UMTRI-90-2

CARDFILE EVALUATION

Dawn L. Massie
Kenneth L. Campbell

March 1990
FINAL REPORT

The University of Michigan
Transportation Research Institute

Ann Arbor, Michigan 48109-2150

The research reported herein was conducted under general research funds provided by the Motor Vehicle Manufacturers Association and General Motors Corporation. The opinions, findings, and conclusions expressed in this publication are not necessarily those of the MVMA or General Motors.

Technical Report Documentation Page

1. Report No. UMTRI-90-2	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Cardfile Evaluation		5. Report Date March 1990
		6. Performing Organization Code
7. Author(s) Dawn L. Massie and Kenneth L. Campbell		8. Performing Organization Report No. UMTRI-90-2
9. Performing Organization Name and Address Transportation Research Institute The University of Michigan 2901 Baxter Road Ann Arbor, Michigan 48109		10. Work Unit No.
		11. Contract or Grant No. 9168
12. Sponsoring Agency Name and Address Motor Vehicle Manufacturers Association Detroit Michigan and General Motors Research Laboratories Warren, Michigan 48090		13. Type of Report and Period Covered Final Report
		14. Sponsoring Agency Code
15. Supplementary Notes		

16. Abstract

The CARDfile was developed by NHTSA to aid problem identification and countermeasure development in the field of crash avoidance research. This report describes the results of an evaluation of selected CARDfile data elements. While the CARDfile as a whole may be nationally representative, many of the variables of interest to crash avoidance, such as precrash stability, rural/urban, roadway separation, avoidance attempt, and VIN, have 100% missing data rates for one or more of the CARDfile states. The bias introduced by such systematic missing data is likely to invalidate the representativeness of the CARDfile as a whole. Beyond the obvious missing data problem, other examples are described of missing and incompatible code values. Although five of the six states appear to code precrash stability, in fact there is not a single category on this variable that is recorded in every state, and no single state has all the categories. Thus the CARDfile distribution cannot reflect the true distribution of these attributes for the aggregate of the six states. The authors conclude that perhaps the most significant disadvantage of the CARDfile is the appearance of uniformity that is created when, for many variables, this is not the case. In order to make valid inferences from the CARDfile, it is necessary to be familiar not only with the coding of the original state files, but also the additional translations introduced to produce the CARDfile. These problems illustrate the advantages that more uniform accident reporting at the state level would provide.

17. Key Words Collision avoidance, state accident data, missing data, bias	18. Distribution Statement Unlimited		
19. Security Classif. (of this report) None	20. Security Classif. (of this page) None	21. No. of Pages 128	22. Price

Executive Summary

The CARDfile is a crash avoidance database developed by NHTSA to assist researchers in problem identification and countermeasure development in this field. The primary focus of the datafile is on precrash events, and particular features of interest include detailed accident configuration categories, data on vehicle movements and orientation prior to the crash, and vehicle make and model information. The data contained in the CARDfile are derived from automated police accident reports from the six states of Indiana, Maryland, Michigan, Pennsylvania, Texas, and Washington. This has resulted in a very sizable database, which comprises over 20% of the annual police-reported accidents for the nation. The state data are converted to a common coding format in CARDfile, with standardized definitions of variables and attributes.

UMTRI evaluated the CARDfile in order to assess the accuracy and completeness of the data, especially with respect to potential bias between states. It was found that many variables and attributes are unavailable for one or more of the six states included in the CARDfile. The result of this missing data is the exclusion of cases by state when particular analyses are conducted. This non-random omission of data introduces bias into the results of these analyses because the states differ in terms of such factors as climate, accident severity, and composition of the motor vehicle population. An additional concern involves discrepancies in the definitions of the data elements used by the six states. The algorithm used to convert the state data into the CARDfile format was not always able to compensate for these differences. While the structure of the CARDfile creates the appearance of uniformity across all six states, in many instances this is not the case.

The extent to which the observed shortcomings of the CARDfile will affect the utility of the database depends on the particular interests of the researcher. The quality of the data is quite good for many of the variables, such as Weather Conditions, Intersection Signalization, and Roadway Alignment. These variables and others have low missing data rates, consistent definitions from state to state, and very few instances of unavailable attributes. Some of the most interesting variables to crash avoidance research, however, are plagued with the most problems in the CARDfile. For example, Roadway Separation is present for only one state, giving the variable a missing data rate of 92% for the file. While Precrash Stability is available for five of the states, none of the attributes are coded for all the states, and no single state has all of the categories. Thus the aggregate CARDfile distribution bears no resemblance to the true Precrash Stability distribution over the six states. A third example is the Vehicle Identification Number (VIN) variable, which is missing for two-thirds of the cases in the CARDfile. VIN is critical for the accuracy check it can provide on such information as the make, model, and year of a vehicle, as well as for identifying vehicles with more specific features of interest. The CARDfile is intended to support the exploration of the relationship between vehicle

design characteristics and crash propensity, and the high missing data rate on VIN severely compromises this application.

The authors consider one of the primary advantages of the CARDfile to be its large size, comprising over 4 million accident and 7 million vehicle records, which facilitates the analyses of rare events. Another is that the combined information from the six states has been shown to be nationally representative in terms of several factors. A third is the convenience and efficiency provided the researcher as compared to working with six individual state files, each with a different format.

These benefits of the CARDfile must be weighed against its disadvantages when considering the appropriateness of its use for any given purpose. When entire states drop out of an analysis because the data are not available, the national representation is likely to be compromised. If certain attributes are coded for the cases from some states but not others, the aggregate result will be distorted. Less obvious, without the undertaking of additional research, are the inconsistencies in code definitions between states. Individual data elements in the CARDfile do not always carry the same meaning from state to state, but this is not well documented.

The convenience provided by the common format of the CARDfile is perhaps its most dangerous advantage. Conversion to the CARDfile format reduces the original state data to their lowest common denominator, sometimes resulting in losses of detail and meaning. Instead of eliminating the need to be familiar with the coding and data processing details of each state, in fact the CARDfile translation preserves this need and adds the necessity of familiarity with the translation process itself. The conversion of state data to CARDfile results in removing this information another step from the original event. The researcher should therefore be aware of the data manipulation involved in producing the CARDfile and attempt to consider the sources and effects of bias introduced into an analysis through missing data and/or coding inconsistencies. Without this awareness, misleading conclusions could easily be obtained.

The creation of the CARDfile was an ambitious and worthwhile undertaking. Many of its difficulties are the result of the different approaches to accident reporting used by the states. Overall, NHTSA was successful in extracting the common information present in the original state files. The quality of the database would benefit tremendously if more uniform accident reporting could be implemented at the state level, particularly with respect to compatible definitions for common data elements.

Contents

Tables

1 Introduction

The Crash Avoidance Research Data file (CARDfile) was developed by NHTSA to aid problem identification and countermeasure development in the field of crash avoidance research. More specifically, it is intended to facilitate the study of relationships between vehicle design characteristics and crash propensity.[1] The primary focus of the CARDfile is on pre-crash, as opposed to post-crash events, so that the movements of the involved vehicles just prior to the collision are emphasized. The crash data contained in the CARDfile are derived from police accident reports from six states: Indiana, Maryland, Michigan, Pennsylvania, Texas, and Washington. The state data are converted to a common coding scheme in CARDfile, with standardized definitions of variables and attributes.

The CARDfile contains three subfiles, one each for accident, vehicle, and driver variables (see Appendix A). The information found in the accident file is common to all vehicles and drivers involved in a particular crash. This subfile includes variables describing the roadway and environmental conditions, the location of the accident, and the general type of crash. Information that is unique to each involved vehicle, such as the pre-crash actions, the vehicle make, model, and type, and the number of injured occupants, is found in the vehicle subfile. The driver subfile includes such information as the age and sex of the driver of each vehicle. Driver characteristics are not a major emphasis of the CARDfile, and consequently there are comparatively few variables in this subfile.

The CARDfile was not specifically designed to be a representative sample of the national crash experience. It has been claimed, however, that conclusions based on analyses conducted with the CARDfile data may be projected to the nation as a whole. One justification cited for this view concerns the sheer size of the database.[2] Since CARDfile contains information on more than 20 percent of the country's annual police-reported accidents, this is expected to reduce sample bias. Another argument that has been made in support of CARDfile's representativeness is that many of the problems likely to be addressed using the database involve characteristics of vehicles, such as make, model, and type, that are thought to show proportionally little variation between states.[3] Finally, comparisons of CARDfile data with NASS (National Accident Sampling System) data[4] and comparisons of the general characteristics of the six CARDfile states with the

---

1. J.J. McDonough and M.L. Edwards, *CARDfile Handbook*. National Highway Traffic Safety Administration, May 1987.

2. M.L. Edwards, *A Database for Crash Avoidance Research*. Warrendale, Pennsylvania: Society of Automotive Engineers, Paper No. 870345, 1987.

3. McDonough and Edwards, *CARDfile Handbook*, p. 2-6.

4. Edwards, *Database for Crash Avoidance*; McDonough and Edwards, *CARDfile Handbook*.

same characteristics for the nation as a whole[5] have suggested that CARDfile is reflective of the nation's crash experience.

A primary objective of our evaluation of the CARDfile is an assessment of the accuracy and completeness of the data, especially with respect to potential bias between states. One concern involves the possibility of missing data causing certain states to be over- or under-represented in particular analyses. Another concern is discrepancies between states in the definitions or availability of variable attributes. A third form of bias would involve errors made in converting the original state files to the CARDfile coding scheme.

Since CARDfile is intended to provide information of use to crash avoidance research, the evaluation focused on variables that are of key interest in this field. These include the pre-crash maneuvers; condition of the environment, roadway, and drivers; type of crash; and severity of the accident.

The version of the CARDfile examined contains data for the three years 1984, 1985, and 1986. The project used the files maintained by the UMTRI Transportation Data Center. Separate files for each CARDfile state are documented in Data Set Codebooks.[6,7,8,9,10,11] This is a very sizable database that includes 4,203,444 accident records and 7,341,293 vehicle and driver records. When possible, the analyses incorporated all of the data records, but due to the prohibitive expense of conducting computer runs on such a large dataset, some of the analyses are based on 5% random sample files that were prepared for this project from the three-year accident and vehicle files. This is noted in the text wherever applicable.

---

5. S. Salvatore, P. Mengert, and R. Walter, *CARDfile Data Base Representativeness, Phase I: General Characteristics including Populations, Vehicles, Roads, and Fatal Accidents*. Cambridge, Mass.: Transportation Systems Center, Project Memorandum No. DOT-TSC-HS802-PM-88-16, August 1988.

6. *CARDfile: Maryland, 1984-1986*. Ann Arbor, Michigan: The Transportation Data Center of the University of Michigan Transportation Research Institute. Data Set Codebook Number 89-5, January 1989.

7. *CARDfile: Michigan, 1984-1986*. Ann Arbor, Michigan: The Transportation Data Center of the University of Michigan Transportation Research Institute. Data Set Codebook Number 89-1, January 1989.

8. *CARDfile: Indiana, 1984-1986*. Ann Arbor, Michigan: The Transportation Data Center of the University of Michigan Transportation Research Institute. Data Set Codebook Number 89-2, January 1989.

9. *CARDfile: Pennsylvania, 1984-1986*. Ann Arbor, Michigan: The Transportation Data Center of the University of Michigan Transportation Research Institute. Data Set Codebook Number 88-20, December 1988.

10. *CARDfile: Texas, 1984-1986*. Ann Arbor, Michigan: The Transportation Data Center of the University of Michigan Transportation Research Institute. Data Set Codebook Number 89-4, January 1989.

11. *CARDfile: Washington, 1984-1986*. Ann Arbor, Michigan: The Transportation Data Center of the University of Michigan Transportation Research Institute. Data Set Codebook Number 89-3, January 1989.

documented in Data Set Codebooks.[6,7,8,9,10,11] This is a very sizable database that includes 4,203,444 accident records and 7,341,293 vehicle and driver records. When possible, the analyses incorporated all of the data records, but due to the prohibitive expense of conducting computer runs on such a large dataset, some of the analyses are based on 5% random sample files that were prepared for this project from the three-year accident and vehicle files. This is noted in the text wherever applicable.

Section 2 of this report discusses possible biases between the states in CARDfile due to missing data, coding discrepancies, and unavailable variables and attributes for a number of CARDfile variables. Sections 3 and 4 provide a similar but more in-depth treatment of the variables for Accident Type, Vehicle Type, Make/Model Code, and Vehicle Identification Number (VIN), since these are of particular interest to crash avoidance research. Section 5 describes the results of comparisons between three of the original state accident files that are maintained by UMTRI with the data contained in the CARDfile. The results of computer runs done as consistency checks between variables and others done to produce two-way tables illustrating the bias of missing data are discussed in Section 6. A brief discussion of the representativeness of CARDfile and a comparison of CARDfile data with FARS data on alcohol use are included in Section 7. The final two sections of the report summarize our findings and discuss the implications for the use of CARDfile in crash avoidance research.

---

6. *CARDfile: Maryland, 1984-1986.* Ann Arbor, Michigan: The Transportation Data Center of the University of Michigan Transportation Research Institute. Data Set Codebook Number 89-5, January 1989.

7. *CARDfile: Michigan, 1984-1986.* Ann Arbor, Michigan: The Transportation Data Center of the University of Michigan Transportation Research Institute. Data Set Codebook Number 89-1, January 1989.

8. *CARDfile: Indiana, 1984-1986.* Ann Arbor, Michigan: The Transportation Data Center of the University of Michigan Transportation Research Institute. Data Set Codebook Number 89-2, January 1989.

9. *CARDfile: Pennsylvania, 1984-1986.* Ann Arbor, Michigan: The Transportation Data Center of the University of Michigan Transportation Research Institute. Data Set Codebook Number 88-20, December 1988.

10. *CARDfile: Texas, 1984-1986.* Ann Arbor, Michigan: The Transportation Data Center of the University of Michigan Transportation Research Institute. Data Set Codebook Number 89-4, January 1989.

11. *CARDfile: Washington, 1984-1986.* Ann Arbor, Michigan: The Transportation Data Center of the University of Michigan Transportation Research Institute. Data Set Codebook Number 89-3, January 1989.

2 State Level Coding Discrepancies and Missing Data

One of the inherent difficulties in creating a database as large and ambitious as the CARDfile comes from attempting to convert different coding schemes to a standard format. Since the six states in the CARDfile each have their own approach to reporting accidents, the options available on the police reports and the manner in which they are coded are not uniform from state to state. Not all of the states record all of the variables of interest to crash avoidance research, and the states differ in their coding of the data so that variable categories intended to be separate options in CARDfile may have already been combined at the state coding level. The results of translating these disparate approaches to a single scheme are higher rates of missing data and inconsistencies in the definitions of code categories.

The accident reporting criteria of the six states are summarized in Table 1 in Appendix B, and below. As the table indicates, Maryland and Pennsylvania have the highest reporting threshold (tow-away) of the six CARDfile states, so one would expect those two state files to contain a greater proportion of more severe accidents than the other four state files. Differences in reporting threshold imply differences between the states with respect to the types of accidents that are available for analysis in CARDfile. This topic is discussed further in the next section on Accident Severity.

State	Reporting Threshold
Indiana	$200 property damage
Maryland	Vehicle towed away
Michigan	$200 property damage
Pennsylvania	Vehicle towed away
Texas	$250 property damage
Washington	$300 property damage

The unavailable information in the CARDfile is a result of data that are not collected at the state level, as well as data that are collected, but coded in such a way that the information cannot be extracted for conversion into the CARDfile format. Missing variables and incompatible code categories are a possible source of bias in the resulting CARDfile. This is of greatest concern when one or more states do not code a given variable at all. CARDfile variables that are entirely missing for one or more states are summarized below.

Missing Variables by State

Missing Variables	State
Roadway Separation	Indiana Maryland Michigan Pennsylvania Texas
Vehicle Identification Number	Indiana Pennsylvania Texas Washington
Land Use (rural/urban)	Pennsylvania Texas
Precrash Stability	Indiana*
Avoidance Attempt	Maryland
Roadway Profile	Michigan

*Variable missing in 1984 and 1985; only the jackknife code is available in 1986.

Each of the variables listed above was included in the evaluation since the entire variable is missing in one or more states. Differences in the accident experience from one state to another create a potential for bias in any analysis using one or more of these variables.

Even when a variable is included in a state file, the absence of individual variable attributes also can alter the resulting CARDfile data in ways that are less obvious. For example, each state has a variable indicating the type of vehicle. However, three states, Indiana, Texas, and Washington, do not distinguish trucks with double trailers from other large trucks. Consequently, even though the CARDfile Vehicle Type variable includes a code for "doubles," the proportion of doubles shown in the CARDfile substantially understates the true proportion for the six states since the doubles are identified in only three of the six CARDfile states. Other variables that are nominally present, but have serious missing attribute problems are Precrash Stability, Component Failure, and Avoidance Attempt. Precrash Stability is clearly the worst variable in the CARDfile from the standpoint of missing attributes. One or more of the four main attributes of this variable are unavailable in every one of the six states, and there is no single attribute that all of the states code. Thus, the CARDfile cannot provide an accurate overall proportion for any of the Precrash Stability codes. Because of their relevance to collision avoidance research, Precrash Stability and Avoidance Attempt are discussed in further detail later in this section.

Tables 2 and 3 in Appendix B provide a complete listing of the variables and attributes that are unavailable in particular states in the CARDfile, noting where possible how the cases pertaining to these unavailable attributes were coded. The information in these tables was compiled from the CARDfile Handbook[12] and the UMTRI Data Set Codebooks cited previously (see notes 6 through 11).

Table 4 summarizes the resulting overall missing data rates for twenty-two variables based on all the cases, from all six states, contained in the 1984-1986 CARDfile. The rate for each variable was determined by the percentage of cases coded as either "Missing" or "Unknown", or as "No Driver Record" for the Driver variables. This takes into account entire variables that are unavailable for particular states, since in these instances all of the cases are coded as "Missing" in CARDfile. As the table indicates, the proportion of missing data varies widely between variables. Many variables have very low missing data rates, some below 1%. Unfortunately, some of the variables of prime concern to crash avoidance research, such as Roadway Separation, Precrash Stability, Avoidance Attempt, and Land Use, are characterized by very high rates of missing data.

Percent Missing Data
for Selected CARDfile Variables

Variable	Percent Missing
Accident File	
Roadway Separation	92%
Land Use (rural/urban)	43%
Roadway Profile	27%
Vehicle File	
Precrash Stability	70%
Vehicle Identification Number	66%
Avoidance Attempt	44%
Restraint Use	35%
Driver Age	12%
Alcohol/Drug Use	11%
Driver Error	11%

While missing data is an obvious source of potential bias in a dataset, differing definitions of code categories, in this case between states, are of equal concern. The "common format" in which the CARDfile data are presented tends to mask these underlying state to state differences. In order to gain insight into these hidden discrepancies, a review was conducted of the various state coding approaches in which the state police reporting forms, instructions to the police, state coding manuals, and CARDfile conversion algorithms were examined. In the remainder of this section the results of this review are discussed in terms of the effects of coding problems on particular variables of interest to crash avoidance research. Tables 5 through 28 in Appendix B accompany this discussion.

12. J.J. McDonough and M.L. Edwards, *CARDfile Handbook*. National Highway Traffic Safety Administration, May 1987.

2.1 Accident Severity

The CARDfile Accident Severity variable indicates the most severe injury sustained by any individual involved in the crash. The attributes for this variable are "Fatal Injury (K)", "Incapacitating Injury (A)", "Non-incapacitating Injury (B)", "Possible Injury (C)", and "Property Damage Only" (PDO). As Table 5 indicates, the state distributions for Accident Severity vary widely, especially with respect to the proportion of property-damage-only crashes. This is probably linked to the different accident reporting requirements in each state (Table 1). Accidents resulting in only property damage are 77% of the cases in Indiana, a state that requires reporting of all accidents involving at least $200 worth of damage, but they make up only 35% of the crashes in Pennsylvania, where accidents not resulting in injury or death must be reported only if an involved vehicle has to be removed from the scene by towing.

A recent report substantiates the link between reporting threshold and accident severity.[13] In 1977, Pennsylvania changed from a $200-damage threshold to the tow-away standard used currently. In the two-year period following this change, the ratio of PDO's to casualty accidents decreased by 75%.[14] Similarly, in 1979 Maryland abandoned their $100-damage limit in favor of the tow-away threshold. The number of reported PDO's decreased by 50% in the two years following the changed reporting requirement.[15]

The differences in reporting threshold from state to state influence more than just the distribution by accident severity. The probability of injury or fatality varies with the type of collision. Pedestrian accidents have a higher probability of injury, whereas rear-end accidents are generally less likely to produce injury. Pennsylvania, with a tow-away reporting threshold, has the lowest percentage of rear-end collisions (11.3% as compared to the overall CARDfile percentage of 18.6%) and the second highest percentage of pedestrian/cyclist collisions (7.3% as compared to 5.8% overall, see Table 17). Thus differences in the reporting threshold affect the distribution of collision types in the state files. The distribution of other variables may also be expected to vary with the reporting threshold. These differences in the reporting threshold have implications for the potential bias of missing data that may exclude whole states from the analysis.

In Table 5A, the property-damage-only crashes have been excluded in order to facilitate comparisons of distributions of casualty accidents. Variations among the states are still apparent, such as the proportion of "Incapacitating Injuries" in Maryland that is over five times that of Indiana (23.6% to 4.5%). Such variability suggests that the states may be using different coding definitions for the injury categories.

---

13. Z.A. Sabra, E.C. Noel, B.V. Chatfield, and R.W. Eck, *Trends in Highway Information*. McLean, Virginia: Federal Highway Administration, Publication No. FHWA-RD-88-055, July 1988.

14. Ibid., p. 14.

15. Ibid., p. 15.

To examine this idea, the state coding schemes were compared for the Accident Severity variable. The attributes and definitions used by Maryland, Michigan, Texas, and Washington appear to correspond exactly to those of the CARDfile variable, so there are no obvious problems with data conversion for these states. Pennsylvania and Indiana code for severity in a slightly different manner, however, which could lead to bias in the data.

Pennsylvania has the greatest proportion of "Possible Injury" cases out of all six states (Tables 5 and 5A). In CARDfile, this attribute is intended for non-evident injuries, such as limping, complaint of pain, and nausea. The Pennsylvania code category that is converted to "Possible Injury" includes as part of its criteria injuries that "can be treated by first aid application whether at the scene or in medical facilities." It would seem from the description that this includes visible injuries, some of which might be more likely coded as "Non-incapacitating Injuries" in the other states.

Indiana codes a "Nature of Injury" variable with attributes indicating particular types of injuries. The degree to which these are incapacitating is not clear for every case, which suggests that errors may arise in converting the data to the CARDfile format. One source of coding error may come from converting the Indiana attribute "Fracture/Dislocation" to "Non-incapacitating Injury" in CARDfile, since such an injury is often of an incapacitating nature. This may be partially responsible for Indiana having the lowest relative frequency of "Incapacitating Injuries" out of the six states (Tables 5 and 5A). Indiana also codes an "Injury Status" variable that was not taken advantage of by the CARDfile conversion algorithm. The use of this variable could increase the accuracy of data conversion into CARDfile, since it indicates the degree of consciousness of the accident victim, and may be useful in distinguishing incapacitating from non-incapacitating injuries.

The overall amount of missing data for Accident Severity is a mere 0.04% of the cases (Table 4). Pennsylvania is the only state with any cases that are unknown for this variable. UMTRI maintains accident files from Michigan, Texas, and Washington, and it was noted that the original files from these three states also contained no missing data for their respective accident severity variables. While Accident Severity is a fundamental variable for which one would expect a low missing data rate, the complete absence of missing data suggests that when the injury information is unknown or missing, the accident is classified as property-damage-only. Thus, the true proportion of injury is likely to be understated.

2.2 Land Use

The CARDfile Land Use variable is intended to describe the character of the land area in the vicinity of the accident site. Its main attributes are "Urban" and "Rural". The primary shortcoming of Land Use is that it is not coded in Pennsylvania or Texas as shown in Table 9, resulting in an overall missing data rate of 42.5% for the entire CARDfile (Table 4).

It is difficult to evaluate coding discrepancies between the states for Land Use because the descriptions of the criteria for the attributes that are contained in the CARDfile Handbook and state coding manuals are not very specific. Both Indiana and Maryland base the rural versus urban dichotomy on whether the accident occurred within or outside the limits of a legal corporation, but there is no indication for either state of the population levels that would be implied. The CARDfile conversion algorithm for the Michigan cases codes all accidents occurring in townships as "Rural" and all those in cities as "Urban". This classification does not necessarily imply anything about the population of the locality either, since many of the townships have populations larger than some of the cities. Washington defines its "Urban" attribute as "an area including, and adjacent to, a municipality or other known place of 5000 or more population".

The state distributions shown in Table 9 reveal that Maryland and Washington have virtually identical proportions of "Urban" accidents in the CARDfile, with the figures for both states roughly 79.5%. This is somewhat surprising given that Maryland has a population density almost seven times greater than that of Washington (461.0 versus 66.3 people per square mile[16]).

However, the population density of the immediate area in which the accident occurred is the primary characteristic that one would like the rural/urban variable to distinguish. The FHWA uses boundaries established by the states around all cities with populations of 5,000 or more to distinguish urban from rural roads. The boundaries are chosen based on the local population density, rather than the municipal boundaries, so that densely populated areas adjacent to a municipality will be included in the "urban" area. For cities of less than 5,000, the municipal boundary is used. These boundaries are the best distinctions between rural and urban for traffic safety related work, and are used by the NHTSA Fatal Accident Reporting System to define accidents in urban areas.

Only Washington seems to use a coding that is similar to the FHWA definition. The differences in the proportions of rural versus urban accidents from state to state may reflect differences in the *definition* of Land Use as well as actual land use differences. There is apparently no way to compensate for these differences in definition when the data are converted to CARDfile format.

2.3 Primary Impact

The Primary Impact variable in the CARDfile describes the object initially impacted in the crash. The state distributions for this variable seem reasonable overall (Table 10), and missing data rates are low. The only unavailable attributes are "Unknown" and "Non-collision" for Michigan. Such cases are included under the "Other" category of Primary Impact in CARDfile because all three attributes form one category according to Michigan's coding format. The only other anomalies suggested by the distributions in Table 10 are unusually high rates of collisions with parked motor vehicles in Maryland and of collisions with fixed objects in Pennsylvania, but an examination of the coding procedures for these two states did not reveal any explanation for these variations, other than real differences between the states.

16. *Information Please Almanac*. 42d ed. Boston: Houghton Mifflin, 1989.

The review of the state coding procedures did produce one possible example of discrepancies in code definitions. Two of the attributes for Primary Impact are "Another Motor Vehicle in Transport" and "Another Motor Vehicle Not in Transport". According to the CARDfile Handbook, the latter is intended to include "parked, driverless, or towed vehicles". For both Maryland and Michigan, the code categories that are converted to "Another Motor Vehicle in Transport" in CARDfile include vehicles stopped or abandoned on the roadway that are not in designated parking areas. It would seem that these cases would be more accurately described by the "Another Motor Vehicle Not in Transport" category.

2.4 Primary Impact Location

This CARDfile variable indicates the location of the primary impact with respect to the traveled portion of the roadway. The main attributes are "On Roadway", "On Shoulder", and "Off Roadway" (Table 11). Neither Michigan or Washington codes the "On Shoulder" attribute. Michigan defines "On Roadway" as "shoulder to shoulder less legal parking areas", which means that its shoulder accidents are included under the "On Roadway" attribute in CARDfile. As an aside, the CARDfile Handbook incorrectly states that these cases are included with "Off Roadway". The Washington state coding manual does not specifically mention how shoulder accidents are handled, so it cannot be determined whether they are included with the CARDfile "On Roadway" or "Off Roadway" cases. "On Shoulder" accidents make up from 3 to 12% of the cases in each of the four states that code the attribute, so the unavailability of "On Shoulder" in Michigan and Washington likely affects only a modest proportion of those cases.

2.5 Relation to Intersection

The main attributes for this variable are "Intersection Related", "Not Intersection Related", and "Driveway Related" (Table 12). The CARDfile Handbook states that the Relation to Intersection variable "indicates whether or not the primary impact occurred within the bounds of an intersection". This implies that the "Intersection Related" code should be used only for accidents that take place at an intersection, not for those where the accident is in some way attributable to the intersection, but not actually located in the intersection. However, this does not accurately describe the coding conversions into CARDfile for any of the states, except Pennsylvania.

The Indiana "Intersection Related" attribute in CARDfile includes crashes that take place within an actual intersection, those that are in close proximity to an intersection, and those that are intersection related. Similarly, that option for Maryland includes both accidents occurring within the limits of an intersection and those caused by events at an intersection. For Texas as well, the state coding options of "Intersection" and "Intersection Related" are both converted to "Intersection Related" in CARDfile.

It is interesting to note that the variable selected from the Michigan state data for conversion to the CARDfile Relation to Intersection variable is Highway Area Type, with the options of "Interchange Area", "Intersection

Area", and "Neither Interchange Nor Intersection Area". "Intersection Area", defined as "normally within 100 feet in any direction from the intersection, but farther if the accident is accountable to the intersection", is converted to "Intersection Related" in CARDfile, which again is a much broader definition than that given in the CARDfile Handbook. The other two codes under the Michigan Highway Area Type variable are converted to "Not Intersection Related" in CARDfile, resulting in the unavailability of the "Driveway Related" attribute. Michigan also codes a variable that indicates whether the accident actually occurred at an intersection, at a driveway access, or not at an intersection. Had this variable been converted to the Relation to Intersection variable in CARDfile, the result would have been no missing attributes, code assignments that match the CARDfile Handbook description, and a decrease in the frequency of "Intersection Related" accidents for Michigan.

In the case of Washington, accidents that it codes as not taking place at an intersection, but related to the intersection, are converted to the CARDfile "Intersection Related" attribute. Furthermore, crashes that occur at an intersection, but under circumstances not related to the intersection, are converted to the CARDfile "Not Intersection Related" attribute. With the exception of these Washington and Pennsylvania cases, the CARDfile "Intersection Related" code seems to consistently include accidents occurring within the bounds of an intersection, in close proximity to the intersection, or that are "intersection related." This broader definition would seem more appropriate for the CARDfile Handbook.

2.6 Roadway Profile

The Roadway Profile variable indicates the vertical alignment of the roadway where the crash took place and has attributes of "Level" and "Grade" (Table 15). The variable is not available for Michigan, which largely accounts for the overall missing data rate of 27% (Table 4). As indicated in Table 15, 99.5% of the Texas CARDfile cases are coded as "Level". While the state of Texas is not marked by a high degree of topographic diversity, the fact that the similarly flat state of Indiana reports proportionately 35 times as many "Grade" cases for Roadway Profile as Texas makes the Texas figures appear improbable. Since Texas accounts for over 32% of the accidents in CARDfile, its apparent coding anomaly may affect the overall distribution of Roadway Profile.

2.7 Roadway Separation

Roadway Separation is intended to indicate the roadway design with respect to the separation of the opposing traffic lanes. This information is of key interest to crash avoidance research since the driving conditions and distribution of collision types vary considerably on divided as opposed to undivided roadways. It is thus unfortunate that only Washington codes Roadway Separation, giving it the highest overall missing data rate, 92%, of all the variables in the CARDfile, as shown in Table 16.

2.8 Precrash Stability

Another critical variable in crash avoidance research is Precrash Stability, which indicates the stability of each vehicle just before the accident. The main attributes are "Tracking, No Skidding", "Skidding", "Spinning", and "Jackknifing". Unfortunately this variable is plagued with difficulties.

One or more of the Precrash Stability attributes are unavailable in all six of the states, and there is no single attribute that all of the states code (Table 21). Before 1986, when the "Jackknifing" option alone became available, none of the attributes of this variable were coded in Indiana. The "Tracking" code is not available for Michigan, and "Skidding" is not recorded for Pennsylvania. "Spinning" is not available for Maryland, Michigan, Texas, or Washington, and "Jackknifing" is not an option for Maryland or Michigan.

Another limitation with Precrash Stability is that Texas codes this information at the accident level, so the pre-crash movements of individual vehicles cannot be distinguished. Therefore, only the single-vehicle crashes in Texas are coded for Precrash Stability, with all of the multi-vehicle cases classed as missing data. The missing data rate for Texas alone for this variable is 88.8%.

An additional potential problem in the coding of Precrash Stability is that in the conversion algorithms for Texas and Washington, the "Tracking" option is used as a default. It is likely that some cases that are actually unknown are included under "Tracking", although it is impossible to determine the number of cases so affected. Texas and Washington do not code a Precrash Stability variable themselves, so the variable is generated in CARDfile by selecting options from other variables in those two states. If a case is not coded under one of those options, it defaults to "Tracking".

The unavailable attributes, varying coding procedures between states, and high proportion of missing data combine to severely restrict the utility of Precrash Stability. The overall missing data rate of 70% for this variable understates the true missing data rate since another 5% are assigned by default to "tracking" in Texas and Washington. Potential bias from missing data is especially a problem, since no one attribute of Precrash Stability may be examined across all the states, and no state may be used for an analysis of all the attributes.

2.9 Avoidance Attempt

The Avoidance Attempt variable indicates for individual vehicles whether an attempt was made prior to the crash to avoid contact with another vehicle, person, or object (Table 22). In terms of code availability (Table 3), Maryland does not code the variable at all, and Indiana has three unavailable attributes, for attempts made to avoid pedestrians/pedalcyclists, vehicles, and unknown objects. The unknown object option is not available for Michigan or Washington either. As was the case for Precrash Stability, in Texas only single-vehicle crashes are coded for Avoidance Attempt, with all other cases appearing as missing data. Finally, Indiana uses the "No

Indication of Avoidance Attempt" option as a default, even though some such cases may be unknown for this variable.

The overall missing data rate for Avoidance Attempt is 44%, the third highest of the CARDfile variables (Table 4). Furthermore, 48% of the cases were coded as "No Indication of Avoidance Attempt" and an additional 6.5% as not applicable (used for parked cars). This results in less than 1.5% of all the vehicles in the CARDfile actually being coded as making some particular avoidance attempt prior to the crash.

2.10 Driver Error

The Driver Error variable shown in Table 27 is used to indicate the primary error made by the driver that brought about the crash. The attributes for this variable are listed in footnote (a) of Table 28. The state coding procedures for this variable were extensively reviewed in order to illustrate the problems confronted when converting different coding schemes to a single standard format. Because of the way the states' variable attributes are categorized, they do not neatly match the CARDfile Driver Error attributes. Table 28 presents the state coding discrepancies on the Driver Error variable and indicates that many of the CARDfile attributes are based on different criteria, depending on the state.

The first part of Table 28 lists those items considered to be driver errors by one or more of the states that were not assigned the same CARDfile Driver Error code for all of the states. For example, the first item listed is "use of alcohol". Cases coded in this manner by the states receive the CARDfile "Other Driver Error" code for the Maryland and Washington cases but the "No Error Indicated" code for the Indiana and Pennsylvania cases. "Use of alcohol" is not included in the Michigan and Texas conversion algorithms for Driver Error.

The second portion of the table lists driver errors, coded only by single states, where a seemingly inappropriate code value was assigned. For example, Maryland cases coded as "blinded by approaching vehicle" receive the "Other Driver Error" code in CARDfile. There are several other examples, listed in both parts of the table, of Maryland cases being coded "Other Driver Error" when they do not seem to have involved any driver error at all. This may account for Maryland having the highest proportion of "Other Error" cases out of all the states, with 37% of its crashes being coded as such (see Table 27).

The Driver Error variable provides an example of the inherent difficulties in converting state data to the standard format of the CARDfile. Presenting the CARDfile Driver Error data by state, as in Table 27, hides the fact that not all of the attributes have the same meaning for all of the states. Many of these coding problems are irreconcilable given the way the data are originally recorded by the states. For many of the instances listed in Table 28, any choice on how to convert the data into CARDfile would have resulted in some data misclassification. On the other hand some of the coding discrepancies, such as the use of the "Other Error" code for Maryland, are apparently the result of mistakes in the conversion algorithm itself and could be resolved.

2.11 Summary

This section has discussed differences in the coded data between the six states in CARDfile. Several factors contribute to the variations in state by state distributions of CARDfile variables, all of which are of concern with respect to potential bias in analyses that use the data. These factors include inherently different crash experiences in each state, different accident reporting thresholds used by the states, and state to state differences in the variables recorded and categorized for coding.

One of the purposes of the CARDfile is to provide a standardized format allowing for the common analysis of data from the six states included in the database. For many of the variables discussed in this section, this goal has been achieved in appearance only. Standardizing disparate coding schemes results in other difficulties. These include missing data, unavailable variables and attributes for particular states, and discrepancies in the coding criteria between the different CARDfile states. While large missing data rates and unavailable variables and attributes may be recognized and therefore taken into account during data analysis, the underlying differences are more subtle and may be missed. Data presented in a common format create the impression that variable attributes have the same meaning throughout the file, when this is not always the case. Taking Driver Error as an example, the variable seems fairly reasonable at first glance since there are only three instances of attributes being unavailable in particular states (Tables 3 and 27), and the missing data rate is moderate at 10.6% (Table 4). What would be missed without extensive additional research, however, is the variation from state to state in terms of what each Driver Error attribute actually implies (Table 28).

This section was not written as an exhaustive account of all instances of missing data and coding discrepancies in the CARDfile but as a description of ten variables that have such problems to a degree where potential bias in analyses using them is a concern. Taken together, the capsule summaries of the difficulties with each variable serve to illustrate the variety of coding problems present in the CARDfile. These range from very high rates of missing data rendering Roadway Separation virtually useless, to state coding differences resulting in none of the Precrash Stability attributes being available in all the states, to variations in reporting requirements affecting state to state distributions of Accident Severity.

It should be stressed that not all of the variables in the CARDfile are plagued by such difficulties. Several of the variables that were examined in detail seem to be free of any major problems; that is, their distributions are reasonably similar from state to state, their missing data rates are low, and they have few if any unavailable attributes. These variables include Light Conditions (Table 6), Weather Conditions (Table 7), Road Surface (Table 8), Intersection Signalization (Table 13), Roadway Alignment (Table 14), and Model Year (Table 19). There are undoubtedly other CARDfile variables that were not examined that are also relatively free of problems involving coding inconsistencies and missing data.

As a final note, several other variables of key interest to crash avoidance research will be treated in a manner similar to that of this section, but in more detail, in the next two sections of this report. These variables include Accident Type, Vehicle Type, Make/Model, and Vehicle Identification Number (VIN).

3 Review of the Accident Type Variable

The CARDfile Accident Type variable is one of key concern to collision avoidance research. Because it is a complex variable that represents a great deal of information, a detailed examination of how it is generated from the original state files is warranted. We reviewed the Accident Type coding procedures for single-vehicle accidents for all of the states and those for multi-vehicle crashes for Maryland and Texas. Only two states were reviewed for the multi-vehicle codes owing to the lengthy and involved nature of the translation algorithms.

The Accident Type variable is coded at both the accident and vehicle levels, with variable 20 used for the former and variable 102 for the latter in the UMTRI files. The distributions of these two variables are shown by state in Tables 17 and 18 respectively. Three broad types of collisions are classified under Accident Type, single-vehicle crashes, two-vehicle crashes, and collisions involving more than two vehicles. The single-vehicle crashes are coded according to the object of primary impact; for example, collisions with a stationary object or collisions with a parked vehicle. The two-vehicle crashes are grouped into categories reflecting the pre-crash paths of the vehicles, their direction of travel, and the general type of accident: a sideswipe for example. These categories are further subdivided at the vehicle level into codes that specifically indicate the pre-crash actions of both vehicles, e.g., one vehicle overtaking the other on the right. All collisions involving more than two vehicles receive the same code, 911, since a more detailed breakdown of such crashes is not possible given the data provided by the states.

3.1 Single-Vehicle Codes

Almost 34% of the cases in the 1984-1986 CARDfile dataset are classified as "single-vehicle" accidents, making them a common class of collisions. There are seven available attributes for single- vehicle crashes under the Accident Type variable, namely "Rollover", "Stationary Object", "Pedestrian/Animal/Pedalcyclist", "Parked Vehicle", "Noncollision", "Specifics Unknown", and "Other." Most of the single-vehicle cases in the CARDfile are assigned one of the specific Accident Type attributes, with only 4.25% of the cases coded as "Specifics Unknown" or "Other." However, it is important to note some accidents including more than one vehicle are assigned to the "single-vehicle" category. For example, all rollovers, even if involving two vehicles, are coded as "single-vehicle" accidents. In these cases, an additional code of 000 is available at the vehicle level for the other vehicle involved in such a "single-vehicle" accident. The CARDfile Handbook gives the example of 000 being used to identify the parked vehicle in a parked vehicle collision. However, the 000 code is applied in other instances as well, and it would be helpful if the Handbook were more explicit in describing the use of this code.

For the Maryland and Washington cases, all parked vehicles involved in collisions receive the 000 code. On the other hand, no parked vehicles are coded 000 in Michigan or Texas because neither state records information on parked cars. Between these two extremes, most, but not all, parked vehicles involved in collisions are coded 000 in Indiana and Pennsylvania. These two states variably record such crashes as involving either one or two vehicles. When the crash is considered to involve two vehicles, the parked one receives the 000 code.

All six states use the 000 code for vehicles involved in two-vehicle rollover accidents but which do not roll over themselves. Presumably this is usually a matter of the rollover vehicle impacting the 000 vehicle, although this is not clear. The 000 code was not used for these vehicles for the Indiana and Maryland rollover cases until 1986. Texas also uses the 000 code for other vehicles involved in fixed object collisions. This might be a case in which a car impacts a fixed object and then another car, with the struck car coded 000. However, neither the Handbook nor examining the conversion algorithm itself makes this clear.

There is apparently an error in the Washington conversion algorithm concerning rollover accidents. For rollovers in which two vehicles were involved, both vehicles are assigned the 000 code, when in fact the actual rollover vehicle should have received the 110 rollover code.

According to the CARDfile Handbook, the rollover code is given priority over the other Accident Type codes, and an accident should be coded as such whether the rollover occurred as the first harmful event or as a subsequent event. For Maryland and Pennsylvania, rollovers occurring as subsequent events receive the rollover code in CARDfile, but this is not possible for Indiana, Michigan, and Washington since those states only code information pertaining to the first harmful event. It would be possible to include subsequent-event rollovers for Texas cases, given the way Texas codes them, but the CARDfile conversion algorithm for Texas only considers rollovers that occur as the first event. Thus more accidents are potentially considered rollovers for Maryland and Pennsylvania than for the other four CARDfile states. Since not all the states record subsequent-event information, it would seem more consistent for CARDfile to include only primary-event rollovers for all six states.

It is only for rollovers that the CARDfile Handbook mentions the topic of multi-event accidents. It is not clear how other subsequent events are normally handled in CARDfile. It would appear that the Accident Type variable pertains only to the first harmful event, except for rollovers in two of the states and for the use of the 000 code, with other subsequent events being excluded from consideration. More explicit documentation concerning the treatment of primary and subsequent events in the CARDfile would be helpful.

Rollovers are always coded as "single-vehicle" accidents. When assigning most of the other single-vehicle codes, the conversion algorithms used for most of the state files restrict the cases considered to crashes involving only one vehicle. However, the Maryland and Washington algorithms consider both one- and two-vehicle crashes in the assignment of "single-vehicle" codes and produce coding errors in the process. In some cases, the same single-vehicle accident code is assigned to both vehicles, where, to be consistent with the other states, the other involved vehicle should have been coded 000. For example, in the case of a car hitting a tree and then impacting another vehicle, the Maryland and Washington algorithms would assign the fixed object code to both vehicles, not just the car that actually collided with the tree. The result is higher frequencies for single-vehicle codes at the vehicle level (variable 102) than the accident level (variable 20) for both Maryland and Washington (see Tables 17 and 18).

An additional coding problem was noted for Maryland concerning the single-vehicle code 116, "Other", which is used in Maryland for collisions with trains. The frequency of this code exactly doubles between the accident and vehicle level because Maryland records all such crashes as involving two vehicles. This is incorrect since trains are not considered motor vehicles in the CARDfile. The frequency of vehicles involved in 116 accidents in Maryland is therefore twice as high as it should be.

3.2 Multi-Vehicle Codes

Due to the complexity of the accident type coding for multiple-vehicle accidents, the coding was reviewed in only two states, Maryland and Texas. The findings are described separately for each state in the remainder of this section.

Maryland

The section of the CARDfile conversion algorithm that generates Accident Type for Maryland incorporates several variables recorded by the Maryland state police. These include the number of vehicles involved in the accident; the movement of the vehicles, for example, accelerating, changing lanes, parking, backing, or making a right turn; the first harmful event, which includes such options as overturned, fixed object, and other motor vehicle in transport; the point of impact and the areas damaged on each vehicle; and the collision type. This last classifies collisions according to the pre-crash orientation and actions of the involved vehicles, for example, "same direction, both vehicles going straight, rear-end" or "opposite directions, both vehicles turning left".

The CARDfile Accident Type variable codes four main types of sideswipes: passing, overtaking right, overtaking left, and changing lanes (Table 17). In the CARDfile, "overtaking" apparently implies a deliberate attempt to pass, usually involving a lane change. "Changing lanes" is coded when the sideswipe actually occurs when one vehicle changes lanes and hits another one. "Passing" sideswipes take place when one vehicle happens to be going past another with no deliberate attempt to overtake or change lanes. Overtaking accidents where the relative positions of the two vehicles are unknown are included under the "passing" option as well.

According to the Maryland police report coding manual, however, "overtaking" and "passing" are synonymous. The CARDfile passing sideswipe cases are derived from the "passing" option of Maryland's movement of vehicles variable, and the overtaking right and overtaking left codes are unavailable for Maryland because there are no corresponding options on the Maryland police report. Presumably cases that CARDfile would consider "overtaking" are combined with "passing" or "lane change" cases by Maryland and are represented under one of those codes in the CARDfile.

Two of the codes for the group of collisions classified under "Initial opposite direction – change trafficway/turn across path" are not available for Maryland. One involves a right-turning vehicle colliding with one that is stopped or going straight, and the other is for the collision of a right-turning and left-turning vehicle. These two configurations are not options under Maryland's collision type variable, and such cases end up being coded in CARDfile as 811, the generic code for two-vehicle collisions.

There are two other Accident Type codes that are unavailable in Maryland because the configurations are not options under Maryland's collision type variable. One involves two vehicles approaching on intersecting paths where one turns right into the other, which is proceeding straight, resulting in the two vehicles being aligned in opposite directions. The other is the collision of two vehicles proceeding on curved and intersecting paths. Both of these types of collisions are coded 811 for the Maryland cases.

Texas

The variables used to generate Accident Type for Texas include the number of vehicles involved in the accident; the vehicle movements/manner of collision, for example, two motor vehicles approaching at an angle, with one straight and the other backing; the first harmful event; the damage scale; the direction of travel; and the other factors variable, which includes such options as vehicle changing lanes or vehicle passing on the left.

It appears that some Texas cases placed under the CARDfile rear-end codes would be more accurately described by some of the other Accident Type codes. All of the accidents classified by Texas as two motor vehicles going the same direction with one going straight and the other stopped are given the rear-end with lead vehicle stopped code in CARDfile. No additional data are considered by the conversion algorithm to verify that such cases are in fact rear-ends, and some of them should probably fall under the "sideswipe—passing" category. Similarly, all cases that Texas codes as two motor vehicles going the same direction with one going straight and the other turning receive the rear-end with lead vehicle turning code in CARDfile. Some of these should probably be coded under the Accident Type code described as initial same direction, change trafficway/turn across path, one vehicle straight or stopped and the other turning.

For the sideswipe collisions, the Texas data enable the distinction between passing and overtaking accidents to be made in a manner consistent with the CARDfile criteria. The changing lanes code is not available for Texas, however. Since Texas codes for this on the accident level, it cannot be determined which vehicle actually changed lanes. By default these cases for Texas pass to code 811, but it would be possible to code them under "sideswipe – specifics unknown". The latter seems preferable since it would preserve the information that they are sideswipe accidents.

For the head-on sideswipe class of accidents, the lateral move/lane change code is unavailable for Texas, again because it cannot be determined which vehicle made the move. Such collisions apparently are combined with the "both straight/passing" cases. It would have been possible instead to code the lateral move cases under "head-on sideswipe—specifics unknown", which seems more appropriate.

The "Initial same direction—change trafficway/turn across path" category has codes for a right-turning vehicle colliding with a straight or stopped vehicle and a for a left-turning vehicle colliding with the same. When these codes are generated for Texas, however, only stopped vehicles are included, not straight-moving vehicles as well. As mentioned above, this is because all of the Texas cases involving two vehicles going in the same direction, one straight and one turning, are coded as rear-ends in CARDfile.

A final note concerns the use of code 911, which is used for all collisions involving more than two vehicles. Since Texas only records vehicle and driver data if the vehicle was involved in the first harmful event or if a driver or passenger was injured, many vehicles involved in multi-vehicle crashes are not recorded for Texas. If frequencies for code 911 are compared between the accident level and vehicle level, the other five CARDfile states all average about 3.2 vehicles per collision, while Texas only averages 2.04 vehicles. Thus, at least a third of Texas vehicles involved in 911 crashes are not represented in the frequencies for this code. This should be kept in mind if 911 frequencies are used at the vehicle level instead of the accident level, since the Texas vehicle counts are artificially low.

3.3 Summary

In this section, various state to state discrepancies in the Accident Type coding procedures, based on a review of the conversion algorithms, have been described. The shortcomings have been emphasized, but in general most of the CARDfile codes appear to be accurate and consistent between the states. Given the complexity of the Accident Type variable and the very different ways that the individual states have of recording and coding this information, this is quite an achievement.

We noted a few differences between states in the derivation of the Accident Type single-vehicle codes that result in inconsistencies in the final data. The states vary in terms of assigning the 000 code to parked vehicles involved in collisions. The CARDfile Handbook does not clearly describe the use of the 000 code or state exactly what constitutes another "involved" vehicle in a single-vehicle crash. It was also noted that vehicles in the Maryland and Washington files that are peripherally involved in single-vehicle accidents are given the same codes as the primary vehicles, whereas this is not done for the other four states.

The coding procedures for the multi-vehicle codes were reviewed only for Maryland and Texas, so we cannot comment on the accuracy of coding these cases for the other four states. Because Maryland does not distinguish between passing and overtaking accidents, the "overtaking sideswipe" codes are not available for Maryland in the CARDfile, with such cases instead falling under the "passing" or "changing lanes" codes. There are four other multi-vehicle codes that are unavailable for Maryland, but this is the unfortunate result of the configurations not being considered as possible options on the Maryland state police reports, not due to any error in the CARDfile translation algorithm itself.

For Texas, two of the rear-end codes appear to be assigned to some cases that are not actually rear-end crashes at the expense of three of the other multi-vehicle codes. The conversion algorithm could take advantage of additional data provided by Texas to correct this. Two other instances were noted that involve codes that are unavailable for Texas where the cases were instead assigned other codes that do not appear to be the most appropriate choices.

Finally, the availability of code attributes in each state for Accident Type should be briefly discussed. While all of the single-vehicle attributes are available across the board, there are several unavailable multi-vehicle codes in each state (Table 2). This is hardly surprising given the great number of attributes listed under Accident Type, but differential availability of the codes between states could affect the outcomes of particular analyses.

Some of the unavailable attributes are of little consequence, since they are unspecific codes and/or they would apply to just a tiny fraction of the cases. An example is the rear-end, direction unknown code not being available for five states. In other instances the code unavailability is more of a problem, such as two of the sideswipe attributes not being available in two of the states and another one not being available in a third state. Analyses involving such cases should consider the possible biasing effects of the differential code availability.

4 Vehicle Type, Make/Model, and VIN Information

One of the purposes of the CARDfile is to allow the relationship of vehicle design to crash propensity to be explored. While the CARDfile does not contain any specific design characteristic variables, it does include certain variables that enable the analysis of design characteristics of interest. The Vehicle Type variable classifies motor vehicles into categories such as passenger car, light van or truck, and medium/heavy straight truck. Vehicle makes and models are coded under another variable, which includes options like Ford Escort and Honda Civic. The ability to identify specific makes and models allows the analyst to identify groups of vehicles with and without a specific design feature. Other variables useful in vehicle design analyses are Model Year and Vehicle Identification Number (VIN). Possible concerns about this body of data in CARDfile include differences in vehicle composition from state to state, the availability of make/model and VIN information, and how this varies among states, among different types of vehicles, and between domestic versus imported vehicles.

4.1 Vehicle Type

The CARDfile Vehicle Type variable classifies vehicles into 19 different types, with two additional categories for missing and unknown data (see Table 20 for a complete listing of Vehicle Type by state). Each state except Pennsylvania has at least one unavailable attribute, and there are instances of certain vehicle types being combined with others (Table 3). For example, the CARDfile Vehicle Type variable distinguishes between a tractor with a semitrailer and a tractor with a double trailer. Not all of the states make similar distinctions, however. In Texas, all tractors with any number of trailers are placed in one category, which is then converted to the tractor with semitrailer option in CARDfile. This means that the CARDfile Vehicle Type attribute "tractor with double trailer" is not available for Texas and that all such vehicles are included under the "tractor with semitrailer" code. Tractors with double trailers are also not identified in Washington and in Indiana before 1986 (Table 3). Thus, the CARDfile vehicle type variable understates the proportion of doubles by perhaps a factor of two.

Coding problems of this nature cannot always be avoided. If states combine code categories in their own files, they cannot then be separated for conversion to the CARDfile format. While the decisions made about handling state to state coding discrepancies when translating to CARDfile's coding scheme may have been reasonable, the underlying problem remains. If a particular type of vehicle is not coded consistently between states, as in the case of tractors with double trailers, the utility of the data will be undermined.

In Table 29 the Vehicle Type variable has been recoded into the three general classes of passenger cars, light trucks, and heavy trucks so that the different state distributions may be compared. The table is based on the complete CARDfile from 1984 through 1986 and summarizes data contained in Table 20. The most notable contrast is that Texas and Washington have relatively more light trucks and fewer passenger cars than the other four CARDfile states. Such differences between states in the types of vehicles involved in crashes might bias the results of analyses conducted on subsets of the six states. Note that the rows of Table 29 do not sum to 100% because not all of the different vehicle types are represented.

As an example of potential bias, suppose an analysis were being conducted on the types of vehicles involved in crashes on roadways with a grade. Since the Roadway Profile variable is unavailable for Michigan (Table 2), all of the Michigan cases would be excluded from the analysis. Because the states vary in their distribution of vehicle types, the result of losing the Michigan data would be to artificially lower the proportion of passenger cars included in the analysis and increase that of light trucks.

Another vehicle type coding error deserves mention before moving on to a discussion of make/model information. Prior to 1987, the layout of the vehicle type section of the Michigan traffic accident report resulted in a significant overstatement of bobtails (tractors without a trailer). Most of the vehicles misclassified as bobtails were actually tractor-semitrailer combinations. The problem was a result of separate boxes on the form for type of vehicle and trailer and occurred when "truck tractor (semi)" was indicated for type but the trailer box mistakenly left blank. Table 20 lists the number of bobtails (which are referred to in CARDfile as "heavy articulated trucks without trailer attached") in Michigan as 7046, but the actual figure is closer to 1500. This should not be a problem with later versions of the CARDfile, since Michigan has revised its accident report form.

The accident report forms for the other CARDfile states were reviewed to determine the potential for similar coding problems with bobtails. Indiana, Pennsylvania, Texas, and Washington each have a distinct "bobtail" option included with all of the other vehicle types, so there is no reason to expect miscoding of bobtails in those states. Like Michigan, however, Maryland has separate vehicle type and trailer blanks on its form, and a bobtail will be indicated when "truck/road tractor" is selected for vehicle type and the trailer box is left blank. Maryland does not even have a "none" option under trailer as does Michigan that would allow for specific coding of bobtails. It is therefore probable that a significant number of tractor-trailer combination vehicles are coded as bobtails in Maryland, thus giving Maryland proportionately the highest frequency of bobtails in the CARDfile (Table 20).

4.2 Make/Model

The six states included in the CARDfile each have their own system of recording makes and models of vehicles. Some use numeric codes while others employ alphabetic designations; some have standardized code values while others do not. In order to provide uniformity between the states, the algorithm used to generate the CARDfile Make/Model variable incorporates the vehicle type, make, and model data supplied by the states to assign cases code values based on the Fatal Accident Reporting System (FARS) make/model coding scheme. The one exception is the Michigan data, since that state records only makes, not models, of vehicles. Since Michigan does record the vehicle identification number, CARDfile employs a VIN-decoding program to derive models for the Michigan cases.

In the FARS make/model coding system, the first two digits of the code represent the vehicle make and the second two digits represent the vehicle model. The system is hierarchical in that certain sets of codes represent more specific information than others. The different levels of codes are briefly described below.

A. **Make known/Model known**—When both the make and model of a vehicle are recorded by the state, the CARDfile conversion program is usually able to assign the appropriate FARS make/model code. Occasional exceptions occur, however, as with the case of Toyota Corollas and Coronas in Maryland and Washington. Both of these states record make/model information using non-standard alphabetic abbreviations, which can result in the ambiguous model code of "CORO" being listed for a Toyota car. The CARDfile conversion algorithm is able to overcome this if the model year for the vehicle is 1983 or later, since Coronas were not manufactured after 1982. The earlier cases cannot be distinguished and are assigned the code of "unknown Toyota passenger car" (see category B below).

B. **Make known/Model unknown/Vehicle Type known**—When the make but not the model is recorded by the state, the CARDfile Vehicle Type variable is used to assign a code representing the make and the type of vehicle. Codes ending in 69 are used for motorcycles of particular makes, e.g., "unknown Yamaha motorcycle"; 79 codes for light trucks, e.g., "unknown Chevrolet light truck"; 89 codes for heavy trucks, e.g., "unknown Ford truck"; 99 codes for passenger cars, e.g., "unknown Cadillac automobile"; and 97 codes for other types of vehicles, e.g., "Volkswagen other vehicle".

C. **Make known/Model unknown/Vehicle Type unknown**—When the make of a vehicle is known but there is no information on model or vehicle type, a code ending in 00 is used to represent an unknown type of vehicle of a particular make. These codes are available only for the major manufacturers, e.g., "unknown Dodge".

D. **Make unknown/Vehicle Type known**—When the vehicle make is not recorded by the state but the type of vehicle is known, one of five codes, one for each of the major vehicle types, is assigned. These each have 99 as the first two digits, e.g., "unknown make, light truck".

E. **Make unknown/Model unknown/Vehicle Type unknown**—Those cases where the make, model, and vehicle type are all unknown are assigned to one of three codes: 1—Missing, 2—Unknown, or 9900—Unknown make, unknown model, unknown vehicle type.

Computer runs that we conducted on a 5% random sample of the 1984-1986 CARDfile indicate that codes representing specific vehicle makes were assigned to 91.1% of the cases, and codes representing specific makes *and* models to 61.2% of the cases. We found these proportions to vary from state to state, however, as indicated in Table 30. The percentages in the table are based on the computer runs using the 5% sample and are listed in terms of 95% confidence intervals for the whole file. The raw counts are derived by multiplying the total number of vehicle cases in each state for the entire 1984-1986 database by the respective percentage.

Except for Pennsylvania, vehicle make is known for about 90% or more of the cases in every state. The rates drop off sharply in all six states for vehicle model, however, especially in Indiana. Knowledge of the particular model is crucial for many analyses of design characteristics, such as those concerning turn signal color or the presence of center high-mounted stop-lights. The low overall rate of known vehicle model automatically results in a great loss of data, and the variation in the rates between states is a likely source of bias.

Because of the hierarchical nature of the Make/Model variable, some codes assigned cases with unknown vehicle models contain more useful information than others. For example, if the particular vehicle type and make are indicated, certain analyses can still be carried out, especially by combining this information with the model year. On the other hand, if not even the vehicle make is known, the code is of little utility.

The CARDfile Make/Model variable was recoded into categories indicating whether or not vehicle make and vehicle type were known, for all cases where vehicle model was unknown. The distributions of cases falling into these categories, by state, are presented in Table 31. From left to right respectively the columns represent the Make/Model categories B, C, D, and E described earlier. As the table indicates, the states vary in terms of how they treat cases where the model is unknown. Michigan and Pennsylvania have noticeably low rates of known vehicle make for these cases. The data in Table 31 are the results of runs conducted on the 5% random sample of the three-year file.

The vehicle type variable was recoded into the general classes of passenger cars, light trucks, heavy trucks, and all other types of vehicles to see if the rates of known makes and models varied with respect to vehicle type. In Table 32, the percentages are again based on a 5% sample of the 1984-1986 CARDfile, with the raw counts and confidence intervals derived in a manner similar to that described for Table 30. Table 32 indicates that passenger cars have the highest proportion of known makes, followed by light trucks, heavy trucks, and other types of vehicles. Differences between classes of vehicles are even more pronounced with respect to model information. While the model is known for 76% of the passenger cars, this rate drops to 34% for light trucks and to even lower percentages for the heavy trucks and other vehicles.

It is not surprising that passenger cars have the highest rates of recording make and model information, since cars are by far the most common type of vehicle contained in the CARDfile, comprising 71% of the cases overall. Other types of vehicles are less common and present coding problems, with the result being a higher missing data rate. To see if some states are more successful than others in recording make/model information within a particular class of vehicle, frequencies and percentages of known makes and models for passenger cars and for light trucks by state were calculated as before and are presented in Table 33.

Since most of the analyses conducted on the CARDfile are likely to involve passenger cars, it is important to be aware of variation between states in recording make/model information of cars. Table 33 indicates that the six states have comparable rates of recording passenger car makes, but the model information varies widely and has a particularly low known rate in Indiana, at about 36%. Therefore, using the passenger car model information in CARDfile has the effect of selecting cars from particular states over others.

Light trucks are an area of increasing research interest in crash avoidance. The overall known make/model rates calculated earlier for light trucks (Table 32) mask a great deal of variability between states (Table 33). In terms of make, the states all have high known rates, except for Pennsylvania where this information is nearly 100% missing. This is because the usual make/model recording procedure in Pennsylvania is to assign specific codes only to passenger cars, coding all other vehicle types as not applicable. For light truck model information, the known rates vary tremendously between states. Only Michigan, at 87%, has a high known rate of light truck models, and all of this information is derived from the vehicle identification numbers.

In order to examine variability between states with respect to the origin of manufacture of vehicles, the Make/Model variable was recoded into domestic and foreign categories by assigning by vehicle make. The distinction was made based on the ownership of the company, so that, for example, American Motors, Chrysler, Dodge, Plymouth, Ford, Lincoln, Mercury, and General Motors were all considered domestic makes, while BMW, Datsun, Honda, Mazda, Mercedes-Benz, Renault, Subaru, Toyota, Volkswagen, and Volvo were all considered foreign. This classification avoids ambiguous situations, such as foreign parts in vehicles that are assembled in this country, by considering only whether the company is of foreign or domestic ownership.

Computer runs were conducted on a 5% random sample of the 1984-1986 CARDfile to ascertain the proportion of foreign cars and light trucks in each state. Cases of unknown make were excluded from the calculations since the origin of manufacture could not be determined. Raw counts were then derived for the entire three-year database by taking the estimated number of vehicles of known make in each state and multiplying by the proportion of foreign vehicles based on the 5% sample. The raw counts and percentages of foreign cars and light trucks are shown in Table 34. The table indicates, for example, that about 9.8%, or 69,204, of the Indiana car cases of known make are of foreign manufacture. The proportion of foreign cars and light trucks varies widely between the states, with Michigan and Indiana reporting the lowest percentages and Washington the highest. This has important consequences because, as will be shown, the degree of specificity of model information coded in the CARDfile is related to the origin of manufacture of the vehicle.

It is possible that some of the states actually have a somewhat higher proportion of imported vehicles than indicated. Taking passenger cars as an example, Salvatore et al.[17] report on the percentage of imports registered in each of the six states. These figures are listed in the second column of Table 35. The third column of the table lists the proportion of foreign passenger cars represented in the CARDfile, and the fourth lists the proportion of CARDfile passenger cars of unknown make. So, for example, in the case of Indiana, 11.1% of its registered cars are of foreign make, 9.8% of its CARDfile cars of known make are foreign, and 2.8% of its CARDfile cars are of unknown make.

Most of the states show a fairly good correspondence between the proportion of registered foreign cars and CARDfile foreign cars, with the exception of Maryland and especially Michigan. For these two states it would appear that either imported cars are substantially under-involved in crashes, or, more plausibly, that vehicle make is much less likely to get recorded for foreign cars than for domestic ones.

To see if there is a difference in the likelihood of vehicle model being recorded depending on whether the vehicle is of domestic or foreign origin, the percentage of cases of known model was calculated by state for the four categories of domestic and foreign cars and light trucks. The percentages shown in Table 36 are based on a 5% random sample of the CARDfile, and the raw counts were estimated from those. The figures are based on cases of known vehicle make, since origin of manufacture cannot be determined otherwise.

Differential recording of model information according to origin of manufacture is of concern because, as was demonstrated earlier, some states have considerably more foreign vehicles than others. As the table indicates, for the CARDfile in general and for most of the individual states, the model information is available for more of the domestic than the foreign vehicles, and this holds for both cars and light trucks. Therefore in model-specific questions, domestic models will be over-represented, and the states with the highest proportion of domestic vehicles, Michigan and Indiana, will be over-represented as well.

4.3 Vehicle Identification Number (VIN)

The importance of including VIN information in a crash avoidance research database should be emphasized here. Not only may the VIN be used as a means of identifying vehicles with particular features of interest, such as engine or transmission type, but it also provides the opportunity for consistency checks on other variables. In the case of the CARDfile, the VIN could be used to verify the information on the variables of Model Year and Make/Model, and it could also be used to derive more specific information such as the model series of particular vehicles.

It is unfortunate that Maryland and Michigan are the only two CARDfile states that record vehicle identification numbers, so that the VIN

17. S. Salvatore, P. Mengert, and R. Walter, *CARDfile Data Base Representativeness, Phase I: General Characteristics including Populations, Vehicles, Roads, and Fatal Accidents*. Cambridge, Mass.: Transportation Systems Center, Project Memorandum No. DOT-TSC-HS802-PM-88-16, August 1988.

variable has a missing data rate of 100% in Indiana, Pennsylvania, Texas, and Washington. For the CARDfile as a whole, the VIN missing data rate is 66.3%. This means that for nearly two-thirds of the cases contained in the CARDfile, VIN cannot be used to verify existing data or to derive more specific vehicle information. Table 37 provides a breakdown of VIN presence according to vehicle type, with the percentages based on a 5% random sample of the CARDfile and the raw counts estimated from those. The percentages in the table are expressed as 95% confidence intervals for the whole three-year file. The table indicates that while the VIN is recorded for about 36% of the passenger car cases in the CARDfile, it is known for only about 29% of the light trucks.

For the CARDfile as a whole, the presence of VIN's appears to be associated with more specific make/model information. Table 38 splits cases according to whether or not the VIN was recorded and lists the proportion and estimated raw counts of known vehicle models for each group. These figures are also based on the 5% sample of the CARDfile, and 95% confidence intervals are provided for the percentages. Table 38 indicates that when the VIN is recorded, there is a greater chance of the model being known as well, and this is overwhelmingly the case for light trucks. The reader should keep in mind that comparisons based on the presence or absence of the VIN are essentially comparisons of Michigan and Maryland versus the other four states. Thus, the comparison presented in Table 38 is partially a reflection of VIN-generated model data (for Michigan) versus state-reported model data.

Similarly, the coding of the VIN seems to be associated with assigning a case to a particular vehicle type, although there is no documentation for the CARDfile indicating that either Michigan or Maryland derive the vehicle type from the VIN. For cases where the VIN is known, Vehicle Type is coded as "Missing" or "Unknown" only 2.9% of the time, as compared with 5.6% for the cases without a VIN. Out of the 131,212 Michigan and Maryland records in the 5% sample file, VIN is unknown for 5.8%. The vehicle model is unknown for all of these unknown-VIN cases. Similarly, 72.6% of the unknown-VIN cases in Michigan and Maryland are unknown for vehicle type, as compared to 2.9% for the known-VIN cases. It may be that when information as specific as the VIN is recorded, other more general information, including the type of vehicle, is more likely to have been obtained as well. On the other hand, more complete vehicle information may be a direct consequence of VIN-decoding procedures used on the Michigan and Maryland data.

4.4 Summary

The variables of Vehicle Type, Make/Model Code, and VIN are critical to identification of vehicles with particular design characteristics of interest. One factor affecting the utility of these variables is their proportion of missing data, which is modest for vehicle make (8.9%) and vehicle type (4.7%), but high for vehicle model (38.8%) and VIN (66.3%). Of the four variables, vehicle model is probably the most useful for subsetting the data according to design characteristics, so its high missing data rate is of particular concern.

An additional concern about the make/model data contained in CARDfile is its accuracy. It is generally recognized that reporting police officers commit occasional errors when recording the make and model of a vehicle because of the detail involved. However, if the VIN is also available, it may be used to validate the make/model information. Since VIN's include a check digit to verify that they have been correctly transcribed, make/model information based on accurate VIN's should be free of errors. Researchers with a serious interest in specific vehicle models usually require the additional accuracy provided by the VIN. It is a significant drawback that only one-third of the cases in CARDfile include the VIN. A possible remedy for this situation would be to replace the four CARDfile states that do not record VIN information with others that do.

Having the VIN available for the cases from all six states in the CARDfile would also decrease the high missing data rate on vehicle model. Currently the model information in CARDfile is VIN-derived only for the Michigan data. It is probably not coincidental that, out of the six CARDfile states, the Michigan cases have the highest overall percentage of known model at 77% (Table 30), that they have by far the highest known model rate for light trucks at 87% (Table 33), and that for foreign vehicles, both passenger cars and light trucks, again the model is known for far more of the Michigan cases than those from the other states. If VIN information were obtained from other states, it could be used to lower the high overall make/model missing data rate.

Universal availability of VIN information in the CARDfile would also reduce the bias between states that results from the current wide variability in the known make/model rate. For passenger cars, the states are comparable in their rate of recording make but vary greatly in terms of model, with this coded for only 36% of the Indiana cases. For the light trucks, the proportion of known makes is roughly equal from state to state, except for Pennsylvania with a 99.3% missing data rate. There is a great deal of variation between states in terms of the known model rate for light trucks.

Another factor influencing the usefulness of the data concerns differences in vehicle composition between states. In terms of vehicle type, there are substantially fewer cars and more light trucks in Texas and Washington as compared to the other four CARDfile states. The origin of manufacture of vehicles varies greatly between states as well. For cases where the vehicle make could be determined, the percentage of foreign cars and light trucks ranged from 4.4% in Michigan to 30.5% in Washington. One concern about these disparities is the bias they impose on the make/model data due to general relationships that exist between that variable and vehicle type and origin of manufacture. Comparing Make/Model to Vehicle Type for the CARDfile as a whole, the proportion of unknown make/model data increases from passenger cars to light trucks to heavy trucks. Another comparison indicated that more specific model information was available for domestic than foreign vehicles.

5 Comparison of Original State Files with CARDfile

UMTRI maintains files from three of the states included in the CARDfile, namely Michigan, Texas, and Washington. This enabled us to compare the distributions of variables between the original files and the translated coding of the CARDfile in order to check the accuracy of the CARDfile conversion algorithms. This was done for a series of variables using the Michigan 1984 cases, Texas 1986 cases, and Washington 1985 cases.

For variables that underwent little or no recoding between the state files and CARDfile, it was possible to compare the frequencies listed in the state codebooks with those in the CARDfile codebooks. For CARDfile variables that were derived from more than one state variable, computer runs were written to convert the state variables into the CARDfile variables so that the resulting frequencies could be compared with those listed in the CARDfile codebooks. The computer runs were based on the conversion algorithms used by CARDfile.

5.1 Michigan 1984

Distributions of the following variables were compared between the 1984 Michigan state file at UMTRI and the 1984 Michigan cases contained in the CARDfile:

Month of Crash	Relation to Intersection
Accident Severity	Intersection Signalization
Light Conditions	Road Alignment
Weather Conditions	Vehicle Type
Road Surface	Precrash Stability
Primary Impact	Avoidance Attempt
Primary Impact Location	Driver Sex

The frequencies were found to correspond for all of the variables except for Vehicle Type. Table 39 lists the frequencies for the attributes of the Vehicle Type variable based on the original Michigan file and the CARDfile. For most of the attributes the frequencies match exactly, but discrepancies exist for several of the categories. The Michigan file contains more passenger cars and straight trucks than CARDfile, while the latter has more light trucks and "other" types of vehicles.

The computer algorithm that we used on the Michigan data for Vehicle Type followed that used by CARDfile; that is, all of the variables and attributes were recoded in the same manner as the CARDfile algorithm. If the CARDfile algorithm was actually used to generate the Vehicle Type data, our figures should have been virtually identical to theirs, with the only difference due to the inexplicable existence of seventy extra cases in the CARDfile version of the Michigan file. A possible explanation for the differences would be CARDfile using VIN information to verify Michigan's coding of vehicle types, resulting in a sizable number of passenger cars and straight trucks being reclassified as light trucks. While this scenario is consistent with the types of vehicles that did and did not show a change in frequencies between the two datasets, there is no indication in any of the documentation examined that this was in fact done.

5.2 Texas 1986

Distributions of the following variables were compared between the 1986 Texas state file at UMTRI and the 1986 Texas cases contained in the CARDfile:

Month of Crash	Relation to Intersection
Accident Severity	Road Alignment
Light Conditions	Alcohol/Drug Use
Weather Conditions	Intersection Signalization
Road Surface	Precrash Stability
Primary Impact	Avoidance Attempt
Primary Impact Location	Driver Sex

There were no major discrepancies for any of the variables between the two files. Because there are 389 more accident cases and 940 more vehicle cases in the Texas file at UMTRI than in CARDfile, there are occasional differences between variable distributions of about one-tenth of one percent.

5.3 Washington 1985

Distributions of the following variables were compared between the 1985 Washington state file and the 1985 Washington cases contained in the CARDfile:

Month of Crash	Relation to Intersection
Light Conditions	Road Alignment
Weather Conditions	Alcohol/Drug Use
Road Surface	Primary Impact
Land Use	Intersection Signalization
Primary Impact Location	Vehicle Type

No discrepancies were noted for any of the variables between the two files. In every instance the distributions matched to within one-tenth of one percent.

5.4 Summary

Distributions were compared for a total of 40 variables between the original state file data maintained at UMTRI and the data contained in the CARDfile. Each comparison included all the cases from a particular state for a particular year. For 39 out of the 40 comparisons, the pairs of distributions were found to match each other. This suggests that the transformation algorithms used to generate the CARDfile function as they were intended. For the one variable where the distributions did not correspond, Vehicle Type in Michigan, it is suspected that CARDfile used a VIN decoding algorithm, or some other additional code conversion procedure, to modify the vehicle type designations assigned by Michigan.

6 Consistency Checks and Illustrations of Bias

Almost all of the previous material has focused on one-way distributions of the CARDfile variables, or two-way distributions of the variables by state. This section presents additional two-way distributions that do not focus on state to state differences. These tables are organized in two groups. The first group presents three consistency checks that compare pairs of variables that contain essentially similar information. These concern vehicle type, accident type, and weather versus road surface condition. The second group of tables is intended to illustrate the bias that results when variables with high missing data rates are used. Land Use (rural/urban) and Roadway Profile are the focus in this group of tables.

6.1 Consistency Checks

In earlier sections, the degree of coding consistency between the six states in CARDfile was discussed in terms of coding definitions and the availability of variable attributes. This section describes the results of three two-way tables run on pairs of CARDfile variables. These provide a check on the internal consistency of the CARDfile coding scheme.

To examine the degree of consistency in coding vehicle type, the variables of Vehicle Type and Make/Model were run against each other using data from a 5% random sample of the three-year file (Table 40). Both of the variables were recoded into the following eight categories of vehicle type: missing/unknown, motorcycle, passenger car, light truck/van, medium/heavy truck, bus, motor home, and other vehicle type. Table 40 indicates that there are proportionately very few instances of cases coded as different vehicle types under the two variables (e.g., as passenger car under Vehicle Type and as light truck/van under Make/Model). There are some minor discrepancies in the coding of vehicle type between the two variables, however.

The Make/Model variable has specific codes available for motor homes and buses, for example, 8580 – Kenworth motor home, or 8785 – Peterbilt bus, but these codes are little used. None of the 275 vehicles classified as motor homes under Vehicle Type were assigned a motor home code under Make/Model, and only 6.7% of the Vehicle Type "bus" cases received a bus code under Make/Model. The majority of the cases for both of these vehicle types were assigned a Make/Model code indicating the vehicle type is missing, unknown, or "other".

The other main difference apparent from Table 40 is that a significant proportion of cases classified as missing, unknown, or "other" under one variable were assigned specific vehicle type codes under the other variable. This is true of 14.3% of such cases under Vehicle Type and 27.9% of those under Make/Model. This is somewhat surprising since the CARDfile conversion algorithm incorporates data from the Vehicle Type variable when assigning the Make/Model codes. If a case is known to be a passenger car, for example, it should receive a make/model code indicating that it is some sort of passenger car, even if the make and model information is missing. One effect of the discrepancies between these two variables is different frequencies for each of the vehicle type categories, depending on which variable is used.

In another consistency check, the Accident Type variable was recoded into categories appropriate for a run against Primary Impact (Table 41). The results of a run on the 5% sample file indicate a reasonable agreement between the two variables. There are over 1500 more cases coded as rollovers under Accident Type than under Primary Impact, but this is mainly because both primary-event and subsequent-event rollovers receive that code under Accident Type in Maryland and Pennsylvania.

Some of the other instances of different codes assigned between the two variables might be legitimate as well. As an example, a vehicle could initially impact a telephone pole before striking a pedestrian, resulting in the accident being coded as "fixed object" under Primary Impact and as "pedestrian/animal/pedalcyclist" under Accident Type. Other cases of different codes are evidently due to errors, however. For example, all cases coded as "another motor vehicle in transport" under Primary Impact should be coded as either "rollover", "two vehicles", or "more than two vehicles" under Accident Type, since all of the other Accident Type codes are exclusively for single-vehicle accidents. Yet there are 1749 such cases under Primary Impact coded as one of the single-vehicle codes under Accident Type.

The third consistency check performed was between the two variables of Road Surface and Weather Conditions, again using the 5% sample file (Table 42). There is a good degree of correspondence between the two variables, especially considering that one would not expect a perfectly isomorphic relationship between weather conditions and road surface. One thing to note from Table 42 is that if one of the two variables is unknown, the other is likely to be as well.

6.2 Bias Arising from Missing Data

Some of the variables contained in CARDfile have very low rates of missing data. Others are characterized by large amounts of missing data, such as when one or more of the states does not code a variable at all. One issue in the evaluation of the CARDfile concerns biasing effects of the missing data. Comparing a variable with a high missing data rate with a relatively complete variable is one means of assessing the bias introduced by missing data.

As an illustration of bias due to missing data, the Land Use (rural/urban) variable was run against Accident Severity and against Accident Type. Land Use is coded by Indiana, Maryland, Michigan, and Washington and is 100% complete for those four states. Because Pennsylvania and Texas do not code the variable at all, however, the overall missing data rate for Land Use is 42.5%. Any analysis conducted with CARDfile data that includes Land Use will automatically exclude all cases from Pennsylvania and Texas. In contrast, the data for Accident Severity and Accident Type are virtually complete, with the two variables having missing data rates of 0.04% and 0.00% respectively.

The results of a two-way analysis running Land Use against Accident Severity are presented in Table 43. The percentages are based on the entire CARDfile for 1984-1986. The three middle columns list the distributions of Accident Severity, first using cases where Land Use is known, then cases where Land Use is missing, and finally all cases. The column on the far right gives the ratio of cases where Land Use is known to all cases, in terms of the percentages for each of the Accident Severity attributes.

For example, the first row of Table 43 indicates that 69.58% of cases where Land Use is known are coded "Property Damage Only" under Accident Severity, 58.31% of cases where Land Use is missing are PDO's, and 64.79% of all cases in the file are PDO's. The ratio of the Land Use known PDO percentage to the PDO percentage for all the cases is 1.07. This indicates that the result of losing the Pennsylvania and Texas cases when Land Use is run against Accident Severity is that PDO accidents are over-represented by 7% in comparison to the entire file. The smallest effect of losing the Pennsylvania and Texas cases is on incapacitating injuries, where the ratio indicates only a 3% over-representation. The widest divergence is for fatal injuries, which are under-represented by 26% for cases where Land Use is known. However, fatal injuries are less than one percent of the total.

The results of a similar two-way are presented in Table 44. This time the Accident Type variable was recoded into various broad categories of collisions and run against the same three sets of cases, according to whether Land Use is coded. It is apparent that the loss of the Pennsylvania and Texas cases changes the distribution of Accident Type. For six out of the ten Accident Type attributes, the ratios indicate agreement to within 10% between cases where Land Use is known and all cases. Two of the other attributes are within 20%, but the remaining two, sideswipes and backing collisions, are over-represented by 33% and 34% respectively for the Land Use known cases compared to the entire file.

In Tables 45 and 46, Land Use is subdivided into its attributes of rural, urban, and missing and run against first Accident Severity and then Accident Type. Both of these tables were constructed from runs made on a 5% random sample of the CARDfile at the accident level. The percentages in the tables are expressed as 95% confidence intervals for the entire CARDfile, and the frequencies are projected estimates based on these percents. Accident Severity and Accident Type were chosen for this illustration because it was expected that the distribution of each would be appreciably different in the rural areas as compared to urban areas. Such differences are illustrated in these two tables. Accidents tend to be more severe in rural areas, probably due to higher average traveling speeds, so that the proportion of fatal and serious injuries are somewhat higher and the proportion of property damage and less severe injuries is lower in comparison to the severity distribution for urban accidents. With regard to Accident Type, there are nearly twice as many single-vehicle accidents in rural areas as compared to urban, and conversely, nearly twice as many intersection accidents in urban areas as compared to rural.

These two tables were prepared to see if one could infer the proportion of rural versus urban accidents in Pennsylvania and Texas, where the rural/urban variable is missing, by comparing the distributions of Accident Severity and Accident Type for the missing Land Use cases with the rural and urban distribution. If the missing cases are distributed more similarly to the rural distribution, for example, then one might infer that Pennsylvania and Texas include proportionately more rural accident experience. The result, however, is seemingly inconsistent, with the severity distribution looking most like the rural accidents, but the collision type distribution looking definitely urban.

Apparently, the factor that has been overlooked here is that Pennsylvania is one of the two states in the CARDfile with a tow-away accident reporting threshold. Consequently, only 35% of the Pennsylvania accidents are property-damage-only. Thus, the differences in reporting threshold would appear to prevent any inferences based on the distribution of Accident Severity. However, it is difficult to understand the seemingly urban character of the distribution of Accident Type for the missing Land Use cases from Pennsylvania and Texas. The separate Accident Type distributions from Table 17 provide some insight. First, the Accident Type distributions are very different in Pennsylvania and Texas. Pennsylvania has the highest proportion of single-vehicle accidents, perhaps as a consequence of a higher reporting threshold. Texas on the other hand, has the lowest proportion of single-vehicle accidents. In view of a relatively low overall population density, this suggests very little rural travel, or an under-reporting of rural accidents in Texas. Based on these tabulations, the essential characteristics distinguishing rural from urban accidents in the four states that code Land Use appear inconsistent with the Texas and Pennsylvania data. The apparent inconsistency may be due to reporting differences in these two states as compared to the other four states.

Roadway Profile is another variable characterized by a higher missing data rate, 27.2%, most of which comes from the variable being unavailable for Michigan. In Tables 47 and 48, Roadway Profile is split into its attributes of level, grade, and missing/unknown and run against Accident Severity and Accident Type using the 5% sample file. Projected frequency estimates and 95% confidence intervals for the entire file were obtained as for Tables 45 and 46. Once again, it is apparent that the loss of the missing data cases alters the relationships between the pairs of variables from what they would be otherwise.

Table 47 indicates that "missing" Roadway Profile cases have appreciably more property-damage-only accidents and less of all the other severity levels (except one) than either the "level" or "grade" attributes. Thus, the omission of Michigan produces a somewhat more severe subset of accidents where Roadway Profile is known. In the run against Accident Type, the "missing" cases are intermediate between the level and grade percentages for six out of the ten attributes (Table 48). This suggests that the loss of missing data from Roadway Profile has less of an effect on Accident Type than the loss of missing data from Land Use.

In one final two-way table, Land Use was run against Roadway Profile using the 5% random sample file (Table 49). This table lists percentages based on the total number of cases for each frequency instead of column percentages as in the previous tables. The confidence intervals and estimated frequencies were obtained as before. The main point to observe from Table 49 is that only 30.2% of the cases are complete for *both* of the variables. Nearly 70% of the cases are missing for either the Land Use or Roadway Profile variables and are, therefore, omitted from the analysis. Again, because the omission of cases is non-random, bias results. Based on Table 9, approximately one-third of the cases with known Land Use are in rural areas, whereas only about one-fourth of the cases in Table 49 with both known Land Use and known Roadway Profile are rural. Similarly, Table 15 shows that about 85% of the cases with known Roadway Profile are coded "level," whereas only 75% of the cases in Table 49 with both known Land Use and known Roadway Profile are coded "level." In this example, the combination of missing data on two variables omits 70% of the CARDfile records and produces a subset that is biased on both dimensions.

7 Representativeness of the CARDfile

While CARDfile was not designed to be a nationally representative database, it has been claimed that it may serve as such on a general level.[18] Certainly the utility of the database for crash avoidance research would be enhanced if this were the case. Assessing the degree of representativeness of the CARDfile is not one of the primary goals of this evaluation, but the topic will be discussed here since the missing data have been shown to introduce bias.

7.1 Other Studies Concerning the CARDfile Representation

There are two separate issues with respect to the representativeness of CARDfile. One concerns the aggregate crash experience of the six states contained in CARDfile compared to the nation as a whole, and the other concerns the actual data that appear in the file. These two measures of representation are not the same.

Salvatore et al.[19] conducted a series of comparisons between the six CARDfile states and the entire nation, focusing on demographics, registered motor vehicles, vehicle miles of travel, roadway characteristics, and fatal accidents. The sources they used were *Population Characteristics* and *Statistical Abstract of the United States* from the U.S. Department of Commerce, *Highway Statistics* and the *Fatal Accident Reporting System* (FARS) database from the U.S. Department of Transportation, and the *National Vehicle Population Profiles* database from R. L. Polk and Company. The main conclusion reached by Salvatore et al. was that the CARDfile states are generally representative of the entire nation in terms of the characteristics examined. They did note several significant differences between the CARDfile states and the nation, which include an overall younger population of licensed drivers in the CARDfile states, fewer cars and more trucks registered in the CARDfile states, fewer imports and more domestic vehicles registered in the CARDfile states, and fewer rural roads and more urban roads in the CARDfile states.

A few points concerning the Salvatore et al. paper should be emphasized. Perhaps the most critical is that none of their comparisons actually involved CARDfile data. They evaluated the crash experience of the six states contained in the CARDfile, but not the database itself. Some

---

18. M.L. Edwards, *A Database for Crash Avoidance Research*. Warrendale, Pennsylvania: Society of Automotive Engineers, Paper No. 870345, 1987, pp.136-137.

19. S. Salvatore, P. Mengert, and R. Walter, *CARDfile Data Base Representativeness, Phase I: General Characteristics including Populations, Vehicles, Roads, and Fatal Accidents*. Cambridge, Mass.: Transportation Systems Center, Project Memorandum No. DOT-TSC-HS802-PM-88-16, August 1988.

confusion over this point apparently exists, since the focus of their evaluation has been subsequently misidentified in the literature.[20]

Second, the particular differences that Salvatore et al. did note should be considered when performing analyses using the CARDfile data. For example, drivers under the age of 19 are over-represented by 5% in the CARDfile states compared to the nation. Since this age group is generally considered to be over-involved in motor vehicle accidents, their over-representation in the CARDfile driving population may affect the types of collisions included in the database and correlations between driver age and other variables in ways that differ from the nation's crash experience.

Third, many variables of particular interest to crash avoidance research were not a focus of Salvatore et al.'s evaluation. For example, none of their comparisons were related in any way to the CARDfile variables of Primary Impact, Precrash Stability, or Avoidance Attempt. Thus, while a generally good correspondence was found between the CARDfile states and the nation for the variables evaluated, this does not necessarily imply anything about the representativeness of the CARDfile states for other variables of concern in the field of crash avoidance.

Returning to the problem of the representativeness of the CARDfile states versus the representativeness of the CARDfile database, a section from Edwards' 1987 paper on the CARDfile (see note 15) serves to illustrate this difference. Included in Edwards' paper is a discussion of road class (for example, interstate, principal artery, major collector, etc.) and a table presenting annual vehicle miles of travel on different classes of roads for the six CARDfile states compared to the entire nation. The data presented, which come from *Highway Statistics*, indicate some differences between the two groups, most of which are minor. What is not emphasized in Edwards' paper, however, is that not only are the data not derived from CARDfile, but there are no variables contained in the database that can serve as indicators of different classes of roads. Data on road class are certainly useful in crash avoidance research, but the most similar such CARDfile variable, Roadway Separation, is coded only in Washington, and thus, has a 92% missing data rate. Although the distribution of vehicle travel by road class in the CARDfile states corresponds to the national experience, it is worthy of note to researchers interested in this characteristic that no classification of the data according to road class is actually available in the CARDfile.

7.2 A Comparison of CARDfile and FARS Data

In order to conduct our own test of how well data contained in the CARDfile mirror the national crash experience, we compared CARDfile data to information contained in a NHTSA technical report concerning the use of alcohol in fatal motor vehicle accidents.[21] The NHTSA data were derived

---

20. E.A. Harwin and H.K. Brewer. *Analysis of the Relationship Between Vehicle Rollover Stability and Rollover Risk Using the NHTSA CARDfile Accident Database.* Washington, D.C.: National Highway Traffic Safety Administration, 1987, p. 3.

21. *Alcohol Involvement in Fatal Traffic Crashes 1986.* National Highway Traffic Safety Administration, Technical Report No. DOT-HS-807-268, January 1988.

from FARS and combine actual blood alcohol concentration (BAC) test results with estimated BAC's for cases where this information was not available. The estimates were based on a model developed by NHTSA. Since FARS comprises data from all fatal accidents from every state, it may be considered a truly nationally representative database.

Alcohol use information is only available for vehicle drivers in CARDfile. It is not based on actual BAC levels but generally derives from the observation of the police officer reporting the accident. For purposes of comparison, the CARDfile Alcohol/Drug Use variable attribute "Indication of Alcohol Use" was considered equivalent to a BAC level of 0.10 or greater in the FARS data. Proportions were calculated by taking the number of "Indication of Alcohol Use" cases out of the number of "Indication of Alcohol Use" and "No Indication of Use" cases combined. In other words, unknown cases for this variable were excluded when calculating percentages. In this section, "alcohol-impaired" will be used to refer to FARS drivers with BAC's of 0.10 or higher as well as to drivers in CARDfile coded as "Indication of Alcohol Use" under the Alcohol/Drug Use variable.

In order to compare the CARDfile and FARS data, the CARDfile cases were restricted to fatal accidents from 1986. Within this subset, all the cases from all six states were considered in the calculations. Overall, the NHTSA paper reports that 25.7% of all drivers involved in fatal accidents in 1986 were alcohol-impaired. The corresponding figure from CARDfile is 28.2%.

Table 50 presents the results of several other comparisons conducted between the FARS data and the CARDfile. The table lists the proportion of alcohol-impaired drivers involved in fatal accidents according to the accident type, sex of the driver, age of the driver, and type of vehicle. For example, the first row of Table 50 indicates that 50.8% of all drivers involved in single-vehicle fatal accidents (not involving pedestrians or pedalcyclists) are alcohol-impaired, according to the FARS data, with a figure of 53.8% for the CARDfile data.

Whereas other comparisons between CARDfile and national data have generally only considered one-way frequencies of particular variables,[22,23] the results listed in Table 50 are based on bivariate frequencies. Thus the correlation of the alcohol use variable with the variables of accident severity, accident type, driver sex, driver age, and vehicle type can be examined. Actual use for research purposes of a database like the CARDfile will likely involve such multivariate analyses.

In general, there is a reasonably good correspondence between the FARS and CARDfile figures for all of the comparisons that were made. The overall incidence of alcohol-impaired drivers is slightly higher for the CARDfile compared to the FARS data, and this is reflected in the results of most of the specific comparisons as well. The slightly different criteria used to define "alcohol- impaired" between FARS and CARDfile might account for this difference. The relative incidence of alcohol use among the different

---

22. Edwards, *Database for Crash Avoidance.*

23. Salvatore, Mengert, and Walter, *CARDfile Data Base Representativeness.*

categories, for example highest among 21-44 year-old drivers, followed by 15-20, followed by over 45, is generally similar between FARS and CARDfile.

7.3 Summary

Based on the small number of comparisons we conducted using FARS and CARDfile data and on the other studies reviewed in this section, it appears that the CARDfile database is generally representative of the nation's crash experience on several variables, despite the fact that it includes only six states. A few qualifying remarks should accompany this statement, however. One is that some significant differences do exist between the characteristics of the six states as compared with the nation with regard to the proportion of younger drivers, the proportion of trucks, the proportion of imported vehicles, and the proportion of urban roads. Second, although the distribution of road class for the six states was shown to be representative, no such variable is actually available in the CARDfile. These differences could affect the results of particular analyses and should be considered when using the CARDfile data.

Perhaps more significant is the effect on representation when the data from one or more states are lost from an analysis due to the unavailability of variables in particular states. One of Salvatore et al.'s concluding observations was the high impact any individual state could have on the aggregate agreement between the six states.[24] Roadway Separation, Land Use, Precrash Stability, Avoidance Attempt, Roadway Profile, and VIN all have missing data for one or more entire states in the CARDfile. The tabulations presented in this report illustrate that the more serious problem of representation in the CARDfile is the bias introduced by the high missing data rates on some variables.

Less apparent is the bias introduced when individual attributes are unavailable or inconsistently defined in some states even though the variable is nominally present. An inconsistent translation of rollover from one state to another produces a biased estimate of this event in the CARDfile. Inconsistencies from state to state bias all of the Precrash Stability attributes. Because the individual data elements are not consistent from state to state, the combined distribution is necessarily biased.

---

24. Ibid., p. 14.

8 Discussion

The goal of this evaluation of the CARDfile has been to systematically examine possible sources of bias and error in the data, especially between the six states contained in the file, as they relate to the use of the file as a tool in crash avoidance research. A variety of approaches was used in the evaluation. The accident reporting criteria, which affect the types of accidents that appear in the CARDfile, were discussed for each state. A series of tables presenting state distributions for particular variables was constructed to allow for the comparison of distributions between states. Significant differences here might indicate differences in coding schemes between states or inherently different crash experiences. Missing data rates for certain variables were calculated, and a list of unavailable variables and attributes was compiled, with instances of combined attributes noted where possible. A review was conducted of the state police and coding forms and the CARDfile conversion algorithm in order to check for more subtle discrepancies in the code definitions applied to each of the six CARDfile states. As a check on the technical accuracy of the CARDfile translation algorithm itself, variable distributions were compared between the original files from three states and the translated coding of the CARDfile. A series of two-way tables was run to illustrate the bias incurred when variables with high missing data rates are included in analyses. Several consistency checks were performed to assess the degree of correspondence between pairs of variables in CARDfile that code for similar information. Finally, a short series of comparisons was conducted between CARDfile and FARS data as a test of the representativeness of the "alcohol-impaired" variable in the CARDfile.

The CARDfile was created to fill a need for a large database, incorporating information from several states all coded in a consistent manner, geared towards variables of key interest to collision avoidance research. Compiling this database was a very ambitious endeavor, primarily because each state has its own approach to reporting and coding motor vehicle accidents. Converting the data from the six different states for the numerous variables contained in the CARDfile required various compromises, undoubtedly after careful consideration. While the CARDfile's standardized format has indeed resulted in a sizable database, the forced uniform coding of the states' disparate approaches has resulted in various problems in the use of the database.

These difficulties include high missing data rates for some variables and unavailable variables and attributes for particular states. Some of the very variables that would be of the greatest utility for crash avoidance research questions have the highest rates of missing data. Roadway Separation is missing for 92% of the cases, Precrash Stability is 70% missing, VIN is missing for 66% of the cases, Avoidance Attempt has a 44% missing data rate, and Land Use is missing for 42.5% of the cases. Furthermore, the presentation of the data in CARDfile in a common format obscures the fact that there are inconsistencies in the variable and attribute definitions between the six states. Unless extensive additional research were undertaken, the typical user of CARDfile would be unaware of these hidden discrepancies and their effect on the analyses performed.

Many of these inconsistencies are unresolvable, given the variables and states included in CARDfile, because of the greatly different approaches used by the states in the first place. However, the situation could be improved in several ways. There are some instances where modifications to the CARDfile conversion algorithm would increase the degree of correspondence of the coding principles applied to each state. For example, the transfer of many of the cases coded as "Other Driver Error" in Maryland to "None Indicated", as detailed in Section 2, would result in greater consistency between states for the Driver Error variable. Also, the coding of rollover could be consistently limited to primary-event rollover in all six states. Second, the documentation of CARDfile could be made more clear, as in the case of the definition of intersection accidents or the use of the 000 code under the Accident Type variable. Many of the variables and attributes described in the CARDfile Handbook would benefit from more detailed definitions, so that the user would know exactly what types of cases were intended to be included. The Handbook's treatment of such problems is inconsistent. Finally, a more drastic approach would be to alter the states and/or the variables included in the CARDfile. Since five of the six states presently included do not code for Roadway Separation, perhaps this variable should be omitted from the file, or other states chosen which would permit a useful road type variable. Similarly, the database would benefit from the inclusion of more states that record VIN, since this variable is currently available for only two of the states. VIN is of major importance in crash avoidance issues where model-specific questions are the primary focus.

The CARDfile contains information on more than 20% of the nation's police-reported accidents. The great number of cases contained in the database is advantageous since it provides large sample sizes so that the results of particular analyses are more likely to be statistically significant. It also allows for the analysis of rare events that may not appear with sufficient frequency in other, smaller databases.[25] However, the CARDfile's large size does not make it immune to problems of bias.

The sampling units used in CARDfile are the six states, and these differ from each other in terms of such factors as severity of accidents, composition of the motor vehicle population, degree of urbanization, and climate. Potential bias is probably introduced from the outset simply because the cases from certain states represent a much greater proportion of the total than others. For example, Texas accounts for 32.2% of the accident cases included in CARDfile, while Washington comprises only 8.4% of the total. Furthermore, when cases are lost from an analysis due to missing data, they are not omitted randomly but instead are included or excluded according to state. This results not only in the loss of a large number of cases but also usually in a change in the distribution of the remaining cases compared to what it would be if the entire file could be used.

---

25. M.L. Edwards, *A Database for Crash Avoidance Research*. Warrendale, Pennsylvania: Society of Automotive Engineers, Paper No. 870345, 1987, p. 134.

It is difficult to assess the implications of these shortcomings on the use of the CARDfile because they depend on the variables needed in a particular analysis, the type of research question being addressed, and the level of detail desired. Perhaps this evaluation should conclude with the cautionary note that for every analysis performed using CARDfile data, the researcher should attempt to carefully consider the sources, degree, and effects of any bias introduced through missing data and coding inconsistencies. In many instances, the consequences are not readily apparent, particularly if missing data are simply excluded from the analysis. Misleading conclusions could result unless there is an awareness of the compromises that were made along the way to producing the final output for each analysis.

In order to illustrate the types of considerations that should be made when performing an analysis with CARDfile data, a recent paper on vehicle rollover stability by Harwin and Brewer will be discussed.[26] Harwin and Brewer's paper examines the correlations between rollover risk and a number of vehicle, driver, and environmental variables. We reviewed their paper, not to critique their approach or findings, but in order to demonstrate how differences in data recording between states may have influenced the choices they made in designing their analysis and possibly affected their results as well. Their research problem is typical of the kind CARDfile is intended to support, and shortcomings with the database similar to those discussed below might be expected to be encountered in other crash avoidance analyses utilizing CARDfile data as well.

Harwin and Brewer calculated rollover risk, defined as the number of rollovers in all single-vehicle accidents, for 40 different models of vehicles, including 9 imported passenger cars, 23 domestic passenger cars, and 8 light trucks or utility vehicles. They then performed linear regression analyses of rollover risk versus their rollover stability factor (ratio of half-track width to center of gravity height) on several datasets defined by particular states and years. They also conducted several stepped multivariate analyses to examine the effect of certain environmental and driver factors on rollover risk.

Given the three general classes of vehicles of interest in Harwin and Brewer's study, that is, light trucks, domestic cars, and foreign cars, the data from two of the CARDfile states would have automatically been excluded from consideration. While domestic passenger car make/model information is relatively complete for all six of the CARDfile states, this is not the case for imported passenger cars or for light trucks. Indiana records make and model for very few imports, and only one of the nine imported passenger car models included in the rollover study would have been available for Indiana. In addition, Indiana has make/model information available for only a small percentage of light trucks compared to most of the other CARDfile states. Pennsylvania as a rule does not code make and model for any light trucks, and none of the light truck models in the study are available for that state. Thus from the outset, two of the six CARDfile states, Indiana and Pennsylvania, had to be excluded from Harwin and Brewer's study due to lack of vehicle make/model information.

---

26. E.A. Harwin and H.K. Brewer. *Analysis of the Relationship Between Vehicle Rollover Stability and Rollover Risk Using the NHTSA CARDfile Accident Database*. Washington, D.C.: National Highway Traffic Safety Administration, 1987.

Maryland, Texas, and Washington are the three states that appear in Harwin and Brewer's analysis, and while all three have high rates of known make/model information for cars and light trucks, the authors would still have been limited to those models routinely coded by all three states. For example, three models of Jeep utility vehicles and two models of Volkswagen passenger cars are included in the study, all of which are coded by all three states. However, it would not have been possible to include the Jeep Comanche or the Volkswagen Fox in the study since none of the three states code either model. While it may have been desirable to include additional vehicle models, particularly light trucks and vans, the possible options were restricted by the availability of information in the CARDfile.

Harwin and Brewer state "In order to draw conclusions of national scope from the CARDfile database, its statistical representativeness to the national accident experience must be established."[27] The authors cite the Salvatore et al. paper[28] to establish the national representation of the CARDfile. As an aside, Harwin and Brewer incorrectly state that Salvatore et al. compared the CARDfile to NASS. While Salvatore et al. examined several data sources for the six CARDfile states, the data from the CARDfile were not actually included. The more important point is that missing data on make/model force Harwin and Brewer to omit a substantial portion of the CARDfile from their analysis. Only data from Maryland, Texas, and Washington are actually used. The bias introduced by omitting three of the CARDfile states render the national representation of the overall CARDfile irrelevant to the subsequent analysis.

For the multivariate portion of their analysis, Harwin and Brewer were limited to those variables coded consistently by the states in question. They initially conducted a survey of the dataset comparing rollovers to single-vehicle accidents in general for nineteen variable attributes prior to selecting those factors to be included in the multivariate regression. It is possible that Michigan was not included in their analysis because data were not available from that state for three of the nineteen attributes.

One of the factors included in the multivariate regression was the percentage of accidents occurring on curves. When the regression was carried out using Maryland and Texas data from 1984 and 1985, this factor was not found to be one of those influencing the risk of rollover. This could be partially due to the fact that nearly 95% of Texas accidents are coded in CARDfile as taking place on a straight roadway. In this subset, Texas cases comprised 77% of the total.

A similar multivariate regression was carried out by the authors using only Maryland cases for the same years, and again roadway curvature was not found to affect the rollover risk. In this case the slightly surprising result may be due to a different explanation. Harwin and Brewer derived rollovers from the CARDfile Accident Type variable, which is designed to include

---

27. Ibid., p. 3.

28. S. Salvatore, P. Mengert, and R. Walter, *CARDfile Data Base Representativeness, Phase I: General Characteristics including Populations, Vehicles, Roads, and Fatal Accidents*. Cambridge, Mass.: Transportation Systems Center, Project Memorandum No. DOT-TSC-HS802-PM-88-16, August 1988.

rollovers occurring as both the first harmful event and as subsequent events. Although not mentioned by the CARDfile Handbook, only two of the states actually include subsequent-event rollovers under Accident Type, with the remaining four coding only primary-event rollovers. In terms of Harwin and Brewer's analysis, the Texas and Washington cases include primary-event rollovers only, while the Maryland cases include rollovers occurring as subsequent events as well. It is reasonable to expect that a different set of factors might influence rollovers occurring as the first event in an accident as compared to rollovers taking place subsequently. In particular, road curvature might be of less relevance to subsequent-event rollovers, and these comprise over two-thirds of the Maryland rollover cases. Harwin and Brewer may not have been aware that subsequent-event rollover is included for Maryland, but not for Texas and Washington. The inclusion of subsequent-event rollovers may explain some of the inconsistency observed by Harwin and Brewer for the regression model based only on the Maryland data.

Among the candidate list of factors is the tracking attribute on Precrash Stability. However, tracking is not actually indicated on the Texas state file as stated by Harwin and Brewer.[29] In Texas, the only applicable attribute on the variable is skidding, since jackknife is not applicable to this vehicle subset. The tracking attribute is generated by the CARDfile algorithm for Texas as the default whenever skidding (or jackknife) is not coded. Thus, cases where skidding was present, but not recorded, are coded as tracking in the CARDfile, since missing data cannot be distinguished in the original state files. Perhaps as a consequence, tracking is indicated for over 90% of the single-vehicle accidents in their preliminary analysis of the candidate factors.

Land Use (rural/urban) was included in the multivariate analysis. As noted by the authors, Texas does not code this variable, so only Maryland data could be used to examine this factor.

In summary, Harwin and Brewer's primary data requirement was a sufficient sample size of police-reported accidents to provide statistical significance of the resulting models. The CARDfile certainly filled this need. However, the shortcomings of the original state files limited many aspects of the analysis. Bias resulting from the exclusion of three states due to missing data negate the national representation of the CARDfile as a whole. Inconsistencies in the coding of important variables for the analysis such as rollover and tracking also appear to have limited the findings. Missing data on Land Use constrained the analysis of this factor even further.

The objective of this discussion of Harwin and Brewer's paper has not been to criticize their work. Overall, their analysis, especially their correlation of rollover risk with the rollover stability factor, appears sound. Rather, the objective was to illustrate the difficulties posed by missing data and coding inconsistencies in the CARDfile in a typical collision avoidance analysis. The pitfalls to which CARDfile exposed Harwin and Brewer illustrate that it is still necessary for the analyst to be very familiar with the coding of the original state files, as well as the translations necessary to put the data in the uniform CARDfile format.

---

29. Harwin and Brewer, *Vehicle Rollover Stability*, p. 6.

9 Conclusions

The CARDfile was developed by NHTSA to aid problem identification and countermeasure development in the field of crash avoidance research. The primary objective was to facilitate the study of relationships between vehicle design characteristics and crash propensity. Accident data from six states, Indiana, Maryland, Michigan, Pennsylvania, Texas, and Washington, were translated to a common format and combined. In the aggregate, the CARDfile contains information on more than 20% of the police-reported accidents in the United States. Furthermore, the aggregate of the six states has been shown to be representative of the national experience on several characteristics.

These conclusions will be developed by considering the gains and losses of the CARDfile. What are the possible advantages to be gained from combining the accident data from six states? First, as shown by Salvatore et al., the aggregate of the six states may be nationally representative whereas the individual states are not. Second, the aggregate file provides a much greater sample size for analyses of rare events. A third advantage is greater convenience and efficiency for the user as compared to working with six individual state files, all in different formats.

With regard to national representation, the CARDfile provides this advantage for a number of characteristics, although not all of the characteristics examined by Salvatore et al. are actually available in the CARDfile. More importantly, this advantage presumes that data from all six states are present and consistently recorded. This is the case for many of the CARDfile variables. However, other variables, such as Roadway Separation, Precrash Stability, Land Use, Avoidance Attempt, and VIN, have large missing data rates that include one or more states. Less obvious are the situations where the variable is nominally present in each state, but individual attributes are missing. The Vehicle Type variable appears quite complete, showing virtually no missing data. However, three states (Indiana, Texas, and Washington) do not distinguish trucks with double trailers from other large trucks. Since these three states contain 54% of the vehicle cases in the CARDfile, the proportion of doubles shown by the CARDfile underestimates the true proportion for the six states, perhaps by about a factor of two. Precrash Stability clearly has the most serious problem with missing attributes. Not a single state codes all of the attributes, and there is not a single attribute that is present in every state. Thus, none of the attributes reflect the correct proportion for the aggregate of the six states. For any analysis using these variables, the bias introduced by the missing data is likely to invalidate the representativeness of the CARDfile as a whole.

With over 7 million vehicle/driver records, the CARDfile certainly achieves the goal of large sample size. The common format facilitates the search of approximately one-fifth of the national accident experience. However, the common format also tends to limit the available information to the lowest common denominator. Furthermore, in the translation, the meaning of the original coding is sometimes lost or distorted. Numerous examples have been cited in this report, such as the coding in Texas on avoidance maneuvers and precrash stability being limited to single-vehicle accidents, and the combining of subsequent-event rollovers with primary-event in Maryland and Pennsylvania but not the other states. When the attributes are inconsistently combined, the result is biased, regardless of whether the inconsistencies arise from the original files or the translation algorithm. Consequently, much of the gain in convenience is offset by losses in detail and meaning.

This leads to perhaps the most dangerous advantage of the CARDfile, its convenience. One might be tempted to think that the existence of the CARDfile eliminates the need to be familiar with the coding and data processing details of each of the CARDfile states. The present authors urge a more cautious approach in which the CARDfile translation is viewed as simply another step removed from the original event, making it even more important to be familiar with every stage of data manipulation. In some cases, such as the Precrash Stability variable, the CARDfile provides only the appearance of common codes, when in fact, every attribute is biased, and the degree of bias varies from one attribute to another within the same variable. Consequently, the CARDfile distribution bears no resemblance to the true distribution of these attributes for the aggregate of the six states. While far from exhaustive, this report provides many examples where the information in the original coding is lost and/or the resulting CARDfile distributions are biased.

The creation of the CARDfile was an ambitious and worthwhile undertaking, and the NHTSA has done a good job overall of extracting the common information available from the original state files. The limiting factor is clearly the lack and incompatibility of the information coded by the states. The problems with the CARDfile clearly illustrate the advantages that would result from more uniform accident reporting at the state level, and particularly the use of compatible definitions for common data elements.

Appendix A: CARDfile Variables

Appendix A:

CARDfile Variables, 1984-1986

ACCIDENT VARIABLES	
Variable Number	Variable Name
1	Case Identifier
2	State Identifier
3	Month of Crash
4	Day of Crash
5	Year of Crash
6	Time of Crash
7	Number of Vehicles Involved
8	Accident Severity
9	Light Conditions
10	Weather Conditions
11	Road Surface
12	Land Use
13	Primary Impact
14	Primary Impact Location
15	Relation to Intersection
16	Intersection Signalization
17	Roadway Alignment
18	Roadway Profile
19	Roadway Separation
20	Accident Type
VEHICLE VARIABLES	
100	Vehicle Number
101	Make/Model Code
102	Accident Type (Vehicle)
103	Number Unknown Injury Severity
104	Number Uninjured
105	Number Possible Injury
106	Number Nonincapacitating Injury
107	Number Incapacitating Injury
108	Number Fatally Injured
109	Model Year
110	Vehicle Type
111	Component Failure
112	Precrash Stability
113	Avoidance Attempt
114	Vehicle Identification Number (VIN)
DRIVER VARIABLES	
200	Driver Age
201	Driver Sex
202	Alcohol/Drug Use
203	Restraint Use
204	Helmet Use
205	Driver Error

NOTE: The variable numbers correspond to those in the data files maintained by UMTRI.

Appendix B: Tables

TABLE 1
Accident Reporting Threshold by State

Reporting Threshold	Indiana	Maryland	Michigan	Pennsylvania	Texas	Washington
Property Damage of $200	X		X			
Property Damage of $250					X	
Property Damage of $300						X
Tow Away		X		X		
Personal Injury	X	X	X	X	X	X
Fatality	X	X	X	X	X	X

Source: T. P. Luce, *Digest of Motor Laws*. 52d ed. American Automobile Association, 1986.

TABLE 2
Accident Variables and Attributes Unavailable in CARDfile

Variable	Excluded Attributes by State					
	Indiana	Maryland	Michigan	Pennsylvania	Texas	Washington
Accident Severity	Missing, Unknown	Unknown	Missing, Unknown			Missing, Unknown
Road Surface			Other (mud, oil) (coded with Unknown)			
Land Use				Variable not available	Variable not available	
Primary Impact			Unknown; Noncollision (both coded with Other)			
Primary Impact Location			On Shoulder (coded with On Roadway)			On Shoulder
Relation to Intersection			Driveway Related			
Roadway Profile			Variable not available			
Roadway Separation	Variable not available	Variable not available	Variable not available	Variable not available	Variable not available	
Accident Type[a]	209, 233, 325, 513, 519	223, 225, 325, 409, 421, 509 513, 519	209, 233, 325	209, 233, 325, 417, 429, 501, 513, 519	209, 211, 227, 311, 315, 429, 519	209, 223, 225, 417, 507, 509

TABLE 2 (continued)
Accident Variables and Attributes Unavailable in CARDfile

[a] Key to code values for the Accident Type variable:

209 = Rear End, Same Trafficway, Same Direction, Direction Unknown

211 = Rear End, Same Trafficway, Same Direction, Specifics Unknown

223 = Sideswipe, Same Trafficway, Same Direction, Overtaking Right

225 = Sideswipe, Same Trafficway, Same Direction, Overtaking Left

227 = Sideswipe, Same Trafficway, Same Direction, Changing Lanes

233 = Other, Same Trafficway, Same Direction, Other

311 = Head-On Sideswipe, Same Trafficway, Opposite Direction, Lateral Move/Lane Change

315 = Head-On Sideswipe, Same Trafficway, Opposite Direction, Specifics Unknown

325 = Other, Same Trafficway, Opposite Direction, Other

409 = Initial Opposite Direction, Change Trafficway/Turn Across Path, One Straight, One Right

417 = Initial Same Direction, Change Trafficway/Turn Across Path, Specifics Unknown

421 = Initial Opposite Direction, Change Trafficway/Turn Across Path, Both Turning

429 = Initial Opposite Direction, Change Trafficway/Turn Across Path, Specifics Unknown

501 = Straight Turning, One Straight, Intersecting Paths, Resulting Direction Unknown, Specifics Unknown

507 = Straight Turning, One Straight, Intersecting Paths, Resulting Same Direction, Turn Right

509 = Straight Turning, One Straight, Intersecting Paths, Resulting Opposite Direction, Turn Right

513 = Curved Paths, Intersecting Paths/Angles, Both Curving/Turning

519 = Other, Intersecting Paths/Angles, Other

TABLE 3
Vehicle and Driver Variables and Attributes Unavailable in CARDfile

Variable	Excluded Attributes by State					
	Indiana	Maryland	Michigan	Penn.	Texas	Washington
Vehicle Type	Tractor with Double Trailer[a]; Other Emergency Vehicle	Fire Truck (coded with Other Emergency Vehicle)	Motor Home (coded with Straight Truck); Other Combination Vehicle; Other Emergency Vehicle		Tractor with Double Trailer (coded with Tractor with Semi-Trailer)	Tractor with Double Trailer (coded with Other Combination Vehicle); Police Vehicle
Component Failure	Exhaust System; Tire Puncture; Defective Signal; Wheel Came Off	Defective Trailer Equipment; Defective Signal; Wheel Came Off	Brakes; Steering; Lights; Exhaust System; Worn Tires; Defective Trailer Equipment; Defective Signal; Wheel Came Off	Exhaust System; Wheel Came Off	Tire Puncture	Exhaust System; Defective Trailer Equipment
Precrash Stability[b]	Variable not available[c]	Spinning; Jackknifing	Tracking; Spinning; Jackknifing	Skidding	Spinning; Not Applicable	Spinning
Avoidance Attempt[b]	Avoid Pedestrian/ Pedalcyclist; Avoid Vehicle; Avoid Unknown Object	Variable not available	Avoid Unknown Object		Not Applicable	Avoid Unknown Object

TABLE 3 (continued) Vehicle and Driver Variables and Attributes Unavailable in CARDfile

Variable	Excluded Attributes by State					
	Indiana	Maryland	Michigan	Penn.	Texas	Washington
VIN	Variable not available			Variable not available	Variable not available	Variable not available
Alcohol/ Drug Use			Drug Use (coded with Alcohol Use)			
Restraint Use						Not Equipped with Restraint
Driver Error	Failed to Signal (coded with Other Error)	Failed to Signal (coded with Other Error)	Right of Way Violation (coded with Disregarded Traffic Sign or Signal)			

[(a)] Indiana did not code for tractors with double trailers prior to 1986.

[(b)] Texas codes this variable only for single vehicle accidents.

[(c)] The "Jackknifing" attribute of Precrash Stability became available for Indiana in 1986.

TABLE 4
Missing Data Rates for Selected CARDfile Variables, 1984–1986

CARDfile Variable	Overall Missing Data Rate
Roadway Separation	92.14%
Precrash Stability	69.91
Vehicle Identification Number (VIN)	66.30
Avoidance Attempt	44.08
Land Use	42.50
Restraint Use	35.18
Roadway Profile	27.21
Driver Age	11.82
Alcohol/Drug Use	11.13
Driver Error	10.59
Driver Sex	7.71
Model Year	5.77
Vehicle Type	4.72
Road Surface	0.79
Light Conditions	0.71
Weather Conditions	0.65
Primary Impact	0.61
Intersection Signalization	0.52
Roadway Alignment	0.51
Relation to Intersection	0.34
Primary Impact Location	0.33
Accident Severity	0.04

NOTE: Rates are calculated based on the percentage of cases coded as either "Missing" or "Unknown", or as "No Driver Record" for the Driver variables, out of all the cases for 1984–1986.

TABLE 5
Accident Severity by State
1984–1986 CARDfile

Accident Severity	Indiana		Maryland		Michigan		Pennsylvania		Texas		Washington		Total	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Property Damage	426908	77.31%	230763	59.60%	804077	71.59%	150017	34.58%	891607	65.92%	220083	62.06%	2723455	64.79%
Possible Injury	48647	8.81	70652	18.25	164798	14.67	176642	40.72	197805	14.62	57030	16.08	715574	17.02
Nonincap. Injury	68459	12.40	46836	12.10	98569	8.78	83544	19.26	191745	14.18	55600	15.68	544753	12.96
Incap. Injury	5581	1.01	36923	9.54	51425	4.58	16859	3.89	61537	4.55	19909	5.61	192234	4.57
Fatal Injury	2581	0.47	1983	0.51	4277	0.38	4941	1.14	9836	0.73	2001	0.56	25619	0.61
Missing	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	0.00	1	0.00
Unknown	0	0.00	0	0.00	0	0.00	1808	0.42	0	0.00	0	0.00	1808	0.04
Total	552176	100.00%	387157	100.00%	1123146	100.00%	433811	100.00%	1352530	100.00%	354624	100.00%	4203444	100.00%

TABLE 5A
Accident Severity by State
(Excludes Property Damage Only Cases)
1984–1986 CARDfile

Accident Severity	Indiana		Maryland		Michigan		Pennsylvania		Texas		Washington		Total	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Possible Injury	48647	38.83%	70652	45.18%	164798	51.65%	176642	62.24%	197805	42.91%	57030	42.39%	715574	48.35%
Nonincap. Injury	68459	54.65	46836	29.95	98569	30.89	83544	29.44	191745	41.60	55600	41.33	544753	36.81
Incap. Injury	5581	4.46	36923	23.61	51425	16.12	16859	5.94	61537	13.35	19909	14.80	192234	12.99
Fatal Injury	2581	2.06	1983	1.27	4277	1.34	4941	1.74	9836	2.13	2001	1.49	25619	1.73
Missing	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	1	0.00	1	0.00
Unknown	0	0.00	0	0.00	0	0.00	1808	0.64	0	0.00	0	0.00	1808	0.12
Total	125268	100.00%	156394	100.00%	319069	100.00%	283794	100.00%	460923	100.00%	134541	100.00%	1479989	100.00%

TABLE 6
Light Conditions by State
1984–1986 CARDfile

Light Conditions	Indiana		Maryland		Michigan		Pennsylvania		Texas		Washington		Total	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Day	343960	62.29%	224223	57.92%	694010	61.79%	236179	54.44%	925752	68.45%	228386	64.40%	2652510	63.10%
Dark/Lighted	95182	17.24	98352	25.40	125113	11.14	91630	21.12	140801	10.41	67852	19.13	618930	14.72
Dark/Not Lighted	74142	13.43	34709	8.97	245187	21.83	78445	18.08	256501	18.96	41380	11.67	730364	17.38
Dawn	9754	1.77	4739	1.22	21729	1.93	10232	2.36	8721	0.64	5340	1.51	60515	1.44
Dusk	17695	3.20	11394	2.94	32874	2.93	17093	3.94	20755	1.53	11666	3.29	111477	2.65
Missing	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
Unknown	11443	2.07	13740	3.55	4233	0.38	232	0.05	0	0.00	0	0.00	29648	0.71
Total	552176	100.00%	387157	100.00%	1123146	100.00%	433811	100.00%	1352530	100.00%	354624	100.00%	4203444	100.00%

TABLE 7
Weather Conditions by State
1984–1986 CARDfile

Weather Conditions	Indiana		Maryland		Michigan		Pennsylvania		Texas		Washington		Total	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Clear/Cloudy	425144	76.99%	298434	77.08%	870623	77.52%	323701	74.62%	1162945	85.98%	274784	77.49%	3355631	79.83%
Rain	73487	13.31	62617	16.17	136088	12.12	74767	17.23	168072	12.43	61131	17.24	576162	13.71
Snow/Ice/Sleet	37282	6.75	8693	2.25	100697	8.97	28996	6.68	9426	0.70	10549	2.97	195643	4.65
Other	5291	0.96	5724	1.48	12287	1.09	5232	1.21	12086	0.89	8160	2.30	48780	1.16
Missing	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
Unknown	10972	1.99	11689	3.02	3451	0.31	1115	0.26	1	0.00	0	0.00	27228	0.65
Total	552176	100.00%	387157	100.00%	1123146	100.00%	433811	100.00%	1352530	100.00%	354624	100.00%	4203444	100.00%

TABLE 8
Road Surface by State
1984–1986 CARDfile

Road Surface	Indiana		Maryland		Michigan		Pennsylvania		Texas		Washington		Total	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Dry	352871	63.91%	268420	69.33%	681634	60.69%	285951	65.92%	1099792	81.31%	227498	64.15%	2916166	69.38%
Wet	118686	21.49	88356	22.82	245715	21.88	99390	22.91	231133	17.09	92131	25.98	875411	20.83
Snow/Ice	63186	11.44	18535	4.79	187032	16.65	46836	10.80	19858	1.47	34995	9.87	370442	8.81
Other	5267	0.95	790	0.20	0	0.00	535	0.12	1747	0.13	0	0.00	8339	0.20
Missing	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
Unknown	12166	2.20	11056	2.86	8765	0.78	1099	0.25	0	0.00	0	0.00	33086	0.79
Total	552176	100.00%	387157	100.00%	1123146	100.00%	433811	100.00%	1352530	100.00%	354624	100.00%	4203444	100.00%

TABLE 9
Land Use by State
1984–1986 CARDfile

Land Use	Indiana		Maryland		Michigan		Pennsylvania		Texas		Washington		Total	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Rural	199282	36.09%	79017	20.41%	426771	38.00%	0	0.00%	0	0.00%	72553	20.46%	777623	18.50%
Urban	352894	63.91	308140	79.59	696375	62.00	0	0.00	0	0.00	282071	79.54	1639480	39.00
Missing	0	0.00	0	0.00	0	0.00	433811	100.00	1352530	100.00	0	0.00	1786341	42.50
Unknown	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
Total	552176	100.00%	387157	100.00%	1123146	100.00%	433811	100.00%	1352530	100.00%	354624	100.00%	4203444	100.00%

TABLE 10
Primary Impact by State
1984–1986 CARDfile

Primary Impact	Indiana		Maryland		Michigan		Pennsylvania		Texas		Washington		Total	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Vehicle in Transport	363137	65.76%	214586	55.43%	717673	63.90%	224806	51.82%	975795	72.15%	234671	66.17%	2730668	64.96%
Vehicle not in Trans.	20408	3.70	81106	20.95	90293	8.04	23998	5.53	90720	6.71	27625	7.79	334150	7.95
Pedestrian	5858	1.06	10304	2.66	13339	1.19	19769	4.56	17241	1.27	4996	1.41	71507	1.70
Pedalcyclist	5124	0.93	4564	1.18	12995	1.16	5515	1.27	10317	0.76	4006	1.13	42521	1.01
Fixed/Sta. Object	92494	16.75	62858	16.24	153183	13.64	139562	32.17	183738	13.58	62770	17.70	694605	16.52
Noncollision	5889	1.07	4571	1.18	0	0.00	5924	1.37	8823	0.65	1214	0.34	26421	0.63
Collision with Train	1243	0.23	124	0.03	845	0.08	427	0.10	1992	0.15	267	0.08	4898	0.12
Rollover	15647	2.83	3089	0.80	34303	3.05	9218	2.12	43693	3.23	14897	4.20	120847	2.87
Other	21288	3.86	5955	1.54	100515	8.95	124	0.03	20211	1.49	4172	1.18	152265	3.62
Missing	2029	0.37	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	2029	0.05
Unknown	19059	3.45	0	0.00	0	0.00	4468	1.03	0	0.00	6	0.00	23533	0.56
Total	552176	100.00%	387157	100.00%	1123146	100.00%	433811	100.00%	1352530	100.00%	354624	100.00%	4203444	100.00%

TABLE 11
Primary Impact Location by State
1984–1986 CARDfile

Primary Impact Location	Indiana		Maryland		Michigan		Pennsylvania		Texas		Washington		Total	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
On Roadway	401648	72.74%	263463	68.05%	853580	76.00%	305436	70.41%	1113328	82.31%	273352	77.08%	3210807	76.39%
On Shoulder	18876	3.42	20812	5.38	0	0.00	25455	5.87	164464	12.16	0	0.00	229607	5.46
Off Roadway	117846	21.34	102882	26.57	269566	24.00	102825	23.70	74738	5.53	81272	22.92	749129	17.82
Missing	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
Unknown	13806	2.50	0	0.00	0	0.00	95	0.02	0	0.00	0	0.00	13901	0.33
Total	552176	100.00%	387157	100.00%	1123146	100.00%	433811	100.00%	1352530	100.00%	354624	100.00%	4203444	100.00%

TABLE 12
Relation to Intersection by State
1984–1986 CARDfile

Relation to Intersection	Indiana		Maryland		Michigan		Pennsylvania		Texas		Washington		Total	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Intersection Related	163762	29.66%	134886	34.84%	655315	58.35%	166184	38.31%	633234	46.82%	154782	43.65%	1908163	45.40%
Not Intersection Rel.	349356	63.27	223977	57.85	467410	41.62	234841	54.13	540290	39.95	156701	44.19	1972575	46.93
Driveway Related	25252	4.57	28294	7.31	0	0.00	32784	7.56	179001	13.23	43141	12.17	308472	7.34
Missing	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
Unknown	13806	2.50	0	0.00	421	0.04	2	0.00	5	0.00	0	0.00	14234	0.34
Total	552176	100.00%	387157	100.00%	1123146	100.00%	433811	100.00%	1352530	100.00%	354624	100.00%	4203444	100.00%

TABLE 13
Intersection Signalization by State
1984–1986 CARDfile

Intersection Signalization	Indiana		Maryland		Michigan		Pennsylvania		Texas		Washington		Total	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Signed	38556	6.98%	23522	6.08%	216907	19.31%	41897	9.66%	198969	14.71%	38036	10.73%	557907	13.27%
Signalled	55777	10.10	55458	14.32	199744	17.78	59534	13.72	234888	17.37	57634	16.25	663035	15.77
No Traffic Controls	61804	11.19	54433	14.06	237771	21.17	62452	14.40	192106	14.20	55804	15.74	664370	15.81
Other	3110	0.56	1473	0.38	343	0.03	797	0.18	7251	0.54	2209	0.62	15183	0.36
Not Applicable	374608	67.84	252271	65.16	467410	41.62	267625	61.69	719291	53.18	199841	56.35	2281046	54.27
Missing	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	466	0.13	466	0.01
Unknown	18321	3.32	0	0.00	971	0.09	1506	0.35	5	0.00	634	0.18	21437	0.51
Total	552176	100.00%	387157	100.00%	1123146	100.00%	433811	100.00%	1352530	100.00%	354624	100.00%	4203444	100.00%

TABLE 14
Roadway Alignment by State
1984–1986 CARDfile

Roadway Alignment	Indiana		Maryland		Michigan		Pennsylvania		Texas		Washington		Total	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Straight	486316	88.07%	325341	84.03%	1064828	94.81%	357689	82.45%	1278599	94.53%	297825	83.98%	3810598	90.65%
Curved	54222	9.82	54871	14.17	55776	4.97	76014	17.52	73931	5.47	56799	16.02	371613	8.84
Missing	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
Unknown	11638	2.11	6945	1.79	2542	0.23	108	0.02	0	0.00	0	0.00	21233	0.51
Total	552176	100.00%	387157	100.00%	1123146	100.00%	433811	100.00%	1352530	100.00%	354624	100.00%	4203444	100.00%

TABLE 15
Roadway Profile by State
1984–1986 CARDfile

Roadway Profile	Indiana		Maryland		Michigan		Pennsylvania		Texas		Washington		Total	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Level	443751	80.36%	264043	68.20%	0	0.00%	289931	66.83%	1345796	99.50%	253080	71.37%	2596601	61.77%
Grade	96787	17.53	116169	30.01	0	0.00	141795	32.69	6734	0.50	101544	28.63	463029	11.02
Missing	0	0.00	0	0.00	1123146	100.00	0	0.00	0	0.00	0	0.00	1123146	26.72
Unknown	11638	2.11	6945	1.79	0	0.00	2085	0.48	0	0.00	0	0.00	20668	0.49
Total	552176	100.00%	387157	100.00%	1123146	100.00%	433811	100.00%	1352530	100.00%	354624	100.00%	4203444	100.00%

TABLE 16
Roadway Separation by State
1984–1986 CARDfile

Roadway Separation	Indiana		Maryland		Michigan		Pennsylvania		Texas		Washington		Total	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Two Way Undivided	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%	228347	64.39%	228347	5.43%
Two Way Divided	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	45876	12.94	45876	1.09
One Way	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	35159	9.91	35159	0.84
Other	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	21030	5.93	21030	0.50
Missing	552176	100.00	387157	100.00	1123146	100.00	433811	100.00	1352530	100.00	5632	1.59	3854452	91.70
Unknown	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	18580	5.24	18580	0.44
Total	552176	100.00%	387157	100.00%	1123146	100.00%	433811	100.00%	1352530	100.00%	354624	100.00%	4203444	100.00%

TABLE 17
Accident Type by State
1984–1986 CARDfile

Accident Type	Indiana		Maryland		Michigan		Pennsylvania		Texas		Washington		Total	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
SINGLE VEHICLE														
Rollover	14590	2.64%	10427	2.69%	35186	3.13%	36131	8.33%	40843	3.02%	16646	4.69%	153823	3.66%
Stationary Object	92654	16.78	56574	14.61	151748	13.51	109803	25.31	181199	13.40	59962	16.91	651940	15.51
Ped./Animal/Cyclist	28107	5.09	18056	4.66	114691	10.21	31704	7.31	40269	2.98	12727	3.59	245554	5.84
Parked Vehicle	42824	7.76	56535	14.60	61935	5.51	18566	4.28	84252	6.23	24346	6.87	288458	6.86
Noncollision	5299	0.96	4403	1.14	1898	0.17	4595	1.06	7942	0.59	991	0.28	25128	0.60
Specifics Unknown	8342	1.51	2559	0.66	32384	2.88	2845	0.66	3475	0.26	1	0.00	49606	1.18
Other	1055	0.19	107	0.03	5600	0.50	702	0.16	1930	0.14	1626	0.46	11020	0.26
Subtotal	192871	34.93%	148661	38.40%	403442	35.92%	204346	47.10%	359910	26.61%	116299	32.80%	1425529	33.91%
MULTIVEHICLE														
Rear End—Same Trafficway Same Direction														
Lead Vehicle Stopped	25033	4.53%	23944	6.18%	102066	9.09%	34260	7.90%	141211	10.44%	37312	10.52%	363826	8.66%
Lead Vehicle Movg. Str.	26619	4.82	17697	4.57	50172	4.47	101	0.02	72979	5.40	12191	3.44	179759	4.28
Lead Vehicle Turning	12767	2.31	4909	1.27	46506	4.14	2291	0.53	91385	6.76	7794	2.20	165652	3.94
Direction Unknown	0	0.00	1138	0.29	0	0.00	0	0.00	0	0.00	0	0.00	1138	0.03
Specifics Unknown	14039	2.54	4468	1.15	36187	3.22	12338	2.84	0	0.00	2824	0.80	69856	1.66
Subtotal	78458	14.21%	52156	13.47%	234931	20.92%	48990	11.29%	305575	22.59%	60121	16.95%	780231	18.56%

TABLE 17 (continued) Accident Type by State

Accident Type	Indiana		Maryland		Michigan		Pennsylvania		Texas		Washington		Total	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Sideswipe—Same Trafficway Same Direction														
Passing	2469	0.45%	1336	0.35%	14940	1.33%	320	0.07%	24345	1.80%	6410	1.81%	49820	1.19%
Overtaking Right	494	0.09	0	0.00	944	0.08	173	0.04	1606	0.12	0	0.00	3217	0.08
Overtaking Left	1268	0.23	0	0.00	1879	0.17	739	0.17	2717	0.20	0	0.00	6603	0.16
Changing Lanes	7067	1.28	9264	2.39	18146	1.62	934	0.22	0	0.00	9129	2.57	44540	1.06
Specifics Unknown	21737	3.94	8907	2.30	2664	0.24	1174	0.27	812	0.06	2394	0.68	37688	0.90
Subtotal	33035	5.98%	19507	5.04%	38573	3.43%	3340	0.77%	29480	2.18%	17933	5.06%	141868	3.38%
Other—Same Trafficway Same Direction														
Other	0	0.00%	938	0.24%	0	0.00%	0	0.00%	10591	0.78%	7742	2.18%	19271	0.46%
Head-on Sideswipe—Same Trafficway Opposite Direction														
Lat. Move/Lane Change	124	0.02%	149	0.04%	5013	0.45%	1877	0.43%	0	0.00%	57	0.02%	7220	0.17%
Both Straight/Passing	467	0.08	10862	2.81	27436	2.44	20585	4.75	33487	2.48	8357	2.36	101194	2.41
Specifics Unknown	26487	4.80	196	0.05	1823	0.16	7207	1.66	0	0.00	178	0.05	35891	0.85
Subtotal	27078	4.90%	11207	2.89%	34272	3.05%	29669	6.84%	33487	2.48%	8592	2.42%	144305	3.43%
Other—Same Trafficway Opposite Direction														
Other	0	0.00%	0	0.00%	0	0.00%	0	0.00%	1983	0.15%	3130	0.88%	5113	0.12%

TABLE 17 (continued) Accident Type by State

Accident Type	Indiana		Maryland		Michigan		Pennsylvania		Texas		Washington		Total	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Initial Opposite Direction—Change Trafficway/Turn Across Path														
One Straight/One Right	83	0.02%	0	0.00%	290	0.03%	108	0.02%	267	0.02%	205	0.06%	953	0.02%
One Straight/One Left	18127	3.28	20404	5.27	58447	5.20	29668	6.84	97629	7.22	16717	4.71	240992	5.73
Both Turning Left	281	0.05	117	0.03	896	0.08	384	0.09	970	0.07	32	0.01	2680	0.06
One Right/One Left	289	0.05	0	0.00	1005	0.09	526	0.12	1710	0.13	518	0.15	4048	0.10
Specifics Unknown	2085	0.38	716	0.18	2824	0.25	0	0.00	0	0.00	316	0.09	5941	0.14
Subtotal	20865	3.78%	21237	5.49%	63462	5.65%	30686	7.07%	100576	7.44%	17788	5.02%	254614	6.06%
Initial Same Direction—Change Trafficway/Turn Across Path														
One Straight/One Right	3034	0.55%	3293	0.85%	537	0.05%	2247	0.52%	1532	0.11%	290	0.08%	10933	0.26%
One Straight/One Left	5612	1.02	4538	1.17	226	0.02	3090	0.71	454	0.03	124	0.03	14044	0.33
Specifics Unknown	2750	0.50	996	0.26	4572	0.41	0	0.00	260	0.02	0	0.00	8578	0.20
Subtotal	11396	2.06%	8827	2.28%	5335	0.48%	5337	1.23%	2246	0.17%	414	0.12%	33555	0.80%
Straight Turning—One Straight or Stopped—Intersecting Path/Angles														
Direction Unknown	1562	0.28%	67	0.02%	1786	0.16%	0	0.00%	41581	3.07%	441	0.12%	45437	1.08%
Same Direction/TL	1952	0.35	3231	0.83	16522	1.47	2509	0.58	17499	1.29	4	0.00	41717	0.99
Same Direction/TR	1992	0.36	2314	0.60	12316	1.10	1983	0.46	19744	1.46	0	0.00	38349	0.91
Opposite Dir./TR	740	0.13	0	0.00	8777	0.78	1497	0.35	6464	0.48	0	0.00	17478	0.42
Opposite Dir./TL	4690	0.85	7744	2.00	36499	3.25	11435	2.64	45902	3.39	29	0.01	106299	2.53
Subtotal	10936	1.98%	13356	3.45%	75900	6.76%	17424	4.02%	131190	9.70%	474	0.13%	249280	5.93%
Straight Paths—Intersecting Path/Angles														
Both Straight	91279	16.53%	29575	7.64%	113609	10.12%	34511	7.96%	177620	13.13%	23509	6.63%	470103	11.18%

TABLE 17 (continued) Accident Type by State

Accident Type	Indiana		Maryland		Michigan		Pennsylvania		Texas		Washington		Total	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Curved Paths—Intersecting Path/Angles														
Both Curving/Turning	0	0.00%	0	0.00%	5549	0.49%	0	0.00%	5821	0.43%	1398	0.39%	12768	0.30%
Other—Intersecting Path/Angles														
Other	0	0.00%	0	0.00%	212	0.02%	0	0.00%	0	0.00%	35542	10.02%	35754	0.85%
Backing														
One Vehicle Backing	34284	6.21%	21246	5.49%	33159	2.95%	2400	0.55%	25828	1.91%	4529	1.28%	121446	2.89%
Two Vehicles Backing	1702	0.31	421	0.11	1633	0.15	17	0.00	1211	0.09	226	0.06	5210	0.12
Subtotal	35986	6.52%	21667	5.60%	34792	3.10%	2417	0.56%	27039	2.00%	4755	1.34%	126656	3.01%
Parking														
En/Lv Parking Space	5543	1.00%	9330	2.41%	8873	0.79%	1088	0.25%	11340	0.84%	3056	0.86%	39230	0.93%
Other Multivehicles														
Collisions with 2 Veh.	15693	2.84%	19828	5.12%	53170	4.73%	18420	4.25%	55981	4.14%	29210	8.24%	192302	4.57%
Collisions with > 2 Veh.	29036	5.26	30868	7.97	51026	4.54	37583	8.66	99691	7.37	24661	6.95	272865	6.49
Subtotal	44729	8.10%	50696	13.09%	104196	9.28%	56003	12.91%	155672	11.51%	53871	15.19%	465167	11.07%
Other														
Missing	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%
TOTAL	552176	100.00%	387157	100.00%	1123146	100.00%	433811	100.00%	1352530	100.00%	354624	100.00%	4203444	100.00%

TABLE 18
Accident Type by State (Vehicle Level)
1984–1986 CARDfile

Accident Type (Vehicle Level)	Indiana		Maryland		Michigan		Pennsylvania		Texas		Washington		Total	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
SINGLE VEHICLE														
Rollover	14590	1.48%	11228	1.55%	35223	1.85%	36156	4.94%	40843	1.74%	16709	2.57%	154749	2.11%
Stationary Object	92654	9.38	57004	7.89	151748	7.97	109803	15.00	181199	7.72	61618	9.49	654026	8.91
Ped./Animal/Cyclist	28107	2.85	18100	2.51	114691	6.02	31704	4.33	40269	1.72	12894	1.99	245765	3.35
Parked Vehicle	42824	4.34	56535	7.83	61935	3.25	18566	2.54	84252	3.59	24345	3.75	288457	3.93
Noncollision	5299	0.54	4421	0.61	1898	0.10	4595	0.63	7942	0.34	1033	0.16	25188	0.34
Specifics Unknown	8342	0.84	2626	0.36	32384	1.70	2758	0.38	3475	0.15	1	0.00	49586	0.68
Other	1055	0.11	214	0.03	5600	0.29	702	0.10	1930	0.08	1674	0.26	11175	0.15
Subtotal	192871	19.53%	150128	20.79%	403479	21.19%	204284	27.90%	359910	15.34%	118274	18.22%	1428946	19.46%
MULTIVEHICLE														
Rear End—Same Trafficway Same Direction														
Lead Veh. Stop—RV	25033	2.54%	23944	3.32%	102066	5.36%	34260	4.68%	141211	6.02%	37312	5.75%	363826	4.96%
Lead Veh. Stop—LV	25033	2.54	23944	3.32	102066	5.36	34260	4.68	141211	6.02	37312	5.75	363826	4.96
LV Moving Str.—RV	26619	2.70	17697	2.45	50172	2.64	101	0.01	72979	3.11	12191	1.88	179759	2.45
LV Moving Str.—LV	26619	2.70	17697	2.45	50172	2.64	101	0.01	72978	3.11	12191	1.88	179758	2.45
Lead Veh. Turn—RV	12767	1.29	4909	0.68	46506	2.44	2291	0.31	91385	3.89	7794	1.20	165652	2.26
Lead Veh. Turn—LV	12767	1.29	4909	0.68	46506	2.44	2291	0.31	91385	3.89	7794	1.20	165652	2.26
Direction Unk.—RV	0	0.00	1138	0.16	0	0.00	0	0.00	0	0.00	0	0.00	1138	0.02
Direction Unk.—LV	0	0.00	1138	0.16	0	0.00	0	0.00	0	0.00	0	0.00	1138	0.02
Specifics Unk.—Both	28078	2.84	8936	1.24	72374	3.80	24676	3.37	0	0.00	5648	0.87	139712	1.90
Subtotal	156916	15.89%	104312	14.44%	469862	24.68%	97980	13.38%	611149	26.04%	120242	18.53%	1560461	21.26%

TABLE 18 (continued) Accident Type by State (Vehicle Level)

Accident Type (Vehicle Level)	Indiana		Maryland		Michigan		Pennsylvania		Texas		Washington		Total	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Sideswipe—Same Trafficway Same Direction														
Passing	4938	0.50%	2672	0.37%	29880	1.57%	640	0.09%	48688	2.07%	12820	1.98%	99638	1.36%
On Right—Veh. Str.	494	0.05	0	0.00	944	0.05	173	0.02	1606	0.07	0	0.00	3217	0.04
On Right—Veh. Ovtk.	494	0.05	0	0.00	944	0.05	173	0.02	1606	0.07	0	0.00	3217	0.04
On Left—Veh. Str.	1268	0.13	0	0.00	1879	0.10	739	0.10	2717	0.12	0	0.00	6603	0.09
On Left—Veh. Ovtk.	1268	0.13	0	0.00	1879	0.10	739	0.10	2717	0.12	0	0.00	6603	0.09
Chg. Lanes—Veh. Str.	7067	0.72	9264	1.28	18146	0.95	934	0.13	0	0.00	9129	1.41	44540	0.61
Chg. Lanes—Veh. Chg.	7067	0.72	9264	1.28	18146	0.95	934	0.13	0	0.00	9129	1.41	44540	0.61
Specifics Unknown	43474	4.40	17814	2.47	5328	0.28	2348	0.32	1624	0.07	4788	0.74	75376	1.03
Subtotal	66070	6.69%	39014	5.40%	77146	4.05%	6680	0.91%	58958	2.51%	35866	5.53%	283734	3.86%
Other—Same Trafficway Same Direction														
Other	0	0.00%	1876	0.26%	0	0.00%	0	0.00%	21182	0.90%	15484	2.39%	38542	0.53%
Head-on Sideswipe—Same Trafficway Opposite Direction														
Lat. Move—Veh. Mov'g	124	0.01%	149	0.02%	5013	0.26%	1877	0.26%	0	0.00%	57	0.01%	7220	0.10%
Lat. Move—Veh. Str'ght	124	0.01	149	0.02	5013	0.26	1877	0.26	0	0.00	57	0.01	7220	0.10
Both Straight/Passing	934	0.09	21724	3.01	54872	2.88	41170	5.62	66974	2.85	16714	2.58	202388	2.76
Specifics Unknown	52974	5.37	392	0.05	3646	0.19	14414	1.97	0	0.00	356	0.05	71782	0.98
Subtotal	54156	5.48%	22414	3.10%	68544	3.60%	59338	8.10%	66974	2.85%	17184	2.65%	288610	3.93%
Other—Same Trafficway Opposite Direction														
Other	0	0.00%	0	0.00%	0	0.00%	0	0.00%	3966	0.17%	6260	0.96%	10226	0.14%

TABLE 18 (continued) Accident Type by State (Vehicle Level)

Accident Type (Vehicle Level)	Indiana		Maryland		Michigan		Pennsylvania		Texas		Washington		Total	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Initial Opposite Direction—Change Trafficway/Turn Across Path														
One Str./One Right—RT	83	0.01%	0	0.00%	290	0.02%	108	0.01%	267	0.01%	205	0.03%	953	0.01%
One Str./One Right—NT	83	0.01	0	0.00	290	0.02	108	0.01	267	0.01	205	0.03	953	0.01
One Str./One Left—LT	18127	1.84	20404	2.83	58447	3.07	29668	4.05	97628	4.16	16717	2.58	240991	3.28
One Str./One Left—NT	18127	1.84	20404	2.83	58447	3.07	29668	4.05	97629	4.16	16717	2.58	240992	3.28
Both Turning Left	562	0.06	234	0.03	1792	0.09	768	0.10	1940	0.08	64	0.01	5360	0.07
Both Turning—RT	289	0.03	0	0.00	1005	0.05	526	0.07	1710	0.07	518	0.08	4048	0.06
Both Turning—LT	289	0.03	0	0.00	1005	0.05	526	0.07	1710	0.07	518	0.08	4048	0.06
Specifics Unknown	4170	0.42	1432	0.20	5648	0.30	0	0.00	0	0.00	632	0.10	11882	0.16
Subtotal	41730	4.23%	42474	5.88%	126924	6.67%	61372	8.38%	201151	8.57%	35576	5.48%	509227	6.94%
Initial Same Direction—Change Trafficway/Turn Across Path														
One Str./One Right—RT	3034	0.31%	3293	0.46%	537	0.03%	2247	0.31%	1532	0.07%	290	0.04%	10933	0.15%
One Str./One Right—NT	3034	0.31	3293	0.46	537	0.03	2247	0.31	1532	0.07	290	0.04	10933	0.15
One Str./One Left—LT	5612	0.57	4538	0.63	226	0.01	3090	0.42	454	0.02	124	0.02	14044	0.19
One Str./One Left—NT	5612	0.57	4538	0.63	226	0.01	3090	0.42	454	0.02	124	0.02	14044	0.19
Specifics Unknown	5500	0.56	1992	0.28	9144	0.48	0	0.00	520	0.02	0	0.00	17156	0.23
Subtotal	22792	2.31%	17654	2.44%	10670	0.56%	10674	1.46%	4492	0.19%	828	0.13%	67110	0.91%

TABLE 18 (continued) Accident Type by State (Vehicle Level)

Accident Type (Vehicle Level)	Indiana		Maryland		Michigan		Pennsylvania		Texas		Washington		Total	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Straight Turning—One Straight or Stopped—Intersecting Path/Angles														
Dir. Unk.—Turning Veh.	1562	0.16%	67	0.01%	1786	0.09%	0	0.00%	41581	1.77%	441	0.07%	45437	0.62%
Dir. Unk.—Veh. Str.	1562	0.16	67	0.01	1786	0.09	0	0.00	41580	1.77	441	0.07	45436	0.62
Same Dir.—LT Veh.	1952	0.20	3231	0.45	16522	0.87	2509	0.34	17499	0.75	4	0.00	41717	0.57
Same Dir.—Veh. Str.	1952	0.20	3231	0.45	16522	0.87	2509	0.34	17499	0.75	4	0.00	41717	0.57
Same Dir.—RT Veh.	1992	0.20	2314	0.32	12316	0.65	1983	0.27	19744	0.84	0	0.00	38349	0.52
Same Dir.—Veh. Str.	1992	0.20	2314	0.32	12316	0.65	1983	0.27	19744	0.84	0	0.00	38349	0.52
Opp. Dir.—RT Veh.	740	0.07	0	0.00	8777	0.46	1497	0.20	6464	0.28	0	0.00	17478	0.24
Opp. Dir.—Veh. Str.	740	0.07	0	0.00	8777	0.46	1497	0.20	6464	0.28	0	0.00	17478	0.24
Opp. Dir.—LT Veh.	4690	0.47	7744	1.07	36499	1.92	11435	1.56	45902	1.96	29	0.00	106299	1.45
Opp. Dir.—Veh. Str.	4690	0.47	7744	1.07	36499	1.92	11435	1.56	45901	1.96	29	0.00	106298	1.45
Subtotal	21872	2.22%	26712	3.70%	151800	7.97%	34848	4.76%	262378	11.18%	948	0.15%	498558	6.79%
Straight Paths—Intersecting Path/Angles														
Both Straight	182558	18.49%	59150	8.19%	227218	11.94%	69022	9.43%	355240	15.14%	47018	7.24%	940206	12.81%
Curved Paths—Intersecting Path/Angles														
Both Curving/Turning	0	0.00%	0	0.00%	11098	0.58%	0	0.00%	11642	0.50%	2796	0.43%	25536	0.35%
Other—Intersecting Path/Angles														
Other	0	0.00%	0	0.00%	424	0.02%	0	0.00%	0	0.00%	71084	10.95%	71508	0.97%
Backing														
One Veh. Back—Stopped	34284	3.47%	21246	2.94%	33159	1.74%	2400	0.33%	25828	1.10%	4529	0.70%	121446	1.65%
One Veh. Back—Backing	34284	3.47	21246	2.94	33159	1.74	2400	0.33	25828	1.10	4529	0.70	121446	1.65
Two Vehicles Backing	3404	0.34	842	0.12	3266	0.17	34	0.00	2422	0.10	452	0.07	10420	0.14

TABLE 18 (continued) Accident Type by State (Vehicle Level)

Accident Type (Vehicle Level)	Indiana		Maryland		Michigan		Pennsylvania		Texas		Washington		Total	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Backing														
Subtotal	71972	7.29%	43334	6.00%	69584	3.66%	4834	0.66%	54078	2.30%	9510	1.47%	253312	3.45%
Parking														
En./Lv. Space—Parking	5543	0.56%	9330	1.29%	8873	0.47%	1088	0.15%	11340	0.48%	3056	0.47%	39230	0.53%
En./Lv. Space—Other	5543	0.56	9330	1.29	8873	0.47	1088	0.15	11340	0.48	3056	0.47	39230	0.53
Subtotal	11086	1.12%	18660	2.58%	17746	0.93%	2176	0.30%	22680	0.97%	6112	0.94%	78460	1.07%
Other Multivehicles														
Collisions with 2 Veh.	31386	3.18%	39656	5.49%	106352	5.59%	36776	5.02%	105993	4.52%	58418	9.00%	378581	5.16%
Collisions with > 2 Veh.	92276	9.35	100259	13.88	161831	8.50	121999	16.66	203492	8.67	78984	12.17	758841	10.34
Subtotal	123662	12.52%	139915	19.37%	268183	14.09%	158775	21.68%	309485	13.19%	137402	21.17%	1137422	15.49%
PARKED VEHICLE RECORD														
Parked Vehicle	41713	4.22%	56535	7.83%	1026	0.05%	22256	3.04%	3417	0.15%	24488	3.77%	149435	2.04%
Other														
Missing	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%
TOTAL	987398	100.00%	722178	100.00%	1903704	100.00%	732239	100.00%	2346702	100.00%	649072	100.00%	7341293	100.00%

TABLE 19
Model Year by State
1984–1986 CARDfile

Model Year	Indiana		Maryland		Michigan		Pennsylvania		Texas		Washington		Total	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Known	938996	95.10%	661087	91.54%	1783000	93.66%	705942	96.41%	2226694	94.89%	602197	92.78%	6917916	94.23%
Missing	2	0.00	51289	7.10	0	0.00	45	0.01	0	0.00	2	0.00	51338	0.70
Unknown	48400	4.90	9802	1.36	120704	6.34	26252	3.59	120008	5.11	46873	7.22	372039	5.07
Total	987398	100.00%	722178	100.00%	1903704	100.00%	732239	100.00%	2346702	100.00%	649072	100.00%	7341293	100.00%

TABLE 20
Vehicle Type by State
1984–1986 CARDfile

Vehicle Type	Indiana		Maryland		Michigan		Pennsylvania		Texas		Washington		Total	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Motorcycle/Moped	12554	1.27%	9837	1.36%	21378	1.12%	16437	2.24%	33856	1.44%	11339	1.75%	105401	1.44%
Passenger Car	724969	73.42	521441	72.20	1379877	72.48	567576	77.51	1561523	66.54	446865	68.85	5202251	70.86
Car w/ Trailer	1533	0.16	961	0.13	4055	0.21	480	0.07	1835	0.08	422	0.07	9286	0.13
Light Van/Truck	129225	13.09	91040	12.61	289680	15.22	98739	13.48	550520	23.46	134253	20.68	1293457	17.62
Van/Tr. w/ Trailer	1953	0.20	1227	0.17	6216	0.33	1011	0.14	7894	0.34	1935	0.30	20236	0.28
Med./Heavy Str. Tr.	24539	2.49	14909	2.06	19709	1.04	10088	1.38	21217	0.90	11487	1.77	101949	1.39
M/H Str. Tr. w/ Trail	706	0.07	617	0.09	1266	0.07	448	0.06	698	0.03	1252	0.19	4987	0.07
Art. Tr. w/ S/F Trail	24353	2.47	8138	1.13	21388	1.12	16973	2.32	54310	2.31	9485	1.46	134647	1.83
Art. Tr. w/ Dbl. Trail	56	0.01	656	0.09	1838	0.10	243	0.03	0	0.00	0	0.00	2793	0.04
Art. Tr. w/o Trailer	2449	0.25	3775	0.52	7046	0.37	1003	0.14	2047	0.09	772	0.12	17092	0.23
Other Comb. Vehicle	749	0.08	122	0.02	0	0.00	5827	0.80	1954	0.08	828	0.13	9480	0.13
Fire Truck	462	0.05	0	0.00	712	0.04	271	0.04	424	0.02	113	0.02	1982	0.03
Ambulance	562	0.06	731	0.10	784	0.04	215	0.03	619	0.03	114	0.02	3025	0.04
Other Emer. Vehicle	0	0.00	799	0.11	0	0.00	101	0.01	210	0.01	294	0.05	1404	0.02
Police Vehicle	4120	0.42	4671	0.65	6364	0.33	2489	0.34	6552	0.28	0	0.00	24196	0.33
Transport Bus	2026	0.21	4030	0.56	3760	0.20	2099	0.29	4922	0.21	1921	0.30	18758	0.26
School Bus	2514	0.25	2988	0.41	4827	0.25	1703	0.23	3629	0.15	949	0.15	16610	0.23
Motor Home	1099	0.11	1079	0.15	0	0.00	260	0.04	1751	0.07	1176	0.18	5365	0.07
Other Vehicle	4066	0.41	1534	0.21	5247	0.28	707	0.10	9822	0.42	366	0.06	21742	0.30
Missing	0	0.00	51271	7.10	0	0.00	0	0.00	1	0.00	0	0.00	51272	0.70
Unknown	49463	5.01	2352	0.33	129557	6.81	5569	0.76	82918	3.53	25501	3.93	295360	4.02
Total	987398	100.00%	722178	100.00%	1903704	100.00%	732239	100.00%	2346702	100.00%	649072	100.00%	7341293	100.00%

TABLE 21
Precrash Stability by State
1984–1986 CARDfile

Precrash Stability	Indiana		Maryland		Michigan		Pennsylvania		Texas		Washington		Total	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Tracking	0	0.00%	500847	69.35%	0	0.00%	607402	82.95%	255166	10.87%	111105	17.12%	1474520	20.09%
Skidding	0	0.00	35112	4.86	60660	3.19	0	0.00	5948	0.25	49318	7.60	151038	2.06
Spinning	0	0.00	0	0.00	0	0.00	6558	0.90	0	0.00	0	0.00	6558	0.09
Jackknifing	251	0.03	0	0.00	0	0.00	30	0.00	2736	0.12	373	0.06	3390	0.05
Not Applicable	0	0.00	153525	21.26	192882	10.13	116811	15.95	0	0.00	110419	17.01	573637	7.81
Missing	798923	80.91	0	0.00	0	0.00	0	0.00	2082852	88.76	377857	58.21	3259632	44.40
Unknown	188224	19.06	32694	4.53	1650162	86.68	1438	0.20	0	0.00	0	0.00	1872518	25.51
Total	987398	100.00%	722178	100.00%	1903704	100.00%	732239	100.00%	2346702	100.00%	649072	100.00%	7341293	100.00%

TABLE 22
Avoidance Attempt by State
1984–1986 CARDfile

Avoidance Attempt	Indiana		Maryland		Michigan		Pennsylvania		Texas		Washington		Total	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Avoid Object	3334	0.34%	0	0.00%	11914	0.63%	5011	0.68%	4502	0.19%	2383	0.37%	27144	0.37%
Avoid Ped./Cyclist	0	0.00	0	0.00	2258	0.12	377	0.05	405	0.02	608	0.09	3648	0.05
Avoid Vehicle	0	0.00	0	0.00	51797	2.72	1833	0.25	14062	0.60	6322	0.97	74014	1.01
Avoid Unknown Object	0	0.00	0	0.00	0	0.00	1193	0.16	1588	0.07	0	0.00	2781	0.04
No Avoidance Attempt	806314	81.66	0	0.00	1639034	86.10	607011	82.90	243293	10.37	226502	34.90	3522154	47.98
Not Applicable	130552	13.22	0	0.00	192882	10.13	116811	15.95	0	0.00	35400	5.45	475645	6.48
Missing	0	0.00	722178	100.00	0	0.00	0	0.00	2082852	88.76	377857	58.21	3182887	43.36
Unknown	47198	4.78	0	0.00	5819	0.31	3	0.00	0	0.00	0	0.00	53020	0.72
Total	987398	100.00%	722178	100.00%	1903704	100.00%	732239	100.00%	2346702	100.00%	649072	100.00%	7341293	100.00%

TABLE 23
Driver Age by State
1984–1986 CARDfile

Driver Age	Indiana		Maryland		Michigan		Pennsylvania		Texas		Washington		Total	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Known	847725	85.85%	557144	77.15%	1716617	90.17%	672100	91.79%	2136165	91.03%	543491	83.73%	6473242	88.18%
Missing/NDR	135975	13.77	165034	22.85	187087	9.83	40459	5.53	210537	8.97	105579	16.27	844671	11.51
Unknown	3698	0.37	0	0.00	0	0.00	19680	2.69	0	0.00	2	0.00	23380	0.32
Total	987398	100.00%	722178	100.00%	1903704	100.00%	732239	100.00%	2346702	100.00%	649072	100.00%	7341293	100.00%

TABLE 24
Driver Sex by State
1984–1986 CARDfile

Driver Sex	Indiana		Maryland		Michigan		Pennsylvania		Texas		Washington		Total	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Female	315008	31.90%	182372	25.25%	647016	33.99%	216198	29.53%	766906	32.68%	209007	32.20%	2336507	31.83%
Male	559818	56.70	380092	52.63	1256688	66.01	463109	63.25	1405603	59.90	373398	57.53	4438708	60.46
No Driver Record	72859	7.38	0	0.00	0	0.00	40459	5.53	0	0.00	0	0.00	113318	1.54
Missing	0	0.00	159714	22.12	0	0.00	0	0.00	0	0.00	34459	5.31	194173	2.64
Unknown	39713	4.02	0	0.00	0	0.00	12473	1.70	174193	7.42	32208	4.96	258587	3.52
Total	987398	100.00%	722178	100.00%	1903704	100.00%	732239	100.00%	2346702	100.00%	649072	100.00%	7341293	100.00%

TABLE 25
Alcohol/Drug Use by State
1984–1986 CARDfile

Alcohol/Drug Use	Indiana		Maryland		Michigan		Pennsylvania		Texas		Washington		Total	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
No Indication of Use	738942	74.84%	533707	73.90%	1594620	83.76%	628499	85.83%	2236455	95.30%	365283	56.28%	6097506	83.06%
Alcohol Use	42283	4.28	36523	5.06	120150	6.31	63070	8.61	106442	4.54	51778	7.98	420246	5.72
Drug Use	703	0.07	870	0.12	0	0.00	211	0.03	3805	0.16	729	0.11	6318	0.09
No Driver Record	72859	7.38	0	0.00	0	0.00	40459	5.53	0	0.00	0	0.00	113318	1.54
Missing	0	0.00	4	0.00	0	0.00	0	0.00	0	0.00	0	0.00	4	0.00
Unknown	132611	13.43	151074	20.92	188934	9.92	0	0.00	0	0.00	231282	35.63	703901	9.59
Total	987398	100.00%	722178	100.00%	1903704	100.00%	732239	100.00%	2346702	100.00%	649072	100.00%	7341293	100.00%

TABLE 26
Restraint Use by State
1984–1986 CARDfile

Restraint Use	Indiana		Maryland		Michigan		Pennsylvania		Texas		Washington		Total	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Not Equipped	14	0.00%	4475	0.62%	11794	0.62%	33370	4.56%	2420	0.10%	0	0.00%	52073	0.71%
Not in Use	450286	45.60	299586	41.48	635589	33.39	265263	36.23	535873	22.84	218205	33.62	2404802	32.76
In Use	87200	8.83	200603	27.78	943493	49.56	125808	17.18	689033	29.36	150819	23.24	2196956	29.93
Not Applicable	12252	1.24	9837	1.36	21378	1.12	16340	2.23	33856	1.44	11339	1.75	105002	1.43
No Driver Record	72859	7.38	0	0.00	0	0.00	40459	5.53	0	0.00	0	0.00	113318	1.54
Missing	0	0.00	158121	21.90	0	0.00	0	0.00	0	0.00	167942	25.87	326063	4.44
Unknown	364787	36.94	49556	6.86	291450	15.31	250999	34.28	1085520	46.26	100767	15.52	2143079	29.19
Total	987398	100.00%	722178	100.00%	1903704	100.00%	732239	100.00%	2346702	100.00%	649072	100.00%	7341293	100.00%

TABLE 27
Driver Error by State
1984–1986 CARDfile

Driver Error	Indiana		Maryland		Michigan		Pennsylvania		Texas		Washington		Total	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
None Indicated	134230	13.59%	298842	41.38%	903423	47.46%	342357	46.75%	1105490	47.11%	183375	28.25%	2967717	40.42%
Speed Related	29748	3.01	79306	10.98	118663	6.23	34080	4.65	417136	17.78	63375	9.76	742308	10.11
Right of Way Viol.	70434	7.13	45505	6.30	0	0.00	65389	8.93	327426	13.95	63966	9.85	572720	7.80
Disregard Sign/Sig.	20515	2.08	15676	2.17	245843	12.91	53287	7.28	102590	4.37	17889	2.76	455800	6.21
Improper Passing	6984	0.71	4541	0.63	109636	5.76	8290	1.13	27587	1.18	4343	0.67	161381	2.20
Follow too Closely	19781	2.00	7604	1.05	361649	19.00	27012	3.69	93977	4.00	19797	3.05	529820	7.22
Failed to Signal	0	0.00	0	0.00	48355	2.54	399	0.05	6414	0.27	1102	0.17	56270	0.77
Asleep at Wheel	145657	14.75	2116	0.29	12370	0.65	27551	3.76	9817	0.42	57498	8.86	255009	3.47
Other Error	51630	5.23	268558	37.19	58136	3.05	133415	18.22	256262	10.92	54799	8.44	822800	11.21
No Driver Record	72859	7.38	0	0.00	0	0.00	40459	5.53	0	0.00	0	0.00	113318	1.54
Missing	382096	38.70	0	0.00	0	0.00	0	0.00	0	0.00	182928	28.18	565024	7.70
Unknown	53464	5.41	30	0.00	45629	2.40	0	0.00	3	0.00	0	0.00	99126	1.35
Total	987398	100.00%	722178	100.00%	1903704	100.00%	732239	100.00%	2346702	100.00%	649072	100.00%	7341293	100.00%

TABLE 28
CARDfile Coding of Driver Error Variable

Driver Errors Where Code Values Differ by State[a]

State Driver Error	Attribute Value Coded in CARDfile[b]					
	IN	MD	MI	PA	TX	WA
Use of Alcohol	3	11	—	3	—	11
Use of Illegal Drugs	3	11	—	3	—	11
Driver Illness	3	11	10	3	3	—
Physical/Mental Disability	—	11	—	3	—	—
Driver Inattention	10	11	10	—	—	10
Left of Center	11	11	7	11	11	11
Wrong Way on One Way	11	11	11	6	11	—
Improper Lane Usage	11	11	7	11	—	—
Improper Turning	11	11	9	11	11	11
Improper U-Turn	—	11	—	6	—	11
Failure to Signal	11	11	9	9	9	9
Improper Signal	—	11	9	9	9	11
Speed Too Slow	—	4	4	11	—	—
Right of Way Violation	5	5	6	5	5	5
Improper or Unsafe Backing	11	11	11	5	—	—
School Bus Law	—	11	—	5	—	—
Violation Driver License Restrictions	3	—	—	11	—	—
Headlight Violation	3	—	2	11	—	11
Defective Vehicle Equipment	3	11	2	3	—	11
Inadequate Safety Devices	—	11	—	3	—	—
Weather Elements	3	11	—	3	—	—
Road Defects	3	11	—	3	—	—
Foreign Object/Substance on Roadway	3	11	—	3	—	—
Animal Present on Roadway	3	11	—	3	—	—
Pedestrian at Fault	3	11	—	3	—	—
Pedalcyclist at Fault	—	11	—	3	—	—

Inappropriate or Questionable Code Values Assigned

State Driver Error	Attribute Value Coded in CARDfile					
	IN	MD	MI	PA	TX	WA
Leaving Vehicle Improperly Unattended		11				
Blinded by Approaching Vehicle		11				
Right Turn on Red After Stop		11				
Fire		11				
Struck by Object from Moving Vehicle		11				
Other Cause (No Violation)		11				
Cause Unknown		11				
Other Error			2			
Making Improper Entrance to Roadway				6		
Making Improper Exit from Roadway (From or onto driveway or ramp)				6		

[a] Key to code values for the Driver Error variable:

1 = Missing
2 = Unknown
3 = None indicated
4 = Speed related
5 = Right of way violation
6 = Disregarded traffic sign or signal
7 = Improper passing
8 = Following too closely
9 = Failed to signal
10 = Asleep at wheel
11 = Other error
98 = No driver record

[b] "—" indicates the error is not specifically mentioned by that state.

TABLE 29
General Vehicle Classes by State
1984–1986 CARDfile

State	Passenger Cars		Light Trucks		Heavy Trucks		Total	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Indiana	726,502	73.6%	131,178	13.3%	52,852	5.4%	910,532	92.2%
Maryland	522,402	72.3	92,267	12.8	28,217	3.9	642,886	89.0
Michigan	1,383,932	72.7	295,896	15.5	51,247	2.7	1,731,075	90.9
Pennsylvania	568,056	77.6	99,750	13.6	34,582	4.7	702,388	95.9
Texas	1,563,358	66.6	558,414	23.8	80,226	3.4	2,201,998	93.8
Washington	447,287	68.9	136,188	21.0	23,824	3.7	607,299	93.6

TABLE 30
Proportion of Vehicles with Known Makes and Models by State
Based on 5% Sample of 1984–1986 CARDfile

State	Cases with Known Make		Cases with Known Make/Model	
	Estimated Number	Estimated Percent	Estimated Number	Estimated Percent
Indiana	897,545	90.9 ± .3	267,585	27.1 ± .4
Maryland	642,016	88.9 ± .3	417,419	57.8 ± .5
Michigan	1,709,526	89.8 ± .2	1,463,948	76.9 ± .3
Pennsylvania	564,556	77.1 ± .4	449,595	61.4 ± .5
Texas	2,264,567	96.5 ± .1	1,485,462	63.3 ± .3
Washington	608,830	93.8 ± .3	412,810	63.6 ± .5

TABLE 31
Make/Model Classifications by State for All Cases with Unknown Model
Based on 5% Sample of 1984–1986 CARDfile

State	Percentage Make Known/ Vehicle Type Known	Percentage Make Known/ Vehicle Type Unknown	Percentage Make Unknown/ Vehicle Type Known	Percentage Make Unknown/ Vehicle Type Unknown	Total Percentage
Indiana	86.2%	1.2%	0.9%	11.6%	100.0%
Maryland	73.6	0.2	8.6	17.5	100.0
Michigan	53.2	2.4	25.1	19.2	100.0
Pennsylvania	40.8	0.0	55.4	3.7	100.0
Texas	89.9	0.7	0.4	9.0	100.0
Washington	82.1	0.9	3.6	13.4	100.0

TABLE 32
Proportion of Vehicles with Known Makes and Models by Vehicle Type Based on 5% Sample of 1984–1986 CARDfile

Vehicle Type	Cases with Known Make		Cases with Known Make/Model	
	Estimated Number	Estimated Percent	Estimated Number	Estimated Percent
Passenger Car	5,081,249	97.5 ± 0.1	3,955,557	75.9 ± 0.2
Light Truck	1,186,265	90.3 ± 0.2	451,910	34.4 ± 0.4
Heavy Truck	188,851	69.7 ± 0.8	30,075	11.1 ± 0.5
Other	125,044	63.0 ± 1.0	24,612	12.4 ± 0.7

TABLE 33
Proportion of Vehicles with Known Makes and Models by State for Two Vehicle Types
Based on 5% Sample of 1984–1986 CARDfile

State	Passenger Cars				Light Trucks			
	Cases with Known Make		Cases with Known Make/Model		Cases with Known Make		Cases with Known Make/Model	
	Estimated Number	Estimated Percent	Estimated Number	Estimated Percent	Estimated Number	Estimated Percent	Estimated Number	Estimated Percent
Indiana	706,160	97.2 ± 0.2	260,814	35.9 ± 0.5	125,537	95.7 ± 0.5	1,968	1.5 ± 0.3
Maryland	502,028	96.1 ± 0.2	360,980	69.1 ± 0.6	90,883	98.5 ± 0.4	45,026	48.8 ± 1.5
Michigan	1,321,655	95.5 ± 0.2	1,145,896	82.8 ± 0.3	286,723	96.9 ± 0.3	256,542	86.7 ± 0.6
Pennsylvania	562,943	99.1 ± 0.1	447,060	78.7 ± 0.5	698	0.7 ± 0.2	698	0.7 ± 0.2
Texas	1,546,161	98.9 ± 0.1	1,346,051	86.1 ± 0.2	550,038	98.5 ± 0.1	127,318	22.8 ± 0.5
Washington	444,603	99.4 ± 0.1	391,823	87.6 ± 0.4	133,464	98.0 ± 0.3	20,973	15.4 ± 0.9

TABLE 34
Number and Proportion of Foreign Cars and Light Trucks by State Based on 5% Sample of 1984–1986 CARDfile

State	Foreign Passenger Cars		Foreign Light Trucks		Foreign Passenger Cars and Light Trucks	
	Estimated Number	Estimated Percent	Estimated Number	Estimated Percent	Estimated Number	Estimated Percent
Indiana	69,204	9.8%	8,285	6.6%	77,489	9.3%
Maryland	100,510	20.0	10,997	12.1	111,507	18.8
Michigan	64,761	4.9	6,595	2.3	71,356	4.4
Pennsylvania	100,767	17.9	—	—	101,126	17.9
Texas	282,947	18.3	60,992	11.1	343,940	16.4
Washington	146,719	33.0	29,095	21.8	175,814	30.5

NOTE: No figures are listed for foreign light trucks in Pennsylvania since vehicle make is not available for 99.3% of the cases.

TABLE 35
Percent of Registered Foreign Cars versus CARDfile Foreign Cars by State

State	Percent Foreign Registered[a]	Percent Foreign CARDfile	Percent Unknown Make CARDfile
Indiana	11.1%	9.8%	2.8%
Maryland	25.2	20.0	3.9
Michigan	9.4	4.9	4.5
Pennsylvania	17.0	17.9	0.9
Texas	16.8	18.3	1.1
Washington	33.0	33.0	0.6

[a] S. Salvatore, P. Mengert, and R. Walter, *CARDfile Data Base Representativeness, Phase I: General Characteristics including Populations, Vehicles, Roads, and Fatal Accidents*. Cambridge, Mass.: Transportation Systems Center, Project Memorandum No. DOT-TSC-HS802-PM-88–16, August 1988.

TABLE 36
Proportion of Cars and Light Trucks with Known Models According to Origin of Manufacture by State Based on 5% Sample of 1984–1986 CARDfile

State	Passenger Cars of Known Model				Light Trucks of Known Model			
	Domestic		Foreign		Domestic		Foreign	
	Estimated Number	Estimated Percent	Estimated Number	Estimated Percent	Estimated Number	Estimated Percent	Estimated Number	Estimated Percent
Indiana	259,241	40.7%	1,315	1.9%	1,876	1.6%	17	0.2%
Maryland	291,077	72.4	69,955	69.6	42,020	52.6	2,969	27.0
Michigan	1,083,443	86.2	62,494	96.5	249,875	89.2	6,555	99.4
Pennsylvania	378,523	81.9	69,025	68.5	—	—	—	—
Texas	1,136,892	90.0	209,947	74.2	121,145	24.8	5,855	9.6
Washington	274,947	92.3	116,935	79.7	19,100	18.3	1,862	6.4
CARDfile	3,424,123	79.3	529,672	69.3	434,015	40.6	17,258	15.1

NOTE: No figures are listed for light trucks in Pennsylvania since vehicle make is not available for 99.3% of the cases.

TABLE 37
VIN Recording by Vehicle Type
Based on 5% Sample of 1984–1986 CARDfile

Vehicle Type	VIN Present		VIN Absent		Total	
	Estimated Number	Estimated Percent	Estimated Number	Estimated Percent	Number	Percent
All Classes	2,474,016	33.7 ± .2	4,867,277	66.3 ± .2	7,341,293	100.0
Passenger Cars	1,876,153	36.0 ± .2	3,335,384	64.0 ± .2	5,211,537	100.0
Light Trucks	383,598	29.2 ± .4	930,095	70.8 ± .4	1,313,693	100.0

TABLE 38
Proportion of Vehicles with Known Models According to Presence of VIN for Three Vehicle Types Based on 5% Sample of 1984–1986 CARDfile

Vehicle Type	VIN Present: Cases with Known Model		VIN Absent: Cases with Known Model	
	Estimated Number	Estimated Percent	Estimated Number	Estimated Percent
All Classes	1,880,252	76.0 ± .2	2,613,728	53.7 ± .2
Passenger Cars	1,506,551	80.3 ± .3	2,448,172	73.4 ± .2
Light Trucks	301,892	78.7 ± .6	150,675	16.2 ± .3

TABLE 39
Comparison between CARDfile Data and Michigan State File Data for the Vehicle Type Variable Michigan 1984 Cases

Value	Attribute	Michigan File		CARDfile		Change from Michigan File to CARDfile	
		Number	Percent	Number	Percent	Number	Percent
2	Unknown	39,391	6.98%	39,385	6.98%	–6	–0.02%
3	Motorcycle	7,091	1.26	7,091	1.26	0	0.00
4	Car	423,131	74.97	412,631	73.10	–10,500	–2.48
5	Car with Trailer	1,387	0.25	1,188	0.21	–199	–14.35
6	Light Truck/Van	68,778	12.19	80,957	14.34	12,179	17.71
7	Light Truck with Trailer	1,384	0.25	1,708	0.30	324	23.41
8	Straight Truck	7,112	1.26	5,412	0.96	–1,700	–23.90
9	Straight Truck with Trailer	482	0.09	356	0.06	–126	–26.14
10	Tractor-Semitrailer	6,434	1.14	6,434	1.14	0	0.00
11	Tractor-Double Trailer	521	0.09	521	0.09	0	0.00
12	Bobtail	2,096	0.37	2,096	0.37	0	0.00
14	Fire Truck	226	0.04	226	0.04	0	0.00
15	Ambulance	245	0.04	245	0.04	0	0.00
17	Police Vehicle	1,934	0.34	1,934	0.34	0	0.00
18	Transport Bus	1,242	0.22	1,242	0.22	0	0.00
19	School Bus	1,484	0.26	1,484	0.26	0	0.00
21	Other Vehicle	1,450	0.26	1,548	0.27	98	6.76
	TOTAL	564,388	100.00%	564,458	100.00%	70	0.01%

TABLE 40
Coding of Vehicle Type: Vehicle Type Variable vs. Make/Model Variable Based on 5% Sample of 1984–1986 CARDfile

MAKE/ MODEL	VEHICLE TYPE								TOTAL
	Missing/ Unknown	Motorcycle	Passenger Car	Light Truck/Van	Med./Heavy Truck	Bus	Motor Home	Other	
Missing/ Unknown	13,490	302	2,322	770	1,123	231	80	179	18,497
Motorcycle	4	4,964	14	1	0	0	1	18	5,002
Passenger Car	1,204	14	258,327	112	27	4	3	992	260,683
Light Truck/Van	347	2	1	64,431	68	24	7	62	64,942
Med./Heavy Truck	141	0	3	45	12,095	27	0	97	12,408
Bus	3	0	3	1	0	121	0	0	128
Motor Home	0	0	0	0	0	0	0	0	0
Other	2,151	7	101	96	26	1,390	184	1,315	5,270
TOTAL	17,340	5,289	260,771	65,456	13,339	1,797	275	2,663	366,930

TABLE 41
Accident Type Variable vs. Primary Impact Variable
Based on 5% Sample of 1984–1986 CARDfile

PRIMARY IMPACT	ACCIDENT TYPE									TOTAL
	Rollover	Fixed Object	Ped./Animal/ Cyclist	Parked Vehicle	Noncollision	Single Veh./ Specifics Unknown	Single Vehicle/ Other	Two Vehicles	More Than Two Vehicles	
Rollover	5,923	0	5	3	5	0	0	167	21	6,124
Fixed Object	1,472	32,189	7	82	4	60	52	321	167	34,354
Pedestrian	5	0	3,410	8	0	47	0	46	7	3,523
Pedalcyclist	2	0	2,100	0	0	20	0	27	2	2,151
Other (Animal, etc.)	4	161	6,335	3	93	490	239	239	26	7,590
Motor Vehicle Not in Transport	11	0	5	12,891	0	1,403	0	1,178	1,315	16,803
Noncollision	30	0	1	5	1,158	0	13	104	13	1,324
Collision with Train	1	0	0	0	0	0	211	13	4	229
Motor Vehicle in Transport	210	0	204	1,334	0	207	4	122,508	12,210	136,677
Missing	0	0	0	1	0	15	0	98	7	121
Unknown	10	1	191	187	0	222	0	554	38	1,203
TOTAL	7,668	32,351	12,258	14,514	1,260	2,464	519	125,255	13,810	210,099

TABLE 42
Road Surface Variable vs. Weather Conditions Variable
Based on 5% Sample of 1984–1986 CARDfile

WEATHER CONDITIONS	ROAD SURFACE					TOTAL
	Dry	Wet	Snow/Ice	Other	Unknown	
Clear/Cloudy	144,200	14,009	8,641	240	440	167,530
Rain	356	27,684	852	55	26	28,973
Snow/Ice/Sleet	111	1,043	8,510	87	21	9,772
Other	740	1,299	365	9	14	2,427
Unknown	133	49	59	4	1,152	1,397
TOTAL	145,540	44,084	18,427	395	1,653	210,099

TABLE 43
Accident Severity Distributions According to Whether Land Use is Coded
1984–1986 CARDfile

Accident Severity	Land Use Known	Land Use Missing	All Cases	Ratio of Known Cases to All Cases
Property Damage	69.58%	58.31%	64.79%	1.07
Possible Injury	14.11	20.96	17.02	0.83
Nonincapacitating Injury	11.15	15.41	12.96	0.86
Incapacitating Injury	4.71	4.39	4.57	1.03
Fatal Injury	0.45	0.83	0.61	0.74
Unknown	0.00	0.10	0.04	0.00
TOTAL	100.00%	100.00%	100.00%	1.00

TABLE 44
Accident Type Distributions According to Whether Land Use is Coded 1984–1986 CARDfile

Accident Type	Land Use Known	Land Use Missing	All Cases	Ratio of Known Cases to All Cases
Single Vehicle	35.63%	31.59%	33.91%	1.05
Rear End	17.61	19.85	18.56	0.95
Sideswipe	4.51	1.84	3.38	1.33
Head-On	3.36	3.54	3.43	0.98
Turn Across Path	6.18	7.77	6.86	0.90
Intersecting Paths	16.60	20.52	18.27	0.91
Backing	4.02	1.65	3.01	1.34
Parking	1.11	0.70	0.93	1.19
Other Two Vehicles	5.37	4.87	5.15	1.04
More Than Two Vehicles	5.61	7.68	6.49	0.86
TOTAL	100.00%	100.00%	100.00%	1.00

TABLE 45
Land Use vs. Accident Severity
Based on 5% Sample of 1984–1986 CARDfile

ACCIDENT SEVERITY	LAND USE					
	Rural		Urban		Missing	
	Estimated Number	Estimated Percent	Estimated Number	Estimated Percent	Estimated Number	Estimated Percent
Property Damage	527,228	67.8 ± 0.5	1,150,915	70.2 ± 0.3	1,041,437	58.3 ± 0.3
Possible Injury	95,648	12.3 ± 0.3	250,840	15.3 ± 0.3	375,132	21.0 ± 0.3
Nonincapacitating Injury	101,869	13.1 ± 0.3	167,227	10.2 ± 0.2	275,097	15.4 ± 0.2
Incapacitating Injury	47,435	6.1 ± 0.2	65,579	4.0 ± 0.1	76,813	4.3 ± 0.1
Fatal Injury	6,221	0.8 ± 0.1	3,279	0.2 ± 0.0	14,291	0.8 ± 0.1
Unknown	0	0.0	0	0.0	1,786	0.1 ± 0.0
TOTAL	778,401	100.0%	1,637,841	100.0%	1,784,555	100.0%

TABLE 46
Land Use vs. Accident Type
Based on 5% Sample of 1984–1986 CARDfile

ACCIDENT TYPE	LAND USE					
	Rural		Urban		Missing	
	Estimated Number	Estimated Percent	Estimated Number	Estimated Percent	Estimated Number	Estimated Percent
Single Vehicle	410,585	52.8 ± 0.5	447,578	27.3 ± 0.3	560,911	31.4 ± 0.3
Rear End	108,867	14.0 ± 0.4	319,699	19.5 ± 0.3	355,482	19.9 ± 0.3
Sideswipe	25,662	3.3 ± 0.2	83,613	5.1 ± 0.2	32,154	1.8 ± 0.1
Head-On	38,104	4.9 ± 0.2	42,626	2.6 ± 0.1	62,522	3.5 ± 0.1
Turn Across Path	31,883	4.1 ± 0.2	121,322	7.4 ± 0.2	137,548	7.7 ± 0.2
Intersecting Paths	83,983	10.8 ± 0.3	314,780	19.2 ± 0.3	364,414	20.4 ± 0.3
Backing	19,441	2.5 ± 0.2	77,056	4.7 ± 0.1	30,368	1.7 ± 0.1
Parking	2,333	0.3 ± 0.1	24,592	1.5 ± 0.1	12,504	0.7 ± 0.1
Other Two Vehicles	29,550	3.8 ± 0.2	100,008	6.1 ± 0.2	87,531	4.9 ± 0.1
More Than Two Vehicles	27,217	3.5 ± 0.2	108,206	6.6 ± 0.2	141,121	7.9 ± 0.2
TOTAL	777,623	100.0%	1,639,480	100.0%	1,784,555	100.0%

TABLE 47
Roadway Profile vs. Accident Severity
Based on 5% Sample of 1984–1986 CARDfile

ACCIDENT SEVERITY	ROADWAY PROFILE					
	Level		Grade		Missing/Unknown	
	Estimated Number	Estimated Percent	Estimated Number	Estimated Percent	Estimated Number	Estimated Percent
Property Damage	1,646,245	63.4 ± 0.3	251,425	54.3 ± 0.7	820,115	71.7 ± 0.4
Possible Injury	449,212	17.3 ± 0.2	104,182	22.5 ± 0.5	170,428	14.9 ± 0.3
Nonincapacitating Injury	371,314	14.3 ± 0.2	76,400	16.5 ± 0.5	98,368	8.6 ± 0.2
Incapacitating Injury	111,654	4.3 ± 0.1	26,856	5.8 ± 0.3	51,472	4.5 ± 0.2
Fatal Injury	15,580	0.6 ± 0.0	3,704	0.8 ± 0.1	4,575	0.4 ± 0.1
Unknown	1,101	0.0 ± 0.0	463	0.1 ± 0.0	60	0.0 ± 0.0
TOTAL	2,595,106	100.0%	463,029	100.0%	1,145,018	100.0%

TABLE 48
Roadway Profile vs. Accident Type
Based on 5% Sample of 1984–1986 CARDfile

ACCIDENT TYPE	ROADWAY PROFILE					
	Level		Grade		Missing/Unknown	
	Estimated Number	Estimated Percent	Estimated Number	Estimated Percent	Estimated Number	Estimated Percent
Single Vehicle	802,350	30.9 ± 0.3	212,993	46.0 ± 0.7	406,054	35.5 ± 0.4
Rear End	482,968	18.6 ± 0.2	62,509	13.5 ± 0.4	237,913	20.8 ± 0.3
Sideswipe	85,688	3.3 ± 0.1	16,206	3.5 ± 0.2	38,890	3.4 ± 0.2
Head-On	77,898	3.0 ± 0.1	30,097	6.5 ± 0.3	35,458	3.1 ± 0.1
Turn Across Path	194,745	7.5 ± 0.1	25,004	5.4 ± 0.3	70,916	6.2 ± 0.2
Intersecting Paths	516,724	19.9 ± 0.2	48,155	10.4 ± 0.4	196,736	17.2 ± 0.3
Backing	80,495	3.1 ± 0.1	9,724	2.1 ± 0.2	36,602	3.2 ± 0.1
Parking	25,966	1.0 ± 0.1	3,241	0.7 ± 0.1	9,151	0.8 ± 0.1
Other Two Vehicles	135,023	5.2 ± 0.1	22,225	4.8 ± 0.3	60,622	5.3 ± 0.2
More Than Two Vehicles	192,148	7.4 ± 0.1	32,412	7.0 ± 0.3	51,472	4.5 ± 0.2
TOTAL	2,594,004	100.0%	462,566	100.0%	1,143,814	100.0%

TABLE 49
Land Use vs. Roadway Profile
Based on 5% Sample of 1984–1986 CARDfile

ROADWAY PROFILE	LAND USE						TOTAL	
	Rural		Urban		Missing			
	Estimated Number	Estimated Percent	Estimated Number	Estimated Percent	Estimated Number	Estimated Percent	Estimated Number	Estimated Percent
Level	243,800	5.8 ± 0.1	714,585	17.0 ± 0.2	1,635,140	38.9 ± 0.2	2,593,525	61.7 ± 0.2
Grade	105,086	2.5 ± 0.1	205,969	4.9 ± 0.1	151,324	3.6 ± 0.1	462,379	11.0 ± 0.1
Missing/ Unknown	432,955	10.3 ± 0.1	710,382	16.9 ± 0.2	1,921	0.0 ± 0.0	1,147,540	27.3 ± 0.2
TOTAL	781,841	18.6 ± 0.2	1,635,140	38.9 ± 0.2	1,786,464	42.5 ± 0.2	4,203,444	100.0%

TABLE 50
Alcohol-Impaired Drivers Involved in Fatal Accidents

PROPORTION OF ALCOHOL-IMPAIRED DRIVERS BY ACCIDENT TYPE		
Accident Type	FARS 1986	CARDfile 1986
Single Vehicle	50.8%	53.8%
Multivehicle	15.7	18.1
PROPORTION OF ALCOHOL-IMPAIRED DRIVERS BY SEX		
Driver Sex	FARS 1986	CARDfile 1986
Male	28.5%	32.3%
Female	14.9	14.7
PROPORTION OF ALCOHOL-IMPAIRED DRIVERS BY AGE		
Driver Age	FARS 1986	CARDfile 1986
15–20	23.3%	30.9%
21–44	31.5	34.1
45 plus	12.9	13.8
PROPORTION OF ALCOHOL-IMPAIRED DRIVERS BY VEHICLE TYPE		
Vehicle Type	FARS 1986	CARDfile 1986
Motorcycles	41.0%	37.1%
Cars	27.5	31.2
Light Trucks/Vans	30.9	34.5
Medium/Heavy Trucks	3.2	5.0