APPENDIX H

ADEQUACY AND LIMITATIONS OF CURRENT DATA SYSTEMS

Marie D. Eldridge
NATIONAL HIGHWAY TRAFFIC SAFETY ADMINISTRATION

January 16, 1975

ADEQUACY AND LIMITATIONS OF CURRENT DATA SYSTEM

Remarks By
Marie D. Eldridge, Director
Office of Statistics and Analysis
National Highway Traffic Safety Administration

AUTOMOBILE COLLISION DATA WORKSHOP CONFERENCE
January 16, 1975 -Rosslyn, Virginia

In the very short time available to me this morning, I'm going to try to give YOU some highlights about the capabilities and limitations of our current data systems. I will also try to indicate where we see improvements within the near future.

However, before talking about the capabilities or the limitations, we really need to ask "capabilities or limitations for what?" So let's briefly talk about the objectives of our accident data systems.

First, we have to classify and count accidents. We need to determine the frequencies of accidents and classify them by their causal mechanisms, by their injury-producing potentials.

Second, we need adequate measurements of accident consequences, injuries, property damage or broader measures such as societal costs, a much neglected area and subject to great controversy but still one on which utlimately our decisions have to rest.

Third, we need to be able to describe, or model, crash injury mechanisms, that is, to relate the causal mechanisms and injury-producing potential to the actual occurrence of crash in-jury. This is particularly important in predicting the effects of proposed safety countermeasures. We have to describe functional relationships between numerous factors which at present are considered separately. All of these things that I have mentioned enter into the process of determining the efficacy and the benefits of existing or proposed safety measures.

Let us consider the criteria by which we should assess our crash data collection systems. It is my view that a comprehensive approach and a comprehensive consideration of all the data requirements that combine to give us the needed information is essential. It just won't do to get vey high accuracy in estimating speeds if at the same time the sample of accidents for which we obtain this information cannot he used to generalize and cannot provide us with the proper support for a rule that will apply to the whole country. High accuracy in one part of the data system can easily be nullified by weakness in another and, to quote an old saying there is no need to put a micrometer on the end of the yard stick.

View Graph II

I have listed in this view graph some of the criteria that we may use in assessing crash data systems. There are many ways of doing this but this may help provide a framework for discussing our present systems.

First of all, there is the quality of the data. We are concerned with its representativeness and in our *ability to generalize* from it to a national crash population. A sample that contains only new cars or only auto fleets is not representative. Frequently, we may have a situation in which sample populations as defined are representative, but in fact, because of missing data or non-returns, we don't get an unbiased sample.

A second criterion is accuracy of information. One of the reasons we are here today is the inaccuracy of certain information that we are now getting in crashes, namely the various speed parameters.

A third criterion is the ability of the system to be responsive and timely. The data need to be collected and processed quickly enough that the information is available before the decision has to be made. The sample sizes have to be large enough that we can have confidence in the decisions based on the results. At the same time we have to concern ourselves with costs arid make tradeoffs between costs and precision. Next there is the breadth or extensiveness of coverage of the information provided by the system in the many parts of our highway safety information matrix. And last but not least the cost efficiency.

View Graph III

If we had a great deal of time we could consider all this at the data item level or individual field level, but even to cover this matrix in any detail will have to be left for possible discussion later in the conference. I will simply mention that under exposure items we have the characteristics of the vehicle occupants and the amounts of driving by various driver types, their characteristics, licensing, training and so on. We have under vehicle exposure the counts of vehicles by type, travel amounts, their conditions, size, etc. The environmental exposure includes such things as traffic density, speed limits, highway types, design and so on. We could go down this matrix cell by cell and fill in the types of things that need to be considered.

The final and very difficult quality has to do with the cost-efficiency of the data systems. When a decision involves a high cost or an extreme inconvenience, a great deal of effort will generally have to go into the data collection and analysis. However, we also wish to keep our data collection efficient in the sense of not collecting information for which there is no need or employing personnel or equipment more skilled or more accurate than is really necessary.

IV. Now let us turn to the capabilities of some of our current data collection systems. Basically, we have two types of systems. The first is based primarily on the state or local traffic and related records systems. The second type involves special investigative work. The state records are kept primarily for purposes other than safety analysis. However, we utilize their records for the Fatal Accident Reporting System, which is essentially a census or 100 percent sample of fatal motor vehicle accidents and for the planned National Accident Reporting System, which will be a probability sample of all accidents, of a given threshold. The accuracy of the information provided through the State traffic record systems varies of course. In some areas of particular interest to us it is quite poor. Speed causal factors and restraint systems usage, for example, may be misreported or unreported frequently. Timeliness is generally not a problem. It usually takes only a few months before an accident is in the file and therefore accessible to us. As far as the quantity of information is conerned, the State Traffic Record Systems are likely to provide us with a large number of cases for the more frequent types of accidents and the items of highest interest, but it is surprising how often in other circumstances we run out of data. The most obvious example is in making comparisons between makes and models of cars. When we get to some types of vehicles that are not on the road in large numbers, we have a very hard time collecting enough accidents to have a useful sample. The breadth of the information provided is generally not adequate. Impact speed for example is reported only in one State; traveling speed in about half the States and not for all accidents even in those States. Restraint system usage is not reported in most States and in many where it is reported, it is not reported for uninjured occupants. In-jury information and causal factors are sketchy. Post crash information, societal cost and property damage are usually not in the file.

It has been generally recognized that we can not obtain adequate information to support the standards by relying solely on these basic records oriented data systems. The second type of accident data collection system - those in which specific data collection efforts are sponsored or paid for by either the Federal Government or some other interested organization in the safety field such as MVMA or the Insurance Institute for Highway Safety. In these systems the investigation is likely to be carried out wholly or in part by professional accident investigators, resulting in substantially more extensive information. NHTSA has under way three types of sponsored studies.

First is the Multidisciplinary Accident Investigation teams. These teams do both on-scene and off-scene in-depth investigations. Teams have been performing clinical in-depth studies of selected accidents in the U.S., primarily on new cars, since 1969. The representativeness of the sample that has been produced up to this time is poor. Different teams have been covering accidents most relevant to their special interest. That situation is gradually changing. The accuracy is generally good. Nevertheless, there is considerable room for improvement. We have no capability for getting a time history of the crash forces and accompanying accelerations except through computer simulation such as the SMAC program. At present we have about 6,000 MDAI cases in the file. Many of these were not the result of on-scene investigation. There is detail on most aspects of the accident with the exception of exposure. As a system for producing statistical information needed for supporting our safety standards, the on-scene in-depth investigations cannot be regarded as cost effective. The average cost per case is about $2,000. The cost decreases to about $800 per in-depth case if the on-scene investigation requirement is Eliminated. This does reduce the accuracy of reconstruction of the accident and of course affects the estimate of speed.

At a somewhat lower level of detail NHTSA has developed a system in conjunction with MVNA to collect a probability sample of towaway involvements of new cars in five selected regions of the country primarily for the purpose of evaluating active and passive restraint systems.

Data are assembled from the police report, a doctor's report, photographs, a brief vehicle investigation, and driver interviews performed by field technicians. Data items collected are restricted to those needed for the statistical analysis of restraint systems effectiveness. This is an example of what we may term a Level II study. We expect to make national estimates based on post stratification. The accuracy of the information in the selected data items' should be good, nearly as good as what comes from the multi-disciplinary in-depth investigations. The quantity will be adequate to match the needs for estimating safety belt effectiveness. Because of the small numbers it is not likely to give us what we need for estimating air cushion effectiveness, very soon. As far as the breadth of the file is concerned, it is designed for calculation of crash injury rates and evaluation of restraint systems effectiveness. It does not address exposure or accident causation. Speeds and occupant contact points are not determined. The cost is around $100 per case.

A third type of sponsored system is basically a bilevel investigation or one in which there is a supplementary investigation carried out by police with NHTSA or other funds added to take care of added costs. We have under development a system for sampling pedestrian and bicyclist accidents in several hundred localities. The system is designed to answer questions at the level of detail that we nced to determine gross behavior and provide some good input for counter-measures. The data to be collected is primarily concerned with the nature and location of pedestrian and cyclist accidents as well as certain other items affecting visibility which would not normally be collected in the state accident reporting system. The cost per case is expected to be high primarily because of the relative rarity of pedestrian and bicyclist accidents and because in order to get an adequate probability sample that will properly represent rural areas, it is necessary to include localities with a very low frequency of accidents. The set-up time in preparing to get the supplemental investigations done in small localities is the same as it is in large localities, but the data rate is low and the total cost is increased disproportionately.

V. As we look ahead to potential improvement in the capabilities of our current systems that may be in sight we are really moving in two directions. The first is to create a national accident sampling system based on a probability sample. We have a contract under way with the Highway Safety Research Institute at Ann Arbor to help develop this system that will include some of the current investigative efforts but provide for sufficient. control of the selection of accidents that we will get a sample from which we can generalize to national crash populations.

The second area in which we anticipate improvements is in determining crash dynamics. These efforts, are of courses of paramount importance to this workshop. The work with the crash recorder is being covered by my colleague, Lynn Bradford. The other approach, SMAC, the computer simulation of the accident dynamics will be dealt with by our representatives from Calspan but I would like to say a few words about our experience with it.

This program uses vehicle rest positions and impact damage to calculate impact velocities, the velocity change during the crash, acceleration pulse and predicted damage. The goal is to reconstruct the accident crash dynamics in sufficient detail that inputs needed by our standards makers are available. The use of the SMAC program may permit us to get, at a reasonable cost, an adequate representative sample of crashes once our national accident sampling program is up and running. However, it should be pointed out that the crash recorder is clearly a very valuable tool in developing necessary refinements to the SMAC program. Ideally, and this is a moderate size "if", IF the crash recorder and the SMAC work hand in hand well enough, we can succeed in reducing considerably the numbers of crash recorders required down stream. Precise calibration of the SMAC program will enable us to use Level 2 data for crash dynamics at a reasonable cost. Currently the cost per case, using the SMAC program is $150.

In the short time available to me I have had to gloss over lightly much of the work related to crash data. Three members of my staff are here to provide detailed back-up and to join in any subsequent discussion of these points. They are Don Mela, Dr. Charles Kahane and Dr. Charles Moffatt. Before finishing these very brief remarks, I want to repeat a point I made earlier. We need to consider all relevant aspects of the data systems in a comprehensive fashion before making decisions on any of them separately. While we may not be able to devote that amount of detail to all aspects of the data systems in this conference I think that at least the major aspects should be considered before coming to any conclusions or decisions.

OBJECTIVES OF ACCIDENT DATA SYSTEMS

CLASSIFY AND COUNT ACCIDENTS

MEASURE ACCIDENT CONSEQUENCES

DESCRIBE CRASH INJURY MECHANISMS

CRASH DATA SYSTEMS CRITERIA

QUALITY

- Representativeness
- Accuracy
- Recency
- Responsiveness to Ad Hoc Demands

QUANTITY

- Precision
- Costs

BREADTH

COST-EFFICIENCY

DATA BREADTH

	HUMAN	VEHICLE	ENVIRONMENT
EXPOSURE			
CAUSAL FACTORS			
CRASH EVENTS			
CONSEQUENCES			

CURRENT DATA COLLECTION SYSTEMS

A. STATE/LOCAL RECORDS SYSTEMS

1. FATAL ACCIDENT REPORTING SYSTEM (FARS)
2. NATIONAL ACCIDENT REPORTING SYSTEM (NAF

B. INVESTIGATIVE FILES - SPONSORED STUDIES

1. MULTIDISCIPLINARY ACCIDENT INVESTIGATION (MDAI)
2. RESTRAINT SYSTEMS STUDY
3. BI-LEVEL STUDIES

NEW AREAS

1. NATIONAL ACCIDENT SAMPLING SYSTEM
2. MEASUREMENT OF ACCIDENT DYNAMICS