part may not be a sufficient condition to good data, it is most certainly a necessary one.

POLICE ACCIDENT DATA: POSSIBLE SOLUTIONS TO SOME TROUBLESOME ISSUES

Charles V. Zegeer, Goodell-Grivas, Inc.

In continuing the discussion of police data problems, this paper will attempt to present some possible solutions that have been found in use in various states. The information used in the paper was taken primarily from research conducted for the National Academy of Sciences under NCHRP Project 20-5, "Use of Data Processing and Accident Location Systems for Highway Accident Analysis."

As a first step in understanding possible solutions to problems with police data, it is necessary to understand what data problems exist and to categorize these problems in a meaningful manner. For discussion purposes, the problems discussed in this paper will be categorized into the following four groups:

1. Location-related problems.
2. Problems associated with the data on police accident report forms.
3. Problems associated with developing and utilizing computerized data bases (accident data as well as roadway data).
4. Problems associated with conducting project evaluations.

Location-Related Problems

Accurate accident location is a key element in most highway research studies. Numerous reference methods exist and are used by various states and include the milepost method, reference post method, coordinate method, link-node method and others. Some of these methods, when used properly, can help solve many of the accuracy problems that have been cited above. Some agencies have made great strides in obtaining accurate locational information by investing the necessary time and efforts, such as working closely with police agencies, field posting of referencing signs and using detailed route logs and reference maps by office coders as done in California, to carefully locate and reference individual accident sites.

Locational accuracy is also being enhanced by the use of computerized highway networks, which are computer files containing the route names or numbers and linear distance information. An example of a successful computerized network is the Michigan Accident Location Index (MALI), which provides fast and accurate traffic accident information for all public roadways within the state. Many large cities use a Dual Independent Map Encoding/Geographic Base File (DIME/GBF) system, which was originally developed by the U.S. Bureau of the Census for coding census data, but has been applied to accident location, such as the system tested in Rhode Island. The file commonly consists of not only street names and segment lengths, but also x and y coordinates for each node, geographic area codes, block numbers, zip codes, addresses and other detailed information to enhance locational accuracy.

Problems Associated with Quality of the Data on the Forms

As stated in the preceding paper by Mak, one of the primary problems with research data is the quality of the data collected on the forms. For example, the standard data item "accident cause" is perhaps one of the worst data items that exist. It could be one of the more important in terms of accident causation studies. For example, in a study of rear-end accidents, one invariably finds that the accident cause is "following too closely," an obvious but not very enlightening finding. The problem that exists is that states generally collect too much "unused" data on their forms. A study of one state's accident form and related research indicated that it only uses about 7 percent of the data that are collected by the police in that state for highway safety or research purposes. To help solve this problem, the researcher should never ask the police to collect "all the data you could ever want," but rather should selectively pick which items will be used. By reducing the number of items collected, efforts can then be made to ensure better quality for those important data elements. In addition, where special data are needed, researchers could utilize supplemental data forms which can be put in place, used for short periods of time, and then removed from the data collection requirements.

Problems Associated with Computerized Data Bases

There is a growing need in every state to merge accident data, traffic data (volumes, speeds, etc.) and roadway data (geometrics, roadway obstacle data, etc.), which are often located in separate files. This merging process is important for two basic reasons. First, a computerized merge is needed since the researcher very often needs to be able to choose or select a limited number of specific data items from different files for use in a given analysis. Thus he or she only needs to "match" certain accident data items with selected characteristics items and, since the entire record is very seldom needed in any analysis, the length of the record makes it very unwieldy and inefficient. Second, the state often needs to be able to merge separate files to produce routine, periodic calculations of accident rates or other data summaries to be used in required reports.

While most states can merge data concerning the primary roadway system, very few have systems that can merge data related to secondary or local road systems. Several states have little or no capabilities to merge their computer accident file with their roadway or traffic file. Perhaps one of the more difficult and costly types of data to collect and extract from any file are inventory data related to specific highway "hardware" (i.e., bridges, poles sign posts, guardrail, etc.). Often the researcher not only needs to know the number of a specific type of hardware that is present per mile of roadway, but also needs such specifics as the distance of obstacles from the roadway, the obstacle type, whether the pole is breakaway or not, the type of breakaway, the type and condition of the crash cushion, etc.

To collect such roadway data, some states have gone to an on-the-road sampling system where the road is actually driven by a team of observers who make counts of various hazards and highway hardware that are present on the roadside. An alternative method that might save both time and money and would use an existing system would involve the use of the photologging systems that already exist in many agencies. The photologging system can be sampled and, while sitting in his office, the data collector can "drive" the section of roadway obtaining the data that are needed for the inventory. In summary, while computer merge problems are not easy to overcome, the fact that most states have overcome these to some extent indicates that computer software systems do exist to make this possible. Also, numerous statistical software packages are currently in use to provide a wide range of statistical analyses (i.e., SPSS, DART, RAPID, SAS and others).

Problems Associated with Evaluations

Any discussion of police data problems should end with a discussion of why the data that are collected are not used in better evaluations. Even with their problems, the existing data could be used to give researchers answers to many of the questions that we face. Review of the states' systems indicate that perhaps the main reason for the lack of good evaluations is the fact that evaluation is usually given a low priority by most highway agencies. In some cases, this low priority may stem from the fact that the engineer is not anxious to find out what projects have "failed" in terms of accident reduction. Second, there are problems with utilizing the data, since most states do not maintain a computerized data base that is readily suitable for performing evaluations. Currently, however, some states have developed or modified software packages that will allow for systematic and economical project evaluations using computerized systems. The Michigan system, for example, makes the use of the previously described before/after with comparison group design quite simple, in that characteristics data are stored by homogeneous sections and, after certain sections are treated, comparison sections can be drawn from the remaining untreated pool by the computer itself, matching on certain roadway and traffic variables.

In summary, better police data and better use of police data can result from careful planning on the part of the engineering researcher. For better location information, the researcher should look to better reference systems, to increased and improved training for the police officers and office data coding techniques and, possibly, to computerized highway networks. In relation to the problem of improving the quality of the data elements themselves, the researcher should understand what data items are actually required and then select the best available data bases to meet those needs. The collection of roadway data may be better achieved through photologging or other techniques, compared with costly field surveys. The researcher should always attempt to simplify and standardize the definitions used and insist on using valid statistical techniques and experimental designs. In any merged systems, extra attention must be placed on handling special locations, such as interchanges, bridges and gore areas. In addition, increased emphasis should be placed on improving edit programs such that the computer itself can detect some of the errors in the data. In terms of routine collection of accident and roadway data, the cost of collecting the data must be weighed against the benefits to be gained from it before the collection of any variable on a wide-scale basis. Finally, the researcher must decide how best to use the data that exists in answering the question at hand, and after the research is completed, it should be disseminated widely so that we can learn from each other's mistakes and successes.

POTENTIAL USE OF THE NATIONAL ACCIDENT SAMPLING SYSTEM (NASS) IN STUDYING ACCIDENTS INVOLVING ROADSIDE OBJECTS

Russell A. Smith, National Highway Traffic Safety Administration

In any discussion of the potential use of the NASS system in studying accidents involving roadside objects, one must have some understanding of the system itself in terms of the accidents that are to be sampled, the accident elements that are available and the data collection methods used. NASS itself is a sample of all police-reported accidents. It is designed to provide nationwide estimates of selected accident statistics and to support design and evaluation of safety countermeasures. The sample is designed to provide these national estimates; however, the sampling error may be unreliably large for certain estimates. When this occurs, special studies can help overcome small sample sizes and can help collect special data not otherwise collected.

The data are collected by small teams of trained persons (2-4), most of whom are college graduates. Data collected include information collected at the accident site, inspection of involved vehicles, interview of drivers and information from police reports, medical records of injured persons and driver history from license files. Photographs are taken by the team and are retained for future study. The data are subjected to an extensive review and edit procedure to ensure quality and consistency.

The sample design is based on a two-stage selection of accidents. The first stage involves the selection of "lists" of police reported accidents by choosing cities or counties among groups of 1280 such "lists." These lists are called Primary Sampling Units (PSU). Because we do not have a master national list of accidents, we assume that the relative size of these lists and, consequently, their probability of selection is proportional to the population in this associated county or city. Currently, 75 PSU's are planned for NASS including large central cities, suburban areas, cities, towns and rural areas. The second phase of the selection involves selection of accidents within the list of a city or county. This is done by random selection within groups or strata defined by accident type (pedestrian, truck, motorcycle and passenger car) and by injury severity (fatal, A-injury or minor injury and property damage). Each of these accident selections is based on a known probability of selection whose inverse is the "weighting factor" for the selected accidents. These weighting factors are used to expand the sample to national estimates. The selection of a given accident occurs approximately four days after the accident itself has occurred.

This scheme can be demonstrated to work well for certain phenomena including number of involved persons by level of injury, age, sex and vehicle type, etc.; number of involved vehicles by vehicle type and size; injury rates to occupants by crash type (percent of serious injury in head-on accidents).

The data are also useful in determining incidence of injury associated with seat belt use, helmut use, contact with vehicle components such as steering wheels or windshields, reported alcohol use, prior driving records described by convictions, and other factors. Because there is no associated exposure or "opportunity to crash," data available to compare with the accident data being collected, use of these data in studying accident prevention questions is somewhat limited. However, the data will be useful in studying such measures when joint efforts in NHTSA and FHWA to collect companion exposure data reach fruition.

Of interest to this group is the use of the