MICHIGAN ACCIDENT DATA IN PLANNING APPLICATIONS

Marion J. Bowman
Charles P. Compton
John A. Green

The University of Michigan
Transportation Research Institute
November, 1982

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.

4. Title and Subtitle	5. Report Date
Michigan Accident Data in Planning Applications	November, 1982
	6. Performing Organization Code
	8. Performing Organization Report No.
7. Author(s) Marion J. Bowman, Charles P. Compton, John A. Green	UM-TRI-82-44
9. Performing Organization Name and Address UM-Transportation Research Institute 2901 Baxter Road, Ann Arbor, Michigan 48109	10. Work Unit No.
	11. Contract or Grant No. MTR-81-002A
12. Sponsoring Agency Name and Address Office of Highway Safety Planning 111 S Capitol Avenue, Lower Level Lansing, Michigan 48913	13. Type of Report and Period Covered Final Report 5/1/81 - 10/31/82
	14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

This report documents an evaluation of the Michigan Accident Data File in terms of the quality of the data contained in it, the procedures involved in the data collection, the scope of potential users of the data, and the utility of the data to such potential users. The quality of the data is discussed first in terms of missing data rates on specific variables and the consistency of the data across variables. The procedures used to collect and digitize the data are then reviewed.

Recommendations to improve the utility of the data and to facilitate the data collection procedures are presented.

17. Key Words	18. Distribution Statement
	unlimited

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	43	

TABLE OF CONTENTS

Section 1
Introduction

In the process of evaluating the traffic-safety-associated problems in a given area for the purpose of developing a highway safety master plan, many highway safety agencies rely heavily on data gathered by the state police or by local police authorities. Thus, the motor vehicle traffic accident data collection system is the source of information widely used for research and evaluation purposes.

Because of this dependence on data in state safety planning and research efforts, it is important that the quality and information content of the data be adequate for the analysis demands that are placed on it.

Historically, accident data systems were designed to collect information both for assisting law enforcement agencies in prosecution and for use in highway planning. Data files were generally maintained by local police or government agencies. Because the form of the data varied from one local jurisdiction to another, it was of limited value in determining statewide traffic accident characteristics. With the advent of modern, high-speed digital computers, however, there has been a striking centralization of data collection efforts at the state level over the past ten to fifteen years. As a result, significant numbers of broadly-based accident reports have become available for analysis. The federal government, operating through the National Highway Traffic Safety Administration (NHTSA), has been aware of this research potential, and is encouraging state governments to develop highway safety plans that are based on actual accident experience.

As might be expected, such use of the data has uncovered problems that were never anticipated in the original design. For example, some data elements that were of little use to police agencies are of great importance to analysis efforts. In addition, high missing data rates or inadequate precision in coding can lead to inconclusive or inapplicable results.

A highway safety plan for a populous state such as Michigan may be instrumental in allocating significant amounts of money for enforcement or construction purposes. Consequently, it behooves the state to ensure that the data is utilized to provide the best possible indication of the accident trends and characteristics that is available.

This report presents the results of a review of the Michigan Police data with respect to data collection procedures and data quality as measured by missing data rates and inter-variable consistency checks.

The results of the review indicate in general, an extensive, well executed procedure for data quality control, and missing data rates that are, in most cases, perfectly acceptable for data analysis. Some recommendations for changes in the way missing data codes are recorded are presented in Section 5.

Section 2
Missing Data Evaluation

This section presents an analysis of the missing data in the Michigan Accident Data as it may affect current or future users of the data. There are two ways missing data can cause a problem for analysts: 1) either the missing data rate is so high that insufficient information is available for analysis, or 2) characteristics of the missing data cases differ from those cases for which data are available on variables of interest. The first step, therefore, is to determine if the missing data rates on any collected data elements are unreasonably high. Appendix A contains a list of these variables and their missing data rates. Then the distribution of missing data cases on other variables can be examined to determine if the missing data cases are randomly distributed on non-missing variables.

After an examination of the missing data, groups of variables from the different format types which contain similar data can be checked for consistency to determine if coding or computer derivation of variables is producing misleading or incorrect information.

The following sections deal with each of the case record formats separately. MALI (Michigan Accident Location Index) records are excluded, as cases are not generally entered into the accident file until the MALI information is coded as completely as possible. The MALI information is generally most useful for the determination of traffic engineering problems or other accident site characteristics specific to a location. Thus, data analysts do not generally access this information unless they are attempting to map specific problem locations.

In the review which follows the missing data rates are determined as a simple one-way distribution of all cases for

the code levels of the variable of interest. Accidents or traffic-units which are not specifically applicable for a certain variable or variables (e.g. Model Year for pedestrians) are not filtered out of this missing data analysis unless otherwise indicated. The purpose in this study is to determine the missing data rate of each variable independently. To make the data set most useful to the widest range of potential analysts, each variable, when possible, should be able to stand alone and present valid information for all cases. Thus, if Model Year is not applicable for some traffic unit types (pedestrians), there should be a code level 'not applicable' to indicate this.

2.1 Accident Record - Format 1

The accident record, Format 1, contains general information about the accident site, environmental conditions, and summary information classifying the accident type and the vehicles and occupants involved. The missing data rates are discussed first for the variables collected in the field or coded from the field form and then for those variables generated by computer from field data. The accident record contains some MALI location information which will not be reviewed for missing data for the same reason the MALI record itself will not be reviewed.

The missing data rates for the accident record variables which are recorded in the field or from the written narrative are quite low. Most variables have a missing data rate below one percent.

The only variables with a high missing data rate, "Highway Area Code" and "Accident Analysis -- How", both contain a catch-all missing data category which includes valid data. The code 'unknown' for "Highway Area Code" is also used for intersection types for which there is no

specific code. This differentiation, between a known area type for which there is no specific code and a completely unknown area type, is lost because of the use of 'other' and 'unknown' as the same code. A similar problem exists for the "Accident Analysis -- How" variable, but further examination of this variable may be found in the section of this report dealing with cross variable consistency checking.

The use of one code value for 'other', 'not applicable', and 'unknown' is quite common in the Michigan Accident Data File in all the format types. This practice leads to an inflated missing data rate for individual variables and a loss of differentiation between data which is missing and data which does not easily fit a category (and may warrant further examination by a researcher). This specific problem will not be mentioned for every variable, but in many cases it is a problem. See the end of this report for specific recommendations concerning this practice.

Internally derived variables are found on the accident format record. Internally derived variables are generally computer generated and use the values of one or several field data variables to generate codes for a given case. Thus, internally derived variables should have no less missing data than the variables they are based on. Other types of internally derived variables are counters of certain aspects of the case (such as the number of injured occupants).

Four derived variables on the accident record, all describing some aspect of numbered state and federal roads, have a 'missing data' rate of 66.2%. Unfortunately, the 'unknown' and 'not applicable' codes for these variables are the same so that a simple univariate does not tell what

percentage of these are really local roads (the meaning of the 'not applicable' code).

The missing data rates for most of the derived accident variables are quite low, as are the variables from which they are derived. "NSC Accident Circumstance", with 13.4% missing data, has the highest missing data rate. This variable, however, attempts to code the likely cause of the accident into twelve general categories and, thus, it is not surprising that a fair proportion of accidents do not fit into any one category. More categories would probably reduce the missing data rate (it is hard to say how much because this variable also lumps the 'unknown' and 'other' categories together). However, this would defeat the the purpose of using the National Safety Council coding scheme which is presumably comparable to data collected elsewhere.

The "Drinking In Accident" variable, containing 7.1% missing data, uses the driver information for each vehicle involved to determine drinking in the accident. The efforts of the Michigan State Police to reduce the missing data rate for the drinking variable is evident by the low percentage of missing data for this variable. In comparison, the State of Washington had 38% missing data on their "Driver Sobriety" variable in 1980.

The "Accident Analysis Subscript", which is a combination of "Accident Analysis--How", "Accident Analysis--Where", and "Accident Type", has many codes which allow for one of the three components to be unknown but still provide useful information. The percentage of partial unknowns must be something less than the sum of the unknowns for its three components (about 17%). This rate of missing data seems quite reasonable for such a complex variable.

2.2 Traffic Unit Record - Format 2

The Traffic Unit Record is completed for each vehicle as well as pedestrians, pedalcyclists, and other 'traffic units' involved in the accident. Because of the various types of traffic units being documented, the complexity of the coding problem for the Traffic Unit Record is greatly increased. Some of the variables in the traffic record are not applicable for some traffic unit types (i.e. "Model Year" for pedestrians), while some are applicable to all types (i.e. "Vehicle Make"). The analysis of the missing data for the variables not applicable for all vehicle types is confounded when the not applicable code is combined with the missing data code and requires first filtering out the cases which are not applicable. For example the 10.3% missing data on "Vehicle Type" is reduced by about 2% if the non-motor vehicles are filtered out (and by another 4% if the 'other' category were not included with the unknowns).

The missing data rates, in general, are not any higher than might be expected for a mass accident file (particularly when almost 20% of all accidents are not investigated at the scene). It is surprising that 9.4% of driver ages are missing since this information is recorded on the driver's license. A rather high rate of missing residence information is also surprising, but all states not immediately adjacent to Michigan are lumped in the 'other' and 'not known' category. This could be important information to analysts if out-of-state drivers are of particular research concern.

"Vehicle Make" is one of the more important variables in the Traffic Unit record. Several variables are derived from, or have code values based on, the coding of "Vehicle Make". The absence of a unique unknown code results in 22,299 (4.2%) 'vehicles' coded as 'other motor vehicles or

unknown'.

The "Motorcycle Driver/Passenger Helmet Use" variables display inconsistent coding in two ways. First, the occupants of cars, trucks, farm and construction equipment, other vehicle types, and pedestrians and pedalcycles are in some instances coded as using helmets. This is not really a problem except that the variable is intended for use of motorized cycle occupants only. Additionally, and perhaps more importantly, the 'none used' category is coded 79.1% of the time and is regularly used for all vehicle types. The 'not coded' category is used infrequently, but also used for occupants of motorized cycles. This mixing of the code levels makes their meaning unclear and potentially misleading.

Internally derived variables for the traffic unit record are just a recoding of field collected data except for the summary restraint variables. The "NSC Vehicle Type Subscript" and "Vehicle Type Subscript" represent the third and fourth version of a vehicle type variable. This data is somewhat redundant in that most analysis programs allow the user to recode variables to suit a particular need. The added cost of carrying these variables is negligible, however, since they are generated by the computer.

The restraint use variables are summaries of restraint use and there is no missing data on these variables. "Total Passengers with Unknown Restraint Use" is a summary of the missing data for each vehicle. Eleven percent of the cases in the file had missing restraint information on at least one occupant. This variable, therefore, did not have any missing data, per se, but does indicate the rate of lost information about restraint usage (11%).

2.3 Occupant Record - Format 3

In the past the Michigan Accident Data File contained one record for each injured occupant of a motor vehicle. Beginning with the 1980 data set, there is one record for each occupant of a vehicle involved in a crash. However, uninjured occupants have only restraint information coded. This results in occupant records with 'unknown' data for most of the occupant variables (e.g. age, sex, and position). Fully 80% of the occupant records now found in the file are for uninjured occupants. Unfortunately, the coding scheme of the occupant variables was not revised to account for the inclusion of these occupants. Particularly with respect to occupant sex, no unique code was provided to indicate unknown or not coded and all unknowns are coded as male. Hopefully this practice will be modified in the future.

In general the missing data rates for the occupant variables, with the uninjured occupants excluded, is quite reasonable. The "Helmet Use" variable in the occupant record suffers the same coding problem as does the "Helmet Use by Driver" variable on the vehicle record; the use of 'not used' and 'not coded' is not standardized.

2.4 Cross Variable Consistency Check

This section deals with comparisons of variables with the same or similar information as well as checks on the values of variables whose coding is dependent on the values of other variables.

A breakdown of those cases with accident type coded as 'other' or 'unknown' indicates that with the 'other' category included, 1635 cases are of 'unknown' accident type. The 'Accident Analysis-How' variable, when filtered to include only these 1635 cases indicates that in 405 cases

an occupant fell from his vehicle and in 143 cases an occupant was injured in his vehicle. These are not 'unknown' accident types but are included in the 'unknown' category as are all other accidents which don't fit the accident typing scheme.

Similarly, a one-way distribution of 'Accident Analysis-How' indicates 52,459 cases coded 'not pertinent', 'other' or 'unknown', but, in a two-way cross tabulation of 'Accident Type' and 'Accident Analysis-How' it turns out 50,313 cases are 'not pertinent'. This changes the apparent missing data rate of 'Accident Analysis-How' from 16.7% to 0.7%. It would be better from an analyst's standpoint to permanently separate these codes to give a more clear picture of the real missing data for the 'Accident Analysis-How' variable (0.7%).

There are four vehicle-type variables carried in the Michigan Accident File. Two, 'Vehicle Type' and 'Vehicle Make', are recorded on the UD-10 by the investigating officer. 'Vehicle Type-Subscript' is a recoding of the 'Vehicle Make' variable and thus is equivalent to it (with many fewer code levels). 'NSC Vehicle Type' is a combination of a vehicle type variable and the 'Trailer Type' variable. Table 1 is a comparison of 'Vehicle Type' from the UD-10 and 'Vehicle-Type Subscript'. Station wagons (which includes 'carryalls') and Jeeps, (as coded on 'Vehicle Type') are variously coded as cars, trucks, or others on 'Vehicle-Type Subscript'. Buses (school or commercial), as coded on 'Vehicle Type', are variously coded as cars, trucks, station wagons, jeeps, and others on 'Vehicle Type'. Two problems are evident with the coding of vehicle type. First, the 'Vehicle Make' variable, from which 'Vehicle-Type Subscript' is generated, contains a mixture of vehicle makes (i.e. Ford, GM, etc.) and vehicle

uses (i.e. ambulance, police car, construction equipment, etc.). This leads to a varied coding practice in the field as evidenced by Table 1. For example, two vehicles, coded as cars less than 1500 pounds, are also coded as school buses. Very few cars were ever made which are that small, and the few on the roads of Michigan are certainly not used as school buses. Perhaps a 'Vehicle Use' variable is called for.

Table 2 is an example of lost data because of the use of 'Other' and 'Unknown' as the same code. For example, the use of 'Other', 'Not Known', or 'not a motor vehicle' as a code value for trailer type results in 291 passenger cars which either have a trailer type that does not fit the trailer codes available or had unknown trailer use. In many modes of analysis which could be used on this data these vehicles would have to be considered not towing a trailer because this code value is also the 'not applicable' code when the traffic unit is an non-motor vehicle.

Analysis of the Hazardous Action and Police Action variables indicates an inconsistency in the coding of these variables which is most evident for pedestrians and pedalcyclists. Of the 5604 pedestrians and 4390 pedalcycles and other non-motor vehicles, 99.4% and 91.8% respectively were coded as 'cited for hazardous violation'. However, 37.6% of pedestrians and 31.7% of non-motor vehicles were coded as having no contributing circumstance and 52.2% of pedestrians and 44.9% of non-motor vehicles were coded as committing no hazardous action. Since a version of this variable is carried on the accident record and is intended to indicate if anyone in the accident has been cited for a hazardous action the vehicle-specific citation variable should indicate the citation status of the traffic unit on whose record it resides.

Table 1

Cross Tabulation of Vehicle-Type Subscript
and Vehicle Type
1980 Michigan Accident Data

Vehicle Type (UD-10)	Vehicle-Type Subscript (Internally Generated)										
	Pass. Car	Truck	Motor cycle	School bus	Comm. bus	Farm Equip.	Const. Equip.	Other Unk.	Ped.	Pedal cycle	Other Road
Car<1500lbs	1468	0	0	2	0	1	2	117	0	0	0
Car<2499lbs	42913	0	0	0	0	2	2	195	0	0	0
Car<3500lbs	135125	0	0	1	3	3	0	516	0	0	0
Car>3500lbs	209298	0	0	25	26	2	1	1515	0	0	0
Sta. Wgn	3619	150	0	13	34	0	1	36	0	0	0
Jeep	1828	190	0	0	0	0	0	39	0	0	0
Pickup	0	61196	0	19	20	1	10	398	0	0	0
Str. Truck	0	7117	0	281	160	14	113	493	0	0	0
Trac.(semi)	0	6803	0	1	0	3	5	131	0	0	0
Other/unk	10416	2494	6971	1206	1234	308	394	21719	5604	4377	13

Table 2

Cross Tabulation of Trailer Type and Traffic Unit Type
1980 Michigan Accident Data

Vehicle Type Sub.	Trailer Type						
	None	Utility Trailer	Single Bottom Truck	Double Bottom Truck	House Trlr	Other Unknown Non-M.V.	Towed Vehicle
Car	403242	807	0	0	75	291	252
Truck	71015	1429	4433	461	84	174	354
M.C.	6970	0	0	0	0	1	0
School Bus	1548	0	0	0	0	0	0
Commer. Bus	1477	0	0	0	0	0	0
Farm Equip.	261	59	0	0	0	14	0
Const. Equip.	508	11	2	0	0	5	2
Misc. Other Unknown	24691	153	96	16	4	168	31
Non-M.V.	0	0	0	0	0	9994	0

As stated in the introduction, one of the ways missing data may be a problem in a data set such as the Michigan Accident file is if, for some variable of interest, the distribution of cases missing data is not the same as those cases for which data is present. For vehicles which are of an unknown or other type there may be a bias problem. For the entire file 33.4% of vehicles (including pedestrians) are coded as having been cited for a hazardous violation. Of vehicles of other or unknown type the percentage of hazardous citations is 77.7%. This would suggest that there is something 'different' about vehicles of unknown type or those for which no vehicle code is applicable. When analysis is done on the Michigan File, many of these cited

drivers may be excluded because these cases are in the 'unknown' category for 'Vehicle Type' (although they may actually be of an 'other' type).

Section 3
Data Collection Procedures

The Traffic Services Division of the Michigan Department of State Police handles all traffic accident reports generated by police agencies throughout the state of Michigan. Through a step-by-step process, each Official Traffic Accident Report (UD-10) makes its way eventually from the originating police officer to a computerized accident data file used by managers and analysts.

The arriving source document in the "incoming mail" department can be one of five types, described in the following sections:

3.1) An original non-fatal Official Traffic Accident Report with property damage (P.D.) or personal injury (P.I.)

3.2) A supplemental document (additional information to a previously submitted UD-10)

3.3) A corrected UD-10

3.4) A previously submitted UD-10 but now resubmitted and corrected

3.5) An original Official Traffic Accident Report containing a fatality to one or more of the occupants of a vehicle or vehicles involved in a crash.

All five incoming data types eventually are entered into a computerized data entry system. The step-by-step progress of each of the five types of data follows. Data from all five types, once entered into the computer, will be traced in the remainder of this section.

Much of the information presented here relative to the data system was developed from an on-site review and from a

Traffic Accident Processing System Flow Chart developed by the Traffic Services section of the State Police. The chart was extremely helpful and was used to clarify the case-flow operation of traffic data as it currently exists within the department.

3.1) Non-Fatal Accident Report

Once the Official Traffic Accident Report (UD-10) has been completed by a policing department in the state, a copy is forwarded to the Traffic Services Division of the Michigan Department of State Police where it enters the system for eventual computerization. Initially, the UD-10 is scrutinized to determine if driver license action is indicated. If so, a copy of this report is sent to the Secretary of State's Office for handling. Each of these UD-10s indicates property damage or personal injury sustained in the crash. The UD-10 is then microfilmed. If property damage to a trunkline highway is indicated, the original document is Xeroxed and a copy is also sent to the Highway Department for initiation of a damage repayment restitution claim. A determination is also made at this point on whether or not the UD-10 reflects a traffic accident or a non-traffic accident. (A non-traffic accident is the result of a traffic unit involved in an off-the-road crash such as a field, private road, private parking lot, etc.) If the crash is determined to be a non-traffic accident and a non-Michigan State Police Report, the source document (UD-10) is discarded. (All Michigan State Police non-traffic accident reports are separately filmed and stored in a specific film area.)

Should the UD-10 report prove to be a valid personal injury (non-fatal) or a property damage traffic accident, it is then carefully checked for errors in location data and errors or missing data in accident information. Those

UD-10s containing missing location information are forwarded to the liaison officer of the Michigan State Police for resolution by telephone. If the liaison officer cannot resolve the discrepancy or complete the missing information by telephone, the UD-10 is returned to the submitting department for completion or correction and will be re-submitted to the liaison officer once it has been corrected. (See Section 3.3 on corrected copies.) All snowmobile crashes are tabulated, whether or not they occur on a road or in a field. If the snowmobile crash is not highway related, a pseudo highway number is assigned as the location code and the report is not coded by the MALI - Michigan Accident Location Index. (Snowmobile crashes involving roads, highways, streets, etc., of course are coded by MALI along with all other traffic accidents.)

Assuming the UD-10 has no location data errors, it is then coded by MALI coders and checked once again for errors in accident data. If errors or blanks appear in the accident data, it once again is referred to the liaison officer for resolution and may make its way back to the originating department for resolution and corrected resubmission. If no accident data errors or missing accident information are unearthed, the UD-10 will be coded for key entry into the data entry system.

3.2) Supplemental Documents

Supplemental documents also appear in the incoming mail. Such supplemental documents generated by a Michigan State Police Post are microfilmed and stored in the separate supplement file. Most frequently, the supplemental information concerns those accidents in which a fatality occurred and these data are microfilmed in the separate fatal file (see Section 5) as well as microfilmed for storage with original source Michigan State Police crash

documentation for eight years or with all other police agency original source documentation for one year. All non-fatal non-Michigan State Police investigation supplemental documents are discarded.

3.3) Corrected UD-10

Correction reports (UD-10s) initiated by the original submitting department occasionally arrive with the incoming mail. If the corrected report reveals no new vital information (location or accident information) the corrected document is discarded unless it is Michigan State Police generated, in which case it is microfilmed and stored in their separate film storage area. Should the corrected copy reveal highway damage (where it previously did not), a copy is sent to the highway department for highway repair restitution. Key-entry personnel are alerted to delete the original UD-10 information from the data base and the corrected copy then goes to the review area (prior to MALI coding) for error checking, missing data checks and possible liaison-officer application.

3.4) Resubmitted UD-10

Occasionally, documents are resubmitted. These resubmissions are reports that had been returned by the liaison officer to the originating department because he was unable to resolve discrepancies (errors or blanks) by telephone. These resubmissions (or "returns") are dispatched to the liaison officer immediately. He reviews the resubmitted UD-10 report once again for location or accident data errors, and once all data are determined to be correct, the copy is either sent for MALI coding or to the accident coding section depending on the reason for its return to the originating department.

3.5) Fatal Accident Report

The Fatal Accident Reporting System (FARS) was established in 1975 by the National Highway Traffic Safety Administration as a national census file of all fatal accidents occurring in the United States. One or more FARS analysts are located in each state and are responsible for the coding, correction, submission, filing and updating of every crash occurring in the state involving a fatality. The incoming accident report which involves a fatality is immediately referred to the FARS analyst once microfilming of the of the source document is completed. During microfilming, the document is reviewed for highway damage. If property damage to a highway is indicated, a copy is sent to the Highway Department for handling of claims. Highway related fatal snowmobile crashes are included in FARS data. Prior to fatal crash data being coded by MALI location coders, the liaison officer checks all fatal crash information for highway and accident errors and missing data. Once errors are resolved, the accident is ready for coding and additional error checks.

FARS analysts assign a separate state fatal number and then submit copies to the Secretary of State's office for driver license notation and the State Highway Department for road information maintenance and/or road repair. The case report is then coded by State accident data coders and is then key entered into the computer file containing all accident data. The FARS analyst has numerous other duties and responsibilities with these data and maintains a separate hard-copy file of all fatal crashes coding information from these crashes into the FARS format. These data are then submitted via terminal to the National Highway Traffic Safety Administration.

3.6) Survey of External Agencies

In an attempt to survey data-flow methodology, several local police agencies were interviewed to observe actual procedures in operation.

While each police agency operates independently and within the constraints of local budget allocations, there are similarities in the handling of data among all who were interviewed. UMTRI asked all respondents the same questions and responses were generally consistent. Each agency trains patrolmen in the use and preparation of the UD-10 at the time the patrolman goes through his/her initial training period. All agencies indicated that it was rare that additional information was to be collected and no agency interviewed had been made aware of "funeral procession" or "slush or road spray" having been added to the coding scheme in Lansing. It is doubtful that this information then, would be collected by the officer in the field. One local agency indicated that "news" filtered to the local level slowly and usually took three to nine months to reach them from Lansing.

Error checking is done by traffic supervisors in each local police agency. All agencies use the U.S. mail to send completed forms to Lansing, with some sending them each day (or as the accident form is completed) and others sending them in once a month. Revisions and corrections are sent with the regular submissions.

No agency claimed to ever have had a call from the liaison officer regarding incorrect data or discrepancies but have averaged one to two UD-10s per year being returned for corrections. Supplementary information concerning accidents is not usually sent off to Lansing except in the case of a fatal crash. Agencies, however, retain supplementary documentation in their own files for future

reference.

While all agencies receive feedback from MALI output, several agencies feel MALI reports are too late and out of date to be of interest to them. The consensus seemed to be that MALI should be updated more frequently to be of value to local jurisdictions.

Departments were asked their opinion of a direct on-line entry system for their use in inputting the UD-10 data to the state, and all felt that a hard-copy file would still have to be maintained. Each replied that insurance companies were the prime users of these data, and all were concerned with the cost of such a system. Smaller departments felt that such an on-line system would cause additional work loads for the few available personnel. In addition, at a time when funding has been curtailed necessitating reduction in police enforcement personnel, the expenditure of additional dollars for equipment seems unsound.

Section 4
Representative Data Users

A list of representative users of the accident data was derived from the computer listing of persons who have accessed the Michigan Data file through the Michigan Terminal System at the University of Michigan. These users were contacted by mail and asked to respond with suggestions, criticisms, or positive attitudes regarding their experiences accessing the data. The list of users involved six organizations: the State of Michigan Office of Highway Safety Planning, the State of Michigan Department of State, the National Highway Traffic Safety Administration, the University of Michigan, Ford Motor Company, and Cromack Engineering, Inc. Individuals at each of these organizations were contacted along with several large metropolitan police departments and their responses are attached in appendix B. (The responses received from the metropolitan police agencies reflected their usage of the MALI data, inasmuch as they concentrate their efforts on locations indices rather than accident data.)

Section 5
Recommendations

This section details the recommendations this review of the Michigan Accident Data system has generated. The first section deals with the data file variables and their documentation, the second deals with the interactive procedures involved in creating and maintaining the data file.

5.1 The Michigan Accident Data File

The primary recommendation resulting from our review of the variables and missing data in the data file is that more distinction be made between data which is missing and data which does not fall into the available categories on a given variable. The commonly used practice of combining 'other', 'not applicable', and 'not known' into one code level increases the complexity of analysis of the data and increases the likelihood of incorrect or misleading results. This problem could become even greater as the data becomes available to a wider range of analysts with divergent backgrounds and minimal familiarity with the data base.

The data set contains several variables documenting the vehicle type, some are collected on the accident form in the field and others are computer generated from the field variables. Consistency between all the vehicle type variables is important and some inconsistencies are noted in the text of this report. Effort should be made to determine how these inconsistencies arise and how they can be eliminated.

5.2 Data Collection and Turn-Around Feedback

The Michigan data collection procedures and data content appear to rank highly when compared with other

states. Information necessary for evaluation seems readily available and easily accessible to analysts, managers and police agencies. Nevertheless, feedback to small police departments could perhaps be improved by increasing reporting turn-around time. With current budget constraints and decreased staff at the state level, this does not appear to be practical or feasible at the present time. Perhaps current equipment budgets could be revamped to include inexpensive computer terminals supplied to local departments for use in accessing state data through the University of Michigan's terminal system (or directly into Lansing's computer data bank). Then each local department might determine its own needs and access relevant data accordingly.

Morale and relationships improve considerably (not to mention increased work performance) when workers or assistants are made aware of the utilization of their specific input to an overall project. The University of Michigan Transportation Research Institute could assist the Michigan State Police in this endeavor with a pilot study of small police agencies within a one or two county area. Such a study might involve contacting all such agencies, determining their needs, supplying inexpensive terminals, and training personnel in the use of data, to easily access information desired via the Michigan data resident at UMTRI. In addition, messages regarding changes to the data collection can then be displayed to each department as it signs on to the system, and a rapid information relay can be achieved.

In addition, a possible cost-saving item and method to improve rapid response lies in the area of modification to the Highway Department of property damage accidents. Should a police department at the local level have an accident

which involves damage to a trunkline highway fixture, that police department could Xerox the UD-10 and send it directly to the Highway Department in Lansing, eliminating time delay and personnel costs in Traffic Services. This would allow for immediate attention by the Highway Department to maintenance and repairs at a time when highway repairs have become a prime target for attention.

Appendix A

List of Variables and Missing Data Rates

Appendix A

The data elements in the tables which follow appear in the order in which they are found on the data tape supplied by the Criminal Justice Data Processing Center and in the tape layout documentation supplied with the tape. The code values and their meanings are supplied with the tape documentation and thus define the data set for analysts not involved in the data collection or computerization. It is important, therefore, that this documentation accurately reflects not only the magnetic tape layout but also, to the extent possible, the coding practices which result in the definition of each code value for each variable. The data set is only as accurate as this documentation.

Variable Name	Miss Data Rate	Specific Missing Data Code
Accident Format		
County Code	-	no unknown code
City/Township	-	no unknown code
Month	-	no unknown code
Day	-	no unknown code
Time of Day	0.5	'not known'
Weather	0.4	'other' and 'unknown' same code
Light	0.4	'not known'
Road Surface	0.9	'other' and 'unknown' same code
Road Defect	0.0	'other' and 'unknown' same code
Traffic Control	0.1	'other' and 'unknown' same code
Highway Area Type	-	no unknown code
Highway Area Code	32.4	'local' and 'unknown' same code
Road Alignment	0.0	'not known'
Accident Location	-	no unknown code
Accident Type	0.5	'other' and 'unknown' same code
Acc. Analysis-Where	0.0	'not known'
Accident Analysis-How	16.7	'other','not pertinent', and 'unknown' same code
Special Accident Tag	-	no unknown code
Investigated By	-	no unknown code
Construction Zone	-	no unknown code

Variable Name	Miss Data Rate	Specific Missing Data Code
	Accident Format	
Highway Control Sect.	66.2	'local road or unknown'
Miles in Control Sect.	-	no unknown code
Highway Dept. District	-	no unknown code
MSP Post Area	-	no unknown code
Year of Accident	-	no unknown code
Day of Week	-	no unknown code
Highway Class	66.2	'county road, city street or unknown'
Highway Number	66.2	'local road -type Unknown'
Accident Analysis-Subscript	0.0	multiple codes with partial unknown data
Severity of Acc.	-	no unknown code
# Moving Vehicles	-	no unknown code
# Persons Killed	-	no unknown code
# Persons Injured	-	no unknown code
Drinking in Acc.	7.1	'not known'
Population Subscript	-	no unknown code
Accident-Type Sub.	0.5	'other non-collision or unknown'
Highway-Class Sub.	66.3	'service drive, local road, or unknown'
NSC Acc-Circumstance	13.4	'other or not stated'
Two-Vehicle-Acc. Sub.	0.0	'not stated'
Enforcement in Acc.	-	no unknown code

Sub.=Subscript

Variable Name	Miss Data Rate	Specific Missing Data Code
	Traffic Unit Format	
Age of Driver	9.4	'not known'
Driver Residence	6.8	'other' and 'unknown' same code
Driver Sex	-	no unknown code
Driver Injury	-	no unknown code
Year Manufactured	7.8	'other' and 'unknown' same code
Vehicle Make	4.2	'other' and 'unknown' same code
Vehicle Type	10.3	'other','unknown',and 'non-m.v.'
Driver Intent	0.3	mixed codes depending on TU type
Hazardous Action	3.7	'other' and 'unknown' same code
Visual Obstruction	0.2	'other' and 'unknown' same code
Object Hit	1.9	'other' or 'not applicable' same
Vehicle Condition	1.9	'unknown' and 'non-m.v.' same code
Trailer	2.0	'other','unknown','non-m.v.' same
Situation	-	no unknown code
Direction of Travel	0.5	'not known'
Impact Code	2.7	'unknown' and 'non-m.v.' same code
Vehicle Drivable	4.3	'unknown' and 'non-m.v.' same code
Drinking/Drugs	11.3	'not known'
Test Result	0.3	'influence not known'
Traffic Unit Num.	-	no unknown code
Total Occupants	0.1	'not known'
Police Action	0.0	'not known'
Hit/Run Driver	-	no unknown code
Damage Severity	8.2	'unknown'
Pass. Restraints:		
Driver	10.8	'use or action unknown'
Front center	0.2	'use or action unknown'
Right Front	0.9	'use or action unknown'
Left Rear	0.1	'use or action unknown'
Center Rear	0.1	'use or action unknown'
Right Rear	0.2	'use or action unknown'
Driver Helmet Use	19.7	'not coded'
Pass. Helmet Use	94.1	'not coded'
Truck Cargo Type	13.0	'unknown or not stated'
Cargo Spillage	9.5	'not known'
Fuel Leak/Fire	-	no unknown code

Variable Name	Miss Data Rate	Specific Missing Data Code
Traffic Unit Format		
Vehicle-Type Sub.	4.8	'ambulance, police equip., snowmobile, dune buggy, go-kart, other or unknown'
Age-MV-Driver Sub.	9.3	'not stated'
Ped-Intent Sub.	0.7	'not known'
NSC Veh Type Sub.	6.7	'other, unk, or not a motor veh'
Total Pass. with:		
Rest. Not Used	-	no unknown code
Rest. Used	-	no unknown code
Unknown Rest. Use		11.0% cases with unknown use

Variable Name	Miss Data Rate	Specific Missing Data Code
		Occupant Format
Age	80.9	'not known'
		1.6% injured occupants only
Seat Position	0.3	'other' or 'unknown'
		1.8% injured occs only
Sex	0.0	unknowns coded as male
Injury Severity	-	no unknown code
Location	0.1	'other' or 'unknown'
		0.7% injured occs only
Rest. Use		
	8.6	'unknown'
		7.1% unknown for inj occs
	2.4	'not available for bus, mc, etc'
Helmet Use	83.8	'info. not coded'
		16.8% not coded for inj occ
		MSP Internally Derived Variables
Location-of-Injured Subscript	0.1	'Other or Unknown'
	0.7	other/unk for inj occ
Age-of-Injured Subscript	80.9	'not known'

Variable Name	Miss Data Rate	Specific Missing Data Code
Cases with Accident Type Unknown		
Accident Type	0.5	(1635) 'other' and 'unknown'
		(405) 'fell from veh.' -HOW
		(143) 'injured in veh.' -HOW
	0.3	(1087) actually 'unknown'
Accident Analysis-How	16.7	(52,459) 'not pertinent','other',or 'unknown'
	16.0	(50,313) 'not pertinent'
	0.7	(2146) 'other' or 'unknown'

Appendix B

Responses to Questionaire

STATE OF MICHIGAN

WILLIAM G. MILLIKEN, GOVERNOR

DEPARTMENT OF TRANSPORTATION

TRANSPORTATION BUILDING, 425 WEST OTTAWA PHONE 517-373-2090
POST OFFICE BOX 30050, LANSING, MICHIGAN 48909

JOHN P. WOODFORD, DIRECTOR

August 30, 1982

Ms. M. J. Bowman
Research Associate
Highway Safety Research Institute
Huron Parkway and Baxter Road
Ann Arbor, Michigan 48109

Dear Ms. Bowman:

This is in response to your recent letter regarding the adequacy of the Michigan Traffic Accident Report (UD-10). The present report was updated in 1978 and we are, in general, satisfied with the data provided by the report. However, in a future update, we would suggest consideration of the following comments:

1. Due to 1974 Michigan legislation, we have one-year to commence litigation to collect for damage to state property. The present procedure for obtaining copies of the reports relating to state or government damage quite often cannot be obtained soon enough for recovery claims. We would suggest that the responding police agency submit a copy of the report directly to the appropriate state department when state or government property is damaged. This would insure our awareness of that circumstance.

2. Under road surface condition, a fifth category would be added to include a combination wet, snowy, and icy situation. The police officer often checks off more than one box in order to record this condition. However, only one box can be entered into the computerized accident files. The fifth category could be entered to cover the actual surface condition.

3. The police officer often fails to record or clearly record whether the accident occurred during the a.m. or p.m. hours. We would suggest the use of military time for that code or require the officer to shade the appropriate box on the report.

If you have further questions, please contact me.

Sincerely,

Maurice E. Witteveen
Engineer of Traffic and Safety

Ford Motor Company
Room 2078
Automotive Safety Center

20000 Rotunda Drive
Dearborn, Michigan 48121
Mailing Address:
P.O. Box 2053
Dearborn, Michigan 48121-2053

August 13, 1982

Ms. M. J. Compton-Bowman
Research Associate
Highway Safety Research Institute
The University of Michigan
Huron Parkway and Baxter Road
Ann Arbor, Michigan 48109

Dear Mickey:

The Ford Motor Company Safety Research activity is an active user of Michigan accident data and is pleased to respond to your request for suggestions on the data generated from the Michigan Accident Reports (UD-10).

Two areas of improvement would greatly enhance the utility of Michigan traffic accident data without the addition of new questions to the UD-10 forms. These are expanded documentation and more detailed vehicle make-model coding.

Better documentation of existing data elements would improve our ability to understand and interpret the data file. What documentation/ definitions are provided for the police officer, data coder, and outside user? Can UM-ADAAS users have access to more comprehensive documentation?

More specific make-model coding is particularly important for passenger cars, e.g., in studies of car involvement by size. Detailed vehicle information could be derived from VINs (Vehicle Identification Numbers) if they were computer coded. VINs could be transcribed from the UD-10, or UD-10 reported license plate numbers could be passed against Michigan vehicle registration files to determine VINs for in-state vehicles. Individual VINs can then be decoded by computer programs like the R. L. Polk VINa program to determine vehicle make, model, body, model year, etc.

We hope this response will be helpful in your review and we appreciate the opportunity to comment. Please feel free to contact me regarding further aspects of these comments.

Sincerely,

J. C. Marsh

CITY OF ANN ARBOR MICHIGAN

100 North Fifth Avenue, P.O. Box 8647, Ann Arbor, Michigan 48107

August 4, 1982

Office of The Chief of Police

M. J. Compton-Bowman
Highway Safety Research Institute
Huron Parkway and Baxter Road
Ann Arbor, Michigan 48109

Dear Ms. Compton-Bowman,

Reference your letter of July 9, 1982 requesting information on the UD-10 "wish list".

The M.A.L.I. report that is generated from data collected via the UD-10 is a great tool. We had a grant partially funded through your efforts and started our work with the M.A.L.I. report.

Our only criticism of the M.A.L.I. is that it does not cover all addressable intersections in the City and that the data is slow in getting into the system. We are not recommending new data be collected.

This Department has a mini-computer and some of the same information that is available through M.A.L.I. is available "on line" for our jurisdiction.

Thank you very much for your interest in including us in your upgrading effort. Please let us know of any other ways that we can assist you.

I also want to thank you for the tour and informative presentation you put on for myself and members of my staff.

Sincerely,

William J. Corbett

William J. Corbett
Chief of Police

WJC:cac

WASHTENAW COUNTY SHERIFF'S DEPARTMENT

2201 HOGBACK ROAD • P.O. BOX 8645 • ANN ARBOR, MICHIGAN 48107 • (313) 971-8400

THOMAS R. MINICK
SHERIFF

KIRK A. PROFIT
UNDERSHERIFF

July 20, 1982

Highway Safety Research Institute
Huron Parkway & Baxter Rd.
Ann Arbor, MI 48109

ATTENTION: M.J. Compton

Gentlemen:

With regard to your inquiry concerning the Washtenaw County Sheriff Department's use of M.A.L.I. data, the following has been our experience.

M.A.L.I. indexing has been a valuable tool in our analyzation of accident history. As our patrol jurisdiction encompasses the police activities of several other departments, it is convenient to capture raw accident data from one centralized location.

Our data selection requests have been handled most professionally by a highly competent staff at the M.A.L.I. center in Lansing. Although there are some time delays for data requests, I have been able to obtain vital information on an immediate basis.

I will suggest more emphasis be given to selective enforcement with possibly so called "hot print-outs" for areas which reveal high accident frequencies. If the computer could identify "hot areas" at the immediate point in time of development and this information could be promptly forwarded to appropriate agencies, much analyzation and exploratory time could be saved.

Additionally, a better clarification on computer sheet headers would be helpful. At times, the print-outs can be confusing, especially when there lacks a had-been drinking accidents only designation.

In conclusion, I believe M.A.L.I. is a valuable tool necessary for the continuation of Michigan's successful enforcement programs. Any cut back in funds and/or operations would only adversely effect the public's well-being.

Sincerely,

Sgt. Carl Rinna
Traffic Services Division
Washtenaw County Sheriff's Dept.

CR:jw

City of Grand Rapids Michigan
DEPARTMENT OF POLICE

CHIEF OF POLICE
WILLIAM G. HEGARTY

333 MONROE AVENUE, N. W.
GRAND RAPIDS, MICHIGAN 49503

TELEPHONE
616-456-3405

July 19, 1982

M.J. Compton - Bowman
Research Associate
Highway Safety Research Institute
Institute of Science and Technology
Huron Parkway and Baxter Road
Ann Arbor, Michigan 48109

re: U.D. 10 Reports, letter of 7-9-82

Dear Sir:

The Grand Rapids Police Department uses up to 9,000 U.D. 10 reports annually. I would not care to see any changes on the report. For our own use we add the Police Sector and District Number, and we add codes to indicate what type of accident occurred for our own computer printout. I see no reason to impose our requirements on others.

Yours for Highway Safety,

William Hegarty
Chief of Police

Sgt. Samuel R. Johnston

Sergeant Samuel R. Johnston
Traffic Unit, Grand Rapids Police Dept.
333 Monroe Ave. N.W.
Grand Rapids, Michigan 49503
Phone: 616-456-3324

SRJ/lkh

August 30, 1982

Memo to: M. J. Compton-Bowman
from: Alex Wagenaar
Subject: Michigan Data collected on UD-10

My interests and expertise are in the area of Alcohol safety and therefore my use of the Michigan data lies in that area.

In my opinion the Michigan UD-10 HBD box is far superior to other states' recording of suspected drunk driving offenders. Having a box in which to indicate HBD is excellent as opposed to states such as New York where alcohol is only mentioned if it is contributing to the crash.

I would like to eventually see drinking vs drugs suspected as a potential indicator, to separate the drunk driver from the possible drug user. Another possible bit of information, might be "open container" indicated on the form.

MEMO TO: Analyzers of Michigan Accident Data

FROM: Art Wolfe

SUBJECT: Notes on Problems with Some Michigan Variables and Codes (Vehicle Type, Infant Age, Hit and Run Driver, Hazardous Action, Police Action, Helmet Use, and the 1980 Occupant File).

A. Vehicle Type (V63/V104)

(1) From 1972 to 1977 passenger cars were classified by the investigating officer as Fullsize, Intermediate, Compact, Sports, Carryall, or Jeep type. During these years, only about seven percent of the vehicles had missing data on Vehicle Type (V63).

(2) Beginning in 1978, passenger cars were classified by the investigating officer as under 1500 lbs., 1500-2499 lbs., 2500-3500 lbs., over 3500 lbs., carryall, or jeep type. During 1978-1980 missing data on Vehicle Type (V104) has run 9-10 percent.

(3) Both coding schemes are dependent on the officer's best judgment and the resulting vehicle size classifications will not be completely accurate. For example, in 1980, 0.3 percent of the vehicles were classified as weighing under 1500 pounds. However, even this small percentage is higher than would be expected. According to the Annual Automotive News specifications on cars sold in the U.S. there haven't been any cars under 1500 lbs. imported since 1975, and for 1973-75 the only such car was the Lotus Europa (not a very large seller). For 1966-72 there were a few other under-1500-lbs. models (Honda 600, NSU Prinz, Subaru 360, Austin Mini, Fiat 600, etc.), but it is doubtful that they make up even 0.1 percent of the Michigan vehicle population.

(4) For the fifth category, the 1974-77 codebooks say "station bus, carryall, etc.," although the police instructions from 1972 through 1980 only say "carryall." For the 1978-80 codebooks this category was changed to "station wagon, carryall, etc." However, there is no reason to assume that police officers would classify a normal station wagon under "carryall." On the other hand, this category has ranged from 0.7-1.1 percent of the vehicles 1974-80, and one wonders what types of vehicles are included under "carryall." It is clearly more than just the Chevrolet/GMC Suburban and IH Travelall truck-based station

wagons. Perhaps it also includes car-pickups such as the Ford Ranchero and Chevrolet El Camino?

(5) For 1972-77, the seventh vehicle category on the police instructions was "pickup or panel." For 1978 on this was changed to "pickup, panel, or van." However, the 1974-80 codebooks all just say "pickup or panel truck." The percentages in this category have increased from 6.9 percent in 1975 to 11.7 percent in 1980 which seems somewhat more than would be expected despite the substantial increases in pickup and van sales in recent years. For 1977 this category had 9.1 percent of the vehicles, while for 1978 the percentage jumped to 11.2 percent. For the same two years, the "carryall" percent dropped from 1.08 percent to 0.87 percent, suggesting that some vans which might have been put in that category prior to 1978 were put in the "pickup, panel, van" category in 1978 and after.

(6) The eighth category in the 1972-77 instructions was "Straight Truck, Dump, Van, Flat Bed, etc.," while the 1978 instructions have "Stake Truck, Dump, Step Van, Flat Bed, Motor Home, etc." This change in definition was associated with a substantial drop in vehicles classified in this category. Whereas in 1976 it was 3.4 percent and in 1977 it was 3.3 percent, in 1978 it was 1.8 percent and in 1979 it was 1.7 percent. It appears that many small vans must have been classified with straight trucks prior to 1978. Of course, it's also possible that beginning in 1978 some van-type straight trucks have been put in the "pickup, panel, van" category. Unfortunately, "van" has at least two meanings as a vehicle type in the U.S., and it is difficult to know what the officer had in mind when he made his classification.

(7) The ninth category on both the 1972-77 and 1978 instructions says simply "Truck Tractor (semi)." Percentages for this category have varied from 1.2 to 1.6, and there doesn't seem to be any change associated with the change of accident report forms and instructions in 1978. Presumably this category would be used for a truck tractor without an attached trailer, as is implied by the 1974-80 codebook definition "Truck Tractor (semi) or road tractor"--although it is not clear what "or road tractor" adds to the definition.

B. Age Data For Infants (V91/V147)

(1) 1974-77 codebooks begin all pedestrian and injured age variables with category "01. 1 year old." Yet police instructions both in 1972 and 1978 are to enter the person's age as of the last birthday, so for an infant not yet one year old, "00" should be entered. And in fact, there are some "00"s in V91 Driver/Pedestrian Age. Since missing data are coded "99," one hopes that these 00s actually do represent infants under one year old.

(2) In 1978, V147, Driver/Pedestrian Age, has a new category "00. Hit and Run Driver, Driverless Vehicle, or Driver/Pedestrian Less Than 1 Year Old." Presumably by filtering on pedestrian traffic units, one could find infants under one year old in the "00" category. Variables 205, 214, 223, 232, 241, and 250, Ages of the first six injured occupants, also show a "00" category for "less than one year old."

(3) In the 1979 and 1980 codebooks for V147 "hit and run driver" and "driverless vehicle" have been added to "unknown" in the "99" category, and "00" is not listed. There are no "00" pedestrians under age one in the 1980 data. ("00. Less than 1 year old" is listed in the occupant age codes for those years.

C. Hit-Run Driver (V75/V120)

For 1974-1977, the marginals for Variable 75, Hit-Run Drivers, show about four percent of the traffic units as being unapprehended hit and run drivers. For 1978 the same variable, V120, shows a similar 3.7 percent as unapprehended hit and run drivers. However, in 1979 the marginals for V120 show only 0.58 percent as unapprehended hit and run drivers, and in 1980 the marginals show 0 cases of unapprehended hit and run drivers. This is hardly believable and suggests that there is something strange about the 1979 and 1980 data as compared to earlier years.

D. Hazardous Action and Police Action (V80-V81/V125-V126)

(1) On the 1972-1977 accident form there was a box for each traffic unit to describe any hazardous action. This should "indicate the specific violation for each pedestrian, bicyclist, or driver which contributed most to the accident." There was also a summary box at the bottom of the form for indicating when either traffic unit driver had been "cited for a hazardous violation which contributed to the accident" and also when the driver was "cited for a violation which did not contribute to the accident."

(2) The 1978 accident form was changed to enter for each traffic unit a hazardous action number from the selection of hazardous actions on the accident board cover. A box was also added to describe any citation charge for the traffic unit, and the check boxes for any hazardous or other citations were moved from the bottom of the form to a middle row in each traffic unit section. However, the instructions for checking these boxes remained the same.

(3) The code categories for Variable 125, Hazardous Action (V80, 1974-77), and for Variable 126, Police Action (V81, 1974-77), are identical for all seven codebooks (even to the repetition of an "of" where there should be an "or" in the sixth hazardous

action category). However, a comparison of marginals before and after the changed form shows some rather strange differences in the data, especially in the citation data. While the proportion of traffic units with a listed hazardous violation was about 54 percent in 1976 and 1977, it dropped to about 51 percent in 1978 and 1979--a change which may or may not be meaningful. But at the same time the proportion of traffic units shown as receiving a hazardous citation went up substantially from about 22 percent in 1976 and 1977 to about 32 percent in 1978 and 1979. At the same time the proportion receiving a citation for a non-hazardous violation declined drastically, from over two percent in 1976 and 1977 to less than 0.2 percent in 1978 and 1979. Whether these changes are somehow related to the change in accident forms or have some other cause is difficult to determine.

(4) The changes in citations for pedestrians and bicyclists are particularly bizarre. In 1976 and 1977 less than five percent of such traffic units were shown as having been cited for a hazardous violation. In 1978, 1979, and 1980 over 95 percent of pedestrians and bicyclists are shown as having received a hazardous citation--this despite the fact that fewer than 75 percent of pedestrians and bicyclists were coded as having performed a hazardous action. Why this tremendous change took place in 1978 and what "citation" means in the pedestrian/bicyclist context is far from clear.

E. Helmet Use (V137-V138)

(1) The 1978 accident form added a new set of variables having to do with driver, passenger, or nonoccupant helmet use. Helmet yes-no boxes were placed after the age, sex, and injury boxes for each person line. The instructions say to mark the "yes" box when the driver of the vehicle (motorcycle, moped, and motor vehicle) or other person was wearing a helmet. Otherwise the "no" box should be checked. Not surprisingly, there is a lot of missing data on this variable (19.7 percent for drivers in 1980).

(2) V137 is labeled Motorcycle Driver Helmet Use, but this title is misleading because this variable is coded for all drivers, pedestrians, and bicyclists (except bicycle passengers). In 1980, 16 bicyclists and 11 pedestrians (perhaps skateboarders or football players) were coded as wearing helmets. I haven't checked to see if any drivers other than motorcycle drivers were so coded.

(3) Similarly, V138 covers helmet use for all passengers including bicycle passengers--although none of the 89 bicycle passengers in 1980 were coded as wearing a helmet.

F. The 1980 Occupant Data Set (V201-V210)

For some reason the occupant data file in 1980 is not complete. For example, in the Vehicle File, there are 6971 motorcycles, motor scooters, mopeds, etc.; 5604 pedestrians; and 4377 pedalcycles. However, in the Occupant File, there are only 5664 motorcycle drivers; 5384 pedestrians; and 3789 pedalcycle drivers. There are also 1033 motorcycle passengers and 89 pedalcycle passengers. Both drivers and passengers are coded as motorcycle "riders" or pedalcycle "riders" on V202, Occupant Location, but the drivers and passengers may be distinguished on V203, Occupant Position. Category 1 of V202 should read "Driver of a motor vehicle (except a motorcycle, etc.)."