CLASSIFICATIONS OF DRIVING EXPOSURE AND ACCIDENT RATES FOR HIGHWAY SAFETY ANALYSIS

PHILIP S. CARROLL

Highway Safety Research Institute, The University of Michigan, Ann Arbor, Michigan, U.S.A.

INTRODUCTION

EVALUATION of highway safety countermeasures in the future will depend heavily on the availability of accident-rate data which is uniquely classified according to variables of the highway transportation system. The accident rate data is derived as the number of traffic accident involvements in a certain class divided by the corresponding exposure value for the same class. Both accident involvement and exposure may be classified according to various driver–vehicle–road–environment combinations which exist in any part of the highway transportation system over any period of time. The purpose of this article is to derive classifications of driving exposure and accident rates which will be valuable in the future for broad analyses of highway safety problems. Succeeding sections deal with definitions of driving exposure, description of a pilot survey of exposure data, and methodology used in deriving recommended data classifications.

DEFINITIONS OF DRIVING EXPOSURE

Driving exposure is a frequently used term in the highway safety research community, but, as yet, there is no definitive agreement on what it means. Many authors in the field of highway safety have discussed the concept of exposure without attempting a direct definition. Dunlap (1953) implied that exposure is a measure of "the frequency of the existence of ... a situation which may or may not involve an accident". Mathewson and Brenner (1957) recommended a "unit of risk in motor vehicle accident rates". In *The Federal Role in Highway Safety* (1959), the discussion referred to exposure to hazard and the chances of being involved in an accident. Platt (1959) distinguishes between exposure to collision and exposure to damage or injury. Stewart (1960) indicated that driving exposure required information on total driving experience, kinds of driving experience, and the distribution of these kinds of experience, within a given time period. Jacobs (1961) asked how one measures exposure, i.e. "the frequency of occurrence of risk situations (and) circumstances associated with risk situations". Thorpe (1964) defined the exposure to accident of particular groups of driver–vehicle combinations as "total vehicle miles", and assumed it to be proportional to twice the number of two-car accidents for the group minus the number of one-car accidents for the group. Goeller (1968) called exposure over a given driving distance "the number of times that danger occurs", and he related it to vulnerability and, hence, to confrontation. Carr (1969) suggested instead of exposure, a relative risk function that characterizes driver–vehicle combinations in all environmental conditions, especially with respect to location along the roadways. Klein and Waller (1970) considered exposure as the

"population at risk (in terms of passenger or vehicle miles)", used as a denominator in the calculation of an accident or injury rate. Haight (1971) noted that "exposure to accidents" evolved as a concept by analogy to "exposure to disease", and indicated continuing difficulties in giving the concept a precise meaning.

One author who makes it explicitly clear that he is attempting a definition of exposure is DeSilva (1942); he defined exposure as "the number and relative danger of external hazards encountered while driving".

For this discussion, a definition previously proposed by the author (Carroll, 1971) will be used in an attempt to reconcile divergent implications: "driving exposure is the frequency of traffic events which create a risk of accidents".

This basic definition is rather flexible because (1) it admits a wide variety of specific measures, (2) it is not exclusive in space or time, and (3) it may be applied to any element of the system (e.g. drivers or vehicles) either singly or in groups.

As a measure of the "frequency of traffic events", driving exposure must be a cumulative quantity which increases continually with increased driving. At the same time, our definition requires that the traffic events be ones "which create a risk of accident". However, the definition does not require that the traffic events be differentiated as to the severity of the risks they create. In fact, it is cogent to argue that all traffic events (including the most normal driving) create some inherent risk of accident. Thus, exposure is seen as a continuous process, amenable to cumulative measurement.

Although the basic definition is best adapted to a measure of the *amount* of driving, it can also deal with the *nature* of driving as it relates to DeSilva's "relative danger of external hazards". For example, the frequencies of traffic events could be measured in several categories of varying risk (e.g. high speed–low speed, or heavy traffic–light traffic). However, problems of data collection and analysis would increase greatly in such cases. Hence, the basic definition will be acceptable until the time that exposure research deals more directly with the nature of driving.

The most commonly used measure of exposure is driving distance expressed in vehicle miles of travel. Other measures that are used occasionally include driving time, traffic volume, number of registered vehicles, number of licensed drivers, and gasoline consumption. Among all of the proposed exposure measures, driving distance (vehicle miles) is the one that relates most directly to the processes of highway travel and, hence, to the risk of accident.

When we employ units of vehicle miles as the measure of driving exposure, we assume that all driving is equally susceptible to the risk of accident. Every unit of distance traveled is viewed as part of a uniform stream of driving; chains of traffic events merge into a continuous flow, and the relative danger of various traffic events is submerged in importance. Thus, the "frequency of traffic events which create a risk of accident" is represented mathematically as a continuous, linear function of distance traveled (vehicle miles).

PILOT SURVEY OF EXPOSURE DATA

As part of a study for the National Highway Traffic Safety Administration, a national pilot survey of driving exposure was conducted (sample of 7145 licensed drivers), as reported by Carroll *et al.* (1971).

The purpose of the survey was to provide data for analysis to identify unique classifications of exposure (i.e. driver–vehicle–road–environment combinations) and thus to determine variables which should be included in future exposure surveys.

TABLE 1. STATES REPRESENTED IN SURVEY SAMPLE

State	Sample size	State	Sample size
Alaska	*	Michigan	855
Arizona	*	Nebraska	156
California	1240	Nevada	*
Colorado	289	New Mexico	112
Georgia	488	North Carolina	358
Hawaii	*	Ohio	538
Idaho	158	South Carolina	82
Indiana	542	South Dakota	*
Iowa	128	Texas	641
Kentucky	166	Utah	*
Louisiana	231	Virginia	306
Massachusetts	612	Washington	241

* No sampling areas selected from state.

The pilot survey was conducted by means of personal interviews of licensed drivers in 37 licensing offices throughout the country during the first six months of 1970. The selection of this "office interview method" was for quick-response experimental research and does not imply its superiority for operational implementation of exposure surveys in the long run. This survey method provided an unbiased and random source of drivers, easy sampling and implementation, and low cost. All of the other methods considered (home interviews, mail questionnaires, telephone interviews) would have required lengthy and costly sampling procedures with state license records or state registration records.

The sample size of 7145 represented 89 per cent of the interview requests. Random sampling was applied in three stages. First, 32 sampling areas (large counties or county groups) were selected (by population-weighted probability sampling) from the 24 states which required personal appearance of drivers for license renewal. The 24 states (Table 1) are well distributed throughout the country, and are considered to be fairly representative. Second, in those sampling areas with more than one licensing office, a specific office was

TABLE 2. SURVEY VARIABLES

Independent variables	
Driver age*	Percentage driving on rural freeways*
Driver sex*	Percentage driving on rural roads*
Area population	Percentage driving at night*
Drive on job?	Percentage driving on wet pavement*
Income	Number of violations not
Vehicle type*	connected with accidents
Passenger car size*	Socioeconomic scale
Passenger car manufacturer*	Engine knowledge
Vehicle year*	Urbanization index
Vehicle use*	*Dependent variables*
Percentage driving on city street	Miles driven in past 30 days
Percentage driving on urban	Number of accidents in past
freeways*	3 years

* Variables which correspond to items on standard accident reports.

randomly selected for interviewing; in five of the largest cities two offices were selected, making a total of 37 interviewing locations. Third, for each office an interviewing interval was defined on the basis of volume of renewal applications. Thus, the sample subjects were randomly selected and self-weighting. The variables represented by interview questions are listed in Table 2.

ANALYSIS OF SURVEY DATA

The purpose of the statistical data analysis was to determine sets of the independent variables which interactively provide the best predictions of exposure and accident rate. A computer algorithm known as AID (Automatic Interaction Detector) was used for this purpose. The program divides a data sample into two groups by picking the best predictor variable for the whole sample, and identifying the two sets of value levels (of the selected variable) which produce a minimum remaining unexplained variability in the dependent variable. The program then repeats the process for each of the groups identified in the
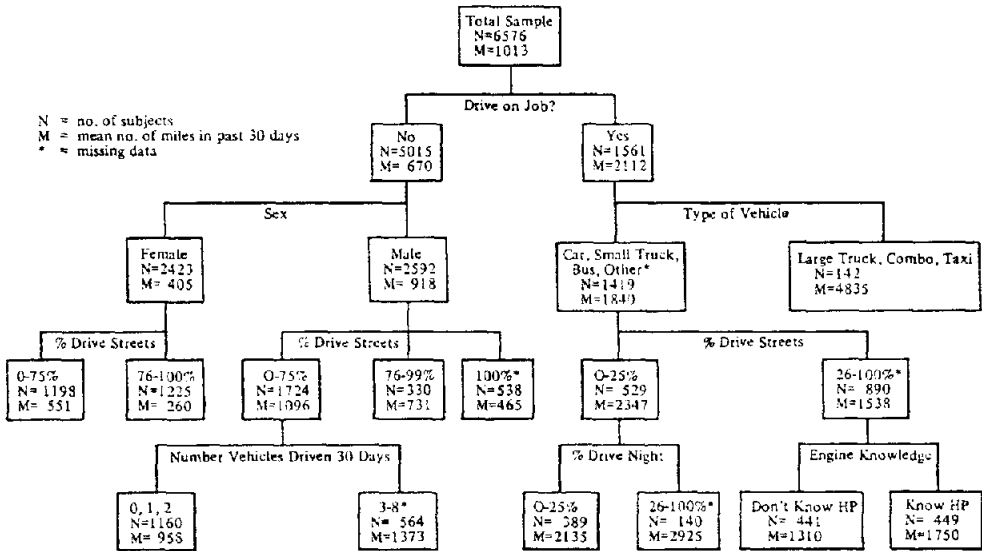

FIG. 1. Initial analysis of exposure classes.

preceding stage of analysis. Repetitions of the process continue for each new subgroup until the subgroup size reaches a specified minimum. The result is a hierarchy of data groups or classes (driver–vehicle–road–environment combinations) in order of their importance as predictors of the dependent variable, e.g., exposure in vehicle miles.

Figure 1 shows an AID chart produced from a computer run in which the dependent variable was *Miles driven in the past 30 days* as a measure of exposure, and in which all 21 independent variables of Table 2 were potential predictors. The total sample consisted of 6576 subjects ($N = 6576$) who responded to the primary dependent-variable question (miles in last 30 days). Their mean mileage was 1013 miles ($M = 1013$). The strongest predictor variable was *Drive on Job* (yes or no). The result of this first split was a group of 1561 subjects who drove on the job ($M = 2112$) and a group of 5105 drivers who did not

drive on the job ($M = 670$). The remaining splits used the predictor variables *Type of Vehicle, Sex, Percentage Driving on City Streets, Number of Vehicles Driven in Past 30 Days, Percentage Driving at Night, and Engine Knowledge.*

The eleven groups at the ends of the branches identify unique classes of driving exposure, i.e. driver–vehicle–road–environment combinations which define the total sample with minimum unexplained variability in exposure. The class with highest exposure consists of those drivers who drive on the job, and drive vehicles in the categories of large trucks, tractor–trailer combinations, and taxis. The class with lowest exposure consists of those drivers who don't drive on the job, are female, and drive mostly on city streets. The variable *Percentage Driving on City Streets* is the only one to appear on both of the main branches of the chart. Six of the classes are defined by four variables, four classes are defined by three variables, and one class is defined by only two variables. Thus, the eleven unique exposure classes are clearly defined in fairly simple terms. Further studies of the interactions among the variables defining these classes may be pursued to gain a better understanding of the nature of driving exposure.
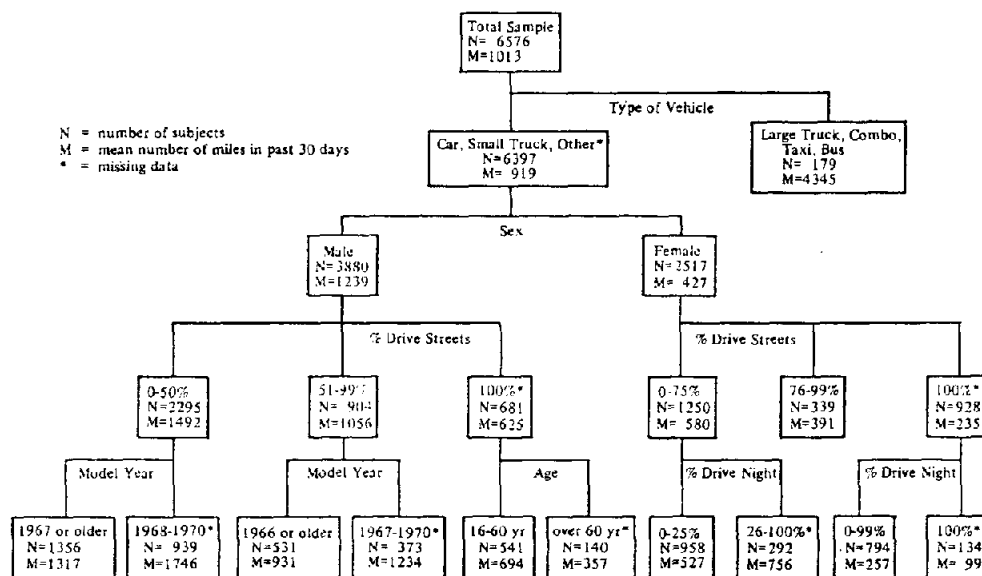

FIG. 2. Exposure classes, only accident-report variables as potential predictors.

Three of the variables in Fig. 1 (*Drive on Job, Number of Vehicles Driven, Engine Knowledge*) do not correspond to any items presently included in standard accident reports. Thus, there is no way available at present to perform analysis of mass accident data in the eleven classes so that it may be compared with exposure data in the corresponding classes. Nevertheless, the hierarchy of Fig. 1 may be used to define areas of future analysis on exposure data *per se*, e.g. causes of exposure differences among classes.

In order to determine unique exposure classes for which accident rates may be determined, another AID run was made where only the 12 independent variables which correspond to accident-report items were included as potential predictors. The resulting hierarchy is shown in Fig. 2. The strongest predictor was *Type of Vehicle* and the next two were *Sex*

and *Percentage Driving on City Streets*, thus confirming the results of Fig. 1 when *Drive on Job* is eliminated. *Percentage Driving at Night* also repeats in Fig. 2, while *Model Year* and *Age* replace the three non-accident-report variables from Fig. 1. The class with highest exposure is that of large trucks, tractor–trailer combinations, taxis and buses, and the class with lowest exposure is that of females driving small vehicles on streets at night. Ten of the twelve classes in Fig. 2 are defined by four variables, one class is defined by three variables, and one class is defined by only one variable. Although accident rates may be calculated for all of the classes, they will not necesarily identify the most critical differences among classes.

In order to determine the most unique accident-rate classes (those exhibiting the largest differences in accident rate between classes), it was necessary to modify the data sample. Only those cases were retained which had responses to both dependent variables, i.e. exposure (in 30 days) and accidents (in 3 yr). Thus, the sample size was reduced to 5629. An accident rate was computed for each case by dividing the number of accidents by the exposure estimate (miles in 30 days), extrapolated to 3 yr; thus a third dependent variable was derived. AID runs were produced for each of the three dependent variables, as shown in Figs. 3–5.

Figure 3 is very similar to Fig. 2, differing primarily in that Driver Sex reverses order with Vehicle Type as the best predictor of exposure. The similarity confirms that the reduced sample was unbiased when 947 cases were removed which did not have responses to the "accidents in 3 yr" variable. The highest exposure at the first level is accumulated by male drivers (62 per cent of the sample, 83·6 per cent of the total mileage). The highest average exposure (2917 miles/month) is attributed to males driving non-passenger cars, mostly on roads other than city streets (only 0–25 per cent on streets). The lowest average exposure (215 miles/month) is attributed to females driving exclusively on city streets in vehicles 5 yr old or older. Vehicle Make and Driver Age are also indicated as partial predictors of exposure.

Figure 4 presents the hierarchy of classes which are the best predictors of accident experience. At the first level, young drivers 16–25 have double the accidents as older drivers. In all age ranges, males have nearly double the number of accidents involving females. Comparison with Fig. 3 shows that the best predictors of exposure and of accidents are quite different; though some of the predictor variables are the same, their interactions are different. Because the structures of the AID charts in Figs. 3 and 4 do not coincide, the mean values of accident frequency and mileage cannot be combined to represent unique accident-rate classes.

Figure 5 presents the hierarchy of classes which are the best predictors of mean accident rate. Comparisons with Figs. 3 and 4 show again that although some of the predictor variables are the same, their interactions are different.

Originally, it was expected that the hierarchy of Fig. 5 would determine the accident-rate classes which should be used for future highway safety evaluations. However, the approach was changed when a discrepancy was noted in accident-rate values. The mean accident rate for the total sample in Fig. 5 was 35·0 accidents per million miles, but the quotient of all accidents in the sample (1776) and total exposure (5·92 million miles) extrapolated to 36 months is only 8·4 accidents per million miles. While the former value is a mean of individual accident rates, the latter is a group accident rate. The discrepancy is due to the unusual distribution of individual accident rates, viz. a large number of "zero" rates and a small number of excessive rates. The discrepancy was so severe that it was felt that the hierarchy
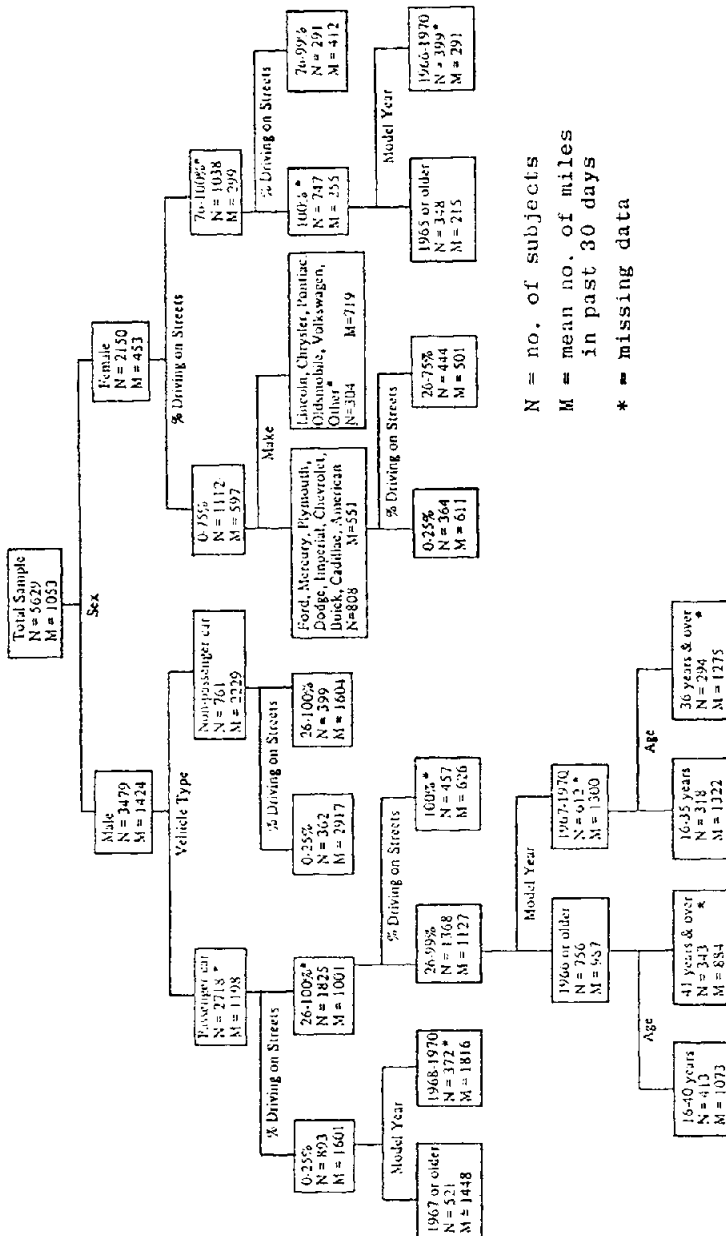
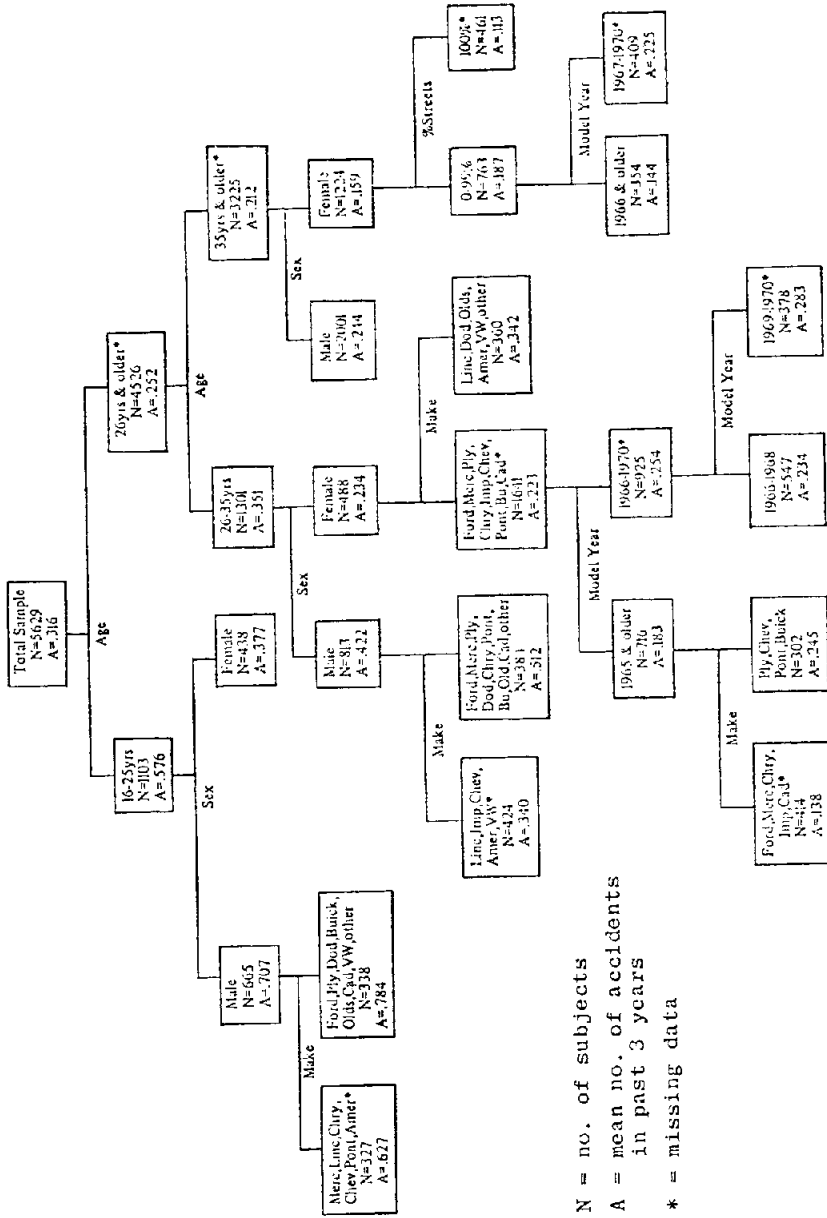FIG. 3. Exposure classes from reduced sample.

PHILIP S. CARROLL
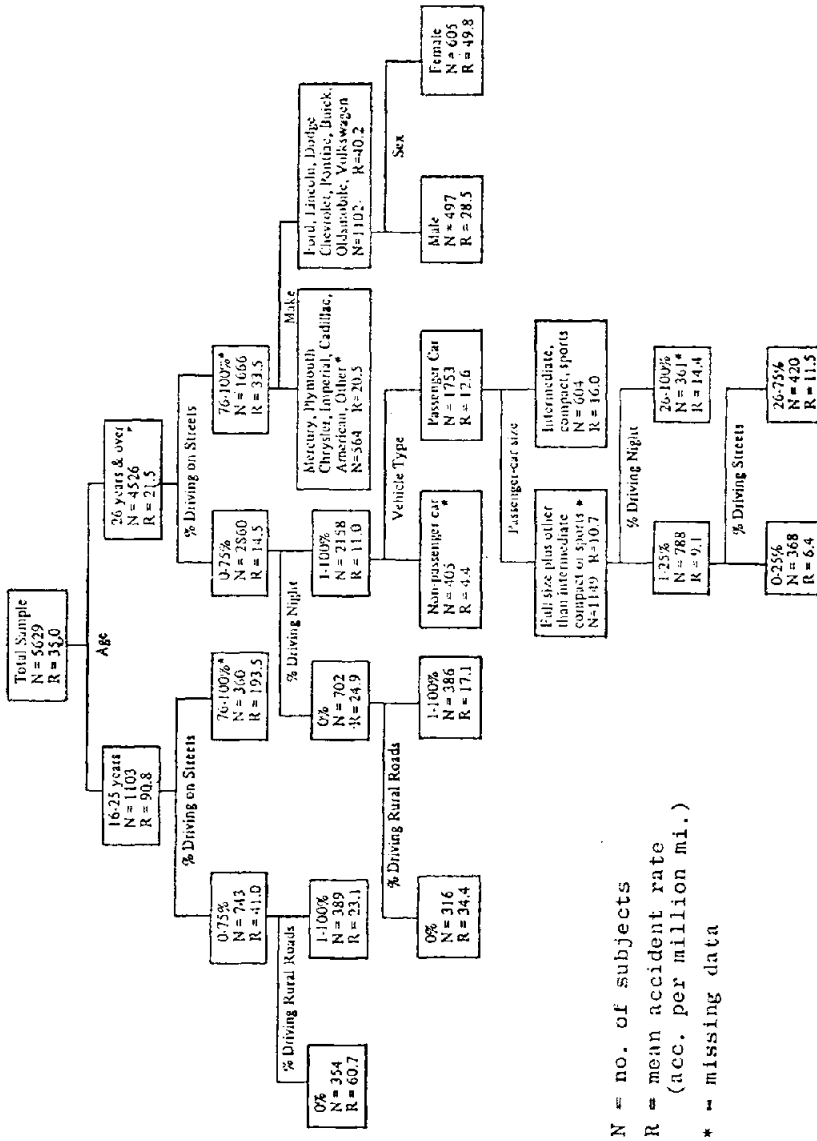

FIG. 4. Accident classes from reduced sample.

Fig. 5. Accident rate classes from reduced sample.

N = no. of subjects

R = mean accident rate
(acc. per million mi.)

* = missing data

of Fig. 5 could be misleading. Consequently, a "group accident-rate difference" method was devised to find unique accident-rate classes, i.e. the best predictors of group accident rates.

The first step is to determine the total number of accidents and total mileage for each level of each independent variable in the whole data set. For each variable, the levels are grouped in all possible two-group combinations which have logical meanings (e.g. old drivers and young drivers would not be grouped together). The group accident rates are determined for each group in the various combinations (total accidents for the group divided by total mileage for the group). The relative difference in group accident rate is determined for each combination. The relative difference (actual difference between the two groups divided by their average) is a better indicator of uniqueness than actual differ-ence. For all 12 variables, one of them will have a combination with a relative difference higher than all the other combinations in the other variables. That variable is selected as the desired "first splitting variable" and the sample is divided into the two groups with highest relative difference in accident rate. The process is repeated as in an AID run until the subgroup size reaches a specified minimum. The smallest groups at the end of the process are unique classes of driver–vehicle–road–environment combinations which have the maximum relative homogeneity with respect to group accident rate.

Figure 6 shows the accident-rate classes defined by the group accident-rate difference method. The sample for this run included 91 cases of zero exposure which were previously removed because their individual accident rates could not be computed. At the first level, the highest accident rate is for females, nearly twice the rate for males. Thus, though the first predictor variable (Driver Sex) is the same as for exposure classes in Fig. 3, its relation-ship to the dependent variable is reversed (males have higher exposure but lower accident rate). The Vehicle Type and Percentage Driving on Streets variables also have very similar interactions in Figs. 3 and 6, but again with reversed relationships between exposure and accident rate.

The highest accident-rate group identified by Fig. 6 is young males (16–20) driving pas-senger cars whose rate is three times the national average. Other high rate groups are young females, and middle-age females driving on city streets. The lowest accident-rate group is males driving vehicles other than passenger cars mostly on roads other than city streets (e.g. truck drivers on freeways). Another low rate group is older males driving passenger cars on roads other than city streets (e.g. salesmen or family heads on vacation trips). There is a factor of nearly 10 between the highest and lowest rates.

In future exposure surveys based on trip logs, it will be possible to classify each trip precisely as to road type and environment, rather than by using the independent "percent driving" variables. Other adjustments may be made in Fig. 6 without changing the unique-ness of classes. For example, the age groups for males and females can be made consistent, and the "road type" split can be made at the same level on all branches of the chart. Thus, a recommended hierarchy of accident-rate classes for future highway safety analyses is derived in Fig. 7. All of the variables from Fig. 6 are included except for Vehicle Manu-facturer, which is not recommended because of uncertainties about meaningful groupings of manufacturers. The Day/Night and Model Year variables are added at the bottom on branches which have larger groups, and which show interaction with these two variables in previous hierarchies.

The recommended hierarchy includes 18 final classes of driver–vehicle–road–environ-ment combinations which will be fairly uniform in their proportions of total exposure.

N = no. of subjects

R = group accident rate
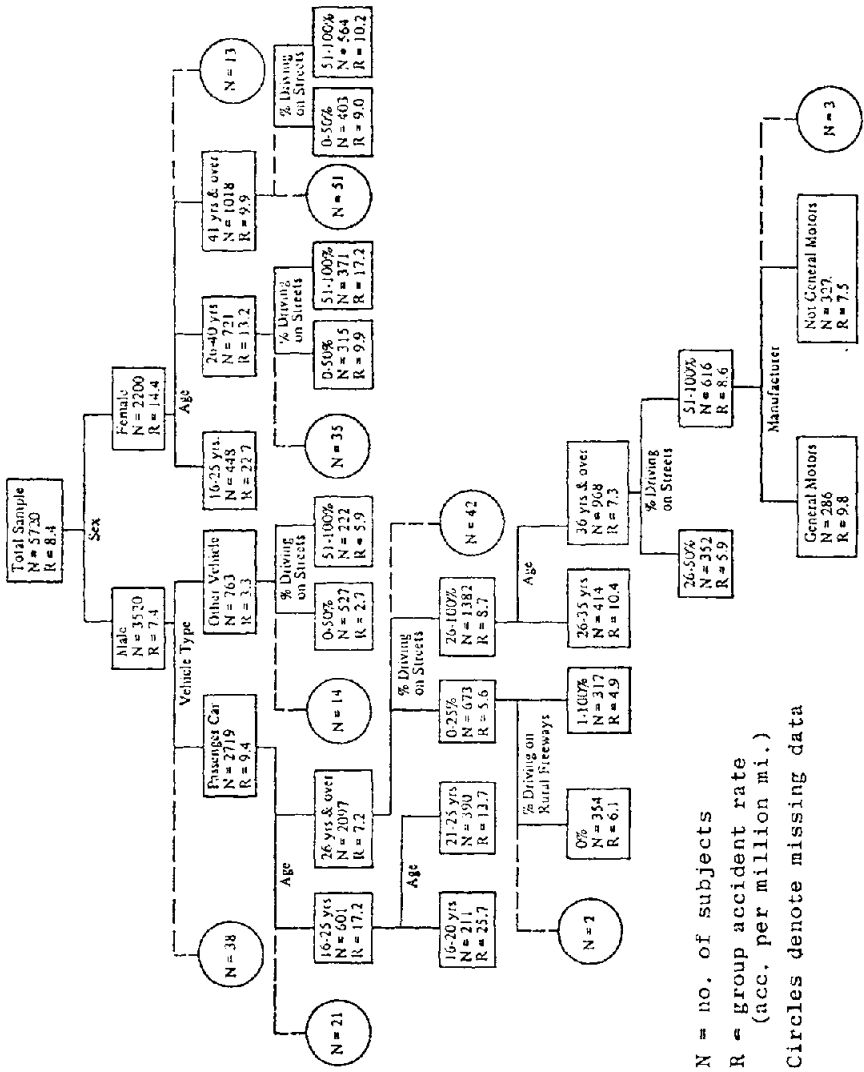   (acc. per million mi.)

Circles denote missing data

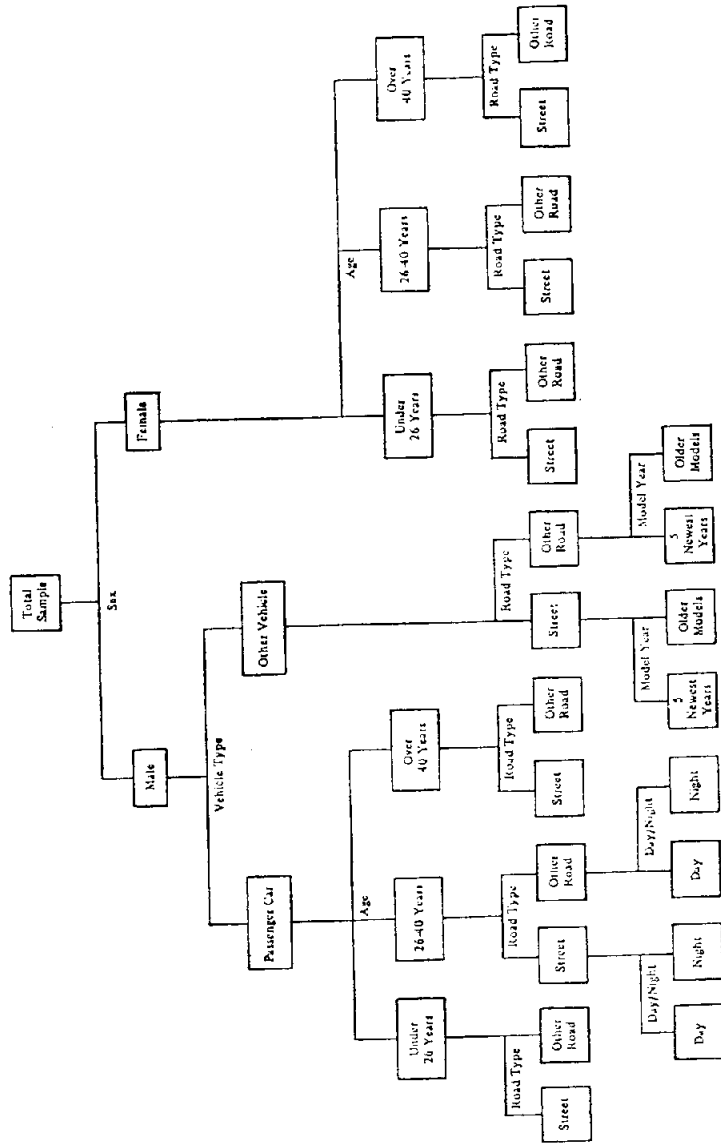Fig. 6. Accident-rate classes based on group accident-rate differences.

FIG. 7. Recommended hierarchy of accident-rate classes.

These classes are the ones for which future exposure data should be collected and for which trends in accident rates should be particularly noted.

REFERENCES

CARR, B. R. (1969) A statistical analysis of rural Ontario traffic accidents using induced exposure data. *Accid. Anal. & Prev.* **1**, 343–358.

CARROLL, P. S. (1971) The meaning of driving exposure. *HIT-Lab Rep.* (April). Highway Safety Research Institute, The University of Michigan, Ann Arbor.

CARROLL, P. S., CARLSON, W. L., McDOLE, T. L. and SMITH, D. W. (1971) *Acquisition of Information on Exposure and on Non-Fatal Crashes*, Vol. I. Highway Safety Research Institute, The University of Michigan, Ann Arbor.

DESILVA, H. R. (1942) *Why We Have Automobile Accidents.* Wiley, New York.

DUNLAP and ASSOCIATES (1953) *An Analysis of Risk and Exposure in Automobile Accidents.* Prepared for the Office of the Surgeon General, Department of the Army.

GOELLER, B. F. (1968) Modeling the traffic safety system. RAND Corp., Santa Monica, California. (1969) *Accid. Anal. & Prev.* **1**, 167–204.

HAIGHT, F. A. (1971) Indirect methods for measuring exposure factors as related to the incidence of motor vehicle traffic accidents. Prepared for National Highway Traffic Safety Administration.

JACOBS, H. H. (1961) Conceptual and methodological problems in accident research. *Behavioural Approaches to Accident Research.* Association for the Aid of Crippled Children, New York.

KLEIN, D. and WALLER, J. A. (1970) *Causation, Culpability and Deterrence in Highway Crashes.* Prepared for the Department of Transportation, Automobile Insurance and Compensation Study.

MATHEWSON, J. H. and BRENNER, R. (1957) Indexes of motor vehicle accident likelihood. *HRB Bull.* **161**, 1–8.

PLATT, F. N. (1959) *Operations Analysis of Traffic Safety.* Ford Motor Company.

STEWART, R. G. (1960) Driving exposure: what does it mean? how is it measured? *Traffic Saf. Res. Rev.* **4**, 9–11.

The Federal Role in Highway Safety (1959) House Document No. 93, First Session 86th Congress.

THORPE, J. D. (1964) Calculating relative involvement rates in accidents without determining exposure. *Aust. Road Res.* **2**, 25–36.

Abstract—Analysis of driving exposure and accident rates will be important in the future for evaluations of highway safety countermeasures. Each should be uniquely classified according to variables of the highway transportation system, i.e. driver–vehicle–road–environment combinations, in order to allow analyses which deal with the most significant differences in exposure and accident rate. Concepts of driving exposure are discussed, and a working definition is presented.

A pilot survey of exposure was conducted, based on a random national sample of 7145 licensed drivers. Data was analyzed using the Automatic Interaction Detector (AID) algorithm, and hierarchies were produced of best predictors of accidents, exposure, and accident rate. The best predictor of exposure was whether or not a person drives on the job. When variables are limited to those appearing on accident reports, Driver Sex and Type of Vehicle were the best predictors, followed by Percent Driving on Streets, Model Year of Vehicle, and Driver Age. The best predictor of accidents was Driver Age, followed by Driver Sex and other variables in different order than the exposure hierarchy.

Résumé—Dans l'avenir, l'analyse de l'exposition à l'action de conduire et la proportion d'accidents sera importante pour les évaluations des contre-mesures concernant la sécurité sur les autoroutes. Chaque évaluation devrait être classifiée uniquement en conformité avec les variables du système de transport sur l'autoroute, par ex: des combinaisons conducteur-véhicule–route–environnement, afin de permettre des analyses qui s'occupent avec les différences les plus significatives d'exposition et proportion d'accidents. On discute les considérations sur les risques de conduire et on présente une définition clarifiante.

On a entrepris une étude-pilote sur l'exposition, basée sur un échantillon national de 7145 conducteurs avec permis. Les données ont été analysées en utilisant l'algorisme par Détecteur d'Interaction Automatique (AID), et on a établi des hiérarchies de meilleures prévisions d'accidents, risques et proportion d'accidents. La meilleure prévision concernant l'exposition

fut dans la catégorie des conducteurs professionels. Quand les variables sont limitées à celles apparaissant dans les rapports sur les accidents, le sexe du conducteur et le type de véhicule étaient les meilleurs indices de prévision, puis le pourcentage de personnes conduisant sur les rues, l'âge de la voiture et l'âge du conducteur. La meilleure prévision des accidents était en liaison avec l'âge du conducteur, puis avec le sexe du conducteur et autres variables, dans un ordre différent de l'hiérarchie de l'exposition.

Zusammenfassung—Analyse der Aussetzung beim Fahren und der Unfallrate werden bei der Auswertung von Gegenmaßnahmen für Strassensicherheit in Zukunft von Wichtigkeit sein. Beide sollten speziell auf Grund der Veränderlichen des Straßentransportsystems klassifiziert werden, d.h. Fahrer-Fahrzeug-Straße-Umweltkombinationen, um Analysen möglich zu machen, welche sich mit den Hauptunterschieden in Aussetzung und Unfallrate befassen. Es werden Begriffe der Fahraussetzung besprochen und es wird eine Bearbeitungsdefinition dargestellt.

Eine Probeschätzung für Aussetzung wurde durchgeführt, welche sich auf eine Zufallsprobe von 7145 Führerscheinfahrern im Lande gründete. Daten wurden mit dem Automatic Inter-action Detector (AID) Algorithmus analysiert, und es wurden Rangordnungen der besten Vorzeichen für Unfälle, Aussetzung und Unfallrate produziert. Das beste Anzeichen für Aussetzung war der Fall, ob eine Person zur Arbeit fährt oder nicht? Wenn die Veränder-lichen auf Vorkommen in Unfallberichten beschränkt sind, sind Geschlecht des Fahrers und Art des Fahrzeuges die besten Voranzeichen, danach kommt der Prozentsatz für Fahren in Straßen, Modelljahr des Fahrzeuges und Alter des Fahrers. Die beste Vorausbestimmung für Unfälle lag im Alter des Fahrers, danach das Geschlecht des Fahrers und andere Veränderliche in anderer als der Aussetzungsrangordnung.