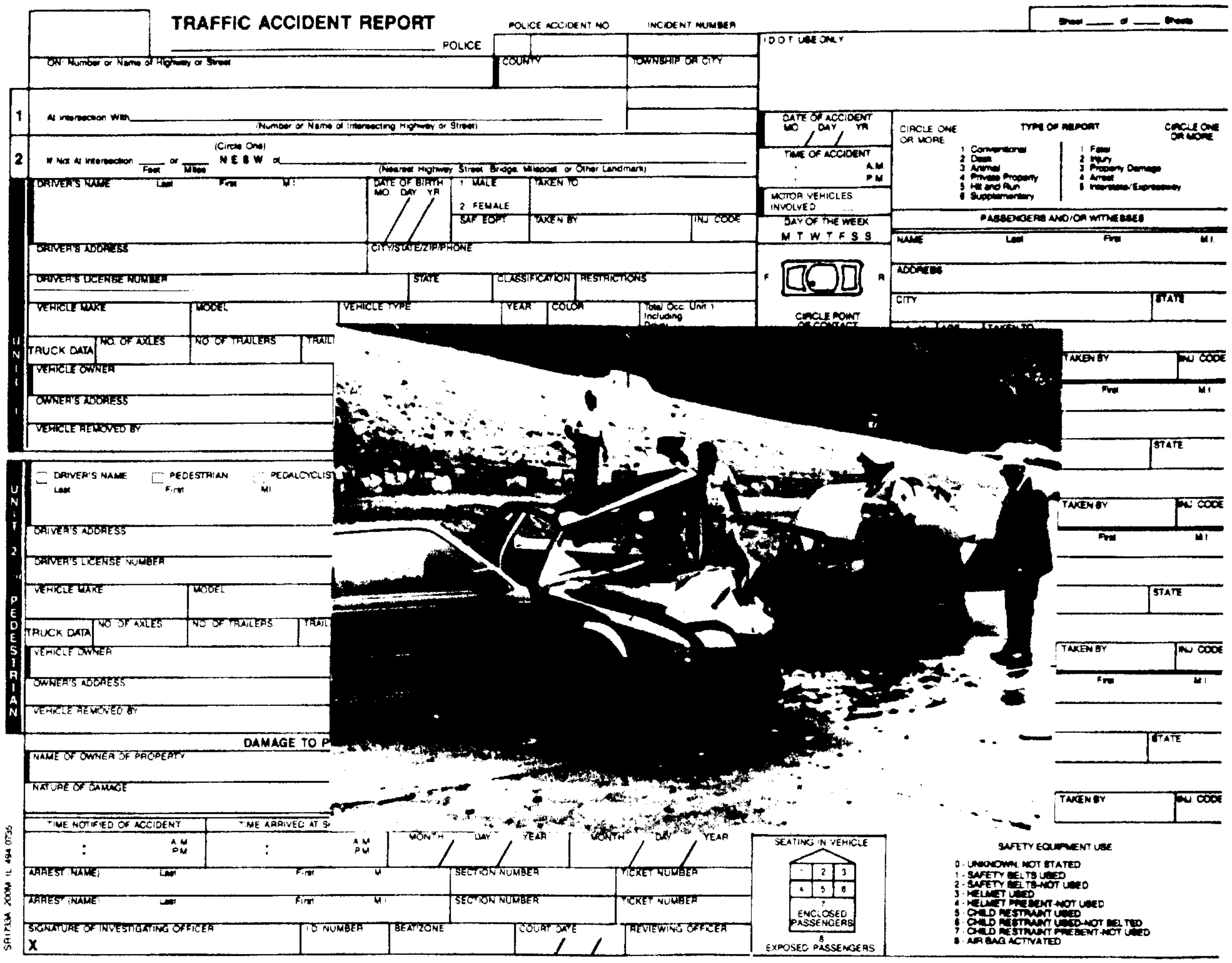
- Council, F.M., and Paniati, J.F. (1990). The Highway Safety Information System. *Public Roads*, (54), 234-240.



The Highway Safety Information System

by Forrest M. Council and Jeffrey F. Paniati

UNC/HSRC-90/12/1

Introduction

Highway engineers and administrators continually face decisions concerning highway safety—decisions ranging from the safety impacts of proposed programs or policies to the design of a section of highway. Often, decisions are made based primarily on engineering judgment or on accident data alone. Consistent, rational decisions, however, result when sound analysis methods are applied to good data bases. These data bases must include not only accident information, but also information concerning roadway geometrics at accident and non-accident locations, traffic volumes specific to vehicle types, intersection configurations and traffic controls present, and hardware and obstacles that appear on the roadsides. Data bases must be computerized and linkable to allow variables from several files to be combined rapidly and inexpensively.

The Need

Both Federal and State Governments need to measure the level of highway safety, although their specific emphases differ somewhat. At the Federal level, the Federal Highway Administration (FHWA) is concerned with formulating new policies and verifying existing ones concerning the impacts of roadway geometrics, hardware, and traffic guidance treatments on the safety of the Nation's highway system. At the State level, the major emphasis is on selecting alternative designs for construction, identifying hazardous locations for treatment, deciding among possible safety treatments, and making safety improvements on reconstruction projects.

Both Federal and State Governments generally measure safety in terms of accident frequency or rate as related to or caused by certain factors which can be modified by highway engineers. These factors include geometrics, traffic control, guidance systems, and roadside hardware. State and Federal Governments require quick access to high-quality accident, roadway, and traffic data.

Reprinted from: December 1990 • PUBLIC ROADS
Vol. 54, No. 3

To meet this need, many States have, for the past decade, been moving toward high-quality data in linkable computer files. Until recently, however, the FHWA had no such system available for its own use. Although certain national safety data bases exist, they are accident-based: that is, they provide detailed information on specific vehicles and drivers, but no information about the highway system and its characteristics.

What the FHWA needs is a location-based system which provides specific information on both failures (i.e., locations where accidents occur) and successes (i.e., accident-free locations). Without this information, it is virtually impossible to determine the factors resulting in success or failure—such differentiation is the essential nature of safety research.

A Solution

In 1983, the FHWA initiated a study with the Texas Transportation Institute to assess the ability of existing data bases to meet FHWA highway safety analysis needs. (*1*)[1] The study examined a wide range of data bases and found that no single data base could meet *all* highway safety data needs. However, it did conclude that integrated State accident, roadway design, and traffic volume data files could provide much of the needed information at a reasonable cost. Special studies could be used to collect additional data, not on State files, on an as-needed basis. Based on this concept, the FHWA and the University of North Carolina Highway Safety Research Center (HSRC) have acquired State data for a new highway safety data base—the Highway Safety Information System (HSIS).

The HSIS uses raw data already collected by a select group of States. These data undergo a series of quality control checks, are prepared in a standard format, and are merged. The merged data are then used in analyses.

The HSIS was not designed to combine the data from all participating States into one large data base. There is no common system of variable definitions applied across all States. There are differences in variable names for similar variables, and large differences in the category labels for the same variable. For example, an accident-type variable can have only a few basic categories in some States and as many as 20 different categories in others. Thus, to combine different States' data for the same variable would mean moving to a lowest common denominator definition in which a great amount of data information and specificity would be lost.

Consequently, the HSIS was developed to maintain the integrity of each State's mergeable files. This type of system provides several advantages for problem analysis:

- Each State's data set can be examined to determine which possesses the most appropriate data variables, categories, sample sizes, and linkages.
- State-by-State analyses can be done using the appropriate data for each State.
- The results can be compared across States to check for consistency and/or differences.

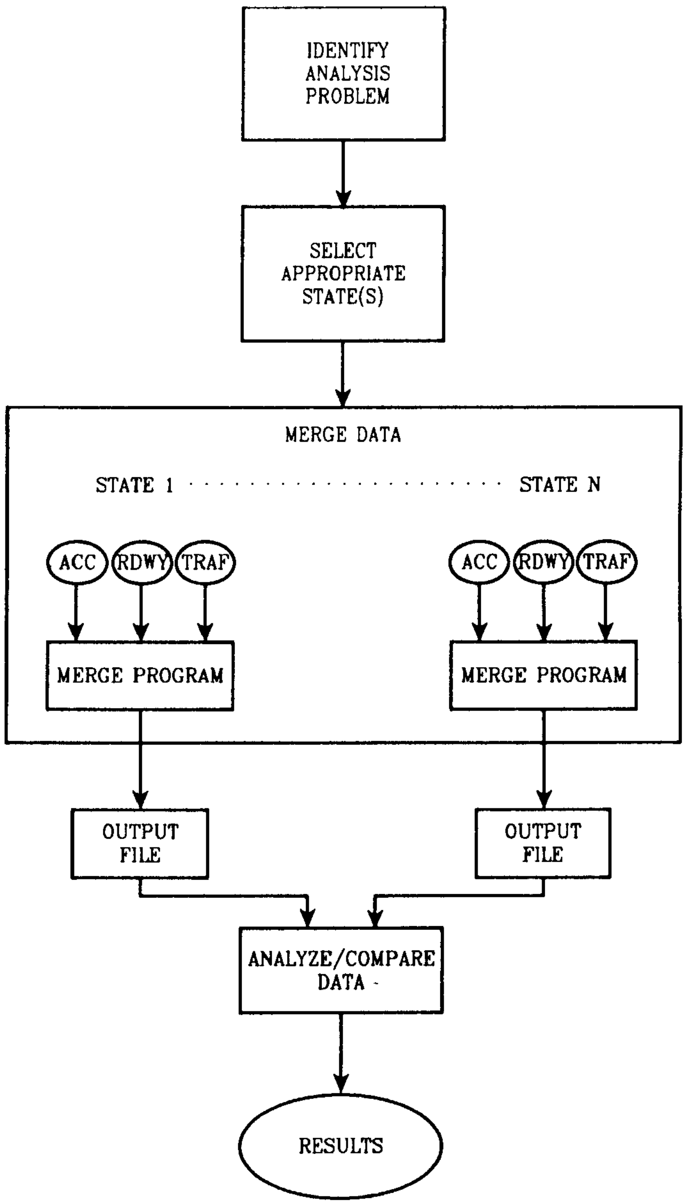


Figure 1.—Flowchart of the HSIS problem analysis process.

[1]Reference identified on page 240.

Consistent results could be generalized to other locations in the Nation. Differing results could lead to closer examination of why these differences exist. This in turn could provide more insight as to where research results would be most accurately applied. The results could be customized by location and/or related administrative guidelines or regulations issued for various areas of the country. Figure 1 provides a flow chart of the problem analysis process.

The Participating States

After selecting the system design, system developers devoted significant effort to the choice of the five States to be included in the HSIS. The selection process included a review of information on data quality and quantity for all 50 States, telephone interviews with several States, and site visits to the most promising ones. The primary criteria used in the selection process were the range, quantity, and quality of data variables collected.

Range of data variables

Working with FHWA staff, the HSRC developed a list of projected critical future analysis issues. This list was supplemented by input on safety research needs from safety researchers in the National Cooperative Highway Research Program, the Transportation Research Board, and other safety analysts. When completed, the list of critical variables was compared to variables available in the States.

Quantity of data variables

To assess the quantity of data available, the HSRC collected information from the potential States on the number of records in each of the files of interest, including:

- The number of accidents per year that could be linked with roadway inventory information.
- The number of miles of roadway included in the State's computerized roadway inventory system.
- The number of miles of roadway for which traffic count information was routinely collected, updated, and computerized.
- The sample size of data in any other special files (e.g., intersection or guardrail inventories).

Quality of data variables

Because a history of data use by the State and/or outside research agencies was felt to be a strong indicator of potential data quality, the HSRC project staff relied on each State's own assessment of its data quality as well as the project staff's knowledge of past research activities using the State's data. Significant weight was also placed on whether a State had developed *and used* a computerized data-merging system. Although several States attach some form of roadway information to accident data in order to identify high-hazard locations, only a few States have true data-merging systems. Such systems allow users to identify locations based on roadway inventory variables and then attach accident and traffic data to these locations. This full-merge capability ensures that a State's data files can be successfully linked. States that have invested the large amounts of funding needed to develop and maintain such a system would be expected to have used their data more often than States without this capability, and the more data are used, the higher the expected quality.

State selection

Eight States were selected for more detailed evaluation. This evaluation included careful study of data quality and quantity and the identification of problems with merging data. As shown on figure 2, Utah, Minnesota, Illinois, Michigan, and Maine were finally

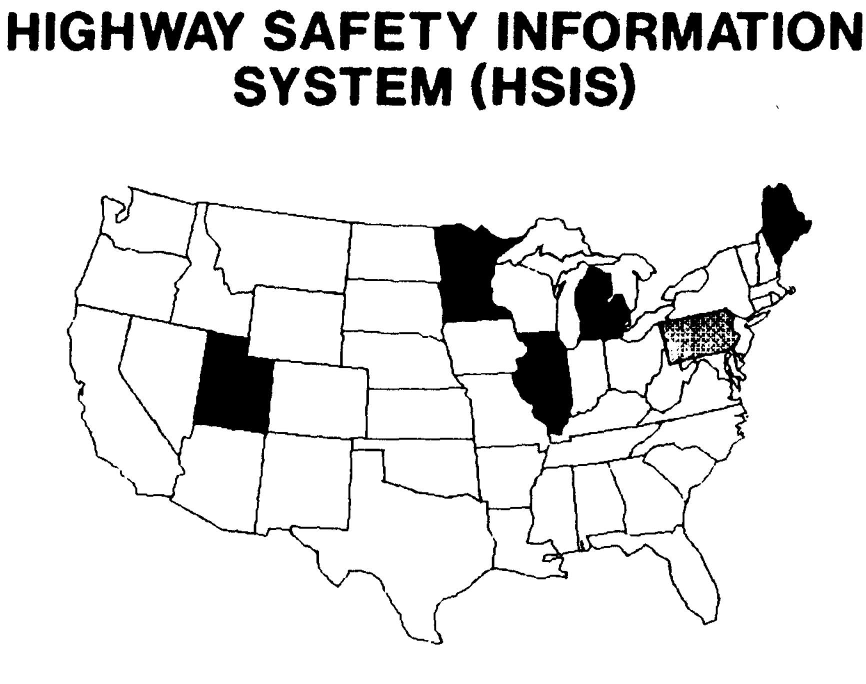


Figure 2.—Map of participating States.

Table 1.—HSIS data quantity

	Accidents/ year	Roadway mileage
Illinois	160,000	16,000
Maine	40,000	22,000
Michigan	145,000	10,000
Minnesota	70,000	60,000
Utah	50,000	50,000
Pennsylvania	150,000	

selected for inclusion in the prototype HSIS. In addition, the accident data for Pennsylvania are currently being captured by the FHWA for future inclusion in the system. Table 1 indicates the quantity of data available in the selected States. The table shows the number of police-reported mergeable accidents per year along with the number of miles of roadway to which these accidents can be linked.

The prototype HSIS was not designed to be a statistically valid sample; it is not nationally representative in terms of providing a random sample of all types of accident and roadway situations. The HSIS does not aim to provide national safety estimates—the National Accident Sampling System already serves this purpose. Rather, the HSIS will acquire quality data on a large number of variables, accident circumstances, and roadway locations for problem analysis.

The States currently in the system are not geographically spread across the United States; however, the roadway sections included do cover terrain types ranging from relatively level terrain with its inherent roadway geometrics to mountainous sections with critical curvature and roadsides. Thus, if an analysis requires certain terrain or types of locations, the analysis file can be restricted to those locations in the HSIS States that meet the specific criteria.

The Data Files

The primary files from each of these States include accident files, roadway inventory files, and traffic files. Certain States provide additional useful information as well. Table 2 shows the files from each of the States.

Data File Processing

So as to produce data files that could be easily manipulated by computer for problem analysis, it was decided early on to convert all raw data files to a Statistical Analysis System (SAS) format. SAS data formats identify each variable by a unique name, enabling the variable to be retrieved by name only rather than by position.

By providing these SAS formats for each variable, the output generated by the SAS program is easily readable. This means that the analyst can use the data directly from the computer without having to refer to a data dictionary for each of the variables being used. It is also much easier to program such operations as cross-tabulations, regression analyses, or frequency counts.

In addition to naming each variable, each of the variable categories was labeled with a brief 16 character description (SAS maximum) and an expanded description extracted from the States' documentation. These descriptions provide the analyst with more information about the variable of interest.

Once the files were formatted and labeled, a series of quality control checks was run for each of the files within each of the States. Single- and multi-variate tables were generated for the variables within each of the files. The HSRC analysts then examined each table to identify variables with unusual amounts of uncoded or unknown data, variables where the data were not consistent, and variables where two data elements that measured essentially the same parameter were inconsistent. When potential problems were found, the State liaison was contacted to determine if these

Table 2.—Files available from HSIS States

	Utah	Minnesota	Illinois	Maine	Michigan
Accident	X	X	X	X	X
Roadway inventory	X	X	X	X	X
Traffic volumes	X	X	X	X	X
Roadway geometrics	X		X		X
Intersection data	X	X			X
Guardrail					X

problems could be corrected or if they would need to be highlighted as a potential problem in future analyses.

The Guidebooks

The aforementioned data conversion, quality control, and consistency check all contributed to the development of HSIS data guidebooks. The detailed guidebooks will make the individual HSIS State files useful in future analysis efforts. The guidebooks provide enough information to allow both analysts and programmers to determine whether a specific analysis effort is possible. The guidebooks list all available variables and, for each, provide detailed definitions of each category, identify potential biases in the data, and supply information on available sample sizes. As future analyses are completed, the guidebooks will be modified to document solutions to some of the problems in the variables and/or highlight additional issues relevant to future research. The guidebooks are bound in looseleaf for easy updating and are organized into four sections as described below.

Section A. basic description

Section A of each guidebook provides an overall description of the individual State's data system and an overview of the types of data residing in each of the files. Details are noted concerning which variables should be used with caution in future analyses and which variables may be more appropriate than others for certain types of activities. The points presented in the section are then summarized and information on key State contact persons is given. For example, in the discussion of the accident file accuracy and coverage for one of the States, the text notes:

> . . .comparison of accident diagram with accident type revealed that the accident diagram variable provides the general nature of the accident without reference to what is involved. For example, for those accidents coded as head-on in the accident diagram variable (which one might assume means head-on with a second moving vehicle), 18 percent were coded as collision with fixed objects and 12 percent

```
ACCDIGM    ACCIDENT DIAGRAM            (SAS Format Name - ACCDIGF)

           01 ='REAR END'              Rear end
           02 ='SIDESWIPE PASSNG'      Sideswipe - Passing
           03 ='LEFT TURN'             Left turn into oncoming traffic
           04 ='RAN OFF RD LEFT'       Ran off road - Left side
           05 ='RIGHT ANGLE'           Right angle
           06 ='RIGHT TURN'            Right turn into cross-street traffic
           07 ='RAN OFF RD RGHT'       Ran off road - Right side
           08 ='HEAD ON'               Head on
           09 ='SIDESWIPE OPPOS'       Sideswipe - Opposing
           10 ='OTHER' Other
           98 ='NOT STATED'            Not stated
           99 ='UNKNOWN'               Unknown

           NOTE:  See discussion in writeup -- this variable does not indicate "what"
                  is struck, only "how" something is struck.  Also, large number of
                  records (20%) coded "Other".
```

```
ACCTYPE    ACCIDENT TYPE               (SAS Format Name - ACCTYPF)

           01 ='COLL OTH VEH'          Collision with other motor vehicle
           02 ='COL VEH OTH RDWY'      Collision with motor vehicle in other roadway
           03 ='COLL PRK VEH'          Collision with parked motor vehicle
           04 ='COLL TRAIN'            Collision with railroad train
           05 ='COL BICYCLIST'         Collision with bicyclist
           06 ='COLL PEDEST'           Collision with pedestrian
           07 ='COLL ANIMAL'           Collision with animal
           08 ='COLL FIXOBJ'           Collision with fixed object
           09 ='COLL OTH OBJ'          Collision with other object
           10 ='OVERTURN'              Overturn
           11 ='FIRE/EXPLOSION'        Fire or explosion
           12 ='SUBMERSION'            Submersion
           90 ='OTHER' Other
```

Figure 3.—Sample of the SAS format information included in the guidebook.

were coded as collision with parked vehicle in the accident type variable. . .. Clearly, if the analyst is interested in what is struck (e.g. another vehicle) in what fashion (e.g. head-on), then some combination of both variables should be used.

Section B. SAS formats

Section B includes SAS format names and category labels for each variable in individual State's files. The variable descriptions include notes regarding any potential consistency, coding, or quality problems. In this way, any problems that might produce biases are highlighted during the planning process before the initial computer runs are made. (See figure 3.)

Section C. single-variable tabulations

Section C of the guidebook provides single-variable tabulations for many key variables in each file. These tables show an estimate of available sample size and indicate data consistency across years. The tables will be updated each year as the new data come in and allow the FHWA to provide quick answers to routine questions. For example, the tables show how many accidents involving utility poles occur each year for every State in the data base and indicate the direction of change in frequency over time.

The tables can also be used to assess the adequacy of the sample size for a particular analysis. For example, an analyst might be asked to investigate accidents involving truck tractors with semitrailers. Figure 4 shows that truck tractors with semitrailers represent only 1.72 percent of the total number of accident-involved vehicles over the 3-year period in this State. However, there are an average of 2,300 of these vehicles in accidents each year, an adequate sample size for many analyses.

Section D. computer programs

Section D includes programs written to process and merge the variables as well as programs which combine files to calculate basic accident rates. These merging programs provide a framework which can be later modified to conduct similar analysis efforts.

TYPE_VEH (ACV-VEHIC-TYPE TYPE OF VEHICLE)
ACCYR (ACD-YEAR-OCCURRED)

FREQUENCY / PERCENT / ROW PCT / COL PCT	85	86	87	TOTAL
AUTOMOBILE	99951	95115	94335	289401
	24.61	23.42	23.23	71.26
	34.54	32.87	32.60	
	71.99	70.72	71.03	
AUTO WITH TRAIL	260	253	191	704
	0.06	0.06	0.05	0.17
	36.93	35.94	27.13	
	0.19	0.19	0.14	
TRUCK/TRK TRACT	3021	2660	2297	7978
	0.74	0.65	0.57	1.96
	37.87	33.34	28.79	
	2.18	1.98	1.73	
TRU/TRAT W/SEMI	***2514***	***2262***	***2193***	6969
	0.62	0.56	0.54	***1.72***
	36.07	32.46	31.47	
	1.81	1.68	1.65	
TRU/TRAT W/TWIN	35	30	45	110
	0.01	0.01	0.01	0.03
	31.82	27.27	40.91	
	0.03	0.02	0.03	
TRU/TRAT W/OTHER	392	329	326	1047
	0.10	0.08	0.08	0.26
	37.44	31.42	31.14	
	0.28	0.24	0.25	
PICKUP TRUCK	13239	14826	15272	43337
	3.26	3.65	3.76	10.67
	30.55	34.21	35.24	
	9.54	11.02	11.50	
TOTAL	138839	134499	132808	406146
	34.18	33.12	32.70	100.00

Figure 4.—Sample of single-variable table from the guidebook.

A Cooperative Effort

Key to the successful development of the HSIS was the continued cooperation of the participating States. The State liaisons provided the raw data files as well as consulting expertise and input on individual data variables when data quality issues were raised. The liaisons attended an initial workshop during which they advised the FHWA on how the system could best be developed using their data and what problems and issues needed to be overcome given the basic proposed design. Later, the State liaisons reviewed and commented on their own State's guidebook. They also took copies of the guidebooks back to their States for further review.

To this point, the FHWA has received the primary benefit from this cooperative effort—a usable, roadway-based safety data system. However, the project has also provided useful information to the participating States, including information on data inconsistencies and potential problems with individual data elements and SAS-formatted files and programs. Since some States are now moving to SAS formatting, they will use the HSIS SAS files and programs to reduce the amount of work needed. In the future, it is anticipated that these States will be key players in providing potential problem analysis to the FHWA. Any reports prepared from these data will be given to the States for their own use.

Conclusion

The HSIS provides the flexibility to analyze a large number of safety problems, ranging from basic problem identification issues to multivariate modeling efforts to predict future accidents from roadway characteristics and traffic factors. The HSIS is a major highway safety analysis tool for the FHWA and other highway analysts and researchers. A computerized multi-State data base of accident, roadway inventory, and traffic information, the HSIS is user friendly in terms of formatting and flexible in terms of the numerous ways in which its files can be manipulated and merged for a specific analysis problem. The HSIS will be an important companion to existing national data bases and will provide information and analysis capabilities not previously available. As with all data bases, however, its ultimate value will depend on the research and analysis in which it is used: the challenge now rests with the users.

Further discussion of the types of analyses which can be conducted with the HSIS will be included in the next issue of Public Roads. The article will cover results from completed HSIS analysis efforts, other potential applications for HSIS data, the status of the FHWA's HSIS demonstration project, potential enhancements to the system, and information on how interested users can access HSIS data.

Reference

(*1*) K.K. Mak and L.I. Griffin. *Assessment of Existing Data Bases for Highway Safety Analysis*, Publication No. FHWA/RD-85/117, Federal Highway Administration, Washington, DC, November 1985.

Forrest M. Council is deputy director of the University of North Carolina Highway Safety Research Center. He is currently the principal investigator on the study involving the maintenance, enhancement, and analytical use of the Highway Safety Information System (HSIS) for the Federal Highway Administration (FHWA). Since 1968, his work at the research center has involved roadway safety using State data files. He served as principal investigator for the project in which the HSIS was developed.

Jeffrey F. Paniati is a highway research engineer in the Informantion and Behavioral Systems Division, Office of Safety and Traffic Operations Research and Development, FHWA. Mr. Paniati is the manager of the Highway Safety Information System. He is a 1985 graduate of the FHWA Highway Engineer Training Program and is a professional engineer registered in the Commonwealth of Virginia.