The main thrust of this presentation was to illustrate that DART is a tool that can be used efficiently with accident and related exposure data. The use of such a tool can upgrade the skills of the analyst and uncover areas of data collection that need improvement. The challenge is to broaden the awareness, acceptance, and use of tools such as DART for highway safety management.

USE OF RAPID FOR PROBLEM IDENTIFICATION AND EVALUATION

David B. Brown, Auburn University, Alabama

The RAPID (Records Analysis for Problem Identification and Definition) system is a user-oriented computer system specifically designed to process state accident data. As opposed to a standardized report generator, RAPID enables the user to specify informational needs. By responding to simple queries, the user is guided through the process, interacting directly with the computer. RAPID has two specialized software modules: (a) ACT, which automatically generates priorities among cities within population subgroupings for any user-defined subset of accidents, and (b) AIM, which finds high-accident concentration areas on the roadway for any user-specified accident type. RAPID also draws on the resources of the Statistical Package in the Social Sciences (SPSS), automatically furnishing the appropriate SPSS labels, codes, and ranges, as well as all control statements and format specifications.

No computer knowledge is required to use RAPID. Although RAPID uses a standard statistical package to produce output, the user does not need to understand how to assemble statistical control statements. This is handled entirely by the RAPID system.

RAPID has been installed for Alabama, South Carolina, Kentucky, Tennessee, and Delaware. It is a portable package that can be applied to any state's accident data as well as to other types of data where statistical processing is required. It is available on commercial time-sharing systems if state resources do not permit in-house installation.

RAPID provides the user with the following capabilities:

- Create a subset, including variables, of the master data base according to any logical specification. Subsets requested could include all pedestrian accidents, alcohol-related accidents in a given county or city, or motorcycle accidents between milepost 235 and 240 on Interstate 85.
- Obtain labeled univariant frequency distributions for the variables chosen to be included in subsets. The production of total statewide univariant distribution for all variables falls within this capability.
- Obtain labeled histograms of frequency distributions.
- Obtain fully labeled bivariant (crosstab) analyses for any of the subset variables.
- Perform up to eight levels of multivariant analyses for any subset produced. For example, in the three-level analysis, a crosstab of accident time of day by day of the week could be produced for all severity levels.
- Obtain a correlation table for all combinations of subject variables.
- Find high accident locations according to user-specified criteria. (Locations are specified by road codes and mileposts.) The interactive nature of this task enables the user to try any number of alternative criteria in order to obtain the number and type of high accident locations the user needs to work with.
- Obtain univariant distributions of any or all variables for the locations found to be high accident locations. (The same capability also exists for any other location specified by the user.) A separate report is produced for each high accident location and for each accident. This condenses the information for ease of review before location investigations.
- Obtain any of the reports specified by the above capabilities for any or all of the high accident locations.
- Create a logically restricted subset from any previously created subset. Because the user can obtain many different logical restrictions (for example, composite geographical areas) from a subset without rereading the master data base, which is usually stored on tape, computer time can be greatly reduced.
- Integrate demographic information and thereby establish priorities among political subdivisions for various accident types. For example, RAPID produces priority lists for cities by population grouping according to the number of motorcycle accidents divided by any one of several demographic indexes, such as population, miles driven, or number of registered motorcycles.
- Obtain further statistical analyses (RAPSTAT), including analysis of variance, breakdown analysis, regression analysis, scatter diagrams, and a variety of student's t-test options.
- Obtain accident report numbers for any subset of accident records so that hard copy for particular types of accidents can be retrieved.

The RAPID system can be best explained by tracing the data from the origin to the final output report.

When a pedestrian accident occurs, an officer in the field records the accident on a standardized form, which is sent to a central point for data entry. Along with thousands of other records, it becomes part of the state's accident records data base, which is generally stored on tape.

The accident records data base is generally not constructed with problem identification in mind. In fact, it contains virtually all of the "codable" elements from the accident records. Many of these are not required for problem identification work, and they are generally not in a form compatible with problem identification. For example, the pedestrian's actual age is probably coded on the tape, whereas age intervals (e.g., 0-4, 5-7, 8-9, 10-15, 16-21, etc.) would be much more useful for problem identification and cross tabulation. In addition, certain calculations and other data manipulation might be required to satisfy the requirements of problem identification. For these reasons, the data base must be cleaned up before it can be used for problem identification. This may be done once a year for the data base compiled from the previous year. The program that reformats and puts the data elements into their proper intervals is known as the BASE program.

The BASE program is then run to create a new tape, the RAPID master data base, which is totally compatible with RAPID formats and objectives. RAPID can work on any properly formatted data base. The arrangement, number, and type of variables are totally flexible and may be specified by the user during the development of the BASE program. Since the new, properly formatted tape is too large to generate statistical reports efficiently, a subset of the RAPID master data base can be created on highway speed direct access storage whenever processing is required. This subset can be either retained for repeated use (cataloged) or used immediately and deleted.

Any number of subsets of the RAPID master data base may be established. User commands specify the variables and the logic. For example, the pedestrian accident record might become part of subsets for a statewide pedestrian analysis and for analysis of all accidents in the city.

Residing in a small subset, the data are now ready for quick processing through any of the RAPID processing options—frequencies, histograms, crosstabs, multivariant analyses, correlation analysis—or through any of the RAPSTAT options. Data may also be processed through the other RAPID specialized software options.

If a user wants to do many logical restrictions without going back to the RAPID master data base stored on tape, he or she can logically restrict from a previously created RAPID disk subset by using slightly modified commands. The process is quicker than creating the first subset. The result is referenced in the RAPID documentation as a restricted subset, which is processed immediately by the system and then deleted. An unlimited number of these restricted subsets can be created and processed simultaneously from any given subset.

The philosophy under which RAPID was developed is quite simple: to free the user from all unnecessary operations without sacrificing computer efficiency. There are many trade-offs among user flexibility, computer efficiency, and simplicity. Quite often an overemphasis on one will lead to a critical sacrifice of the other. By understanding what is actually taking place within RAPID as well as the reason for the current RAPID design, the user can better understand and employ the full resources at his or her disposal.

ACCIDENT DATA: A LIMITED TOOL FOR EVALUATION

A. James McKnight, National Public Services Research Institute

Accident statistics used for highway safety program evaluation have been criticized as long as they have been compiled. The charges leveled against them are that they are neither representative nor comprehensive nor accurate. They are also inadequate; other data are needed before accurate conclusions may be drawn.

Accident statistics do not include all of the accidents that occur. They are not supposed to. Minimum thresholds of property damage and injury are used in all accident reporting systems to keep the system from being swamped with statistics on minor accidents that would be of no real benefit to the practitioner or scientist.

The problem is that a large number of the accidents that are supposed to be reported are not. Drivers surveyed on their accident experience almost invariably list more reportable accidents than are shown on their official records. Only a third of all insurance claims appears on state motor vehicle records, even though police are called to the scene about three-fourths of the time. Can countermeasures directed at a population of accidents be legitimately evaluated through a sample of those accidents?

Data from accident reports are not comprehensive. They are limited to the number of variables that can be used to describe the accident. Police reports are limited by the many other duties the police must perform at the accident scene. Drivers' reports are limited by the amount of information the police can request without losing a driver's cooperation. Information provided in accident reports is often inaccurate. Few police have enough training in accident reconstruction to determine what really happened. Those that have the training often lack the time necessary to gather and analyze the available data. Data sources are often unreliable. Most of the information concerning speeds and direction, for example, comes from the people involved. Driver reports, both those given orally to the police and those submitted in written form, are frequently distorted by misperception, inability to recall, and simple bias.

Data other than accident data are needed to evaluate the impact of highway safety programs. Other factors such as exposure or outside causes may be responsible for changes in the number or severity of accidents.

When the effect of these factors cannot be controlled experimentally by the way the program is conducted, they must be controlled statistically through the use of data that describe their nature and magnitude. However, vehicle and driver records are kept for on-line, operational use—not for compiling statistics.

From the criticism, it might seem that accident data were inadequate to assess the impacts of highway safety programs. Actually, accident data have proven sufficiently representative, complete, and accurate to provide some measure of the impact of highway safety programs on the real accident experience of people, vehicles, and roads.

The problem arises when, in the evaluation, impact is not found. The effects of most safety programs are marginal; only rarely does a safety program achieve results that could be called dramatic. As we move from changes in the vehicle and highway to changes in the way people drive, we are lucky to find countermeasures that make a difference of more than a few accidents per thousand drivers.

The smaller the impact, the more precise the measure must be. Of the many programs that have produced no discernible impact, a substantial share could have been shown to be cost-effective had a more precise measure of impact been used. The same is true where outside factors are involved. A true impact may be masked by differences in exposure and other accident-related factors that could be identified and controlled with better data.

Despite their shortcomings, accident data are the best available criteria for evaluating program impact. Accidents define safety; for administrators and legislators they are the most convincing evidence of impact. Accidents are also the only common denominator in comparing programs with different immediate objectives and are the criteria most readily expressed in the dollar terms needed for cost/benefit analyses.

The issue is not whether accidents are acceptable criteria for evaluating highway safety programs; rather, it is what can be done to improve their reliability as a measure of program impacts. Some suggestions include (a) limiting the data, (b) making better use of driver reports, (c) consolidating files, and (d) collecting exposure data.

LIMITING THE DATA

We need to recognize that the agencies we rely on have functions other than serving as data pipelines. We have to do a better job of accommodating our requests to what they are able to provide.

The job of the traffic police is to keep the street safe. In an accident, they must protect the accident scene, take care of the injured, and see that damaged vehicles are cleared away so that traffic can start moving again. Serving as an arm of research and evaluation is the least of their concerns, and their priorities are not going to change. If we want reliable accident information, we must accommodate police responsibilities, not add to them.

One way to do this is to reduce the amount of information requested. For evaluation, it is most important to know who, what, where, and how bad. If we can collect this information reliably, we can assess the involvement of the people, vehicles, and roads toward which our programs are directed. Other information, such as whether the sun was out, what direction cars were traveling, or where the