Variability in Rural Accident Reporting

DANIEL S. TURNER AND EDWARD R. MANSFIELD

A research project was conducted for the Alabama Highway Department to assess the accident-reporting consistency of jurisdictions across the state. During the first phase of the research, a literature review was conducted, variables were selected for regression analysis, preliminary regression studies were begun, and a manual evaluation was conducted on the five-year accident pattern. Statistical investigations revealed strong relations between the number of accidents and several predictor variables for rural data and countywide data. The strongest single-variable model used population as the independent term and had an R^2 greater than 0.93. The strongest multiple-variable model used eight independent terms and had an R^2 greater than 0.99. Evaluation of the five-year accident pattern for individual Alabama cities and counties disclosed that 28 percent of them had erratic reporting trends. Significantly, county irregularities were not as severe as those for cities. Almost one-seventh of all cities had major discrepancies in the number of accidents reported over a five-year period. Traffic engineers and others performing safety studies must exercise care to ensure that study data reflect the character and quantity of accidents that actually occurred in the study area. Overall, the initial phase of the project was successful in documenting the existence of discrepancies in Alabama traffic accident data and in establishing strong regression relations between the number of accidents and predictor variables. Future phases of the project will address regression of city data, define reasons for discrepancies, and recommend improvements.

Historical accident data are a significant source of information used by engineers to establish safety programs and to implement safety countermeasures. These data are becoming increasingly important as safety programs receive more emphasis. This paper outlines a portion of a research project undertaken for the Alabama Highway Department to determine the consistency of the traffic accident data base.

PROBLEM STATEMENT AND RESEARCH PLAN

Reliability in the reporting of traffic accident data is obviously desirable but is not always present. In Alabama, individual jurisdictions are charged with investigating and reporting collisions. Both the quantity and quality of data appear to vary from location to location. This research was undertaken to assess the consistency of reporting, to devise a technique for predicting the number of accidents in a given location, and to identify jurisdictions where the number of accidents reported did not conform to the expected number.

The primary research technique was a regression analysis of the number of reported accidents for each Alabama city, followed by a confidence band analysis to isolate jurisdictions that did not do an adequate job of reporting. The initial portion of the project was directed toward determining the adequacy of the contemplated regression techniques.

This paper deals with the specific work steps in the first phase of the project--analysis of rural area reporting and of year-to-year consistency.

LITERATURE REVIEW

A literature review was conducted for several purposes:

1. To document the nature of the existing problem,
2. To identify previous research of a similar nature, and
3. To identify and designate variables for the statistical analysis.

The literature review is described in detail in a paper by Willis and others elsewhere in this Record. Three main reasons for consistency problems in accident data reporting were identified:

1. Variations in threshold accident reporting values (as Table 1 indicates, Alabama's threshold property-damage value of $50 or more is low in comparison with the criteria of other states),
2. Failure to investigate accidents properly, and
3. The secondary importance given to accident reporting in comparison with the many other duties of law enforcement personnel.

Table 1. Accident reporting thresholds by state.

State	Amount of Property Damage ($)	Other Criterion	State	Amount of Property Damage ($)	Other Criterion
Alabama	50		Montana	100	
Alaska	500		Nebraska	250	
Arizona	300		Nevada	250	
Arkansas	100		New Hampshire	300	
California		Injury	New Jersey	200	
Colorado		All	New Mexico	100	
Connecticut	250		New York		Injury
Delaware	250		North Carolina	200	
Florida	100		North Dakota	300	
Georgia	100		Ohio		All
Idaho	100		Oklahoma	100	
Illinois		All	Pennsylvania		Towaways
Indiana	200		South Carolina	100	
Iowa	250		South Dakota	250	
Kansas	200		Tennessee	200	
Kentucky		On request	Texas		Inoperable vehicle
Louisiana	100		Utah	200	
Maine	200		Vermont		All
Maryland		All	Virginia	100	
Massachusetts	200		Washington	300	
Michigan	200		West Virginia		All
Minnesota	100		Wisconsin	200	
Mississippi	50		Wyoming	250	
Missouri		Fatality			

Note: Data obtained from the International Association of Chiefs of Police. No information was available for the District of Columbia, Hawaii, Oregon, and Rhode Island.

SELECTION OF REGRESSION VARIABLES

During the initial portion of the investigation, a conference was conducted to select variables for the regression analysis. Representatives from the Accident Investigation and Surveillance Branch of the Alabama Highway Department and members of the project research staff prepared the list of items given in Table 2.

One of the primary considerations was the availability of the various data items. Specific variables were dismissed from further consideration if they were not readily available from conventional sources. The reason for dismissal was that, even if a variable was an excellent predictor, traffic engineers would refrain from using the prediction equation if it were not convenient to obtain data values for the variables.

It became apparent that many of the most desirable data items were not available for individual cities. However, these variables were applicable to counties and were found to be readily available. At this point, the data were placed in three categories: county, city, and rural. The county classification was used for variables applicable on a countywide basis, including both urban and rural areas. Examples include the number of vehicle registrations and the number of driver licenses.

The city classification was restricted to data applicable to incorporated cities in the state. One example is the census data used to establish the population. The final classification, rural, was used to handle data items that were only applicable to areas outside incorporated cities. For example, the rural population of a specific county would be the county population minus the population of all incorporated areas. Variables in each file are given in Table 3.

Table 2. Desirable data items for regression analysis.

Data Item	Description
Population	Census data
Paved highway	Miles of paved highway by classification
Land area	Square miles in city limits
Population density	Persons per square mile
Land use	Square miles by land use (urban, agricultural, etc.)
Employment activity	Number of jobs by category (manufacturing, etc.)
Gasoline tax	Allocation of state gasoline tax to cities and counties
Law enforcement	Number of law enforcement officers
Vehicle travel	Vehicle miles for each city
Vehicle registrations	Automobiles in a city
Driver licenses	Drivers in a city
Accidents	Number of traffic accidents

Data Strengths and Weaknesses

The obvious shortcoming of the city and rural data was the limited number of independent variables for use in the prediction equation. It is fortunate, however, that the population variable was in both sets. This was the single most desirable variable for use as a predictor because it was the easiest variable to obtain on a widespread basis.

The accident data associated with the rural data file had possessed a greater degree of reliability than the other two files. Most of these data were gathered by troopers from the Alabama Department of Public Safety, which has statewide programs for officer training and traffic accident investigation that should create a high degree of accuracy and uniformity in the reporting of accident data. On the other hand, the cities are subject to local policies and emphasis, so the quantity and quality of accident data can vary considerably from jurisdiction to jurisdiction within the state.

The county data set contained the greatest number of variables and thus offered the greatest opportunity to identify any accident prediction relations. A majority of the desirable variables, as given in Table 2, were found to be available in the county data file.

Variables Used in Analysis

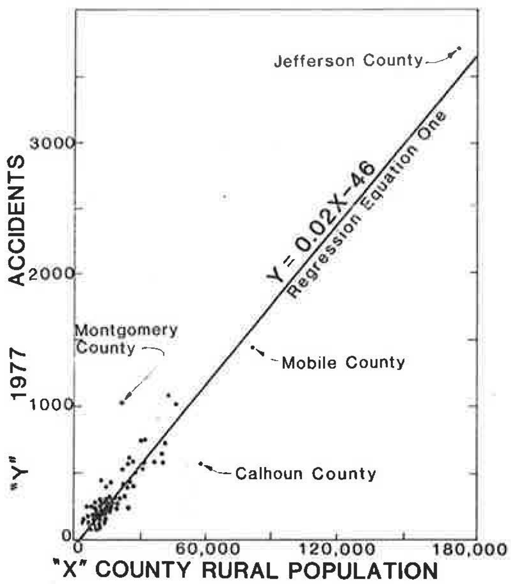
The final objective for the regression was to identify cities with irregular accident reporting characteristics. Unfortunately, the city data file was not as large or as versatile as desired for a rigorous statistical analysis. To overcome this, a procedure was formulated to use the strengths of all three data sets. The procedure consisted of the following steps:

1. The strongest accident data (the rural file) would be used in the initial investigation of the relation between accidents and population.
2. The most complete data set (county file) would be used to investigate a wide range of variables to determine the strongest possible prediction technique and to determine whether population alone was sufficient to use for prediction purposes.
3. A prediction method would be developed for

Table 3. Variables used in regression analysis.

	Data File			
Data Item	County	City	Rural	Source
Population				
1970-1977	X		X	Alabama Municipal Data Book, 1980 (12) and Information Bulletin, Directory of Mayors and Commissioners in Alabama (13)
1980	X	X		Alabama Municipal Journal (14)
Law enforcement officers				
Uniformed		X		Crime in Alabama: 1979 (15)
Civilian		X		Crime in Alabama: 1979 (15)
Traffic accidents, 1975-1979	X	X	X	Urban and Rural Accident Statistics (16)
Total miles of paved highway	X			Alabama County Data Book, 1980 (17)
Miles of state and federal route	X			Alabama County Data Book, 1980 (17)
Miles of Interstate highway	X			Alabama County Data Book, 1980 (17)
Miles of county road	X			Alabama County Data Book, 1980 (17)
Square miles of land area	X			Alabama County Data Book, 1980 (17)
Urban and rural land	X			Alabama County Data Book, 1980 (17)
Urban, agricultural, and other land	X			Alabama County Data Book, 1980 (17)
Automobile registrations, 1978	X			Alabama County Data Book, 1980 (17)
Driver licenses, 1978	X			Total licenses issued (renewals and new applications), 1977, 1978, 1979, and 1980, Alabama Department of Public Safety
Gasoline tax allocation, 1979	X			Gasoline tax distribution spread: Oct. 1, 1978-Sept. 30, 1979, Alabama Treasurer's Office

Figure 1. Rural data with regression equation.



the city data file. Population would be used as the independent variable if warranted by the findings of the first two steps. Should the initial step reveal that population was not a strong predictor, efforts would have to be renewed to find suitable variables.

The aim of the three-step investigative process was to determine whether the population variable was sufficient to use for predicting accidents, or if further variables would have to be added to the data set before proceeding with the research.

REGRESSION OF RURAL DATA

An objective of regression analysis is to fit an equation to the data that best explains the functional relation between the dependent variable and some set of regressor variables. The criterion for determining the appropriate equation is to minimize the sum of the squared differences between the actual observed value and the predicted value of the dependent variable. By observing the patterns of the individual differences for one or more models, it is possible to determine the transformations needed in order to better meet the necessary assumptions and to determine the appropriate model.

Regression Work Steps

The analysis of the data consisted of several steps. Initially, the data were edited in order to identify erratic or inconsistent observations. The second step consisted of determining the most complete model that best explained the functional relation between the number of accidents and various regressor variables that characterize the rural area. Finally a reduced model was sought that was less complex (or contained fewer regressor variables) than the complete model but did not excessively sacrifice predictive ability.

Step One

Erratic data points, called outliers, may occur for several reasons. Clerical errors could exist, a value of a variable may have been recorded or keypunched incorrectly, or the data source may have been in error. It is also possible that the area might have been atypical when compared to other areas of similar characteristics. During the regression analysis, outliers were identified by using various plots of residuals, such as the residual versus each regressor variable, the predicted Y-values, and the "deleted residuals" (18). In addition, statistics such as "Cook's distance measure," the "deleted residual t-statistic," and leverage factors were used for spotting outliers (19-21).

Step Two

The data were analyzed for variations of fundamental assumptions for regression analysis. These assumptions were that the residuals were normally and independently distributed, with a constant variance for each set of values of the regressor variables. In addition, if the regressor variable was multicollinear, alternative estimation procedures were examined.

Step Three

After the identification and correction of clerical errors, successive analyses were performed to determine the "best" equation to fit the data. A few unusual localities were removed to determine their effect on the remainder of the locations; then variables were systematically added, removed, or combined. During this entire phase of the analysis, several measures of effectiveness were examined and tabulated to identify the equation that best fit the data.

Step Four

The last phase of the analysis consisted of simplifying the equation that had previously been identified as best fitting the data. This involved removing those variables that made only marginal contributions to the success of the regression equation. The multiple work steps just described were not applicable in all cases. There were times when simpler methods were quite appropriate and the detailed analysis described here was not necessary.

Population as a Predictor

The initial step was to plot the variables to determine the presence or absence of patterns and to locate erroneous data points. The data were found to lie in a linear band with only a small amount of scatter, as shown in Figure 1. Four counties were identified as falling away from the rest of the data. Montgomery County had a higher than average number of accidents, and Calhoun had a lower than average number. Two locations, Mobile and Jefferson, fit the linear pattern but were displaced from the remainder of the data due to large rural populations. All four of the outlier points have been identified in the figure.

Because no errors were apparent and the data formed a linear pattern, a computer-assisted analysis was performed to determine the best linear equation. The least-squares regression technique was used to produce the following formula:

$$Y = 0.020X - 46 \quad (1)$$

where Y is the number of accidents and X is the population. The generalized formula has been superimposed on Figure 1 as regression Equation 1.

The two values used to indicate how effectively the regression equation fit the data were definitely quite strong. The value for R^2 was 0.92, which is exceptional for accident data, and the standard error was 138.6. Part of the strength of these measures was due to the displacement of Jefferson County. A single remote data point significantly influences the curve fit for a scattered group of points. Jefferson is the only data point on the right end of the curve, and its high numerical value exerted a large influence on the measures of effectiveness.

To determine the effects of the Jefferson County rural area, this data point was removed and the analysis was repeated. The resulting formula was similar to the initial regression results:

$$Y = 0.016X + 7 \tag{2}$$

where Y and X retain their former definitions. For regression Equation 2, the R^2 was 0.78 and the standard error was 117.5. Even though the R^2 was lower than the initial regression, it is still considered strong and could be used for predictive purposes. This was particularly true in this case because the standard error was even smaller.

The examination of rural population as an accident predictor was fruitful. Equations were developed that positively linked the two variables. Equation 2 was the most appropriate analysis tool and proved strong enough to be used for predictive purposes.

REGRESSION OF COUNTY DATA

Computer Runs

The second regression study was thorough, encompassing all of the variables given in Table 3. Other variables were formed by combining or factoring existing variables. The same regression techniques previously discussed were used for the county data. Many computer analyses were conducted, with variables added or deleted between runs, with and without outlier points. The results of each run guided the scope of the following run. A series of more than 25 refinements was conducted, and each refinement had multiple steps. The most prominent of the runs are discussed in the following paragraphs.

Results of the Analysis

Variables in the initial computer run included the calculated value of population density (population/land area) and its square to account for curvature. The fit was very good, and the predictor could be considered to be accurate. The R^2 was 0.9955 and the standard error was 285.

Even though the initial run was strong, attempts were made to improve the predictive equation. Several subsets of the variables were examined. The R^2 values did well for these runs, but the standard error tended to increase as variables were deleted. When population alone was used, the standard error increased to 459.

It became apparent that the largest counties were not being fit well. Several runs were made to determine the most effective method of handling the five counties with large populations. One county was not fit well for any run, whereas another dominated any equation in which it was included due to its large size. As a result, runs were made omitting the five largest counties. The resulting R^2 was 0.93 and the standard error was 209.

During additional runs, subsets of other variables were examined. Several cases were noted in which variables were interchangeable. For example, driver licenses and gasoline tax allocations provided the same explanatory power for the number of accidents; either could be used when coupled with roadway mileage and county area.

The R^2 values for the 25 runs were closely grouped from 0.9955 to 0.9257. The standard error ranged from 459 to 203, the smaller values being associated with the less complex models. All of these measures of effectiveness were quite strong.

The best single-variable predictor model was based on population, and the measures of effectiveness (R^2 = 0.9257 and standard error = 203) were quite strong. In the following equation, Y is the number of accidents and X is the county population:

$$Y = 0.033\,69X - 337.42 \tag{3}$$

In summary, it was shown that the number of accidents in a county could be estimated by using the characteristics of the county. In particular, population was found to be the best single-variable predictor available. The other variables helped the model, but the equation did well even without them.

EVALUATION OF FIVE-YEAR RECORD

During preparation of the data, several cases of erratic year-to-year reporting were noted. To evaluate data consistency, accident histories for each city and county rural area in the state were evaluated. The research technique involved a manual screening of accident records for the five most recent years. An analysis of coefficients of variation was conducted to verify the screening process. The objective was to identify those jurisdictions with erratic accident reporting patterns.

Classification Criteria

To quantify any discrepancies noted during the review of accident data, subjective criteria were formulated and placed in three categories. The various criteria for each category are as follows:

1. Category 1--(a) One unusual year in a five-year period, (b) a major change in the number of accidents, and (c) a minor change in the number of accidents;
2. Category 2--(a) A highly abnormal year in a five-year period, (b) two or more unusual years in a five-year period, (c) major changes in the number of accidents, (d) an accelerated decrease in the number of accidents, and (e) a city with few previous accidents suddenly reporting a significant number; and
3. Category 3--(a) A highly erratic accident pattern, (b) severe changes in the number of accidents, and (c) an obvious and drastic error in the reporting of the number of accidents.

Category 1 was reserved for the mildest types of erratic accident histories. In general, a city with some unusual occurrence in accident reporting would be placed in this category whether the city had any control over the erratic reporting or not. A city could receive a category 1 accident history rating due to one year with an unusually large number of accidents even though random chance rather than the city's reporting procedures caused the erratic pattern. Not all of the cities on the category 1 list could be termed deficient in reporting practices.

Category 2 applied to cities with more erratic accident histories than cities in the first category. Although it was possible that such deviations were the result of random chance, it was much more likely that improper reporting caused the problem. The criteria for category 2 in the list above indi-

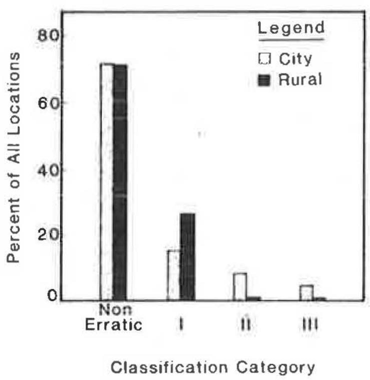
Table 4. Summary of Alabama locations with erratic accident histories.

	Category 1		Category 2		Category 3		All Categories	
Location	No.	Percent	No.	Percent	No.	Percent	No.	Percent
Cities	64	15.0	36	8.5	21	5.0	121	28.6
Rural areas	18	26.9	1	1.5	0	0.0	19	28.4

Table 5. Examples of erratic year-to-year accident reporting (category 3).

	No. of Accidents				
City	1975	1976	1977	1978	1979
A	1150	78	3	1481	1595
B	285	51	78	173	6
C	36	6	1	61	55
D	4	5	68	476	341
E	1	1	213	132	227

Figure 2. Percentage of Alabama cities with erratic accident histories.



cate more severe abnormalities than the criteria for category 1.

Category 3 cities experienced the most severe deviations and the most erratic patterns of accident reporting. The patterns were so unusual and pronounced that they were almost certainly due to variances in reporting practices. This type of pattern is obvious from the number of accidents occurring in consecutive years.

Results of the Evaluation

The 67 county rural areas and all 423 cities were subjected to manual review based on the criteria outlined above. Because the criteria were subjective, two independent reviews were conducted to offset any bias on the part of the reviewer. A summary of the findings is presented in Table 4, and a sample of highly erratic year-to-year reporting by some cities (category 3) is given in Table 5.

In Table 4, note that the city and rural data files had almost exactly the same percentage of erratic locations (28.6 percent for city and 28.4 percent for rural). This would seem to indicate that a certain amount of error may be associated with both files. It could also be interpreted to mean that the random nature of traffic accident occurrence exerted a powerful influence on all data. Further examination tends to discredit the second conclusion.

The erratic pattern for rural areas was concentrated in category 1. Only one location was rated as high as category 2. Thus, the deviations from uniform accident reporting could be considered mild for the rural data. City ratings were much different from those observed for rural locations (see Figure 2). A significant percentage received ratings worse than category 1. One out of every 20 cities exhibited the most severe variations and was placed in category 3. This indicates a very serious problem. Apparently, reported data for these cities do not realistically represent the number or the character of accidents that actually occurred. Safety studies that use such data could produce misleading results.

A significant conclusion can be drawn from the preceding analysis. Traffic engineers must exercise care when using Alabama accident data. Several years worth of accident data should be examined to ascertain the true accident pattern before performing a study on data for a single year at a specific location.

Locations of Erratic Data

The cities that were found to report data erratically are shown in Figure 3. The category 3 (most severe) locations are denoted by large, dark circles. There is not a clear overall pattern to the spots.

Category 2 cities are denoted by triangles and category 1 cities are shown by small squares. In several locations, these symbols are clustered or are spaced along a particular route; however, no pattern could be discerned. Several attempts were directed toward isolating a relation between location and severity of erratic reporting. These attempts were not successful.

The major conclusion that can be drawn from Figure 3 is that erratic reporting of accidents appears to be widespread in all categories. Further investigation would be necessary to determine whether the deviation at specific locations is due to fluctuations in traffic volume or local law enforcement policies.

Effect of City Size

A brief investigation was conducted to determine whether city size contributed to deficient reporting practices. The majority of Alabama cities are small: one-third have populations under 500; one-half have less than 1000. It would seem that the large number of small cities would contribute to erratic reporting.

Cities with erratic accident reporting and cities with nonerratic reporting were compared as to size. The two were virtually the same, which indicates that the cities with erratic reporting were a representative sample of all Alabama cities. No relation between accident reporting and city size could be established from the data gathered for this project.

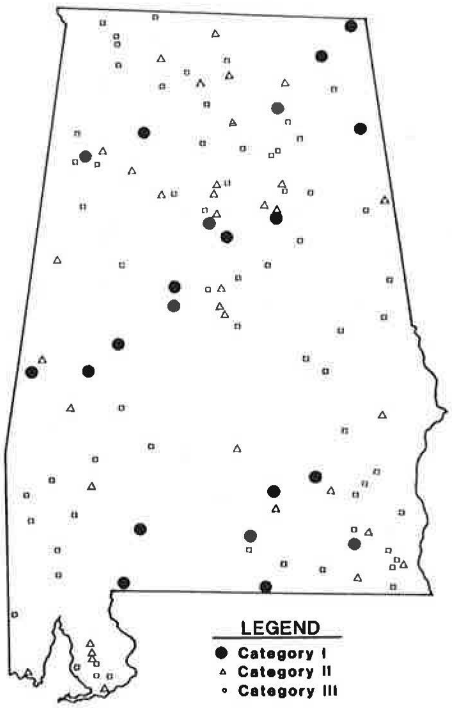
CONCLUSIONS DRAWN FROM YEAR-TO-YEAR REPORTING

Accident data for a five-year period were compiled for each Alabama city. Each city's record was examined for signs of irregular reporting by using a subjective scale to establish the degree of erratic behavior. The following conclusions were reached:

1. Approximately one-fourth of all Alabama cities displayed erratic patterns in the number of reported traffic accidents during the period 1975-1979. Five percent of all Alabama cities showed serious discrepancies in the number of accidents reported over the five-year period.

2. During the 1975-1979 period, 8.5 percent of

Figure 3. Locations of Alabama cities with erratic accident histories.



all Alabama cities showed serious accident reporting discrepancies.

3. Approximately one-fourth of Alabama county rural areas displayed erratic accident reporting patterns during the period.

4. Rural area accident reporting was less erratic than city accident reporting. Although the percentage of jurisdictions with erratic patterns was the same for both groups (28 percent), there were no severe discrepancies in the rural area classification.

5. Traffic engineers and others performing safety studies must be very careful in using accident data for a specific location because one-seventh (5 percent + 8.5 percent) of all Alabama cities have seriously erratic accident histories. It is recommended that several years of data be checked to ensure that data were reported uniformly and that they accurately represent a specific location.

6. Erratic reporting of traffic accidents does not seem to be strongly linked to the size or location of Alabama cities.

SUMMARY

The literature review indicated that many researchers have found bias and inconsistency in traffic accident data. None of the previous studies documented the consistency of reporting from location to location or from year to year.

Desirable variables were identified for use in the regression analysis. The majority of these variables were not readily available for Alabama cities, so three data files were developed: county, rural, and city. Because there were few variables in the city data set, the other two files were subjected to extensive analysis techniques. Strong relations were detected for both data files. Population was shown to be the best single-variable predictor model, giving excellent values for the measures of effectiveness.

Five years of accident data were screened to evaluate consistency in year-to-year reporting of accidents. About one-fourth of all Alabama cities and one-fourth of the rural areas were found to report erratically. The inconsistency in rural reporting was mild, but the reporting of at least one-seventh of all cities was seriously erratic. The reporting in these cities was so inconsistent that it might seriously bias the results of any safety studies in which it was used.

At the close of the initial phase of the project, there were two significant findings. Strong regression relations had been identified between population and traffic accidents. This indicated that statistical methods might be successful in identifying jurisdictions that do not do an adequate job of accident reporting. The second finding was conclusive proof of erratic year-to-year reporting. For the first time, the magnitude of the inconsistencies was documented.

Succeeding portions of the research project are to be directed toward regression of city data, identification of reasons for deficiencies, and recommendations for improving future reporting of accident data.

ACKNOWLEDGMENT

Funding for this project was provided by the Alabama Highway Department and the Office of Highway and Traffic Safety. We wish to thank those agencies, and especially Cecil Colsen and John McCarthy, for their technical input and constructive suggestions during the preparation of this paper. The assistance of John D. Antrim, C.O. Willis, Jr., Marcy Wentowski, and Nell Vice is also gratefully acknowledged.

The contents of this paper reflect our views, and we are responsible for the facts and the accuracy of the data presented. The contents do not necessarily reflect the official view or policies of the Alabama Highway Department or the Office of Highway and Traffic Safety. The paper does not constitute a standard, specification, or regulation.

REFERENCES

1. F.M. Council and others. Accident Research Manual. Highway Safety Research Center, Univ. of North Carolina, Chapel Hill, Rept. DOT-FH-11-9424, Feb. 1980.
2. K. Pfundt. Three Difficulties in Comparison of Accident Rates. Accident Analysis and Prevention, Vol. 1, No. 3, 1969.
3. E.G. House and others. How Complete Are Driver Records? An Analysis Based on Insurance Claim Crashes. Highway Safety Research Center, Univ. of North Carolina, Chapel Hill, 1974.
4. Helmets Cut Crash Odds for North Dakota Cyclists. Highway Loss Reduction Status Report, 1979.
5. J.K. Kihlberg and K.J. Tharp. Accident Rates as Related to Design of Rural Highways. NCHRP, Rept. 47, 1968.
6. J.P. Bull and B.J. Roberts. Road Accident Statistics: A Comparison of Police and Hospital Information. Accident Analysis and Prevention, Vol. 5, No. 1, 1973.
7. D. Mitchell and C. Lynn. Use of the Fatal Accident Reporting System (FARS) in Evaluating Characteristics of Virginia Traffic Fatalities. Proc., Human Factors Society, Santa Monica, CA, Oct. 1978.
8. J.P. Gerrard and W.W. Mosher. Analysis of Reliability of Accident Information Obtained from Off-Scene Sources: Part 1. Institute of

Transportation and Traffic Engineering, Univ. of California, Los Angeles, Rept. Rr-68-63, 1968.
9. R. Henson, C. Cannell, and S. Lawson. Effects of Interviewer's Style and Question Form on the Validity of Reporting Automobile Accident Information. Highway Safety Research Institute, Univ. of Michigan, Ann Arbor, HIT Lab Reports, Vol. 3, No. 12, Aug. 1973.
10. P.J. Cooper. Effects of Increased Enforcement at Urban Intersections on Driver Behavior and Safety. TRB, Transportation Research Record 540, 1975, pp. 13-22.
11. J. Byun, W.R. McShane, and E.J. Cantilli. Changing Baseline in Transportation Safety: An Assessment of Some Key Factors. TRB, Transportation Research Record 709, 1979, pp. 6-10.
12. Alabama Municipal Data Book, 1980. Alabama Office of State Planning and Federal Programs, Montgomery, 1981.
13. Information Bulletin, Director of Mayors and Commissioners in Alabama. Alabama League of Municipalities, Montgomery, Aug. 1980.
14. The Alabama Municipal Journal. Alabama League of Municipalities, Montgomery, Vol. 38, No. 8, Feb. 1981, pp. 14-18.
15. Crime in Alabama: 1979. Uniform Crime Reports Division, Alabama Criminal Justice Information Center, Montgomery, 1980.
16. Urban and Rural Accident Statistics. Accident Identification and Surveillance Section, Alabama Highway Department, Montgomery, 1975-1979 Computer Summaries.
17. Alabama County Data Book, 1980. Alabama Office of State Planning and Federal Programs, Montgomery, 1981.
18. R.F. Gunst and R.L. Mason. Regression Analysis and Its Applications. Marcel-Dekker, Inc., New York, 1980.
19. O.F. Dunn and V.A. Clark. Applied Statistics: Analysis of Variance and Regression. Wiley, New York, 1974.
20. S. Chatterjee and B. Price. Regression Analysis by Example. Wiley, New York, 1977.
21. E.R. Mansfield, R.F. Gunst, and J.W. Wester. An Analytic Variable Selection Technique for Principal Component Regression. Journal of the Royal Statistical Society, Series C, Vol. 26, 1977, pp. 34-40.

Publication of this paper sponsored by Committee on Traffic Records.

Application of Microcomputer Technology to Local Accident Problem Identification

DAVID B. BROWN AND CECIL W. COLSON

The potential for implementing a microcomputer-based problem identification system in small to medium-sized cities is explored in terms of the City Accidents RAPID Evaluation (CARE) system. The benefits of such a system are examined. One primary benefit is overall data improvement for all applications. The capabilities of CARE are explained in terms of a user-oriented menu-driven operating system. Example outputs are presented along with the methodology for their generation. Finally, some technical specifications are provided to illustrate considerations required for actual installation.

Problem identification is an essential part of the design of an optimal safety system at all levels (1). NHTSA has recognized the criticality of performing systematic problem identification and has incorporated this as a requirement for each state highway safety plan (HSP) (2,3). But problem identification is also essential at the local level for local countermeasure implementation. In fact, the closer to the source of the problem the process of problem identification and evaluation is performed, the more effective it will be. For example, if local law-enforcement officers knew the locations in their city where accident rates are high as well as the times and types of accidents at those locations, they would then be in a position to implement selective enforcement countermeasures.

The benefits of having a local problem identification capability are obvious. Being able to obtain information for specific problem subject areas (such as accidents involving alcohol or pedestrians) or specific locations gives the local decisionmaker the information needed to develop an optimal allocation of resources. A few years ago, it was not economically feasible to provide direct on-line query capability to a small town. Now, however, with the advances made in microtechnology, every city and town of any reasonable size can take advantage of the tools that have been developed. One of the major benefits of distributed problem identification will be the tremendous increase in accuracy of the records themselves as local police realize the important role that accident records can play in countermeasure development.

City Accidents RAPID Evaluation (CARE) (4) is a microcomputer-based system that enables city officials to quickly retrieve information stored in their accident records. The users of CARE need no formal training in computer hardware or software since no knowledge of computers is required. The various options of CARE are incorporated into menus that thoroughly guide the user in obtaining the desired output. By following the directions given on these menus, all output required can be readily obtained at the terminal and/or on a printer.

CARE is patterned after Records Analysis for Problem Identification and Definition (RAPID) (5), a system developed for statewide accident problem identification that has been installed in Alabama, Kentucky, South Carolina, Tennessee, and Delaware. The differences between CARE and RAPID are as follows:

1. CARE is designed to operate on its own dedicated hardware, a microcomputer in the $10 000-$15 000 price range, whereas RAPID requires a large system because of the large subsets necessary for statewide application.