ENVIRONMENTAL DETERMINANTS OF TRAFFIC ACCIDENTS: AN ALTERNATE MODEL

James C. Snyder*, Georgia Institute of Technology

This study is concerned with identification and quantification of environmental determinants of traffic accidents and with the construction of a conceptual model of traffic accidents based on environmental factors. Dependent variables include accident numbers and rates (number of accidents per million vehicle-miles of travel). Independent variables include physical characteristics of the road, the road frontage (adjacent land use), and physical and social characteristics of the region. Data are derived from a sample of 135 road segments, each 2 miles long, in Oakland County, Michigan. A wide range of environmental characteristics are represented. Automatic interaction detection, multiple classification analysis, and multiple regression techniques are used to construct a series of predictive models. Analysis indicates that the number of accidents on a road segment is best predicted from traffic volumes and accident rates, whereas accident rates are best predicted from the type of road, the intensity of road frontage development, and the percentage of population between 16 and 24. Inspection of the formulated models suggests a conceptual macromodel that is different from traditional models of traffic accidents.

•MUCH of the previous research on traffic accidents has focused on the road itself, with some consideration of roadside characteristics (1, 2). Also, much of the work has dealt either with particular road sections or types or with large cross-sectional areas. Many of the basic relationships have been defined (such as the positive relationship between accident rates and traffic volume) although quantitative results have varied among studies.

The focus here, however, was on the development of a conceptual model of traffic accidents based on environmental factors across a regional geographic area with a wide range of environmental characteristics. The hypothesized general model is shown in Figure 1.

STUDY AREA AND DATA

Oakland County, Michigan, was selected as the study site. The county has an area of approximately 900 square miles and a population of approximately 900,000. The county is totally urban in the southeast, the intensity of which diminishes through suburban development to a totally rural character in the northeast.

A stratified, systematic, unaligned sample of 135 road segments, each 2 miles long, was selected. Measurements were then taken for each roadway, the land use adjacent to the roadway, and the spatial area around the segment to a distance of 3 miles. The areal measurements were derived from spatially indexed data (i.e., census districts) and weighted by a distance decay factor based on a distribution of travel distances.

The accident statistics came from the Oakland County accident file and included 13,498 reported accidents from 1968 to 1970. The resultant file contained 135 road

*This report summarizes research conducted by the author while he was with the Highway Safety Research Institute, University of Michigan.

Publication of this paper sponsored by Committee on Traffic Records.

segments, each with a number of dependent and independent variable measurements. The variables used include accident rate, accident number, road type, road volume (ADT), number of intersections, percentage of developed frontage, percentage of commercial frontage, percentage of residential frontage, number of land use category changes, percentage of regional land developed, vehicle density, employment density, population density, residential density, value of homes, rent, and percentage of the population between 16 and 24. In the figure, the road, the road frontage, and the region are the overall environment in which travel occurs. The travel activity, constrained by a set of environmental factors, produces some accident set. The underlying causal model logic is this: Travel activity has certain attendant characteristics; that is, it has an origin and destination, a mix of drivers and vehicles, a mix of dynamic parameters such as speed, vehicle density, and traffic volume, and a mix of physical parameters such as the roadway, lighting, sound, and visual appearance—all of which constitute a complex set of interacting factors. These factors affect the driver and vehicle and thus the accident set.

Although many studies have focused on some subset of these factors, this study attempts to look at the broader determinants of these factors. More specifically, the study addresses the construction of a series of mathematical predictive models of accident numbers and rates in a spatial, geographic context. The resultant models are static and descriptive and were derived from cross-sectional, spatially indexed, empirical data. Inspection of these models suggested an overall conceptual model.

ANALYSES

Methods

Bivariate relationships were examined via correlation matrices and bivariate regression plots. Second, automatic interaction detection (AID) was used to explore the structure of the data and to reveal interactions between variables. Then variables were selected for entry to multiple classification analysis (MCA) and finally to multiple regression analysis.

Accident Numbers and Rates

Initial analysis revealed two points of interest. First, the correlation between accident numbers and accident rates improved with increasing traffic volume. That is, variations in numbers of accidents and in traffic volumes, on which rates are calculated, produced less variation in rates where numbers and volumes were large. Therefore, rate prediction on low-volume roads was not successful. Numbers of accidents, however, were successfully predicted on all roads.

Second, significant intercorrelations existed within types of predictor variables (e.g., road, road frontage, and regional) because several variables used were surrogates for the same underlying factor. For example, percentage of developed frontage, percentage of commercial frontage, and number of intersecting roads each related to the intensity of road frontage activity. Selecting variables for future study may well rely on convenience of data collection rather than on a search for the best predictor within a class of predictors.

The first AID analysis of all 135 road segments showed that the type of road was the best overall predictor of accident rates. The AID tree is shown in Figure 2.

The AID analysis uses analysis of variance techniques to subdivide the sample into a series of subgroups, which maximizes the ability to predict values of the dependent variable. The program operates by finding the dichotomy, based on an independent variable, that produces the lowest within-group sum of squares for the dependent variable. This bifurcation accounts for more of the variance of the dependent variable than any other split. Each subgroup is further split in a similar manner until preselected criteria are met (e.g., minimum N for a subgroup). Each box of the AID tree gives the name of the independent variable, variable categories or values, number of cases, and value of the dependent variable for that group.

In this case, the first split was made on the type-of-road variable (categories in-

Figure 1. Conceptual traffic accident model.

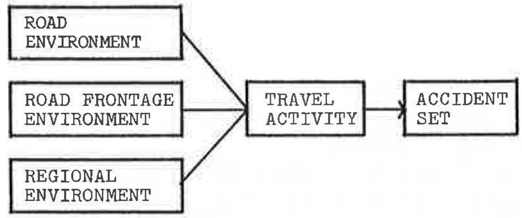


Figure 2. AID analysis of accident rates.

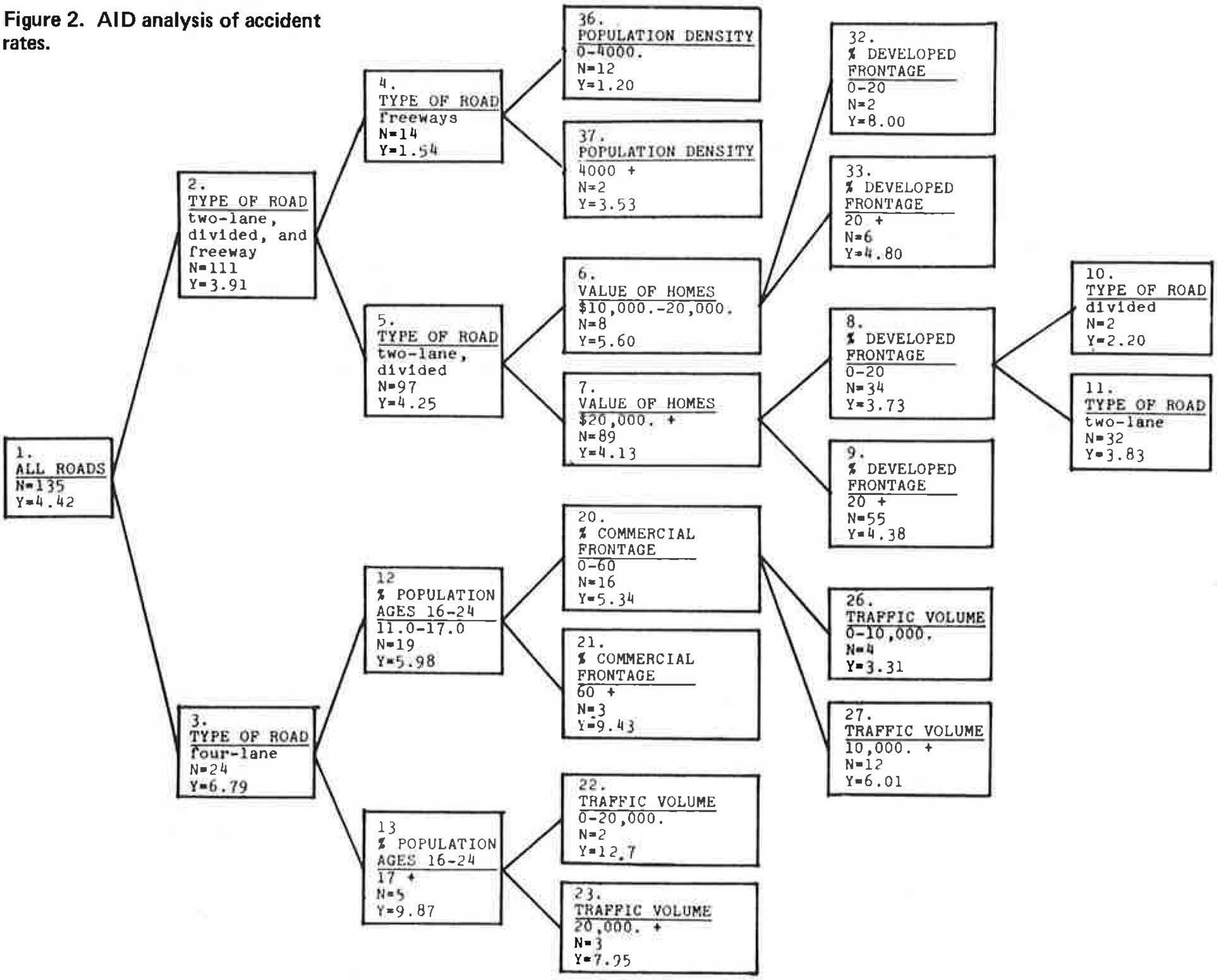
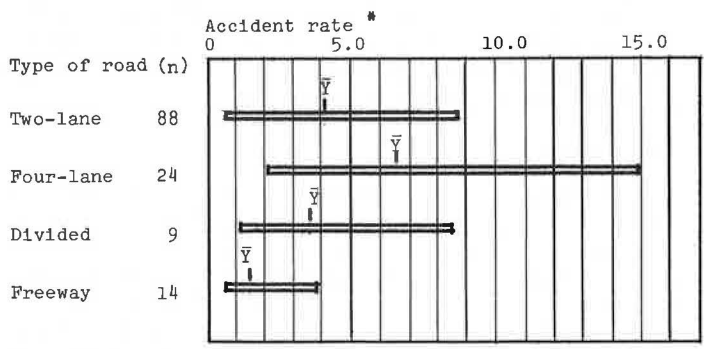


Figure 3. Means and ranges for accident rates by road type.



* accident rate = number of accidents per one million vehicle-miles

clude two-lane, four-lane, divided, and freeway), indicating that the type of road was the best overall predictor of accident rate. Each road type grouping was then split on different predictor variables.

This asymmetry indicated interaction among road type and the other independent variables. That is, different sets of environmental factors were affecting each road type. Figure 3 shows the means and ranges for accident rates for each of the road types.

When an AID analysis was conducted with number of accidents as the dependent variable, three closely competing predictor variables were found: type of road, percentage of area developed, and percentage of commercial frontage. Interaction between the type of road and other predictors was again existent. Figure 4 shows the means and ranges for number of accidents for each road type. Each road type was analyzed separately in order to avoid complex interaction terms.

Two-Lane Roads

Accident rates varied widely on low-volume two-lane roads as expected, and thus accident rate prediction was unsuccessful. The numbers of accidents, however, were best predicted from traffic volume and measures of road frontage activity. The AID tree (Fig. 5) split first on traffic volume, second on percentage of developed frontage, and third on the number of intersecting roads. However, the latter two were closely competing, intercorrelated measures of road frontage activity, and thus the tree was essentially symmetrical.

The conceptual model called for the inclusion of regional effects, but regional intensity measures such as population and vehicle density tended to be intercorrelated with road frontage measures. One of the qualitative measures, the percentage of population between 16 and 24, had a positive correlation with number of accidents and no significant intercorrelation with other independent variables. Thus those three variables were entered into the MCA (Tables 1 and 2).

This analysis produces a model of the form

$$Y_{ijk} = Y + a_i + b_j + c_k + \ldots + e_{ijk}$$

where Y_{ijk} is an individual case-dependent variable value, i, j, and k are categories on successive predictors to which the case belongs, and a_i, b_j, and c_k represent adjustments to Y, the grand mean for the dependent variable. Hence, the effect of predictor A is a_j. Thus, one simply finds the three variable categories for a particular case and makes the appropriate adjustment to the grand mean to arrive at the estimated dependent variable value. Thus, for a particular road segment, the predicted number of accidents would be the mean number of accidents for that road type plus adjustments for each of the independent variable categories. The unadjusted deviation considers only the effect of that one independent variable, whereas the adjusted deviation considers the effect of that variable given the effects of the other independent variables. The η statistic is the correlation ratio and indicates the ability of the predictor to explain variation in the dependent variable. η^2 indicates the proportion of the total sum of squares explainable by the predictor. The β statistics are analogous to the η statistics but are based on adjusted means rather than raw means.

The multiple regression is given in Tables 3 and 4. The R^2 is higher in the regression models because there is no loss of information with continuous data, whereas the MCA divides the data into subgroups.

Other Roads

The same types of analyses were conducted for accident rates and numbers for each of the other road types. Accident rates for four-lane roads were best predicted from percentage of developed frontage and percentage of population between 16 and 24. Accident numbers were best predicted from traffic volume, percentage of commercial frontage, and percentage of population between 16 and 24. The rate prediction models differed from the number prediction models in that the former did not include traffic volume as a predictor.

Figure 4. Means and ranges for number of accidents by road type.

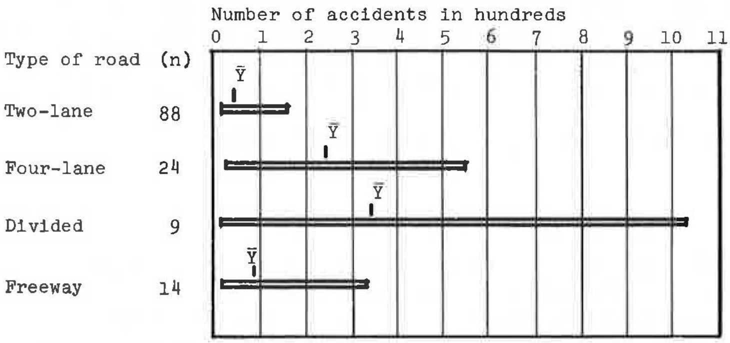


Figure 5. AID analysis of number of accidents on two-lane roads.

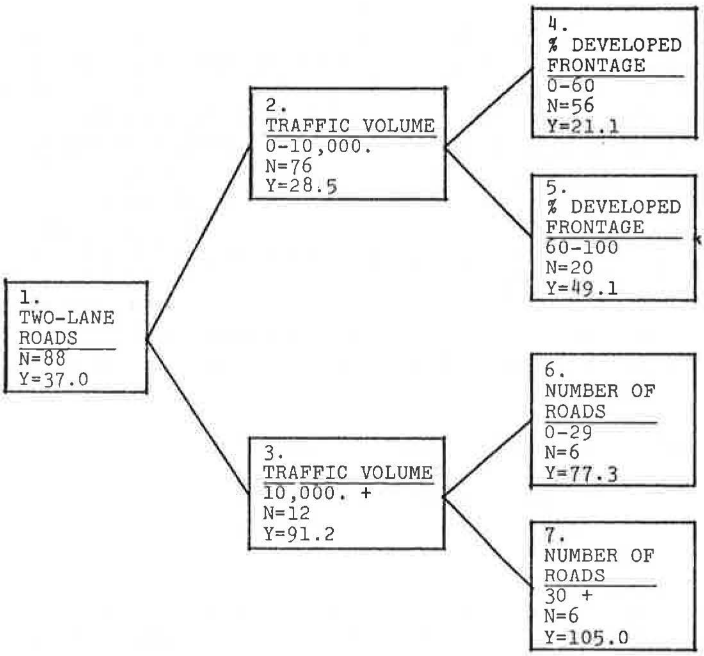


Table 1. MCA of number of accidents on two-lane roads.

Variable	η	η^2	β	β^2
Traffic volume	0.812	0.660	0.576	0.331
Percentage of developed frontage	0.695	0.483	0.310	0.096
Percentage of population between 16 and 24	0.464	0.216	0.263	0.069

Model building for divided and freeway segments was limited by a small sample number. The accident rate on divided roads was best predicted from the percentage of commercial frontage and on freeways by the regional population density. Prediction of numbers was not necessary because, on high-volume roads, the rates and traffic volumes are relatively stable and numbers can be computed directly from those statistics.

Table 5 gives a summary of all of the MCA and regression models developed in this study.

ALTERNATE MODEL

The general model of traffic accidents based on environmental factors appears to have been substantiated. The models exhibit operationally acceptable R^2 values. The inclusion of regional scale variables contributed in two ways: importance in terms of additional specification of models and substitutability of those terms for more traditional measures. The former contribution can be measured in β values, which, in this case, ranged from 0.073 to 0.429. In the latter case, several bivariate models used regional variables as the best predictor. The availability and convenience of areal data may justify additional substitution in operational situations.

Inspection of the models, as a group, leads us to consider the general relationship among accident numbers, rates, and traffic volume. Accident numbers tended to increase with increasing levels of traffic volume. A positive relationship was exhibited for all four road types, with bivariate regression R values of 0.845, 0.702, 0.831, and 0.735. Also, accident numbers tended to increase with increasing accident rates except on two-lane roads where the variability was high because of small numbers of accidents and low traffic volumes. Bivariate regression plots of accident rates and numbers for the remaining three road types exhibited R values of 0.750, 0.961, and 0.928. Accident rates, however, exhibited no significant relationship with traffic volumes in this study (R of -0.033 for all roads, -0.155 for two-lane roads, and 0.149 for four-lane roads).

All of this leads to the suggestion of a conceptual model that is different from the traditional traffic volume-accident rate model. This alternate model is based on two relationships:

$$\begin{aligned} \text{Number of accidents} &= \text{f (traffic volume, accident rate)} \\ &= \text{traffic volume} \times \text{accident rate} \\ \text{Accident rate} &= \text{f (type of road, road frontage} \\ &\quad\ \text{environment, regional environment)} \end{aligned}$$

The relationships are shown in graphic form in Figure 6.

Traffic volume and accident rate determine the number of accidents on a road segment in a simple multiplicative relationship. The type of road, road frontage characteristics, and percentage of the population between 16 and 24 determine the accident rate. Traffic volume does not directly affect the accident rate, but volume is associated with the variables that affect the accident rate. For instance, highly developed road frontage activity and heavy traffic volume tend to occur together. And traffic volume is, of course, closely associated with the type of road. This basic set of relationships is modified for two road types. First, road frontage is not an important variable for freeways because these roads have limited access. Second, the importance of the 16 to 24 age group is not exhibited on divided roads and freeways, not because it does not exist, but because the longer trip distances on those roads reduce the effectiveness of a 3-mile radius areal population measurement. For these reasons, regional variables do not appear in the divided road model, and residential density is the best predictor for freeways.

CONCLUSIONS

This study has attempted to use a wide range of environmental variables in the prediction of traffic accidents over a wide range of road types. Although cross-study

Table 2. Deviations for variables in MCA of number of accidents on two-lane roads.

Independent Variable	Number of Cases	Class Mean	Unadjusted Deviation From the Grand Mean	Adjusted Deviation From the Grand Mean
Traffic volume (ADT)				
<5,000	53	18.4	-18.6	-12.7
5,000 to 9,999	23	51.6	14.5	8.6
10,000 to 14,999	9	89.3	52.2	43.4
15,000 to 19,999	1	124.0	86.9	35.4
20,000 to 24,999	2	84.5	47.4	26.0
Percentage of developed frontage				
0 to 19	34	15.7	-21.3	-10.3
20 to 39	26	36.0	-1.0	0.6
40 to 59	14	53.2	16.1	6.2
60 to 79	8	59.3	22.2	15.6
80 to 99	6	95.6	58.5	20.6
Percentage between 16 and 24				
11.0 to 12.9	13	24.6	-12.3	1.1
13.0 to 14.9	52	33.5	-3.5	2.5
15.0 to 16.9	16	37.3	0.2	5.0
17.0 to 18.9	4	83.7	46.6	26.0
19.0+	3	89.3	52.2	30.7

Note: Grand mean number of accidents = 37.0; N = 88.

Table 3. Results of multiple regression analysis of number of accidents on two-lane roads.

Variable Number	Variable Name	Mean	Standard Deviation	Range Min	Range Max
12	Number of accidents	37.0	32.2	1.0	164.0
10	Traffic volume	5,014.7	4,266.0	173.0	20,000.0
3	Percentage of developed frontage	30.3	26.2	0.0	98.0
12	Percentage between 16 and 24	14.3	2.0	11.0	23.0

Table 4. B, β, and significance levels for regression analysis.

Item	Road Volume	Percentage Developed	Percentage Between 16 and 24
B	0.00516	0.262	0.024
β	0.692	0.214	0.073
F ratio	84.97	8.49	1.413
p	≤0.01	≤0.01	≤0.05

Note: In the overall regression, R = 0.94, F = 231.2, and p ≤ 0.01. R^2 = 0.89, N = 88, and constant term = 0.0.

Figure 6. Alternative model.

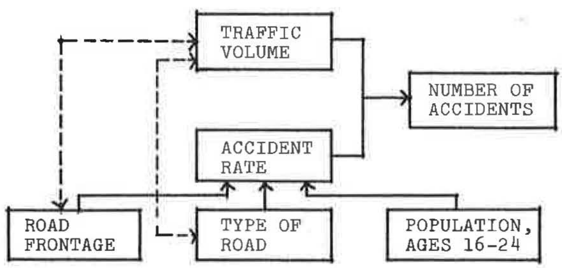


Table 5. Summary of statistical models.

Road Type	Dependent Variable	N	Road Variable	β	Road Frontage Variable	β	Regional Variable	β	MCA R^2	Regression R^2
Two-lane	No. of accidents	88	Traffic volume	0.692	Percentage of developed frontage	0.214	Percentage between 16 and 24	0.073	0.742	0.890
Four-lane	Accident rate	24			Percentage of developed frontage	0.525	Percentage between 16 and 24	0.429	0.234	0.890
Four-lane	No. of accidents	24	Traffic volume	0.644	Percentage of commercial frontage	0.226	Percentage between 16 and 24	0.119	0.780	0.890
Divided	Accident rate	9			Percentage of commercial frontage					0.880
Freeway	Accident rate	14					Population density			0.690

comparisons are difficult at best, it appears that the models developed here are at least as successful as previous attempts in terms of variance explained. It appears that the inclusion of regional variables is justified and that the underlying conceptual model is at least tentatively supported.

The use of such models has been documented in numerous previous studies and need not be elaborated here. However, the operational tasks of problem area identification, factor identification, and the like may in some cases find marginal benefit in using these models because of the relative ease of collecting the independent variable data.

In conclusion, although this approach is basically sound, much additional work is warranted in this general area.

ACKNOWLEDGMENTS

The author is now an assistant professor of city planning at the Georgia Institute of Technology. The work reported here was conducted at the Highway Safety Research Institute, University of Michigan, and was part of a PhD dissertation with co-chairmen Donald Cleveland and John Nystuen, with the assistance of William Pollock of HSRI.

REFERENCES

1. Kihlberg, J. K., and Tharp, K. J. Accident Rates as Related to Design Elements of Rural Highways. NCHRP Rept. 47, 1968.
2. Traffic Control and Roadway Elements—Their Relationship to Highway Safety, Rev. Ed. Highway Users Federation for Safety and Mobility, published in 12 separate chapters, 1968 to 1971.