

December 19, 2002

Ms. Annette M. Sandberg  
Deputy Administrator  
Federal Motor Carrier Safety Administration  
Room 8202  
400 7th Street, SW  
Washington, DC 20590

Dear Ms. Sandberg:

The Committee for Review of the Federal Motor Carrier Safety Administration's Truck Crash Causation Study held its fourth meeting on August 20–21, 2002, in Washington, D.C. The enclosed meeting roster lists the members, government staff, guests, and Transportation Research Board staff in attendance.

TRB formed the committee in 2000 at the request of FMCSA to provide advice on study methods. This is the committee's fourth letter report. The others were submitted on November 15, 2000, March 9, 2001, and December 4, 2001.

On behalf of the committee, I thank the staff members of FMCSA and the National Highway Traffic Safety Administration for their presentations and responses to committee questions. The committee believes that this continuing exchange of views and ideas will help the project achieve its objectives.

The committee recognizes that the Truck Crash Causation Study is potentially of great importance to highway safety. It is the first study to conduct on-scene investigations of a large and nationally representative sample of truck crashes and the first to employ crash investigators as well as truck safety inspectors for data collection. The goal of FMCSA and NHTSA is to ensure that this study is a landmark in understanding the causes of truck crashes and that it leads to actions that reduce the number and costs of truck crashes. The committee's advice is intended to contribute to the achievement of this goal.

At the latest meeting, the committee reviewed questions arising from a recent review of crash files that had been conducted by a task force of five committee members (John Billing, Michael Belzer, Anne McCartt, James McKnight, and Frank Wilson). The task force visited FMCSA in

Washington in July to review crash case files and reported on their observations at the meeting. In addition, the meeting included presentations and discussions on FMCSA's plans for data analysis in preparation for its report to Congress.

The committee's discussions are summarized in this letter, and several recommendations to FMCSA are presented. Our comments and recommendations are in three areas: requirements for presenting the data in public access files, data quality, and analysis planning. A section summarizing all of our recommendations follows the sections on these three topics.

DOT may have already undertaken some of the recommended activities. Nonetheless, to ensure that they are not overlooked, we identify below all the actions that we believe are urgently needed.

## **DATA PRESENTATION**

The task force report and past committee discussions have identified several areas for attention that relate to the organization, content, and documentation of the public database that will be released at the conclusion of the study. The task force reviewed the database in its most complete form, as it will be maintained by FMCSA as the primary record of the investigations. Public versions will have different content. The committee's recommendations in this area are presented below. Some of these amplify recommendations in our earlier letters, in particular, our recommendations regarding presentation of information from interview responses on pages 5 and 7 of our December 4, 2001 letter and our recommendation regarding the importance of complete and accessible documentation on page 5 of our March 9, 2001 letter.

### **Documentation and Other User Aids**

The task force's experience highlighted the inevitable difficulty of using such a complex database and the dangers of misunderstanding and misinterpretation. Public data files derived from the study must be accompanied by complete, carefully prepared, and user-friendly documentation. Developing these materials will be a demanding task. In addition, readily available assistance to users from competent DOT staff will be necessary. At the meeting, NHTSA staff indicated that they are aware of the importance of this task and pointed out that they have long experience in providing support for public research databases (for example, the FARS database of fatal highway crashes). Nonetheless, it is not too early for FMCSA to begin planning the documentation and support arrangements and considering the costs of these activities.

### **Recording of Sources**

Documentation must make unambiguously clear to users the source of each item in the database. The source of an item that is the same in all cases can be explained in manuals, but sources that can vary from case to case must be identified in the record of each case. Elements of the assessments contained in each crash record (e.g., critical event, critical reason, and related

events) in general should be documented by identifying their sources and by narrative explanation of the analyst's reasoning.

Lack of clear identification of sources of information would lead to problems for future users of the database. As one example, if speed is identified as a factor in a crash, it should be clear to the user whether this conclusion was supported by direct evidence of speed (e.g., from measurement of skid marks or statements of witnesses) or whether speed was implicated solely as a deduction from the circumstances of the crash (e.g., coding speed as a factor because a crash involved a rollover on an exit ramp).

### **Case Summary and Point of Entry**

The task force used the crash summary narratives written by the crash investigator as the "point of entry" to the database, that is, the first data element scrutinized to determine whether a case was of interest. NHTSA staff explained at the meeting that these narratives, prepared by the primary investigator in the field, were not intended to serve as the point of entry and that other data elements will be more appropriate for this purpose in the completed database.

The crash summaries that the task force reviewed varied greatly in content, were sometimes not carefully written, and sometimes contained apparent errors. Since we believe that users of the data will naturally refer to these summaries if they are in the public file, we recommend that, if they are to be made public, FMCSA and NHTSA training and oversight of field staff ensure that the summaries follow a standard format and are reasonably accurate, complete, and comprehensible. Examples provided in NHTSA's field coding manual are excellent and could be used by field investigators as models.

The Crash Assessment Form is the key record of analysts' conclusions regarding critical reasons and related factors for each case. These assessments probably will be central to addressing the congressional mandate to examine causation. The committee suggests that the Crash Assessment Form be expanded somewhat beyond the format and content in the cases that the task force reviewed. The form should provide a detailed narrative identifying all facts used in the assessment and include the basis for the decisions made concerning the critical event and reason. In particular, it should explain judgments made by the assessor when sources of information conflict. Again, the examples provided in the coding manual are good models.

Project staff presented to the committee a description of a new "Overview Form," a one-page summary form under development, which is to be part of the computerized database and which would allow the user to easily ascertain the essential aspects of each crash. We believe that this is a very promising step toward facilitating use of the database. We do not believe that we need contribute to the detailed development of definitions of information to be presented in this form, since it should evolve in accordance with the needs of users of the database, but we would be happy to comment on the form as work on it nears completion if requested to do so.

### **Presentation of Information from Interview Responses**

In the master database, certain data fields contain unedited interview responses. We understand that if the investigator concludes on the basis of other information that an interview response is

untruthful, this judgment is noted (with a “flag”) in the data file. We also understand that certain information from interviews obtained with the assurance of confidentiality will not be placed in the public file.

If the public file does include any unedited interview responses (that is, if confidentiality restrictions do not exclude all such data), then the source of such data elements must be clear to users, documentation should note possibilities for substantiating interview responses by comparison with other data fields, and DOT should consider retaining in the public database the flags marking implausible responses that appear in the master file.

The committee believes that FMCSA and NHTSA should take advantage of all opportunities for obtaining information from public sources (e.g., public police or court records or other independent sources) to substantiate interview information, in order to minimize the impact of exclusion of confidential interview responses on the utility of the database.

To aid committee discussion of this issue, we request that FMCSA provide information before the next meeting concerning the specific data items that will be included in the public version of the file.

## **DATA QUALITY**

In its second review of the data in July 2002, the five-person task force examined the five crash cases that it initially reviewed in July 2001 and approximately 25 other completed cases. The 30 cases are all early ones, mostly among the first half-dozen conducted by each of the investigative teams. They were provided to the task force because later case files were not yet complete. It is reasonable to expect that the proficiency of the investigators will increase markedly with experience and that later cases will be much more complete and accurate than the first ones. Also, discussions at the meeting suggested that some of the impressions the task force formed of the case data may have derived from misunderstandings of the definitions of data elements or of the structure of the database. However, in the cases the task force reviewed, the apparent frequency of missing and miscoded data is a source of concern, especially since these cases have been completed and have undergone quality-control checks.

Missing data and other data quality problems could seriously diminish the value of the study if they are not diagnosed and addressed as soon as possible. Therefore, the committee believes it is essential that FMCSA and NHTSA begin a systematic and quantitative analysis of rates of missing data, data quality, and the causes of missing data and coding errors and take action to reduce these problems wherever they are found to be significant. The committee recommends the following actions:

1. Automated edit checks should be employed that compare data elements for consistency. For example, information related to truck configuration is coded in several fields in the database. These fields should be compared with each other, and cases containing apparent inconsistencies should be flagged for further examination. NHTSA staff reported that the

data entry process incorporates hundreds of such checks; however, the task force came across some apparent instances of lack of cross-checks.

2. Manual edit checks should be systematized. The full case files contain much uncoded information, including photographs of the vehicles, interview transcripts, and police accident reports. NHTSA and FMCSA staff reported at the meeting that these materials are now used in editing. Such manual checks should be conducted wherever there is reason to believe that they could contribute to data quality. They should be performed systematically, according to written protocols, and records kept of the checks performed and resulting changes to coding. For example, a checklist of comparisons of the photographs with the coded data could be developed, so that the same checks are done on each case. During comparison of uncoded with coded information, analysts should take the opportunity to fill in missing coded data items wherever possible. (For example, it may be possible to estimate truck dimensions.)
3. NHTSA should continuously tabulate missing data rates for each data field. These tabulations may highlight specific data collection problems that can be remedied while collection is still under way. They will also allow FMCSA to plan its analysis, since missing data may render some questions impossible to analyze.
4. Cross-tabulations of missing data, for example, rates by date of collection and by field office, should be prepared to search for sources of problems. FMCSA reported at the meeting that it plans to conduct such analyses. Missing data rates should also be tabulated by such key characteristics as vehicle type, truck type, and time of day. If a data item has a missing data rate for trucks different from that for nontrucks or the rate is different for two truck types, any comparative analyses of the two vehicle types probably will be biased, since missing data are seldom random. This effort can generate feedback to the field and to assessors and identify cases where further investigation may be necessary to fill data gaps.
5. Coding should distinguish among circumstances of missing data, including "not collected," "refused to respond," and "not applicable."
6. FMCSA also should tabulate rates of form completion (for example, of the fraction of multivehicle crashes with interviews with the other driver, or of the fraction of crashes with the Level 1 vehicle inspection completed).
7. For certain high-priority data items where problems with missing data or low reliability are discovered, FMCSA should consider devoting greater effort to obtaining more credible data in the remaining cases. Deciding which data elements deserve more resources depends on FMCSA's judgment about important database applications. For example, fatigue-related information is an area where the task force found questionable data entries and missing data, indicating the difficulty of documenting fatigue. Under present data collection procedures, there is a risk that fatigue data may be too unreliable or incomplete to be useful. After analyzing the quality of fatigue data obtained so far, FMCSA should consider the need for changing standard procedures to devote a greater share of investigative resources to collecting the fatigue-related information items. A second example highlighted in the task force's report to the committee is information related to driver pay and work organization.

The task force review suggests that it will be difficult under present procedures to produce useful data on this topic.

The committee would like to clarify that this recommendation is aimed at ensuring accurate data; the committee is not proposing in-depth accident reconstruction investigations, which it understands is not the methodology that FMCSA has chosen for the study. Decisions to devote greater effort to pursuing certain data items should not be made on a case-by-case basis according to ad hoc considerations, since such a procedure could bias the database.

8. FMCSA and NHTSA should consider developing a systematic way of checking the validity of data elements that depend on the exercise of judgment by analysts. For example, a coder might be given a case that the coder analyzed a year previously as a test of consistency, or the same case might be given to each of the coders to test whether all analyze it the same way. Outcomes of these comparisons, and any resulting changes in coding, should be recorded and reported.
9. FMCSA and NHTSA should compare rates of case completion (with respect to the population of eligible crashes) by data collection center, time of day, and other characteristics to search for potential sources of data bias or procedural problems.
10. Once the extent of missing data and nonresponse problems is documented, NHTSA and FMCSA should develop and document a plan for resolving or correcting for the problems as far as possible. Techniques are available to adjust for bias introduced by missing data in some circumstances. An example is the method used in the NHTSA study “Multiple Imputation of Missing Blood Alcohol Concentration (BAC) Values in FARS” (D. B. Rubin et al., 1998). Where missing data are the result of nonresponse to interview questions, adjusting for bias may require personal (e.g., gender and estimated age) data on respondents and nonrespondents. Interviewers should collect such information. However, it is important to stress that the bias introduced by nonrandom missing data cannot be removed by simply reweighting the responses for under- or overrepresentation.

Since the results of the missing data analysis will be relevant to future committee efforts concerning analysis planning, as described below, we request that FMCSA communicate them to the committee.

## **ANALYSIS PLANNING**

In past meetings, the committee discussed at length the definitions and relative importance of two applications of the database: (a) statistical testing of hypotheses concerning factors associated with increased accident risk and (b) assessment by expert evaluators of the events or conditions precipitating each crash. These discussions were summarized in previous letter reports. FMCSA has told the committee that it plans to present results of both applications in its report to Congress. We expect that the completed database will prove to be useful to government, researchers, and the public for both of these kinds of applications.

In its preceding letter reports, the committee recommended that FMCSA should conduct analysis planning, that is, that it should begin outlining in detail how it would initially employ the database in carrying out each of these two kinds of application, and especially that FMCSA should begin to plan the analyses for its report to Congress. The presentations at the meeting by Dan Blower of UMTRI and Ralph Craft of FMCSA described FMCSA's progress on analysis planning.

Our specific comments on the FMCSA work presented at the meeting are in the following two subsections. In summary, it is the committee's view that, although FMCSA now has taken some sound initial steps in the development of its analysis plan, it would be desirable to develop this plan to an advanced state while substantial data collection still remains to be carried out. In our first letter report 2 years ago, we concluded that "there is a clear need for a thorough analysis plan that documents agency plans for interim and final analyses for the study. . . . Regardless of methodology, data collection must be based on the research questions being addressed and the analysis to be undertaken." Although the study is no longer in the preliminary stage, there would still be benefits from developing a thorough plan now, before data collection is complete. It is not evident that a sufficient level of effort has yet been devoted to this task. Therefore, we recommend that FMCSA consider whether a reallocation of resources is necessary among the tasks of data collection, database design, and analysis planning.

We also note that the congressional charge to DOT provides for review and updating of the study every 5 years. Therefore it is worthwhile to identify possible improvements to the data collection and analysis methodology even if they could not be practically implemented in the present effort.

### **Statistical Assessment of Factors Affecting Crash Risk**

The presentation at the meeting on plans for statistical testing of hypotheses with regard to factors influencing accident risk shows that FMCSA has made progress on this part of the analysis plan. In the presentation, four specific candidate hypotheses about factors affecting crash risk were defined and the statistical method for testing each was outlined. The committee agrees that the preliminary step identified in this plan—examination of the completeness and consistency of the data elements needed for each of the planned statistical analyses—is the correct one, and recommends that FMCSA begin this process.

Also, as stated by Dr. Blower in the presentation, more work is needed on defining the details of each of the four proposed statistical analyses, in particular, determining the appropriate "comparison crash types," which are critical to this form of analysis. Assuming that the four hypotheses presented to the committee are indeed high-priority issues for FMCSA, we recommend that FMCSA develop these detailed analysis plans.

In addition to determining appropriate comparison crash types, obtaining adequate sample size may be a problem for the four proposed statistical analyses or for other similar analyses that FMCSA decides it would like to be able to carry out. Considering that the total sample is not large, that only a subset of cases will be pertinent for any particular hypothesis being tested, and that the magnitude of the problem of missing or unreliable data elements is at present unknown,

sample size could be a serious constraint. We recommend that FMCSA estimate sample size requirements for each of the four proposed statistical analyses, and for any additional ones it devises, and compare the requirements with the content of the database, taking into account rates of missing data, the likely frequency of errors in the data, and the likely size of the effects sought, as indicated by past research. Also, FMCSA should estimate the significance level and power of the statistical tests it plans to conduct, given the available sample size. These comparisons might reveal that reducing rates of missing data would be necessary before a particular statistical analysis could be conducted.

The committee recommends that FMCSA expand this portion of its analysis plan by adopting a comprehensive and strategic perspective, rather than by searching through the data being collected to seek analyses that are feasible. That is, FMCSA should identify a list of high-priority potential risk factors, including factors that it believes are related to important tactics for reducing crash frequency as well as factors concerning which there is greatest interest on the part of Congress, industry, and the public. FMCSA should then determine which of these can be assessed using the database and the planned statistical analysis method, and which would require other approaches.

### **Use of Crash Assessment Results and Report to Congress on Causation**

Along with planning of the statistical analyses, there is a parallel need for planning how the critical event, critical reason, and related factors data will be analyzed, that is, how causation will be examined. The critical reason assessment for each crash case is the element of the study that appears to approach Congress's question about causes of crashes most directly. Examination of the role of the related factors in each crash also is important in deriving a full conception of crash causation. Presentation of this information in the report to Congress will require sensitive and sophisticated discussion and analysis. Our understanding is that FMCSA does not yet have a detailed analysis plan for this effort. Until a plan is prepared, FMCSA cannot be certain that it is collecting and coding the data needed to support this aspect of the study. The committee recommends that FMCSA begin systematic planning of this critical part of the study.

## **SUMMARY OF RECOMMENDATIONS**

The committee's recommendations, extracted from the sections above, are repeated below.

### **Data Presentation**

- It is not too early for FMCSA to begin planning documentation and support arrangements for public use of the database and considering the costs of these activities.
- Documentation must make unambiguously clear to users the source of each item in the database.
- If the crash summaries now in the file are to be made public, FMCSA and NHTSA training and oversight of field staff should ensure that they follow a standard format and are reasonably accurate, complete, and comprehensible.

- The Crash Assessment Form should provide a detailed narrative identifying all facts used in the assessment and include the basis for the decisions made concerning the critical event and reason.
- If the public file includes any unedited interview responses, then the source of such data elements must be clear to users, documentation should note possibilities for substantiating interview responses by comparison with other data fields, and DOT should consider retaining in the public database the flags marking implausible responses that appear in the master file.
- FMCSA and NHTSA should take advantage of all opportunities for obtaining information from public sources (e.g., public police or court records or other independent sources) to substantiate interview information,

## **Data Quality**

- FMCSA and NHTSA should begin a systematic and quantitative analysis of rates of missing data, data quality, and the causes of missing data and coding errors and take action to reduce these problems wherever they are found to be significant.

## **Analysis Planning**

- FMCSA should consider whether a reallocation of resources is necessary among the tasks of data collection, database design, and analysis planning.
- FMCSA should begin the examination of the completeness and consistency of required data elements which was identified as the first step in the planned relative risk analysis described at the August 2002 committee meeting.
- FMCSA should develop the detailed plans for its proposed relative risk analyses.
- FMCSA should estimate sample size requirements for each of the proposed relative risk analyses, and for any additional ones it devises, and compare the requirements with the content of the database, taking into account rates of missing data, the likely frequency of errors in the data, and the likely size of the effects sought, as indicated by past research. Also, FMCSA should estimate the significance level and power of the statistical tests it plans to conduct, given the available sample size.
- FMCSA should expand its analysis plan for statistical assessment of factors affecting crash risk by adopting a comprehensive and strategic perspective.
- FMCSA should begin systematic planning of how it will use the critical event, critical reason, and related factors data to examine crash causation.

## **PLANNED COMMITTEE ACTIVITIES**

The committee believes that it may be able to assist FMCSA in the two components of analysis planning by preparing two background papers. The content of these papers was discussed in open session at the August, 2002 meeting. The committee will use the papers as a resource to help it formulate its advice on the study more clearly and completely. Before formulating its final recommendations, the committee would like to be able to consider in some detail the analysis planning needs of the study. The papers are to aid the committee in this respect. The

committee's intent is not to present a plan to FMCSA as a substitute for FMCSA's own efforts; rather, the committee's next report will illustrate the form or direction that planning could take.

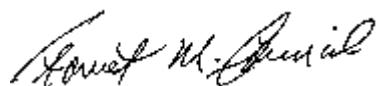
In the first background paper, the committee will list and rank critical policy questions related to truck safety. The committee plans to have a consultant review the literature and draft a list for committee review, and will develop a priority ranking of these issues on the basis of such factors as the probability of successful treatment and potential for reduction in accident losses. Then, with consultant assistance, we will consider which of these questions can be expressed as hypotheses that can be tested by the relative risk method (described in the presentation of Dr. Blower at the meeting) that FMCSA plans to employ, and which of them can be studied using the large-truck crash causation database.

In the second background paper, the committee, with consultant assistance, will attempt to provide guidance on a possible methodology for the crash causation analysis effort. The paper will consider how the critical reasons, related factors, and other data items in the current database could be used to explore causes in a way useful to DOT in developing programs to reduce accident frequency. The paper will also consider whether modifications are needed in the coding or presentation of data to support the examination of crash causation. Part of this effort will involve review of similar efforts in other fields.

Finally, because not all high-priority truck-safety questions will be answerable in the current study and because Congress has directed FMCSA to update the study in the future, the committee will begin to examine data collection and analysis methods that might be appropriate in a follow-up effort. Several alternatives have been raised in committee discussion in each of our meetings, including methods of automated data collection that conceivably could yield much more complete and accurate information than would traditional methods. We plan to further consider this option and discuss it among ourselves and with FMCSA at our next meeting.

FMCSA staff informed the committee that they are considering publishing a progress report on the study in December 2002. The committee will meet again after it has reviewed the progress report and has completed drafts of the background papers. This meeting is scheduled for March 2003.

Sincerely,



Forrest Council  
Chairman  
Committee for Review of the Federal Motor Carrier  
Safety Administration's Truck Crash Causation Study

Enclosure

MEETING ATTENDANCE  
August 20 and 21, 2002

Committee Members

Forrest M. Council, Chair  
Michael H. Belzer  
John R. Billing  
Kenneth L. Campbell  
James W. Dally (NAE)  
Lindsay I. Griffin III  
Anne T. McCartt  
A. James McKnight  
Raymond C. Peck  
Lawrence A. Shepp (NAS, IOM)  
Steven J. Vaughn  
Frank R. Wilson

TRB Staff

Joseph Morris

Government Staff and Guests

Terry Shelton, FMCSA  
Ralph Craft, FMCSA  
Dale Sienicki, FMCSA  
Joe Carra, NHTSA  
Seymour Stern, NHTSA  
Gregg Radja, NHTSA  
Kirsten Theriault, NHTSA  
Gary Toth, NHTSA  
Nancy Bondy, NHTSA  
Jim Page, Veridian Corporation  
Don Hendricks, Veridian Corporation  
Steve Mavros, KLD Associates  
Richard Ketterer, KLD Associates  
Richard Reed, Accident Research and Analysis  
Dan Blower, UMTRI  
Nicholas Alexandrou, Volpe Center  
Representative of Transport Topics  
Peter Kissinger, AAA Foundation for Traffic Safety  
Bella Dinh-Zarr, AAA