONS

The second step in developing a typology of intersections based on traffic intensity characteristics is to determine homogeneous groups of intersections based on the scores of the intersections on the principal components. The particular cluster analysis method used in the present study is non-hierarchical with an objective function minimized via a hill-climbing algorithm. This objective function involves the determinant of the pooled within-cluster cross-products matrix, a quantity known as the generalized variance. Minimization of this quantity is equivalent to minimization of the ratio of the determinants of the within-cluster and total variances, a likelihood ratio statistic known as Wilks' lambda (Everitt, 1980, p.66; Tatsuoka, 1971, p. 85). Restarts of the algorithm from different initial configurations are used to detect the possibility of a local optimum.

The cluster analysis results for the three-leg intersections are listed in Table 8. The objective function, converted in terms of the usual log-likelihood ratio (multiplied by minus two) is seen to improve significantly when the number of clusters is increased from two to three (Figure 3). There is another significant improvement when the number of

TABLE 8

CLUSTER ANALYSIS RESULTS--THREE-LEG INTERSECTIONS

Number of clusters	W *	$\lambda = \frac{W}{T}$	$-2 \ln \lambda$
1	1.6975×10^{7}	1.0	0.00
2	1.9076×10^{6}	1.2238×10^{-1}	4.20
<u>3</u>	<u>5.1153×10^{4}</u>	<u>3.0134×10^{-3}</u>	<u>11.61</u>
4	2.3471×10^{4}	1.3827×10^{-3}	13.17
5	1.2948×10^{4}	7.6277×10^{-4}	14.36
6	1.6255×10^{3}	9.5758×10^{-5}	18.51
7	9.0112×10^{2}	5.3085×10^{-5}	19.69
8	6.1889×10^{2}	3.6459×10^{-5}	20.44
9	5.3280×10^{2}	3.1387×10^{-5}	20.74
10	3.0641×10^{2}	1.8051×10^{-5}	21.84

* W = determinant of pooled within-cluster cross-products matrices;
T = determinant of total cross-products matrix = W at one cluster.

clusters is increased from five to six, but the three-cluster solution is preferred in view of sample size considerations for further analyses. The 257 three-leg intersections are divided among the three clusters in the proportions: (1) 72 or 28.0%, (2) 162 or 63.0%, and (3) 23 or 8.9%.

The three clusters can be interpreted by inspecting the means for each cluster on the three principal components and by inspecting the means on the original seventeen variables. The latter statistics are listed in Table 9, together with the cluster means for the three accident rate variables. Only the three accident rates do not have cluster means which are significantly

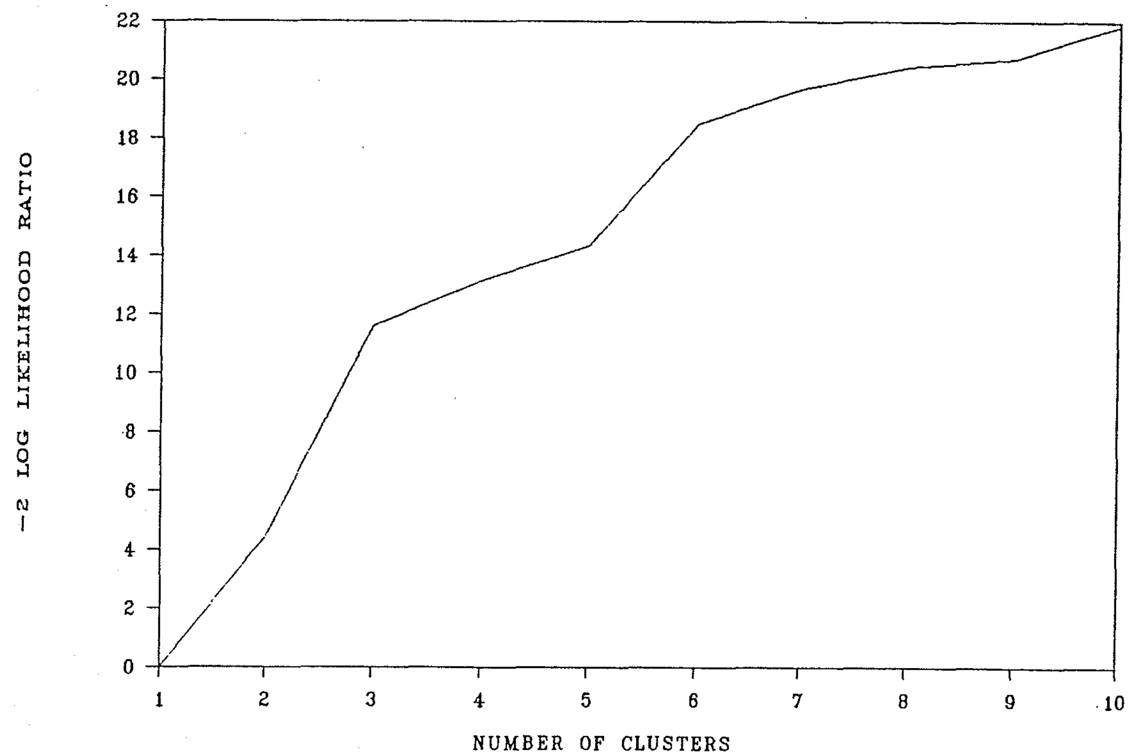

FIGURE 3

PLOT OF CLUSTERING CRITERION FOR THREE-LEG INTERSECTIONS

different from one another as determined in a one-way analysis-of-variance. This result indicates that the patterns in the interrelationships among the intensity variables cannot be used to predict accident rates; further analysis steps are indeed necessary.

TABLE 9

CLUSTER MEANS FOR THE TRAFFIC INTENSITY AND ACCIDENT RATE VARIABLES FOR THREE-LEG INTERSECTIONS

Variable	Grand mean	Cluster mean 1 (N = 72)	2 (N = 162)	3 (N = 23)
C_1 (x 10^{-3})	6.41	7.03	5.79	8.87
C_2 (x 10^{-3})	6.15	7.26	5.27	8.87
C_3 (x 10^{-3})	3.33	3.93	2.83	4.96
C_{13} (x 10^{-7})	2.76	3.27	2.00	6.56
C_{23} (x 10^{-7})	2.76	3.81	1.82	6.10
C_{123} (x 10^{-7})	5.52	7.07	3.82	12.66
B_1 (x 10^{-2})	6.28	8.49	6.16	0.17
B_2 (x 10^{-2})	5.93	8.46	5.47	1.30
B_3 (x 10^{-2})	2.49	0.00	3.93	0.22
B_{13} (x 10^{-5})	1.90	0.00	3.01	0.00
B_{23} (x 10^{-5})	2.08	0.00	3.31	0.00
B_{123} (x 10^{-5})	3.98	0.00	6.32	0.00
$B_{23}C_T$ (x 10^{-10})	4.05	0.00	6.43	0.00
$B_{13}C_T$ (x 10^{-10})	3.34	0.00	5.30	0.00
B_1C_{13} (x 10^{-11})	2.02	3.41	1.61	0.52
B_2C_{23} (x 10^{-11})	2.07	3.92	1.44	0.78
$B_{12}C_2$ (x 10^{-10})	7.09	15.96	4.15	0.00
ACC_T	2.89	3.15*	2.78*	2.87*
ACC_I	0.67	0.58*	0.74*	0.48*
ACC_B	0.25	0.24*	0.28*	0.00*

* Indicates means <u>not</u> significantly different from one another at the p = .05 confidence level as determined by F-tests.

The first cluster is distinguished by zero levels of bicycle intensity on the non-priority leg of the intersections, and consequently by zero values for the interaction terms involving that intensity variable. In contrast, the intersections of the first cluster have high levels of bicycle traffic on the priority legs, and consequently high levels on the last three bicycle-car interaction variables. The second cluster is distinguished by relatively low car traffic intensity on the non-priority leg but relatively high bicycle intensity on that same leg; car and bicycle intensities on the priority legs are approximately average. Finally, cluster-3 intersections exhibit the highest average car intensities on all three legs but the lowest bicycle intensities on the priority legs. The clustering in general appears to capture the phenomenon of negatively correlated car and bicycle traffic intensities.

The cluster analysis results for four-leg intersections are listed in Table 10, and the log-likelihood ratio statistic is plotted in Figure 4. In contrast to three-leg intersections, there are no apparent natural numbers of clusters of four-leg intersections. Figure 4 shows that the five-cluster solution is superior to either the four- or six-cluster solutions, but the three- and seven-cluster solutions are also candidates for selection. The three-cluster solution is preferred on the basis of sample size criteria: it is desirable to have large clusters for further analysis. The sample sizes of the three clusters are: (1) 93 or 53.4%, (2) 56 or 32.2%, and (3) 25 or 14.4%.

Table 11 gives the cluster means for the twenty-one traffic intensity and three accident rate variables. The hypothesis of equal means across the three clusters is accepted at the p = .05 confidence interval for only two of the intensity variables (both involving bicycle traffic) but for two of the three accident rate variables. The first cluster is characterized by the

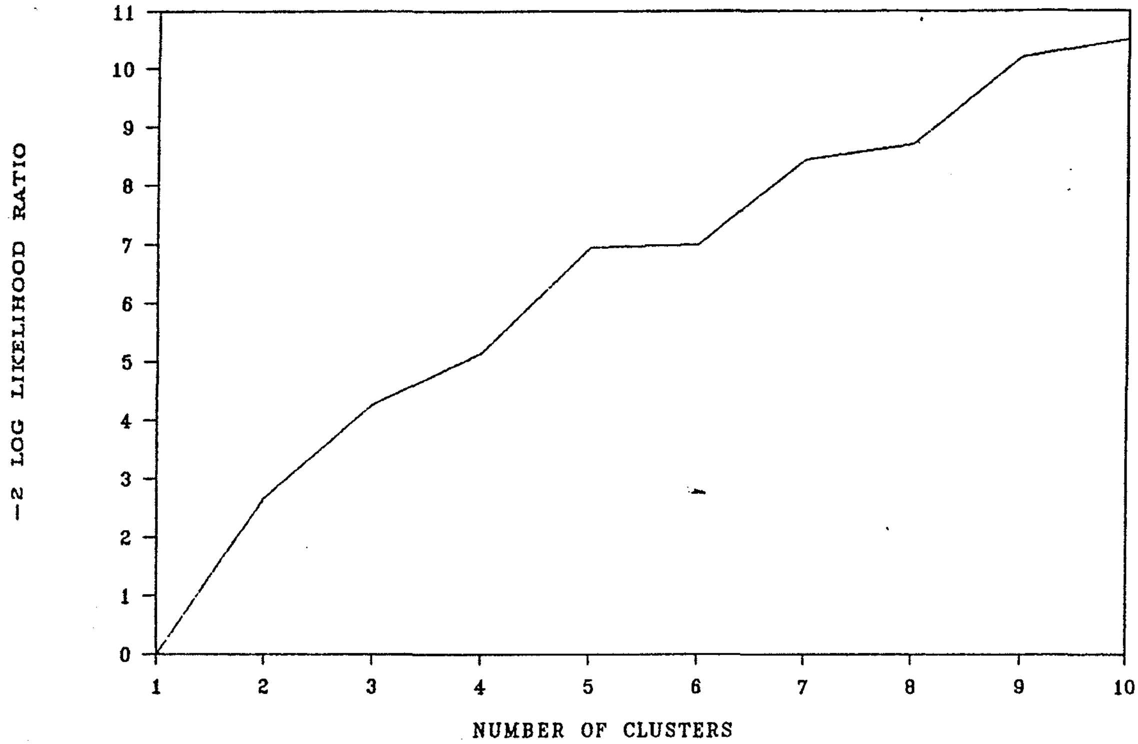

FIGURE 4

PLOT OF CLUSTERING CRITERION FOR FOUR-LEG INTERSECTIONS

lowest levels on almost every one of the variables, particularly on the bicycle-car interaction variables. The second cluster is distinguished by high car and bicycle traffic intensity on the two priority legs of the

TABLE 10

CLUSTER ANALYSIS RESULTS--FOUR-LEG INTERSECTIONS

Number of clusters	W *	$\lambda = \frac{W}{T}$	$-2 \ln \lambda$
1	9.1664×10^{8}	1.0	0.00
2	2.4301×10^{8}	2.6511×10^{-1}	2.66
3	1.0907×10^{8}	1.1899×10^{-1}	4.26
4	6.9916×10^{7}	7.6274×10^{-2}	5.15
5	2.8311×10^{7}	3.0886×10^{-2}	6.95
6	2.7552×10^{7}	3.0058×10^{-2}	7.01
7	1.3433×10^{7}	1.4655×10^{-2}	8.45
8	1.1801×10^{7}	1.2874×10^{-2}	8.71
9	5.8164×10^{6}	6.3453×10^{-3}	10.12
10	4.7913×10^{6}	5.2270×10^{-3}	10.51

* W = determinant of pooled within-cluster cross-products of pooled within-cluster cross-products matrices;
T = determinant of total cross-products matrix = W at one cluster.

intersections (denoted by the subscripts 1 and 3), but relatively low car and bicycle traffic intensity on the non-priority legs (subscripts 2 and 4). Finally, cluster three is distinguished by relatively equal levels of car traffic on the priority and non-priority legs and very high bicycle traffic on the non-priority legs. Clusters two and three have the highest accident rates.

TABLE 11

CLUSTER MEANS FOR THE TRAFFIC INTENSITY AND ACCIDENT RATE VARIABLES FOR FOUR-LEG INTERSECTIONS

		Cluster mean		
Variable	Grand mean	1 (N - 93)	2 (N - 56)	3 (N - 25)
C_1 (x 10^{-3})	6.16	3.99	10.41	4.68
C_2 (x 10^{-3})	3.30	2.59	3.41	5.72
C_3 (x 10^{-3})	6.10	4.02	10.02	5.08
C_4 (x 10^{-3})	1.86	1.47	1.43	4.28
C_{12} (x 10^{-7})	2.44	1.25	4.08	3.22
C_{23} (x 10^{-7})	2.45	1.21	3.91	3.83
C_{34} (x 10^{-7})	1.24	0.68	1.57	2.61
C_{41} (x 10^{-7})	1.20	0.68	1.56	2.31
B_1 (x 10^{-2})	4.41	2.42	8.06	3.62
B_2 (x 10^{-2})	3.81	3.04	2.88	8.74
B_3 (x 10^{-2})	4.52	2.66	7.67	4.36
B_4 (x 10^{-2})	3.19	2.64	1.96	8.00
B_{12} (x 10^{-5})	2.17	0.96	3.51	3.28
B_{23} (x 10^{-5})	2.15	1.10	3.00	4.15
B_{34} (x 10^{-5})	1.91	0.83*	2.95*	3.65*
B_{41} (x 10^{-5})	1.81	0.73*	2.97*	3.24*
$B_{12}C_T$ (x 10^{-10})	4.92	1.52	9.32	7.72
$B_{23}C_T$ (x 10^{-10})	4.47	1.10	8.14	8.80
$B_{34}C_T$ (x 10^{-10})	4.81	1.73	7.94	9.24
$B_{41}C_T$ (x 10^{-10})	4.38	1.23	7.67	8.71
$B_{24}C_{1+3}$ (x 10^{-11})	3.70	1.47	2.97	13.62
ACC_T	6.32	4.77	7.91	8.48
ACC_I	1.86	1.54*	1.96*	2.80*
ACC_B	0.46	0.39*	0.63*	0.36*

* Indicates means <u>not</u> significantly different from one another at the p = .05 confidence level as determined by F-tests.

7. TRAFFIC VOLUMES AND ACCIDENT RATES

Relationships between the traffic intensity variables and the accident rate variables were investigated separately for each cluster of intersections. These investigations were performed using non-linear canonical correlation analysis, as implemented in the CANALS computer program developed by the Department of Data Theory of the University of Leiden (Gifi, 1981; Van der Burg, 1983).

Canonical correlation analysis is a generalization of regression analysis to more than one dependent variable. The objective in canonical correlation analysis is to find a weighted average (linear combination) of the independent variables (in this case, the traffic intensity measures) which is maximally correlated with a weighted average of the dependent variables (in this case, the accident rates). The correlation between the independent and dependent variable sets, so maximized, is called a canonical correlation. If appropriate, a second set of linear combinations can be determined that is maximally correlated under the constraint that this second dimension is statistically independent of the first dimension. The possible number of dimensions so determined is equal to the number of variables in the smaller of the two variable sets (here, the three accident rate variables). In conventional canonical correlation analysis all variables must have interval-scale properties or they must be dichotomous.

Non-linear versions of canonical correlation analysis are designed for use with mixed sets of categorical and ordinal variables, as well as numerical variables. In the present application, all traffic intensity variables are treated as ordinal and all accident rate variables are treated as numerical so that all non-linearities are applied to the intensity variables. For each ordinal variable, scores are determined for each variable category that maximize the canonical correlation when the variables are linearly combined. This optimization problem involving both category scores and variable weights requires an iterative solution, and in the CANALS program this is accomplished by means of the principle of alternating least squares (Young, et al., 1976; Van der Burg and De Leeuw, 1983). The same solution method is used in the non-linear principal components analysis algorithm (PRINCALS) used in the first part of the methodology.

A one- or two-dimensional CANALS solution was computed for each cluster of intersections for both three-leg and four-leg intersections. A two-dimensional solution was selected whenever the canonical correlation of the second dimension was close to that of the first dimension; otherwise, a one-dimensional solution was selected (no third-dimension correlations were significantly greater than zero). For each CANALS solution, correlations are computed between each variable and the canonical dimension or dimensions. These correlations indicate how important each variable is in explaining the overall relationship between the traffic intensity measures and the accident rates. (The variable coefficients, or weights in the linear combinations which form the canonical dimensions, are not as informative as the correlations because there are known high colinearities among the variables which affect the coefficients but not the correlations.)

Table 12 shows the results for the three clusters of three-leg intersections. Two-dimensional solutions were found for the first and third clusters, while a one-dimensional solution was found for the second cluster. For the first cluster (of 72 intersections), the first dimension links total and injury accidents mainly to three different car traffic intensity variables: C_{13}, C_{23} and C_2. The second dimension of the solution for the

TABLE 12

NON-LINEAR CANONICAL CORRELATION RESULTS
FOR CLUSTERS OF THREE-LEG INTERSECTIONS

Variable	Correlations with Canonical Variates				
	Cluster 1 (n = 72)		Cluster 2 (n = 162)	Cluster 3 (n = 23)	
	dimensions 1	dimensions 2		dimensions 1	dimensions 2
Explanatory set					
C_1	0.44	-0.16	0.71	0.67	0.26
C_2	0.58	0.11	0.56	0.37	0.36
C_3	0.41	-0.10	0.48	0.38	0.31
C_{13}	0.60	0.02	0.67	0.33	0.44
C_{23}	0.60	0.06	0.67	0.28	0.27
C_{12+3}	0.51	-0.13	0.64	0.50	0.29
B_1	0.25	0.32	0.22	-	-
B_2	1.10	0.34	0.29	-	-
B_3	-	-	0.07	-	-
B_{13}	-	-	0.24	-	-
B_{23}	-	-	0.25	-	-
B_{12+3}	-	-	0.20	-	-
$B_{23}C_T$	-	-	0.43	-	-
$B_{13}C_T$	-	-	0.36	-	-
B_1C_{13}	0.52	0.13	0.64	-	-
B_2C_{23}	0.45	0.35	0.44	-	-
$B_{12}C_2$	0.40	0.39	0.40	-	-
Canonical correlation between variable sets	0.84	0.72	0.88	0.88	0.82
Dependent set					
ACC_T	0.92	0.18	0.97	0.66	0.76
ACC_I	0.91	0.15	0.77	0.99	0.02
ACC_B	0.48	0.66	0.33	-	-

first cluster links the bicycle traffic intensity variables B_1 and B_2 and the bicycle-car interaction variables $B_{12}C_2$ and B_2C_{23} with bicycle injury accident rates.

The one-dimensional solution found for the second cluster of 162 three-leg intersections explains total accident rates well, has moderate explanatory power for injury accidents, and very little explanatory power for bicycle-injury accidents. No additional dimensions were found which could explain bicycle-injury accidents for this cluster of intersections. As expected by the poor explanation of bicycle accidents, none of the bicycle traffic intensity variables are highly correlated with the dimension, but four car intensity variables and one car-bicycle interaction variable have correlations of 0.64 or greater. Of the four car intensity variables, three are interaction terms between intensities on the different legs of the intersections; these can capture the degrees to which various turning maneuvers occur.

The third and final cluster of 23 three-leg intersections had no bicycle-injury accidents, and consequently the analysis was limited to the relationships between the car traffic intensity variables and total accident rates and injury accident rates. A two-dimensional solution was found in which the first dimension explained both injury accidents and total accidents and the second dimension explained only total accidents. The explanatory variables for the first dimension are C_1, car traffic intensity on one priority leg of the intersection and C_{12+3}, an interaction term which potentially captures the scale of left turns from the priority road onto the non-priority road. The strongest independent variable for the second dimension (explaining total accidents) is C_{13}, an interaction term which potentially captures the scale of left turns from the non-priority road onto the priority road.

The CANALS solutions for the three clusters of four-leg intersections are documented in Table 13. All of these solutions are two-dimensional. For the first cluster of 93 intersections, the first dimension explains bicycle-injury accidents, and the independent variables with the highest correlations are B_{23} and B_{12}, interaction terms potentially capturing bicycle turning maneuvers, and $B_{12}C_T$, the interaction between B_{12} and total car traffic intensity. The second dimension explains total accidents and also injury accidents to a modest degree. The strongest independent variables are all car intensity variables and, particularly, car intensity interaction variables.

The first dimension for the second cluster of only 23 intersections has a moderately strong positive correlation with total accidents, a lower positive correlation with injury accidents, but a negative correlation with bicycle-injury accidents. This indicates that intersections of this type that have relatively high rates of total accidents tend to have relatively low rates of bicycle-injury accidents. The strongest independent variables for this dimension are all car intensity variables and car intensity interaction variables. The second dimension explains primarily bicycle-injury accidents, but also total and injury-accidents as well, and many of the independent variables have similarly high correlations. This dimension captures a very general relationship between overall traffic intensity and rates of all types of accidents.

For the final cluster of 25 four-leg intersections, the first dimension explains bicycle-injury accidents. However, the strongest independent variables are car traffic intensities. The second dimension explains both

TABLE 13

NON-LINEAR CANONICAL CORRELATION RESULTS
FOR CLUSTERS OF FOUR-LEG INTERSECTIONS

	Correlations with Canonical Variates					
	Cluster 1 (n = 93)		Cluster 2 (n = 23)		Cluster 3 (n = 25)	
	dimensions		dimensions		dimensions	
Variable	1	2	1	2	1	2
Explanatory set						
C_1	0.12	0.51	0.24	0.39	0.16	0.41
C_2	-0.18	0.06	0.32	0.14	-	-
C_3	0.23	0.54	0.25	0.39	0.09	0.21
C_4	0.06	0.43	-0.13	0.12	0.46	-0.03
C_{12}	0.16	0.40	0.25	0.33	-	-
C_{23}	0.20	0.46	0.26	0.22	-	-
C_{34}	0.14	0.60	0.15	0.22	0.44	0.27
C_{41}	0.11	0.59	0.11	0.29	0.43	0.16
B_1	0.41	-0.21	-0.27	0.26	-	-
B_2	0.24	0.35	0.11	0.29	0.12	-0.06
B_3	0.32	-0.12	-0.08	0.15	-0.24	-0.23
B_4	0.18	0.23	0.07	0.29	-	-
B_{12}	0.44	0.14	-0.06	0.39	-	-
B_{23}	0.46	0.23	-0.00	0.33	-0.25	-0.09
B_{34}	0.34	0.15	0.05	0.28	-	-
B_{41}	0.38	0.21	0.04	0.39	-	-
$B_{12}C_T$	0.43	0.26	-0.05	0.42	-	-
$B_{23}C_T$	0.32	0.18	0.05	0.40	-0.29	-0.07
$B_{34}C_T$	0.41	0.30	-0.06	0.43	-0.20	-0.06
$B_{41}C_T$	0.40	0.16	0.04	0.31	-	-
$B_{24}C_{1+3}$	0.21	0.36	-0.00	0.38	0.25	-0.18
Canonical correlation between variable sets	0.93	0.87	0.98	0.93	0.99	0.99
Dependent set						
ACC_T	0.14	0.96	0.68	0.74	-0.13	0.85
ACC_I	-0.10	0.56	0.31	0.67	-0.05	0.99
ACC_B	0.82	-0.02	-0.35	0.93	0.74	0.58

injury and total accidents, and the strongest independent variable is car intensity on one of the priority legs, C_1.

8. FUNCTIONAL FORMS OF THE ACCIDENT-EXPOSURE RELATIONSHIPS

In the final step of the analysis, accident rates are regressed against traffic intensities. The objective is to estimate the parameters of the functions which best describe the relationships between traffic intensities and accident rate. The results of the analyses in the previous step indicate that such functions are non-linear for many of the variables.

Separate regresssions were estimated for each cluster of intersections. The CANALS results were used to select which measures of traffic intensity should be regressed against which accident rate variables. Also, the category quantifications provided in the CANALS outputs guide the choice of an appropriate functional form.

The results of the regressions of accident rates on selected traffic intensity variables are listed in Table 14 for the three clusters of three-leg intersections. The dependent variable chosen for demonstration purposes is total accidents in each case, and the variable C_{13} was found to be the most effective explanatory variable for all three clusters. This indicates that, for "T" intersections with a through priority road (Figure 1), an interaction term (calculated as the product of the traffic on the non-priority leg and the priority leg clockwise from the non-priority leg) exhibits explanatory power at least as good or better than any single traffic intensity measure and any product of two traffic intensities each raised to a power. (That is, functional form (2) is as effective as functional form (1) for these intersections.)

TABLE 14

LOG-LOG REGRESSIONS OF TOTAL ACCIDENTS (ACC_T)
ON TRAFFIC INTENSITY VARIABLES
BY CLUSTER AND DIMENSION--THREE-LEG INTERSECTION

Cluster	Dimension	Explanatory variable	Adjusted R^2	Constant (T-statistic)	Coefficient (T-statistic)
1	1	C_{13}	0.23	-.365 (-.98)	0.427 (3.77)
2	1	C_{13}	0.29	-.022 (-.13)	0.384 (6.51)
3	2	C_{13}	0.64	-.835 (-1.82)	0.641 (4.65)

The goodness-of-fit measures listed in Table 14 vary substantially among the clusters. However, all R^2 values are significant and are judged to be satisfactory in light of the aggregate nature of the data. The results shown are for ordinary least-squares estimations, rather than weighted least-squares estimates for Poisson-distributed dependent variables, because of the exploratory nature of the case study application.

The regressions of Table 14 are plotted in Figures 5 through 7. The shapes of the relationships are similar, with varying degrees of curvature. They all display a diminishing marginal effect of traffic intensity on total accidents, particularly in the lower range of intensity.

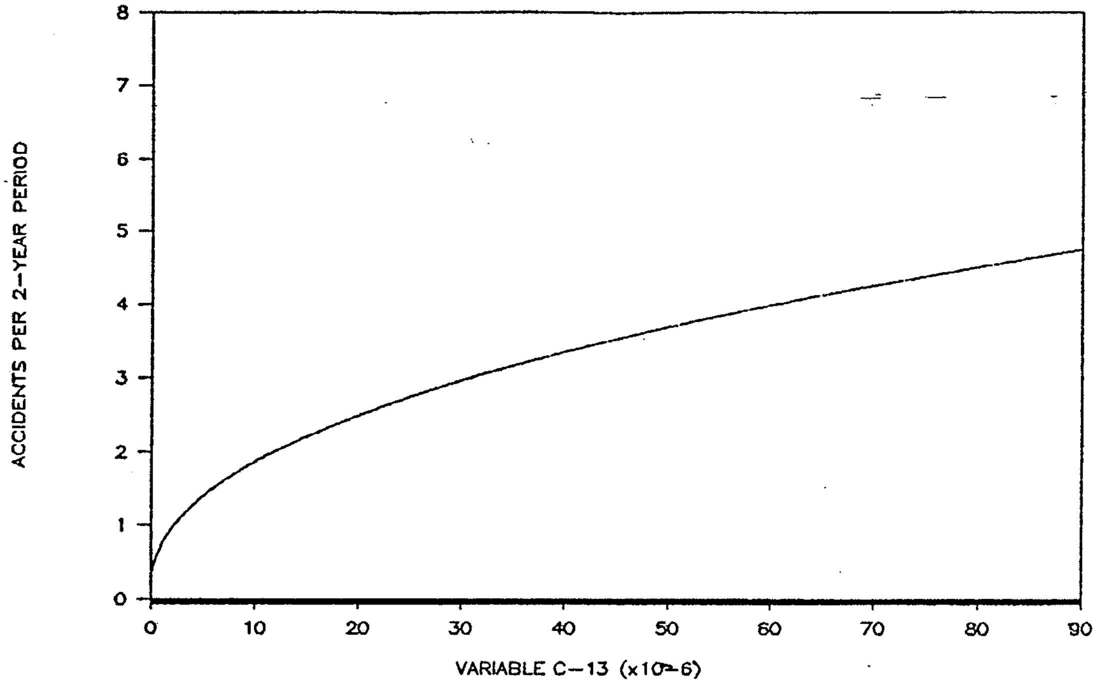

FIGURE 5

REGRESSION EQUATION FOR CLUSTER ONE OF THREE-LEG INTERSECTIONS

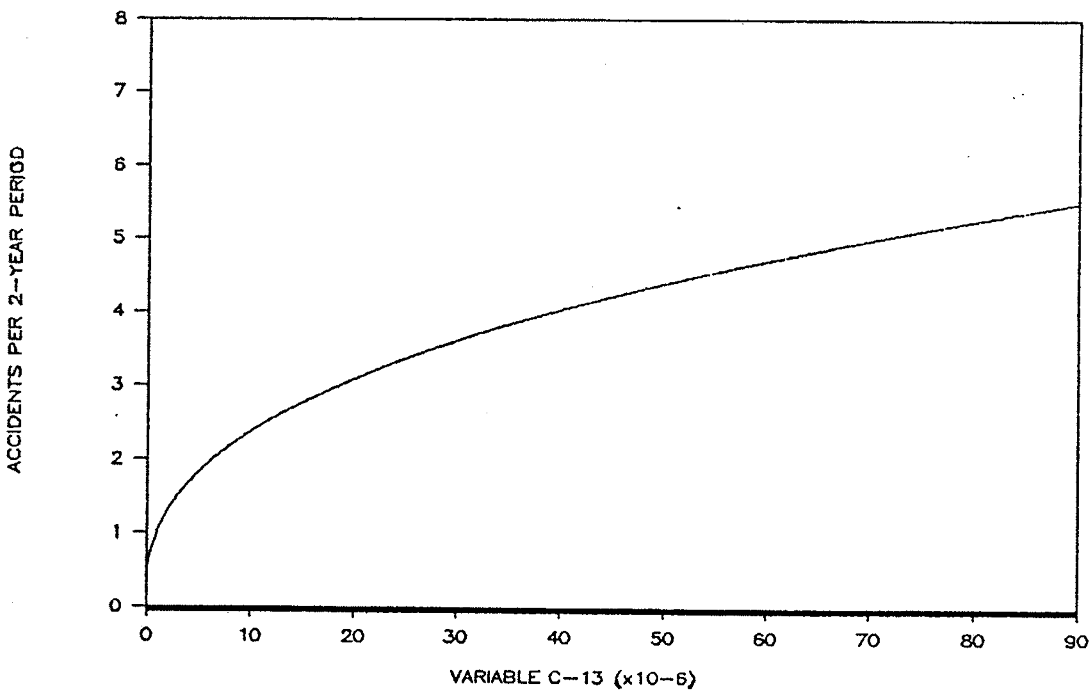

FIGURE 6

REGRESSION EQUATION FOR CLUSTER TWO OF THREE-LEG INTERSECTIONS

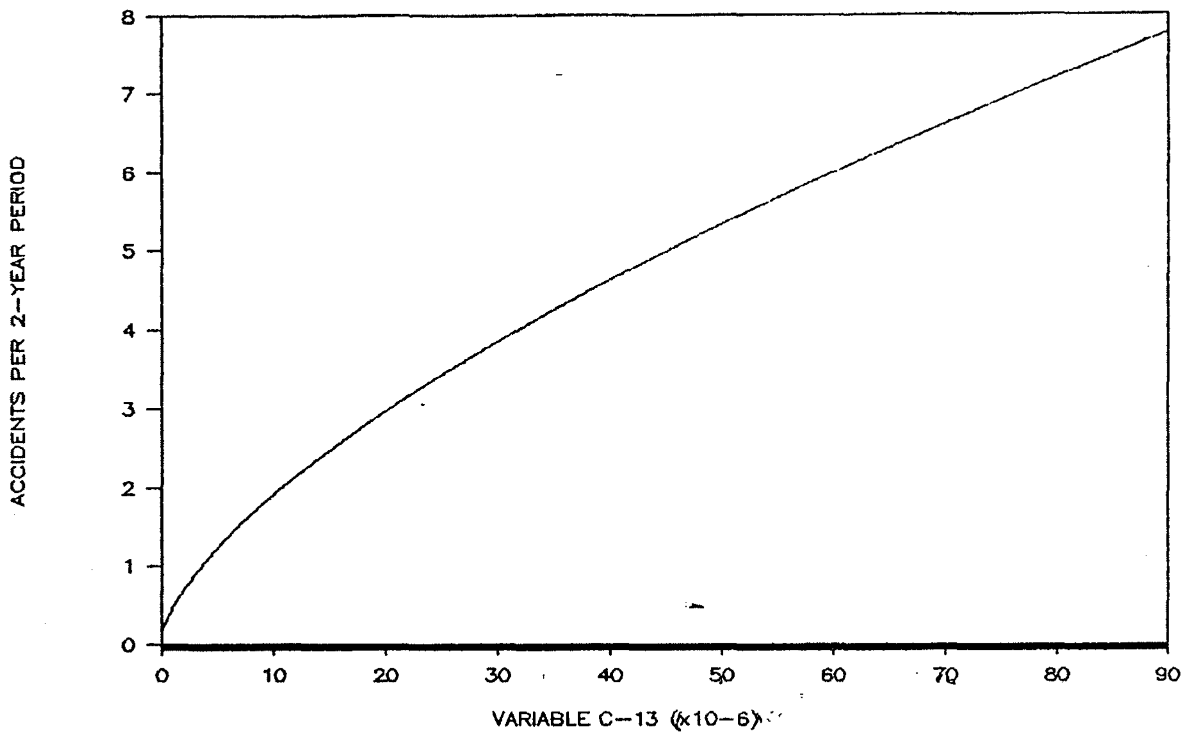

FIGURE 7

REGRESSION EQUATION FOR CLUSTER THREE OF THREE-LEG INTERSECTIONS

Results of residual analyses for each of the three intersection clusters are given in Tables 15 through 17. In each case an F-test is performed to test differences in the mean number of actual accidents minus predicted accidents for each physical design category. Categories with less than five intersections in a cluster were dropped from the analyses. (The design categories are described in Table 3.) There are no significant differences among the categories for any cluster of intersections.

TABLE 15

TEST OF SIGNIFICANT DIFFERENCE AMONG DESIGN TYPES IN TERMS OF MEAN ACTUAL-PREDICTED TOTAL ACCIDENTS--CLUSTER ONE OF THREE-LEG INTERSECTIONS

Design type	Sample size	Regression residual		F-statistic (degrees-of-freedom)	Probability difference due to chance
		mean	standard deviation		
3.1.1	9	2.21	5.43		
3.1.2	12	-0.88	2.58		
3.2	27	2.24	8.34	1.59	0.19
3.4	11	-1.44	1.18	(4.64)	
3.8	10	-1.53	0.89		

TABLE 16

TEST OF SIGNIFICANT DIFFERENCE AMONG DESIGN TYPES IN TERMS OF MEAN ACTUAL-PREDICTED TOTAL ACCIDENTS--CLUSTER TWO OF THREE-LEG INTERSECTIONS

Design type	Sample size	Regression residual		F-statistic (degrees-of-freedom)	Probability difference due to chance
		mean	standard deviation		
3.1.1	10	1.69	6.16		
3.2	25	-0.11	2.59	0.39	0.76
3.4	70	0.34	8.16	(3,146)	
3.8	45	-0.42	1.67		

TABLE 17

TEST OF SIGNIFICANT DIFFERENCE AMONG DESIGN TYPES IN TERMS OF MEAN ACTUAL-PREDICTED TOTAL ACCIDENTS--CLUSTER THREE OF THREE-LEG INTERSECTIONS

Design type	Sample size	Regression residual		F-statistic (degrees-of-freedom)	Probability difference due to chance
		mean	standard deviation		
3.1.1	7	-0.33	2.00	1.48	0.24
3.1.2	10	-4.57	8.99	(1.15)	

Regression results for the four-leg intersections are listed in Table 18. The ordinary least-squares regressions fit less well than for the three-leg intersections, indicating more diversity among the intersections within each cluster. Moreover, a different explanatory variable was found to give the best results for each cluster, indicating considerable diversity among the clusters. For the first cluster, the two-parameter functional form of equation (2) is at least as good as any three-parameter equation (1) forms; this is consistent with the results for the three-leg intersection clusters. However, for the second and third clusters, the special case of functional form (1) with only one significant traffic intensity variable was found. This result is consistent with the conclusion of McGuigan (1981) that product flows are not always justified as exposure measures. Results for all intersection clusters show that this conclusion is valid for certain groups of intersections, but not for others.

TABLE 18

LOG-LOG REGRESSIONS OF ACCTOT ON TRAFFIC INTENSITY VARIABLES BY CLUSTER AND DIMENSION--FOUR-LEG INTERSECTION

Cluster	Dimension	Explanatory variable	Adjusted R^2	Constant (T-statistic)	Coefficient (T-statistic)
1	2	C_{41}	0.19	0.774 (4.28)	0.419 (4.21)
2	1	C_1	0.15	-.045 (-.07)	0.794 (2.98)
3	2	C_3	0.35	0.518 (1.07)	1.07 (3.18)

The regression equations are graphed in Figures 8 through 10. The first cluster exhibits diminishing marginal effects similar to those found for the three-leg intersections, but the regressions for the second and third clusters are approximately linear.

The results of the residual analyses for the four-leg intersections (Tables 19 through 21) showed that there were no differences among the design types. This indicates that traffic intensity variables, in terms of the clustering of intersections and the descriptions of accidents as functions of traffic volumes, account for differences among intersections; the design category adds no additional safety information in this particular case study.

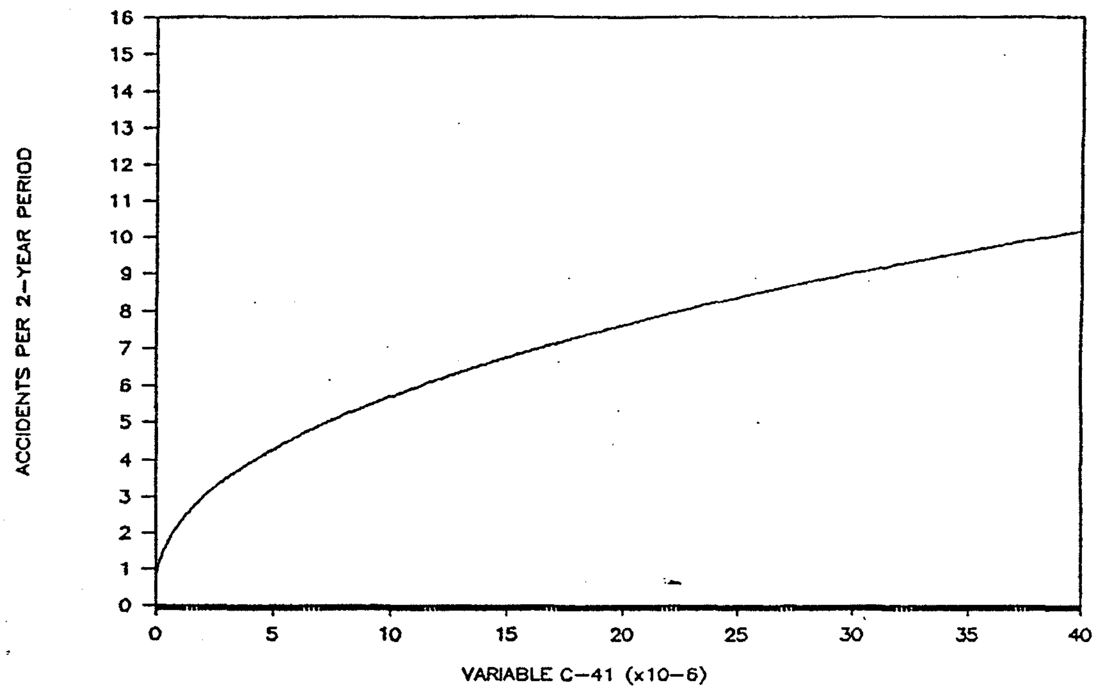

FIGURE 8

REGRESSION EQUATION FOR CLUSTER ONE OF FOUR-LEG INTERSECTIONS

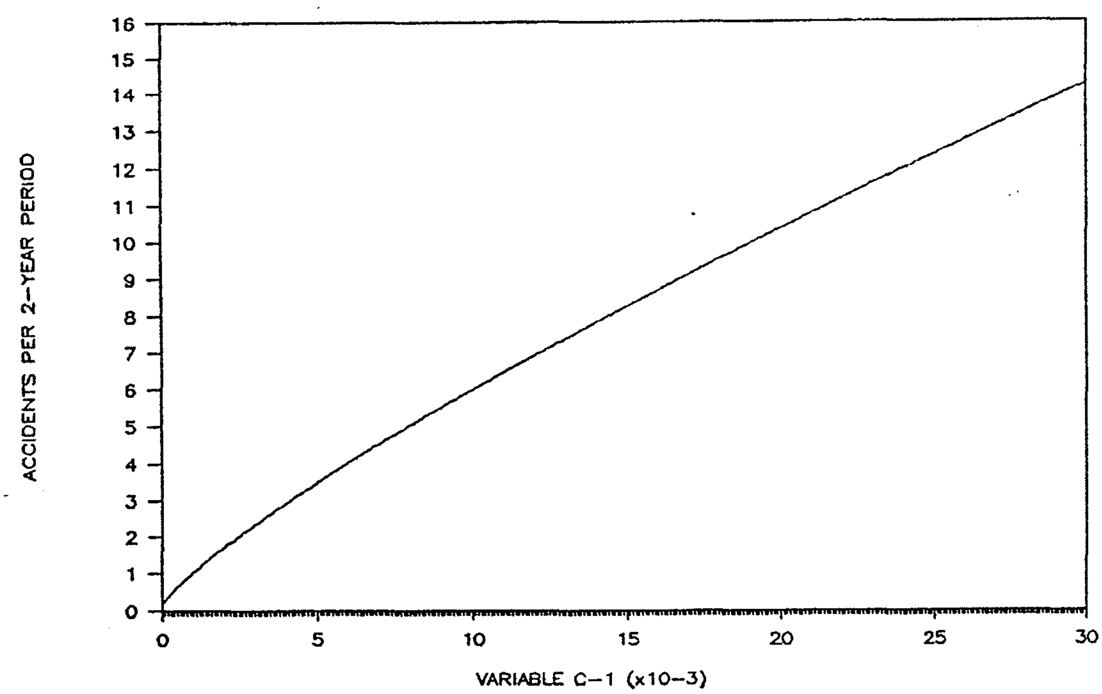

FIGURE 9

REGRESSION EQUATION FOR CLUSTER TWO OF FOUR-LEG INTERSECTIONS

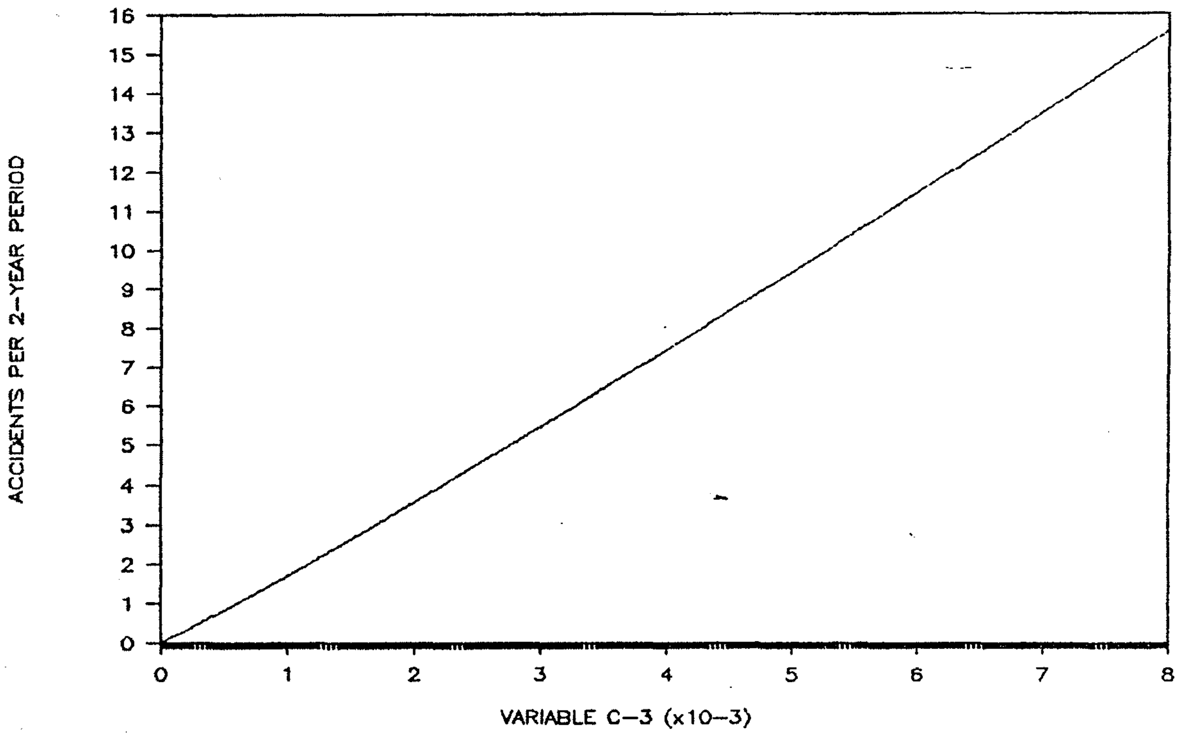

FIGURE 10

REGRESSION EQUATION FOR CLUSTER THREE OF FOUR-LEG INTERSECTIONS

9. CONCLUSIONS

Two general conclusions can be drawn from the case study application of the proposed method. First, treatment of traffic intensity variables as ordinal scales is an effective way of dealing with data of questionable accuracy. Non-linear forms were important in developing a traffic-intensity

TABLE 19

LOG-LOG REGRESSIONS OF ACCTOT ON TRAFFIC INTENSITY VARIABLES
BY CLUSTER AND DIMENSION--CLUSTER ONE OF FOUR-LEG INTERSECTION

Design type	Sample size	Regression residual		F-statistic (degrees-of-freedom)	Probability difference due to chance
		mean	standard deviation		
4.1.1	23	0.28	5.45		
4.1.2	9	-1.29	2.97		
4.2	15	1.72	5.67	0.95	0.44
4.4	25	0.50	2.90	(4.81)	
4.8	14	1.56	1.70		

TABLE 20

LOG-LOG REGRESSIONS OF ACCTOT ON TRAFFIC INTENSITY VARIABLES
BY CLUSTER AND DIMENSION--CLUSTER TWO OF FOUR-LEG INTERSECTION

Design type	Sample size	Regression residual		F-statistic (degrees-of-freedom)	Probability difference due to chance
		mean	standard deviation		
4.1.1	7	1.36	3.75		
4.1.2	13	3.50	8.61		
4.2	11	0.00	2.09	0.96	0.44
4.4	7	7.76	18.70	(4.41)	
4.8	8	0.81	4.99		

TABLE 21

LOG-LOG REGRESSIONS OF ACCTOT ON TRAFFIC INTENSITY VARIABLES BY CLUSTER AND DIMENSION--CLUSTER THREE OF FOUR-LEG INTERSECTION

Design type	Sample size	Regression residual		F-statistic (degrees-of-freedom)	Probability difference due to chance
		mean	standard deviation		
4.1.1	8	-0.31	12.67	0.08	
4.4	8	1.04	3.64	(1.14)	0.78

typology of intersections and in relating traffic intensity variables and accident rates. These functional forms were generally consistent with results found in previous studies (e.g., studies in which power functions were used).

The second general conclusion is that interaction terms between the traffic volumes on various intersection approaches are important explanatory traffic intensity variables. This result is again consistent with the use of flow product terms in previous studies.

Three conclusions specific to the case study can also be drawn. First, there are unique types of non-signalized arterial road intersections in the Netherlands, based on traffic intensity characteristics, and these types exhibit different forms of relationships between accident rates and traffic intensities. Second, for all types of three-leg arterial road intersections, the variable that is the most effective in explaining accident rates measures the interaction between the approach volume on the non-priority leg and the approach volume on the priority leg representing a left turn from the non-priority leg; for four-leg intersections the most effective intensity variable is different for each type of intersection. Third, there were no differences in accident rates among intersection geometric designs that were not explained by traffic intensity differences among the intersections. This final result highlights the importance of accounting for traffic intensity patterns when evaluating design alternatives.

REFERENCES

Bureau Goudappel Coffeng (1982). Analyse van drie- en viertaks kruispunt-kenmerken intensiteiten en ongevallen. Deventer, the Netherlands.

Bennett, G.T. (1966). Junction accidents. Journal of the Institute of Highway Engineers, 13: 117-121.

Breunning, S.M., and A.J. Bone (1959). Interchange accident exposure. Highway Research Board Bulletin 240: 44-52.

Ceder, A., and M. Livneh (1978). Further evaluation of the relationships between road accidents and average daily traffic. Accident Analysis and Prevention, 10: 95-109.

Chapman, R. (1973). The concept of exposure. Accident Analysis and Prevention, 5: 95-110.

Van der Burg, E. (1983). Canals User's Guide. Leiden: Department of Data Theory, University of Leiden.

Van der Burg, E. and V. de Leeuw (1983). Non-linear canonical correlation. British Journal of Mathematical and Statistical Psychology, 36: 54-80.

Erlander, S., J. Gustavsson and E. Larusson (1969). Some investigations on the relationship between road accidents and estimated traffic. Accident Analysis and Prevention, 1: 17-64.

Everitt, B. S. (1980). Cluster Analysis. London: Heineman Educational Books.

Gifi, A. (1981). Non-Linear Multivariate Analysis. Leiden: Department of Data Theory, University of Leiden.

Gifi, A. (1983). Princals User's Guide. Leiden: Department of Data Theory, University of Leiden.

Grossman, L. (1954). Accident exposure index. Highway Research Board Proceedings, 33: 136-143.

Hakkert, A.S., and D. Mahalel (1978). Estimating the number of accidents at intersections from a knowledge of the traffic flows on the approaches. Accident Analysis and Prevention, 10: 69-79.

Howarth, C.I. (1982). The need for regular monitoring of the exposure of pedestrians and cyclists to traffic. Accident Analysis and Prevention, 5: 341-344.

Jorgensen, N.O. (1969). A model for forecasting traffic accidents. Proceedings of the Symposium on the Use of Statistical Methods in the Analysis of Road Accidents. Crowthorne, Berks., England: Road Research Laboratory.

Leong, W.H.J. (1973). Relationship between accidents and traffic volumes at urban intersections. Journal of the Australian Road Research Board, 5: 72-90.

Mathewson, J., and R. Brenner (1957). Indexed of motor vehicle accident likelihood. Highway Research Board Bulletin 161: 23-57.

McDonald, J.W. (1953). Relation between number of accidents and traffic volumes at divided highway intersections. Highway Research Board Bulletin 74: 7-17.

McGuigan, D.R.D. (1981). The use of relationships between road accidents and traffic flow in "black-spot" identification. Traffic Engineering and Control, 22: 448-453.

Surti, V.H. (1965). Accident exposure for at-grade intersections. Traffic Engineering, 36:26.

Surti, V.H. (1969). Accident exposure and intersection safety for at-grade, unsignalized intersections. Highway Research Record 286: 81-94.

Tanner, J.C. (1953). Accidents at rural three-way intersections. Journal of the Institute of Highway Engineers, 2: 56-67.

Tatsuoka, M. M. (1971). Multivariate Analysis. New York: Wiley.

Young, F.W., J. de Leeuw and Y. Takane (1976). Regression with Qualitative and Quantitative Variables: An Alternating Least Squares Method with Optimal Scaling Features. Psychometrika, 41: 505-529.

Zalinger, D.A., B.A. Rogers, and H.P. Johri (1977). Calculation of hazard indices for highway-railway crossings in Canada. Accident Analysis and Prevention, 9: 257-273.