**U.S. Department
of Transportation**

**National Highway
Traffic Safety
Administration**

**NHTSA**

**People Saving People**

http://www.nhtsa.dot.gov

**DOT HS  808 487**

**October 1996**

**Final Report**

# Police Accident Report (PAR)
# Quality Assessment Project

| 1. Report No.<br>DOT HS 808 487 | 2. Government Accession No. | 3. Recipient's Catalog No. |
|---|---|---|
| 4. Title and Subtitle<br>Police Accident Report (PAR) Quality Assessment Project | | 5. Report Date<br>October 1996 |
| | | 6. Performing Organization Code |
| 7. Authors<br>Leonard Greenberg | | 8. Performing Organization Report No. |
| 9. Performing Organization Name and Address<br>Applied System Technologies, Inc.<br>6110 Executive Boulevard, Suite 900<br>Rockville, Maryland 20852 | | 10. Work Unit No. (TRAIS) |
| | | 11. Contract or Grant No.<br>DTNH22-94-D-07087 |
| 12. Sponsoring Agency Name and Address<br>National Center for Statistics and Analysis (NRD-31)<br>National Highway Traffic Safety Administration<br>400 Seventh Street, S.W.<br>Washington, D.C.   20590 | | 13. Type of Report and Period Covered<br>Final Contractor Report<br>September 1995 - March 1996 |
| | | 14. Sponsoring Agency Code |

**15. Supplementary Notes**

Preliminary work for this report was completed under Contract No. DTNH22-92-D-07144

**16. Abstract**

In this study the author assesses the accuracy and completeness with which Police Accident Reports (PARs) are completed by state and local police. Focusing on data elements commonly used in studies carried out by the National Highway Traffic Safety Administration (NHTSA), the author compares data from PARs to data collected in independent reinvestigations of the same crashes by NHTSA's Crashworthiness Data System (CDS) researchers. A total of 939 crashes in 1991 from four CDS study sites were investigated. In all, 24 data elements were studied. For four data elements, very low discrepancy rates were found. For sixteen data elements, discrepancy rates varied by jurisdiction -- high in some, low in others. For four data elements, uniformly high discrepancy rates were found. These four data elements were manner of collision, vehicle identification number, occupant protection system use, and deployment of air bags (however the latter was treated as a separate data element in only one jurisdiction). The study results will be of interest to analysts relying on PAR-reported data. In addition, the author gives various recommendations for the improvement of PAR data quality.

| 17. Key Words<br>Data quality, police accident report, highway safety data, Crashworthiness Data System | 18. Distribution Statement<br>This document is available to the public through the National Technical Information Service, 5285 Port Royal Road, Springfield, VA 22161 | | |
|---|---|---|---|
| 19. Security Classif. (of the report)<br>Unclassified | 20. Security Classif. (of this page)<br>Unclassified | 21. No. Of Pages<br>86 | 22. Price |

Form DOT F 1700.7 (8-72) (facsimile)        Reproduction of completed page authorized

# TABLE OF CONTENTS

EXECUTIVE SUMMARY

The goal of this project was to assess the accuracy and completeness with which certain important data elements are reported in the motor vehicle accident reports routinely filed by state and local police. The data elements studied were selected from the variables most commonly treated in safety analyses conducted at the national level by the National Highway Traffic Safety Administration (NHTSA). A major purpose of the investigation was to help analysts assess the extent to which these data elements can be relied upon, as they contribute importantly to the databases NHTSA uses to support its national studies.

A key NHTSA database is the Crashworthiness Data System (CDS), a component of the National Accident Sampling System (NASS). CDS consists of accident-specific information, generated by teams of researchers located in a sample of geographically disparate areas across the country, on accidents that meet certain criteria: (a) the accident was reported to the state by the police, (b) it resulted in property damage and/or personal injury, and (c) it involved at least one passenger car, light truck, or van that was in-transport at the time of the accident and was subsequently towed. In a typical year, the CDS database, which is carefully quality controlled, consists of some five thousand accident reports, filed independently of and subsequent to those filed by the police.

The concept underlying this investigation was to compare information stored in the CDS database with comparable information contained in the police accident reports (PARs) filed with respect to the same set of accidents. The study sample consisted of all NASS-reported accidents that took place in 1991 in four of the jurisdictions in which NASS research teams were located. Covered data elements for which the PAR- and CDS-reported values were mutually inconsistent were counted as discrepancies; data elements for which CDS had a specific value but whose PAR-reported value was either missing, ambiguous, or "unknown" also contributed to the discrepancy rate.

Major findings of the study were as follows:

a.  Some variables displayed generally low discrepancy rates -- no more than one or two percent -- in all of the jurisdictions studied. Variables for which uniformly low discrepancy rates were noted were as follows:

| Variable | Range of discrepancy rates |
|---|---|
| Month of accident | No discrepancies at all |
| No. of motor vehicles involved | "       "       "       " |
| Vehicle make | 0.0 - 1.4% |
| Occupant seating position | 1.3 - 2.2% |

i

b. Some variables displayed extremely high discrepancy rates -- ten percent or greater -- in all of the jurisdictions studied. Variables for which uniformly high discrepancy rates were noted were as follows:

| Variable | Range of discrepancy rates |
| --- | --- |
| Manner of collision | 12.1 - 27.5% |
| Vehicle identification number | 17.8 - 30.3% |
| Occupant protection system use | 11.6 - 37.4% |
| Deployment of air bags | 19.1%* |

   * Only one jurisdiction treated Air Bag Deployment as a separate data element.

c. Most variables displayed discrepancy rates that varied by jurisdiction -- high in some, low in others. The variables in question were as follows:

| Variable | Range of discrepancy rates |
| --- | --- |
| First harmful event | 1.4 - 5.7% |
| Time of accident | 0.0 - 6.2% |
| Day of week | 0.0 - 3.7% |
| Number of occupant fatalities | 0.8 - 3.0% |
| Speed limit | 1.5 - 8.6% |
| Number of occupants | 2.6 - 6.9% |
| Model | 1.0 - 3.9% |
| Model year | 1.0 - 6.7% |
| Body type | 4.3 - 28.7% |
| Was vehicle towed due to damage? | 1.8 - 4.6% |
| Driver alcohol/drug presence | 3.8 - 15.5% |
| Occupant age | 4.0 - 9.1% |
| Occupant sex | 1.3 - 5.1% |
| Occupant injury classification | 0.0 - 13.1% |
| Occupant transportation to medical facility | 0.0 - 17.8% |
| Was occupant ejected? | 0.0 - 5.0% |

Some of these variations are relatively minor but others are not. Five variables in particular -- Speed Limit, Body Type, Driver Alcohol/Drug Presence, Occupant Injury Classification, and Occupant Transportation to Medical Facility -- showed variations that were of major magnitude. These wide variations were found in each case to result from basic differences among jurisdictions in accident report design and/or reporting practice.

For those variables characterized by high discrepancy rates in certain (or in all) jurisdictions, recommendations were offered -- to the extent relevant -- that might hopefully be expected to improve future data quality.

A further major recommendation of the study was as follows:

> Notwithstanding the desire to decentralize certain data systems to the state and local level, it nonetheless makes sense that where the data in question serve -- as in this case -- a national purpose as well, **steps should be taken by state and local reporting authorities to assure the inclusion of certain minimum data elements, uniformly and consistently reported.** The initiative to assure that this is done should be undertaken by the states with Federal encouragement and support.

# ACKNOWLEDGMENTS

# 1. OVERVIEW

## 1.1 Purpose and Scope

Every police jurisdiction in the United States is responsible for filing, under certain conditions, motor vehicle accident reports. The forms, instructions, and procedures that are used to generate these reports vary, however, from state to state and sometimes within state. This variability, coupled with the inevitability of human error, opens the door to the possibility that the information provided by these reports -- commonly known as police accident reports or PARs -- may not be uniformly accurate or complete.

The issue of PAR accuracy and completeness has taken on added relevance as questions have been raised concerning such matters as the efficacy of occupant restraints, both active and passive, and the presence of alcohol or drugs in motor vehicle accidents. PARs are a useful source of information on these and other topics as well as on basic accident rates, causes, and concomitants. In addition to being used to generate state and local statistics, PARs are used to support nationwide research and policy analyses conducted by the National Highway Traffic Safety Administration (NHTSA) and its National Center for Statistics and Analysis (NCSA). It is important, therefore, that the extent to which PAR-reported information is inaccurate or incomplete be recognized and accounted for in any analyses conducted. The present study -- termed the Police Accident Report Quality Assessment Project -- is a step in that direction; its purpose, in the words of the Statement of Work, is to

> "conduct an analysis of the accuracy and completeness with which police officers complete Police Accident Reports (PARs)."

The sample for the study was restricted to accidents that

(a)    took place in 1991, and

(b)    are reported as well in the computerized database of the NHTSA Crashworthiness Data System (CDS).

The latter requirement was designed to provide an independent data source against which the accident data reported by the police can be compared. CDS is a respected national database consisting of accident reports produced by teams of NHTSA-supported accident researchers located throughout the country. Each team is assigned a designated geographic area, known as a primary sampling unit or PSU; in 1991, the year on which this study is based, there were 24 such areas. The research team in each area is responsible for investigating a sample of the police-reported light motor vehicle accidents -- accidents involving passenger cars, light trucks, or vans -- that took place within its jurisdiction, making use of not

1

only the police report but other sources of information as well: personal interviews, inspection of the scene, inspection of the vehicle(s), medical reports, etc. Accidents are eligible for inclusion in the CDS database (see reference cited on page A-1 of Appendix A) if at least one vehicle of defined interest -- i.e., either a passenger car, light truck, or van -- was (a) in-transport at the time of the accident and (b) towed due to damage. In 1991, the database thus produced consisted of a total of 4,748 accidents.

Of the 24 PSUs in which CDS research teams were located in 1991, four were selected for inclusion in the study. Selection was based on a constrained randomization process designed to assure representation from each of the four major geographic regions used by the Census Bureau (Northeast, North Central, South, and West) as well as from the three defined *population density strata*: urban, suburban, and rural. The four areas selected are identified below:

PSU 8     Large <u>suburban</u> county in the <u>Northeast</u>

PSU 13    Predominantly <u>rural</u> county in the <u>North Central</u> region

PSU 45    Large <u>suburban</u> county in the <u>South</u>

PSU 79    Inner city <u>urban</u> area in the <u>West</u>

Of the accidents reported by the police in these jurisdictions in 1991, some 944 resulted in ultimate preparation of a CDS report, becoming part of the CDS database as well. These 944 accidents define the study sample for this project.

It is important to note that every jurisdiction has its own unique PAR form and associated reporting procedures; some have more than one. In PSUs 8, 13, and 45, only one PAR version was used in 1991. In PSU 79, two versions were used: one by the local police, the other by the state police. This study thus addresses five PSU/PAR versions in all. The versions used in PSU 79 were labeled, for the purpose of the study, "79A" and "79B", the former pertaining to the local police, the latter to the state police.

## 1.2   Other Relevant Background

The Statement of Work for this study identified some two dozen variables of interest, ranging from accident-specific variables such as Date and Time of Accident to occupant-specific variables such as Age, Sex, and Occupant Protection System Use. Vehicle-specific variables such as Make, Model, and Driver Alcohol/Drug Presence were also to be studied.

The first step in the study was conducted under previous contract DTNH22-92-D-07144. The purpose of that effort was to analyze the report forms, reporting instructions, and coding systems employed in each of the study PSUs in 1991, in order to:

(a)   determine the extent to which each of the variables listed in the Statement of Work was addressed both in the report form and in the accompanying instructions,

(b)   determine the extent to which each of the variables has a counterpart in the CDS database to which it can legitimately be compared, and

(c)   develop detailed algorithms that could be given to a programmer to develop comparison statistics showing the level of agreement and/or disagreement between the PAR and CDS data.

The results of this effort are documented in the ASTI publication, "Comparison of Coding Systems for All Target Variables", June 1994. The conclusion reached was that despite major differences in the way certain variables are addressed in certain jurisdictions, and the fact that in every jurisdiction there are some variables that are not covered at all, there was enough coverage in general, and enough comparability between the PAR and CDS variables in particular, to justify taking the next step: creating a computerized PAR file that would permit the PAR and CDS databases to be linked and compared.

The creation of such a file was begun in Task 1 of the present contract (DTNH22-94-D-07087). Following the development of a suitable data entry program entailing anywhere from 16 to 23 variables per PSU/PAR version, data entry was conducted in small increments. First, an initial break-in phase was conducted in which two data entry operators were trained and tested on a small sample of accidents. This was then followed by two separate timing phases in which the average level of effort, in minutes, required to enter the data for a single accident was measured.

Task 1 ended with a production phase in which enough additional reports were entered to bring the percentage entered to roughly 25%. At the end of each phase, the files created by the two operators were compared, using a computerized data comparison program, and a special listing prepared in which all pairs of entries that differed were identified. All such differences were then "adjudicated" by referring back to the original report to determine which of the operators (if either) was correct. Error rates in the vicinity of one percent or less were observed for each of the operators separately with respect to all variables other than Vehicle Identification Number (VIN).[1]

---

[1]   VIN, with a data entry error rate ten times that of the others, was especially troublesome for essentially two reasons: (a) the greater number of characters, and therefore opportunities for error, per entry, and (b) the difficulty encountered in interpreting handwritten characters, i.e., Z's that looked like 2's, G's that looked like 6's, etc.

3

In addition to the checking of discrepancies, a sample of non-discrepant cases was checked to determine if there were any significant likelihood that both operators might have made the identical error on the same data element. Cases of this nature, which would have defeated the purpose of the double entry process employed in this project, were found to be virtually non-existent.

Proceeding in this step-by-step fashion, a single adjudicated file, essentially free from data entry error, was created for each separate PSU/PAR version and data entry phase. A final report documenting this series of activities ("Police Accident Report Quality Assessment Project: Final Report on Task Order 9401") was produced in May 1995.

Task 2 of the present contract, begun in September 1995, extended these activities in two respects:

a.  Completion of computerized PAR database. - The remaining 75% of the accident reports for each PSU/PAR version were entered, compared, and adjudicated. The following data entry error rates were observed:

|                          | Operator 1 | Operator 2 |
|--------------------------|------------|------------|
| Variables other than VIN | 0.4%       | 1.0%       |
| VIN                      | 4.0%       | 13.5%      |

The adjudicated files for each separate phase were then consolidated into a single file, one file per PSU/PAR version. Files were also created for other variables originally reported by the police in a form not conducive to direct data entry (e.g., Make and Model, Occupant Protection System Use in PSU 79A). A short report documenting the data entry activities conducted in both Tasks 1 and 2 was produced in December 1995.

b.  PAR-CDS comparisons. - A PAR-CDS comparison process, necessary for achieving the ultimate goal of this study, then followed. It consisted of three components:

Step 1.  Linkage. - A merged file was created in which the PAR and CDS values for each accident, vehicle, and/or occupant were linked.

Step 2.  Comparison. - The CDS and PAR values for all variables applicable to a given accident, vehicle, and/or occupant were compared and any discrepancies noted.

Step 3.  Report generation. - Summary statistics were generated describing the number, percentage, and character of discrepancies noted for each target variable in each PSU/PAR version.

4

This report documents the preceding series of steps and is organized as follows:

o    Section 2 describes the methodology used in Step 1 to link the PAR and CDS databases.

o    Section 3 presents the summary statistics produced in Step 3. These include:

    (a)   the number and percentage of cases in which the CDS and PAR values agreed or were at least consistent,

    (b)   the number and percentage of cases in which the CDS and PAR values disagreed or were <u>in</u>consistent, and

    (c)   the number and percentage of cases in which no comparison could be made, for one or more of the following reasons:

        -   One value or the other (either CDS or PAR) was missing, i.e., no entry was made.

        -   One value or the other was reported to be "unknown".

        -   The two values, because of the nature of the coding scheme, had no bearing on each other.[2]

        -   The value contained in either the original PAR report or the photocopy was illegible.

        -   The value contained in either the original PAR report or the CDS file was out-of-range.

o    Section 4 is a consolidated summary, including key observations, conclusions, and recommendations, of the information presented in Section 3.

As noted earlier, not all target variables were reflected in every PSU/PAR version nor were all variables reported in a manner conducive to computerized comparison. Appendix A identifies, for each PSU/PAR version, the specific data elements included in the computerized database and any special instructions given to the data entry operators concerning their entry.

---

[2]    The CDS variable "Treatment-Mortality" is an example: it features some codes that deal with treatment, while others deal with mortality. For CDS purposes, this may be adequate but it does present a problem when dealing with the target variable "Injured Transported to Medical Facility". The problem is this: if an occupant dies, the resulting CDS code -- 1 (Fatal) -- says nothing about whether the person had previously been transported to a medical facility for treatment.

## 1.3  Additional Observations

In addition to the PSU- and variable-specific findings presented in Sections 3 and 4, several observations of a broadly generic nature are presented below:

a.    Variations in Report Forms and Reporting Procedures

The variations in report forms and reporting procedures noted among police jurisdictions present two distinct sets of problems:

(1)    Variations of this nature complicate the task of producing valid national estimates. If one jurisdiction fails to report a variable or reports it in a manner different from that of the others, any resulting national estimates are compromised both in terms of sample size and generalizability of results.

(2)    The use of separate report forms and reporting procedures by over fifty jurisdictions to accomplish what would appear to be the same purpose, may not be the most efficient use of state funds. At least two elements of inefficiency are involved:

(a)    Tax revenues devoted to developing state-specific forms and procedures might be better directed toward improving driver education, repairing highways, etc.

(b)    The use of disparate forms and procedures complicates unnecessarily the task of drawing cross-state comparisons and producing nationwide statistics.

While it is too late to modify reporting systems that have evolved over the years, there is promise for the future. As technologically advanced data collection methods emerge, PARs may be revised by the states. As such revisions become necessary, it would behoove NHTSA or some other national traffic safety organization to take the lead in coordinating, supporting, and facilitating the development of standardized data collection instruments, thus avoiding unnecessary duplication of development efforts. The following principle, applicable not only to NHTSA but to other agencies inside and outside the Department of Transportation, is respectfully proposed:

> Reporting systems that serve a common end and are ultimately used to produce nationwide statistics should be (i) common to all states, and (ii) developed by the states acting in concert rather than separately.[3]

---

[3]    The concept of a uniform reporting system on matters of national as well as local concern is hardly unique to the field of highway traffic safety. The Uniform Hospital Discharge Data System, developed in the 1970s, is an example of such a system.

b. __Nature of CDS-PAR Discrepancies__

Where discrepancies exist between the CDS and PAR values for a given variable, one is tempted to conclude that the CDS value -- because of the additional passage of time involved in its preparation as well as the greater range of information sources available -- is necessarily correct. Readers are cautioned that while this is generally the case, it is not __always__ the case. "Discrepancies" sometimes result from differences in the way the respective variables are defined. But even where there are no definitional differences, a few cases were observed in which it was the CDS value, rather than the PAR, that appeared to be in error.

> __Example 1__. - The PAR for a given crash reported the DATE OF ACCIDENT as "August 20, 1991" and the DAY OF WEEK as "Tuesday". The CDS database reports the DAY OF WEEK, derived by software operation on the value entered for DATE OF ACCIDENT, as "Saturday". Since the two items reported by the police are internally consistent (August 20, 1991 was indeed a Tuesday), it is considered more likely that someone in the CDS reporting chain -- either the original researcher or the person responsible for data entry -- erred on one data element (DATE OF ACCIDENT) than the police officer on two.

> __Example 2__. - CDS reported the SEX of a passenger named Thelma as "male". The PAR showed it as "female".

> __Example 3__. - In a jurisdiction that reports time in military fashion, the PAR showed the TIME of a given accident as "2201"; CDS reported it as "1001". Two scenarios are possible: either (a) the accident occurred at 10:01 in the evening and the time was incorrectly entered as "1001" by someone in the CDS reporting chain or (b) it occurred at 10:01 in the morning and the time was incorrectly entered as "2201" by the police. The former type of error is considered to be the more likely.[4]

These observations do not in any way detract from the value of CDS as an important information source. They are presented simply to highlight the fact that not every CDS-PAR "discrepancy" can be automatically taken to denote that it was the PAR value that was incorrect. As an adjunct to this study, fifteen randomly selected discrepancies involving the variable Vehicle Identification Number were investigated by the COTR, Dr. Carl Pierchala. In fourteen cases (see discussion, top of page 48), the discrepancy resulted from an apparent error on the part of the police; in the fifteenth case, there was no way of determining which source was correct.

---

[4]    Errors of this nature were possible in 1991 but would now be caught through a quality control software feature, implemented in 1995, that takes into account the light conditions at the time of the accident.

## 2. METHODOLOGY USED TO LINK CDS AND PAR FILES

For each separate PSU/PAR version, linkage was accomplished in three stages -- first accidents, then vehicles, then occupants. Each of these stages is described below.

Accidents

Accidents were linked in the obvious manner, by Case Identification Number (CASEID). This produced a correct match in every case.

Vehicles

All vehicles involved in the same accident were then linked by Vehicle No. and an output listing produced showing which vehicles had been linked and which had not. By scanning the output, two lines of investigation were pursued:

a.  Unlinked vehicles. - All vehicles that remained unlinked were examined to determine if there were not another vehicle in the other file that had the same (or approximately the same) Vehicle Identification Number (VIN).

b.  Improperly linked vehicles. - The VINs of all vehicles that had been linked were examined to determine if the linkage appeared to be correct. (Since VIN is not reported in PSU 79, judgments in that PSU concerning possibly improper linkage were based on Model Year instead.)

The results of these investigations were as follows:

-   Unlinked vehicles. - In every case where a vehicle had not been linked, it was because a vehicle reported by the police had been legally parked, was otherwise not relevant to CDS, or had fled the scene. This was to be expected: CDS does not require a General Vehicle (GV) record for vehicles that were not in transport at the time of the accident or that fled the scene, whereas the PAR database was designed to include all vehicles regardless of status. The number of PAR-reported vehicles for which there was no corresponding CDS record is shown below:

| PSU/PAR Version | Number of PAR-Reported Vehicles With No CDS Counterpart |
|---|---|
| 8 | 12 (out of 489) |
| 13 | 7 (out of 478) |
| 45 | 13 (out of 386) |
| 79A | 2 (out of 232) |
| 79B | 1 (out of 66) |

8

All vehicles for which there was no CDS record were dropped from the merged file, bringing the PAR and CDS components of the file into numerical accord.

- Improperly linked vehicles. - The specific linkages reached based on Vehicle No. alone proved to be correct in almost every case. There were only two exceptions: one each in PSUs 13 and 45. In each case, the police had assigned a slightly different Vehicle No. sequence than that assigned by CDS. The mismatch was corrected by changing the PAR sequence to match that of CDS.

## Occupants

The occupants of each vehicle were then linked based on Seat Position. The program designed to perform this function made use of the following relationship between PAR and CDS seat position codes:

|  | Corresponding PAR Code | | | | |
| CDS Code | PSU 8 | PSU 13 | PSU 45 | PSU 79A* | PSU 79B |
|----------|-------|--------|--------|----------|---------|
| 11 | 1 | 1 | 11 | LF | 1 |
| 12 | 2 | 2 | 12 | CF | 2 |
| 13 | 3 | 3 | 13 | RF | 3 |
| 21 | 4 | 4 | 21 | LR | 4 |
| 22 | 5 | 5 | 22 | CR | 5 |
| 23 | 6 | 6 | 23 | RR | 6 |
| 31 | - | - | 31 | -- | - |
| 32 | - | - | 32 | -- | - |
| 33 | - | - | 33 | -- | - |

\* In PSU 79A, Seat Position is reported in clear text ("right front", "left rear", etc.). The PAR data entry operators were instructed to code these entries as shown above. In addition, code "XR" was to be assigned to any rear seat passenger whose left/right/center designation had been omitted. If only one rear seat passenger was reported by CDS, the two records were assumed to reflect the same person and were linked.

An output listing similar to that prepared for vehicles was produced showing which occupants had been linked and which had not. Several distinct phenomena were noted:

(1) PAR-reported occupants with no CDS counterpart. - As with vehicles, many occupants could not be linked because they had no CDS counterpart. Again, this was to be expected: CDS normally creates an Occupant Assessment (OA) record only for occupants whose vehicle was towed, whereas the PAR data entry operators were instructed to report all occupants regardless of status. The number of PAR-reported occupants for whom no counterpart record could be found in CDS is shown below:

9

| PSU/PAR Version | Number of PAR-Reported Occupants With No CDS Counterpart |
|---|---|
| 8 | 146 (out of 710) |
| 13 | 89 (out of 775) |
| 45 | 84 (out of 590) |
| 79A | 80 (out of 377) |
| 79B | 29 (out of 107) |

As with vehicles, all such unlinkable occupants were dropped from the merged file.

(2) <u>Anomalies involving seat position</u>. - Some occupants could not be linked, or were improperly linked, because of one or more anomalies involving the variable Seat Position:

    (a) <u>Differences in reported values</u>. - The same passenger, clearly matched by Age and Sex, was sometimes assigned a different position by CDS than by the police. One of these values is clearly incorrect but in most cases, the source of the error could not be determined.

    (b) <u>Missing entries or unknowns</u>. - Occasionally, there was no PAR entry for Seat Position. On rare occasions, the CDS value was "U" (unknown).

    (c) <u>Ambiguity</u>. - Two or more persons in the same vehicle were sometimes assigned the same position code, making linkage based on Seat Position alone impossible. Other cases of ambiguity included (i) occupants whose assigned position code was "other", and (ii) occupants in PSU 79A for whom the left/right/center designation was omitted.

    (d) <u>PAR/CDS coding differences</u>. - Some CDS seat position codes (e.g., those ending in "4" or "5") had no direct PAR counterpart.

Cases in which occupants had not been properly linked -- or could not be linked at all -- were detected (and corrected) by visually scanning the output listing referred to on the previous page, making use of corollary information such as Age and Sex. The number of occupants for whom this form of intervention was required is shown below:

| PSU/PAR Version | Number of Occupants Matched with the Aid of Human Intervention |
|---|---|
| 8 | 16 (out of 564) |
| 13 | 25 (out of 686) |
| 45 | 21 (out of 506) |
| 79A | 11 (out of 297) |
| 79B | 3 (out of 78) |

The final step in linking previously unlinked (or improperly linked)
occupants was accomplished by changing one of the seat position
codes (either CDS or PAR) to match the other. In cases of potential
ambiguity, i.e., where more than one person had originally been
assigned the same position code, it was sometimes necessary to
create new codes in order to accomplish the desired purpose.
EXAMPLE: If two persons sitting in a remote section of the same
vehicle had originally been assigned CDS code 98 ("Other") and no
PAR seat position code had been entered for either, linkage was
accomplished by assigning one person dummy code "91" and the other
dummy code "92". The codes thereby created appear only in the
merged file, not the original CDS and PAR files, and were used only
to facilitate the linkage process, not to convey where the person
actually sat.

Given the importance of Seat Position not only as a mechanism for
linking occupants but as one of the target variables of primary
interest, a finer breakdown of the above-mentioned anomalies is
presented below:

| PSU: | 8 | 13 | 45 | 79A | 79B |
|---|---|---|---|---|---|
| Number of occupants | 564 | 686 | 506 | 297 | 78 |
| Number of anomalies in seat position | 16 (2.8%) | 25 (3.6%) | 21 (4.1%) | 11 (3.7%) | 3 (3.8%) |
| Breakdown by type: | | | | | |
| - Differences in reported values | 6 (1.1%) | 6 (0.9%) | 9 (1.8%) | 2 (0.7%) | 1 (1.3%) |
| - Missing entries or unknowns | | | | | |
| PAR missing | 4 (0.7%) | 7 (1.0%) | 2 (0.4%) | 1 (0.3%) | - |
| CDS unknown | 2 (0.4%) | - | - | - | - |
| - Ambiguity (e.g., more than 1 person in same position; left/right/center designation omitted) | 3 (0.5%) | 6 (0.9%) | 2 (0.4%) | 6 (2.0%) | - |
| - PAR/CDS coding differences | 1 (0.2%) | 6 (0.9%) | 8 (1.6%) | 2 (0.7%) | 2 (2.6%) |

11

## 3. RESULTS OF CDS-PAR COMPARISON PROCESS

After the PAR and CDS files were merged, with all accidents, vehicles, and occupants linked as described in Section 2, the merged file was subjected to a series of comparisons to determine the extent to which the linked values were in agreement. Not all target variables were included in the comparison; Appendix A lists the specific variables applicable in each PSU/PAR version and any special data entry instructions associated with the PAR data element corresponding to each such variable.

This section summarizes the results of these comparisons. The summaries that follow address all of the target variables identified in Appendix A, with the following exceptions:

a.    Seat Position. - No additional analysis was performed with respect to this variable beyond that presented in Section 2.

b.    Damaged Area. - In PSUs 79A and B, this information is uncoded, i.e., does not lend itself to computerized comparison. In the other PSUs, the data element reported is "point of initial impact" rather than, as in CDS, "area of greatest deformation". Since, in any given accident, the area of greatest deformation (as reported by CDS) is not necessarily the point of initial impact (as reported by the police), the fact that the two values differ does not in and of itself denote error. Other complications are involved as well in the interpretation of this variable. Because of funding limitations, it was decided not to pursue the effort required to establish the precise relationship between "point of impact" and "area of deformation".

c.    Extent of Deformity. - There is a disparity in interpretation with respect to this variable as well. CDS uses a nine-point scale to convey the extent of deformity; the various PSU/PAR versions use scales with as few as three points:    "Light/ Moderate/Severe", "Under $200/$200 to $500/Over $500", etc. As noted in an earlier ASTI report:

> A precise relationship between the CDS damage scale and those used in the various PAR jurisdictions has yet to be established. The database to be developed as a result of this task will provide an excellent empirical basis for studying these relationships.    (ASTI, "Comparison of Coding Systems for All Target Variables", June 1994)

Again, however, because of existing funding limitations it was decided not to pursue the effort required to establish a precise relationship between the two sets of scales.

12

For the remaining target variables, the CDS-PAR comparison process took the following form:

(1) Pairs of values for which no direct comparison was possible were set aside as a separate grouping. Reasons for this form of indeterminacy included:

- Missing entries. - One value or the other (either CDS or PAR or both) was missing, i.e., no entry was made.

- Unknowns. - One value or the other was explicitly stated to be "unknown".

- Illegible entries. - The value contained in either the original PAR report or the photocopy was illegible.

- Out-of-range values. - One value or the other was out-of-range.

(2) The remaining pairs of values -- those for which legitimate PAR and CDS values were reported -- were then divided into three categories:

AGREE            Pairs of values that definitely agree or are at least consistent.

DISAGREE         Pairs of values that definitely disagree or are inconsistent.

INDETERMINATE    Pairs of values for which no judgment could be reached because of differences in the two coding schemes involved, i.e., one value has no bearing on the other.[5]

This taxonomy differs slightly from that originally planned. At an earlier stage in this study, it was planned to subdivide agreements and disagreements into two categories: "definite" and "probable". For a number of reasons, however, including the desire to simplify the presentation of data, it was decided to forgo the distinction between "definite" and "probable".

The summaries presented in this section are organized by PSU/PAR version. They show, for each target variable in each jurisdiction, the number and percentage of cases that "agree", that "disagree", and that fall into each of the other categories defined above. Where one or more categories other than simple agreement or disagreement is shown for a given variable, an overall discrepancy rate is formed and each of the line items that contributed to the

---

[5]    An example of a situation of this nature was given in footnote 2.

rate is denoted by asterisk.  For the purpose of this study, an item
was deemed to be "discrepant" if:

(a)  the PAR value was either missing, ambiguous, out-of-
range, or unknown and the CDS value was specific, or

(b)  the two values definitely disagreed or were definitely
inconsistent.

Where an overall discrepancy rate is not shown, the discrepancy rate
is simply the rate of disagreement.  Readers are of course free to
redefine the concept of discrepancy by combining these numbers in
any alternative manner they choose.  On page 17, for example, the
discrepancy rate of 17.8% shown for Vehicle Identification Number
was arrived at by combining the 0.4% of cases in which a PAR entry
was missing while CDS was specific with the 17.4% of cases in which
both values were specific but disagreed.  One might, if one desired,
add to this composite the 1.0% of cases in which a PAR entry was
missing and the CDS value was stated to be unknown.  Given, however,
that the focus of this study is on PAR (rather than CDS) reporting
quality, occurrences of the latter nature -- those in which there
was no specific CDS value against which to compare the PAR value --
were not regarded as discrepancies.  Only where (a) the CDS value
was specific and the PAR value was not, or (b) both values were
specific but were neither in agreement nor consistent, was a
discrepancy deemed to exist.

Note that where the CDS and PAR values are specific but disagree,
the data summaries that follow make no distinction between cases in
which the PAR value was incorrect and those in which the CDS value
was incorrect.  The effort required to conduct such an effort on a
sufficiently broad basis to warrant the presentation of numbers
exceeded the funding limitations of this study.  Several cases were
noted, however, in which it was clear that it was the CDS value, and
not the PAR, that was in error (see earlier discussion on page 7).

The following table summarizes, by PSU/PAR version, the number of
accidents, vehicles, and occupants involved in the study sample:

| PSU/PAR Version | Number of: Accidents | Vehicles | Occupants |
|---|---|---|---|
| 8 | 276 | 477 | 564 |
| 13 | 293 | 471 | 686 |
| 45 | 215 | 373 | 506 |
| 79A | 122 | 230 | 297 |
| 79B | 33 | 65 | 78 |
| Total | 939 | 1,616 | 2,131 |

14

Accident-Level Variables (N = 276):

<u>Time</u>

```
PAR entry missing, CDS unknown                      2      0.7%
CDS/PAR values:   Agree                           270     97.8%
                  Disagree                          4      1.5%*

                          * Overall discrepancy rate      1.5%
```

<u>Month</u>

```
CDS/PAR values:   Agree                           276    100.0%
                  Disagree                          0      0.0%
```

<u>Day of Week</u>

```
CDS/PAR values:   Agree                           274     99.3%
                  Disagree                          2      0.7%
```

<u>Number of Motor Vehicles</u>

```
CDS/PAR values:   Agree or are consistent         276    100.0%
                  Disagree**                        0      0.0%
```

** No essential differences were noted between the PAR and CDS values for this variable. Occasional differences involved matters of definition only. The applicable CDS value is the number of vehicle forms created for inclusion in the CDS General Vehicle File; vehicles not in transport at the time of the accident or that fled the scene are not included. The PAR value, on the other hand, pertains to all vehicles regardless of status. PAR-CDS differences attributable solely to this distinction were not counted as discrepancies. There were seven such cases in this PSU. [NOTE: The same comment applies to PSUs 13 and 45 as well, in which where there were seven and twelve such cases respectively.]

15

## Occupant Fatalities

| | | |
|---|---:|---:|
| PAR entry missing, CDS specific | 4 | 1.5%* |
| CDS/PAR values:   Agree | 272 | 98.5% |
|                   Disagree | 0 | 0.0%* |

* Overall discrepancy rate    1.5%

NOTE:    The CDS Accident File has no separate variable that reports the number of persons killed in a given crash.  For the purpose of this study, that number was derived by scanning the occupant-level variable TREATMENT-MORTALITY for every occupant in every crash.  The numbers thus derived were then compared, accident by accident, to the numbers reported by the police.

There were five accidents in this PSU for which fatalities were reported.  In every case, both the reported value for PAR and the derived value for CDS was "1".  Except for the four cases shown above in which there was no PAR entry at all, there were no cases in which the PAR and CDS values for this variable disagreed.

16

Vehicle-Level Variables (N = 477):

Speed Limit

| | | | |
|---|---|---|---|
| PAR entry missing, CDS specific | | 20 | 4.2%* |
| PAR specific, CDS unknown | | 1 | 0.2% |
| CDS/PAR values: | Agree | 432 | 90.6% |
| | Disagree | 21 | 4.4%* |
| | Indeterminate | 3 | 0.6% |

<p align="center">* Overall discrepancy rate    8.6%</p>

Vehicle Identification Number

| | | | |
|---|---|---|---|
| PAR entry missing: | CDS specific | 2 | 0.4%* |
| | CDS unknown | 5 | 1.0% |
| Illegible PAR photocopy | | 1 | 0.2% |
| CDS/PAR values: | Agree | 386 | 80.9% |
| | Disagree | 83 | 17.4%* |

<p align="center">* Overall discrepancy rate    17.8%</p>

Make

| | | | |
|---|---|---|---|
| PAR entry missing: | CDS specific | 1 | 0.2%* |
| | CDS unknown | 2 | 0.4% |
| CDS/PAR values: | Agree | 473 | 99.2% |
| | Disagree | 1 | 0.2%* |

<p align="center">* Overall discrepancy rate    0.4%</p>

Model

| | | | |
|---|---|---|---|
| PAR entry missing, CDS unknown | | 3 | 0.6% |
| CDS/PAR values: | Agree | 470 | 98.5% |
| | Disagree | 5 | 1.0%* |

<p align="center">* Overall discrepancy rate    1.0%</p>

Model Year

| | | | |
|---|---|---|---|
| PAR entry missing: | CDS specific | 3 | 0.6%* |
| | CDS unknown | 3 | 0.6% |
| CDS/PAR values: | Agree | 469 | 98.3% |
| | Disagree | 2 | 0.4%* |

<p align="center">* Overall discrepancy rate    1.0%</p>

Body Type

```
    PAR entry missing, CDS specific                     13      2.7%*
    CDS/PAR values:   Agree                            340     71.3%
                      Disagree                         124     26.0%*

                          * Overall discrepancy rate           28.7%
```

Towed Due to Damage?

```
    PAR entry missing, CDS specific                     12      2.5%*
    CDS/PAR values:   Agree                            461     96.7%
                      Disagree                           4      0.8%*

                          * Overall discrepancy rate            3.3%
```

Driver Alcohol/Drug Presence

```
    PAR entry missing:  CDS specific                    72     15.1%*
                        CDS unknown                    205     43.0%
    PAR and CDS both unknown                              3      0.6%
    CDS/PAR values:   Agree                            195     40.9%
                      Disagree                            2      0.4%*

                          * Overall discrepancy rate           15.5%
```

18

Occupant-Level Variables (N = 564):

<u>Age</u>

|  |  |  |
|---|---|---|
| PAR entry missing:  CDS specific | 10 | 1.8%* |
| CDS unknown | 2 | 0.4% |
| PAR unknown:  CDS specific | 16 | 2.8%* |
| CDS unknown | 4 | 0.7% |
| CDS/PAR values:  Agree | 496 | 87.9% |
| Disagree by one year | 23 | 4.1% |
| Disagree by > one year | 13 | 2.3%* |
| * Overall discrepancy rate | | 6.9% |

<u>Sex</u>

|  |  |  |
|---|---|---|
| PAR entry missing, CDS specific | 6 | 1.1%* |
| CDS/PAR values:  Agree | 553 | 98.0% |
| Disagree | 5 | 0.9%* |
| * Overall discrepancy rate | | 2.0% |

<u>Injury Classification</u>

|  |  |  |
|---|---|---|
| PAR entry missing, CDS specific | 10 | 1.8%* |
| PAR unknown:  CDS specific | 44 | 7.8%* |
| CDS unknown | 22 | 3.9% |
| CDS/PAR values:  Agree or are consistent | 468 | 83.0% |
| Disagree | 20 | 3.5%* |
| * Overall discrepancy rate | | 13.1% |

<u>Transported to Medical Facility?</u>

|  |  |  |
|---|---|---|
| PAR entry missing, CDS specific | 17 | 3.0%* |
| PAR unknown, CDS specific | 1 | 0.2%* |
| PAR specific, CDS unknown | 37 | 6.6% |
| CDS/PAR values:  Agree or are consistent | 494 | 87.6% |
| Disagree | 15 | 2.7%* |
| * Overall discrepancy rate | | 5.9% |

Occupant Protection System Use

```
PAR entry missing, CDS specific                     9      1.6%*
PAR unknown:  CDS specific                          84     14.9%*
              CDS unknown                           16      2.8%
PAR specific, CDS unknown                           43      7.6%
Illegible PAR photocopy                              1      0.2%
CDS/PAR values:  Agree or are consistent           293     52.0%
                 Disagree                           118     20.9%*
```

                          * Overall discrepancy rate    37.4%

NOTE:    Two separate PAR variables are involved.  "Active Restraint
         Type" covers the matter of availability and "Active
         Restraint Usage" the matter of usage.  The corresponding
         CDS variable is "Manual (Active) Belt System Use", which
         covers both availability and usage.

Air Bag Deployed?

```
PAR entry missing, CDS specific                      5      0.9%*
PAR unknown:  CDS specific                          91     16.1%*
              CDS unknown                            1      0.2%
CDS unknown:  PAR specific                           6      1.1%
              PAR value out of range                 1      0.2%
Illegible PAR photocopy                              3      0.5%
CDS/PAR values:  Agree or are consistent           445     78.9%
                 Disagree                           12      2.1%*
```

                          * Overall discrepancy rate    19.1%

NOTE:    The applicable PAR variable ("Passive Restraint Type") and
         CDS variable ("Air Bag System Deployment") both cover in a
         single data element the matters of air bag availability and
         deployment.  The 12 cases of disagreement were as follows:

         -    In nine cases, CDS said an air bag was "not deployed";
              PAR said there was "none available".

         -    In two cases, PAR said an air bag was "not deployed";
              CDS said there was "none available".

         -    In one case, CDS said an air bag was "deployed"; PAR
              said there was "none available".

         Although the first two sets of responses might in a limited
         sense be regarded as consistent, they were classified above
         as discrepancies since in each case, a more compatible pair
         of responses was possible.

<u>Ejection</u>

| | | |
|---|---|---|
| PAR entry missing, CDS specific | 13 | 2.3%* |
| PAR unknown, CDS specific | 4 | 0.7%* |
| CDS/PAR values: Agree or are consistent | 546 | 96.8% |
| Disagree | 1 | 0.2%* |

* Overall discrepancy rate   3.2%

NOTE:   There were seven cases in this PSU in which an occupant was ejected. In each such case, both the PAR and CDS values were in agreement. The sole case of disagreement involved a PAR entry that was placed in the wrong box, i.e., all entries in that particular report were positioned one box to the left of where they should have been.

Accident-Level Variables (N = 293):

Time

| | | | |
|---|---|---|---|
| PAR entry missing, CDS specific | | 4 | 1.4%* |
| CDS/PAR values: | Agree | 275 | 93.9% |
| | Disagree | 14 | 4.8%* |

\* Overall discrepancy rate 6.2%

Month

| | | | |
|---|---|---|---|
| CDS/PAR values: | Agree | 293 | 100.0% |
| | Disagree | 0 | 0.0% |

Day of Week

| | | | |
|---|---|---|---|
| PAR entry missing, CDS specific | | 1 | 0.3%* |
| Illegible PAR photocopy | | 3 | 1.0% |
| CDS/PAR values: | Agree | 279 | 95.2% |
| | Disagree | 10 | 3.4%* |

\* Overall discrepancy rate 3.7%

Number of Motor Vehicles

| | | | |
|---|---|---|---|
| CDS/PAR values: | Agree or are consistent | 293 | 100.0% |
| | Disagree** | 0 | 0.0% |

** See related comment, bottom of page 15.

22

Vehicle-Level Variables (N = 471):

## Number of Occupants

| | | | |
|---|---|---:|---:|
| PAR entry missing: | CDS specific | 7 | 1.5%* |
| | CDS unknown | 4 | 0.8% |
| | CDS missing** | 1 | 0.2% |
| PAR specific: | CDS unknown | 3 | 0.6% |
| | CDS missing** | 6 | 1.3% |
| CDS/PAR values: | Agree or are consistent | 445 | 94.5% |
| | Disagree | 5 | 1.1%* |

* Overall discrepancy rate    2.6%

**    In CDS, the variable "Number of Occupants This Vehicle" is not required to be reported for vehicles that are not CDS relevant. The seven missing values shown above fell into this category: all involved medium or heavy trucks or truck-tractors. [NOTE: Ordinarily, CDS does not contain missing values. Selected data elements, such as this one, for non-CDS relevant vehicles happen to be an exception.]

## Vehicle Identification Number

| | | | |
|---|---|---:|---:|
| PAR entry missing: | CDS specific | 1 | 0.2%* |
| | CDS unknown | 4 | 0.8% |
| CDS/PAR values: | Agree | 381 | 80.9% |
| | Disagree | 85 | 18.0%* |

* Overall discrepancy rate    18.2%

## Make

| | | | |
|---|---|---:|---:|
| PAR entry missing: | CDS specific | 3 | 0.6%* |
| | CDS unknown | 1 | 0.4% |
| PAR value out-of-range | | 2 | 0.4%* |
| CDS value out-of-range | | 2 | 0.4% |
| CDS/PAR values: | Agree or are consistent | 451 | 95.8% |
| | Disagree | 2 | 0.4%* |
| | Indeterminate** | 10 | 2.1% |

* Overall discrepancy rate    1.4%

**    For certain classes of vehicles (trucks, police equipment, etc.), the PAR instructions for this PSU require that Body Type be reported, rather than Make. There were ten such cases in all. In cases of this nature, no comparison of Make is of course possible.

## Model Year

| | | | |
|---|---|---|---|
| PAR entry missing: | CDS specific | 1 | 0.2%* |
| | CDS unknown | 2 | 0.4% |
| PAR specific, CDS unknown | | 1 | 0.2% |
| CDS/PAR values: | Agree | 463 | 98.3% |
| | Disagree | 4 | 0.9%* |

* Overall discrepancy rate    1.1%

## Body Type

| | | | |
|---|---|---|---|
| PAR entry missing: | CDS specific | 5 | 1.1%* |
| | CDS unknown | 1 | 0.2% |
| PAR value out-of-range | | 1 | 0.2%* |
| CDS/PAR values: | Agree | 440 | 93.4% |
| | Disagree | 24 | 5.1%* |

* Overall discrepancy rate    6.4%

## Towed Due to Damage?

| | | | |
|---|---|---|---|
| PAR entry missing, CDS specific | | 12 | 2.5%* |
| CDS/PAR values: | Agree | 450 | 95.5% |
| | Disagree | 9 | 1.9%* |

* Overall discrepancy rate    4.4%

## Driver Alcohol/Drug Presence

| | | | |
|---|---|---|---|
| PAR entry missing: | CDS specific | 17 | 3.6%* |
| | CDS unknown | 4 | 0.8% |
| CDS/PAR values: | Agree or are consistent | 449 | 95.3% |
| | Disagree | 1 | 0.2%* |

* Overall discrepancy rate    3.8%

NOTE:    In the CDS database, Driver Alcohol/Drug Presence is reported as two separate variables, one applicable to alcohol and the other to drugs. The PAR form for this PSU, however, records only a single variable ("HBD - Yes or No?"). Although "HBD" stands for Had Been Drinking, the instructions clearly state that the variable applies to both alcohol and drugs, i.e., if either is present, the correct response is "Yes". CDS, however, apparently treats it as applying to alcohol only: of the 471 cases in this PSU, in all but three the value assigned to Drug Presence was "Not reported". For the purpose of this study, "Not reported" was taken to imply the absence of drugs.

24

Occupant-Level Variables (N = 686):

Age

```
    PAR entry missing:  CDS specific                    34      5.0%*
                        CDS unknown                     11      1.6%
    CDS/PAR values:   Agree                            593     86.4%
                      Disagree by one year              20      2.9%
                      Disagree by > one year            28      4.1%*

                          * Overall discrepancy rate            9.1%
```

Sex

```
    PAR entry missing:  CDS specific                    24      3.5%*
                        CDS unknown                     10      1.5%
    CDS/PAR values:   Agree                            641     93.4%
                      Disagree                          11      1.6%*

                          * Overall discrepancy rate            5.1%
```

Injury Classification

```
    PAR entry missing:  CDS specific                    27      3.9%*
                        CDS unknown                     10      1.5%
    CDS/PAR values:   Agree or are consistent          639     93.1%
                      Disagree                          10      1.5%*

                          * Overall discrepancy rate            5.4%
```

Transported to Medical Facility?

```
    PAR entry missing:  CDS specific                   111     16.2%*
                        CDS unknown                      4      0.6%
    PAR specific, CDS unknown                            5      0.7%
    CDS/PAR values:   Agree or are consistent          555     80.9%
                      Disagree                          11      1.6%*

                          * Overall discrepancy rate           17.8%
```

25

Occupant Protection System Use

```
    PAR entry missing:  CDS specific                30     4.4%*
                        CDS unknown                 19     2.8%
    PAR unknown, CDS specific                        8     1.2%*
    PAR specific, CDS unknown                        6     0.9%
    PAR value out-of-range                           1     0.1%*
    Illegible PAR photocopy                          3     0.4%
    CDS/PAR values:  Agree or are consistent       532    77.6%
                     Disagree                       87    12.7%*
```

                    * Overall discrepancy rate    19.3%

NOTE:     The applicable CDS variable, as in PSU 8 and the other
          PSUs, is "Manual (Active) Belt System Use". The applicable
          PAR variable, entitled "Restraint by Occupant Position",
          covers all form of restraints, both active and passive.

Air Bag Deployed?

    The applicable PAR variable is the one used above to report
    Occupant Protection System Use. Only one of the codes for this
    variable, however, applies specifically to air bags and even that
    code -- code D ("airbag, activated and non-activated") -- fails
    to tell whether the bag was activated or not.

    In the PAR sample for this jurisdiction, there were only two
    occupants for whom code D was reported. It would be reasonable
    to assume that the only reason that particular code was selected
    (as opposed to the several others available) was that an air bag
    was indeed activated. That appears to be the case: the
    corresponding CDS value in both cases was 1 ("air bag deployed").
    There were five other cases, however, in which CDS reported a
    value of 1 but the police did not report a value of D. In four
    of those cases, the reported PAR value was B ("belt used") rather
    than D; in the fifth case, the PAR photocopy was illegible.

    Beyond this handful of cases, no meaningful comparison involving
    air bags is possible for the other occupants in the PSU 13
    sample.

26

Accident-Level Variables (N = 215):

First Harmful Event

    PAR entry missing, CDS specific                      1      0.5%*
    CDS/PAR values:   Agree or are consistent     212    98.6%
                    Disagree                2      0.9%*

                * Overall discrepancy rate    1.4%

Manner of Collision

    PAR entry missing, CDS specific                      1      0.5%*
    PAR specific, CDS unknown                      2      0.9%
    CDS/PAR values:   Agree                    154    71.6%
                    Disagree              58    27.0%*

                * Overall discrepancy rate   27.5%

NOTE:   The applicable PAR variable is labeled "Type of Collision".
          Although the instruction manual states that this particular
          variable applies only if two or more motor vehicles were
          involved, a specific value ("Head-on", "Rear end", "Side
          swipe", etc.) was nonetheless checked in forty accidents
          where only a single vehicle was involved, over two-thirds
          of the 58 disagreements noted above.  In each such case,
          the corresponding CDS value was "Not a collision".  All
          such cases -- cases in which a PAR entry was made even
          though none was required -- were treated as discrepancies.

Time

    PAR entry missing, CDS unknown                    2      0.9%
    CDS/PAR values:   Agree                    211    98.1%
                    Disagree                2      0.9%*

                * Overall discrepancy rate    0.9%

Month

    CDS/PAR values:   Agree                    215  100.0%
                    Disagree                0      0.0%

Day of Week

    CDS/PAR values:   Agree                    214    99.5%
                    Disagree                1      0.5%

Number of Motor Vehicles

    CDS/PAR values:  Agree or are consistent      215   100.0%
                    Disagree**                      0     0.0%

   **   See related comment, bottom of page 15.

Occupant Fatalities

    PAR entry missing, CDS specific           2     0.9%*
    CDS/PAR values:  Agree                   212    98.6%
                    Disagree**                    1     0.5%*

                    * Overall discrepancy rate    1.4%

   **   See earlier note on page 16.  In only one case within this
       PSU did the reported PAR value and the derived CDS value
       disagree:   the PAR reported no fatalities whereas CDS
       reported one.  The individual in question may have died
       after the police report was completed.

Vehicle-Level Variables (N = 373):

Speed Limit

```
    PAR entry missing, CDS specific                    10      2.7%*
    CDS/PAR values:   Agree                           360     96.5%
                      Disagree                          3      0.8%*

                      * Overall discrepancy rate             3.5%
```

Vehicle Identification Number

```
    PAR entry missing:  CDS specific                    3      0.8%*
                        CDS unknown                     5      1.3%
    PAR specific, CDS unknown                           1      0.3%
    CDS/PAR values:   Agree                           254     68.1%
                      Disagree                         110     29.5%*

                      * Overall discrepancy rate            30.3%
```

Make

```
    PAR entry missing:  CDS specific                    1      0.3%*
                        CDS unknown                     3      0.8%
    CDS/PAR values:   Agree                           367     98.4%
                      Disagree                           2      0.5%*

                      * Overall discrepancy rate             0.8%
```

Model

```
    PAR entry missing:  CDS specific                    6      1.6%*
                        CDS unknown                     4      1.1%
    CDS/PAR values:   Agree                           356     95.4%
                      Disagree**                         7      1.9%*

                      * Overall discrepancy rate             3.5%
```

    **  Typical discrepancies:  "Corona" as opposed to "Corolla",
        "Skylark" as opposed to "Skyhawk", etc.

Model Year

```
    PAR entry missing, CDS specific                     3      0.8%*
    CDS/PAR values:   Agree                           348     93.3%
                      Disagree                          22      5.9%*

                      * Overall discrepancy rate             6.7%
```

29

### Body Type

| | | |
|---|---:|---:|
| PAR entry missing, CDS specific | 6 | 1.6%* |
| Both unknown | 3 | 0.8% |
| CDS/PAR values:  Agree | 354 | 94.9% |
| Disagree | 10 | 2.7%* |

* Overall discrepancy rate    4.3%

### Towed Due to Damage?

| | | |
|---|---:|---:|
| PAR entry missing, CDS specific | 2 | 0.5%* |
| CDS/PAR values:  Agree | 366 | 98.1% |
| Disagree | 5 | 1.3%* |

* Overall discrepancy rate    1.8%

### Driver Alcohol/Drug Presence

| | | |
|---|---:|---:|
| PAR entry missing, CDS specific | 3 | 0.8%* |
| PAR unknown:  CDS specific | 7 | 1.9%* |
| CDS unknown | 10 | 2.7% |
| PAR specific, CDS unknown | 3 | 0.8% |
| CDS/PAR values:  Agree or are consistent | 342 | 91.7% |
| Disagree | 8 | 2.1%* |

* Overall discrepancy rate    4.8%

Occupant-Level Variables (N = 506):

Age

```
PAR entry missing:  CDS specific                      9      1.8%*
                    CDS unknown                       1      0.2%
CDS/PAR values:  Agree                              454     89.7%
                 Disagree by one year                30      5.9%
                 Disagree by > one year              12      2.4%*

                    * Overall discrepancy rate              4.2%
```

Sex

```
PAR entry missing, CDS specific                      8      1.6%*
CDS/PAR values:  Agree                             492     97.2%
                 Disagree                             6      1.2%*

                    * Overall discrepancy rate              2.8%
```

Injury Classification

```
PAR entry missing:  CDS specific                      3      0.6%*
                    CDS unknown                       5      1.0%
PAR ambiguous, CDS specific**                         1      0.2%*
CDS/PAR values:  Agree or are consistent            486     96.0%
                 Disagree                             11      2.2%*

                    * Overall discrepancy rate              3.0%
```

** The ambiguous PAR entry was one in which two separate injury codes were checked.

Transported to Medical Facility?

```
PAR entry missing:  CDS specific                     31      6.1%*
                    CDS unknown                       7      1.4%
PAR specific, CDS unknown                             3      0.6%
CDS/PAR values:  Agree or are consistent            462     91.3%
                 Disagree                             3      0.6%*

                    * Overall discrepancy rate              6.7%
```

31

Occupant Protection System Use

| | | |
|---|---|---|
| PAR entry missing, CDS specific | 12 | 2.4%* |
| PAR specific, CDS unknown | 10 | 2.0% |
| PAR ambiguous, CDS specific** | 1 | 0.2%* |
| CDS/PAR values:  Agree or are consistent | 397 | 78.5% |
| Disagree | 86 | 17.0%* |

* Overall discrepancy rate   19.6%

**    The ambiguous PAR entry was one in which both "Y" and "N" appeared to have been checked.

Ejection

| | | |
|---|---|---|
| PAR entry missing, CDS specific | 10 | 2.0%* |
| PAR specific, CDS unknown | 1 | 0.2% |
| CDS/PAR values:  Agree or are consistent | 480 | 94.9% |
| Disagree** | 15 | 3.0%* |

* Overall discrepancy rate   5.0%

**    Disagreements were subdivided as follows:

a.    In three cases, the police reported that ejection took place; CDS reported that it did not.

b.    In twelve cases, the police reported that ejection did not take place; CDS reported that it did.  Three of those cases, according to CDS, involved total ejection; nine involved partial ejection.

Accident-Level Variables (N = 122):

First Harmful Event

|  |  |  |  |
|---|---|---|---|
| PAR entry missing, CDS specific |  | 4 | 3.3%* |
| CDS/PAR values: | Agree or are consistent | 115 | 94.3% |
|  | Disagree | 3 | 2.4%* |
|  | * Overall discrepancy rate | | 5.7% |

Manner of Collision

|  |  |  |  |
|---|---|---|---|
| PAR entry missing, CDS specific |  | 4 | 3.3%* |
| CDS/PAR values: | Agree | 100 | 82.0% |
|  | Disagree | 18 | 14.8%* |
|  | * Overall discrepancy rate | | 18.1% |

Time

|  |  |  |  |
|---|---|---|---|
| CDS/PAR values: | Agree | 119 | 97.5% |
|  | Disagree | 3 | 2.5% |

Month

|  |  |  |  |
|---|---|---|---|
| CDS/PAR values: | Agree | 122 | 100.0% |
|  | Disagree | 0 | 0.0% |

Occupant Fatalities

|  |  |  |  |
|---|---|---|---|
| PAR specific, CDS unknown |  | 1 | 0.8% |
| CDS/PAR values: | Agree | 120 | 98.4% |
|  | Disagree | 1 | 0.8%* |
|  | * Overall discrepancy rate | | 0.8% |

NOTE: See earlier note on page 16. In only one case involving
this PSU/PAR version did the reported PAR value and the
derived CDS value disagree: the PAR reported no fatalities
whereas CDS reported one. The individual in question may
have died after the police report was completed.

Vehicle-Level Variables (N = 230):

Speed Limit

| | | | |
|---|---|---:|---:|
| PAR entry missing, CDS specific | | 2 | 0.9%* |
| CDS/PAR values: | Agree | 226 | 98.3% |
| | Disagree | 2 | 0.9%* |

* Overall discrepancy rate   1.7%

Number of Occupants

| | | | |
|---|---|---:|---:|
| PAR entry missing: | CDS specific | 6 | 2.6%* |
| | CDS unknown | 2 | 0.9% |
| | CDS missing** | 2 | 0.9% |
| PAR specific, CDS missing** | | 7 | 3.0% |
| CDS/PAR values: | Agree or are consistent | 203 | 88.3% |
| | Disagree | 10 | 4.3%* |

* Overall discrepancy rate   6.9%

** See earlier comment, page 23. The nine missing CDS values shown above all involved vehicles that were either non-CDS relevant -- for whom this data element is not required to be reported -- or that fled the scene.

Make

| | | | |
|---|---|---:|---:|
| PAR entry missing: | CDS specific | 1 | 0.4%* |
| | CDS unknown | 2 | 0.9% |
| CDS/PAR values: | Agree | 226 | 98.3% |
| | Disagree | 1 | 0.4%* |

* Overall discrepancy rate   0.9%

Model

| | | | |
|---|---|---:|---:|
| PAR entry missing: | CDS specific | 3 | 1.3%* |
| | CDS unknown | 3 | 1.3% |
| PAR specific, CDS unknown | | 3 | 1.3% |
| CDS/PAR values: | Agree | 215 | 93.5% |
| | Disagree** | 6 | 2.6%* |

* Overall discrepancy rate   3.9%

** Typical discrepancies: "Crown Victoria" as opposed to "LTD", "Celebrity" as opposed to "Cavalier", etc.

## Model Year

| | | |
|---|---|---|
| PAR entry missing:  CDS specific | 2 | 0.9%* |
| CDS unknown | 1 | 0.4% |
| Illegible PAR photocopy | 1 | 0.4% |
| CDS/PAR values:  Agree | 219 | 95.2% |
| Disagree | 7 | 3.0%* |

                             * Overall discrepancy rate   3.9%

## Towed Due to Damage?

| | | |
|---|---|---|
| PAR entry missing, CDS specific | 6 | 2.6%* |
| CDS/PAR values:  Agree | 222 | 96.5% |
| Disagree | 2 | 0.9%* |

                             * Overall discrepancy rate   3.5%

## Driver Alcohol/Drug Presence

| | | |
|---|---|---|
| PAR entry missing, CDS specific | 8 | 3.5%* |
| PAR unknown:  CDS specific | 1 | 0.4%* |
| CDS unknown | 17 | 7.4% |
| CDS/PAR values:  Agree or are consistent | 204 | 88.7% |
| Disagree | 0 | 0.0%* |

                             * Overall discrepancy rate   3.9%

Occupant-Level Variables (N = 297):

<u>Age</u>

|  |  |  |  |
|---|---|---|---|
| PAR entry missing: | CDS specific | 6 | 2.0%* |
|  | CDS unknown | 9 | 3.0% |
|  | CDS missing** | 1 | 0.3%* |
| CDS/PAR values: | Agree | 264 | 88.9% |
|  | Disagree by one year | 12 | 4.0% |
|  | Disagree by > one year | 5 | 1.7%* |

                          * Overall discrepancy rate      4.0%

** The term "missing", as used here, denotes a missing <u>record</u> rather than a missing data element; as noted earlier, CDS ordinarily does not contain missing values. The situation referred to involves an occupant for whom no CDS Occupant Assessment record was created even though the PAR report showed him to be a passenger in a CDS-relevant vehicle. It was classified above as a discrepancy since it is believed to have resulted from an error on the part of the police; according to the CDS researcher, the individual in question was actually seated in an "alternate" vehicle. The same interpretation of the term "missing" applies to each of the other variables on this page and the next.

<u>Sex</u>

|  |  |  |  |
|---|---|---|---|
| PAR entry missing: | CDS specific | 6 | 2.0%* |
|  | CDS unknown | 5 | 1.7% |
| PAR specific, CDS missing (see above) |  | 1 | 0.3%* |
| CDS/PAR values: | Agree | 281 | 94.6% |
|  | Disagree | 4 | 1.3%* |

                          * Overall discrepancy rate      3.6%

<u>Injury Classification</u>

|  |  |  |  |
|---|---|---|---|
| PAR entry missing: | CDS specific | 7 | 2.4%* |
|  | CDS unknown | 7 | 2.4% |
|  | CDS missing (see above) | 1 | 0.3%* |
| CDS/PAR values: | Agree | 278 | 93.6% |
|  | Disagree | 4 | 1.3%* |

                          * Overall discrepancy rate      4.0%

Occupant Protection System Use

```
PAR entry missing:  CDS specific                      34    11.4%*
                    CDS unknown                       11     3.7%
                    CDS missing (see above)            1     0.3%*
PAR unknown:  CDS specific                            31    10.4%*
              CDS unknown                             11     3.7%
PAR specific, CDS unknown                             46    15.5%
Illegible PAR photocopy                                1     0.3%
CDS/PAR values:  Agree or are consistent            119    40.1%
                 Disagree                             43    14.5%*

                      * Overall discrepancy rate     36.6%
```

Accident-Level Variables (N = 33):

First Harmful Event

    CDS/PAR values:  Agree or are consistent    32   97.0%
                       Disagree                  1    3.0%

Manner of Collision

    CDS/PAR values:  Agree                    29   87.9%
                       Disagree                  4   12.1%

Time

    CDS/PAR values:  Agree                    33  100.0%
                       Disagree                  0    0.0%

Month

    CDS/PAR values:  Agree                    33  100.0%
                       Disagree                  0    0.0%

Day of Week

    CDS/PAR values:  Agree                    33  100.0%
                       Disagree                  0    0.0%

Occupant Fatalities

    CDS/PAR values:  Agree                    31   93.9%
                       Disagree                  1    3.0%
                       Indeterminate            1    3.0%

NOTE:    See earlier note on page 16. Of the 33 accidents involving this PSU/PAR version, five involved fatalities. In three of those cases, the PAR and CDS values were in agreement. In a fourth case -- classified here as indeterminate -- the police reported one fatality while CDS reported none; the indeterminacy results from the fact that the individual involved was seated in a non-CDS relevant vehicle and thus does not appear in the CDS Occupant Assessment File (i.e., the accuracy of this particular observation cannot be confirmed). In the fifth case -- classified here as a disagreement -- CDS reported two fatalities while the police reported one; one of the deceased may have died after the police report was completed.

Vehicle-Level Variables (N = 65):

Speed Limit

```
    PAR entry missing, CDS specific            1      1.5%*
    PAR specific, CDS unknown                  3      4.6%
    CDS/PAR values:   Agree                   61     93.8%
                      Disagree                 0      0.0%*

                      * Overall discrepancy rate      1.5%
```

Make

```
    PAR entry missing, CDS unknown             1      1.5%
    CDS/PAR values:   Agree                   64     98.5%
                      Disagree                 0      0.0%*

                      * Overall discrepancy rate      0.0%
```

Model

```
    PAR entry missing, CDS unknown             2      3.1%
    PAR specific, CDS unknown                  3      4.6%
    CDS/PAR values:   Agree                   59     90.8%
                      Disagree                 1      1.5%*

                      * Overall discrepancy rate      1.5%
```

Model Year

```
    PAR entry missing, CDS unknown             2      3.1%
    CDS/PAR values:   Agree                   63     96.9%
                      Disagree                 0      0.0%*

                      * Overall discrepancy rate      0.0%
```

Body Type

```
    PAR entry missing, CDS specific            2      3.1%*
    PAR specific, CDS unknown                  1      1.5%
    CDS/PAR values:   Agree                   60     92.3%
                      Disagree                 2      3.1%*

                      * Overall discrepancy rate      6.2%
```

39

<u>Towed Due to Damage?</u>

| | | |
|---|---:|---:|
| PAR entry missing, CDS specific | 1 | 1.5%* |
| CDS/PAR values:  Agree | 62 | 95.4% |
| Disagree | 2 | 3.1%* |
| | * Overall discrepancy rate | 4.6% |

<u>Driver Alcohol/Drug Presence</u>

| | | |
|---|---:|---:|
| PAR entry missing, CDS specific | 1 | 1.5%* |
| PAR unknown:  CDS specific | 1 | 1.5%* |
| CDS unknown | 5 | 7.7% |
| CDS/PAR values:  Agree or are consistent | 57 | 87.7% |
| Disagree | 1 | 1.5%* |
| | * Overall discrepancy rate | 4.5% |

40

Occupant-Level Variables (N = 78):

## Age

| CDS/PAR values: | Agree or are consistent | 67 | 85.9% |
| | Disagree by one year | 6 | 7.7% |
| | Disagree by > one year | 5 | 6.4%* |

* Overall discrepancy rate     6.4%

NOTE:    In PSU 79B, drivers who are uninjured are not required to be listed in the section of the report in which certain occupant-specific variables appear. Of the 78 occupants in the sample, 21 were drivers who were not listed in that section and who were therefore presumably uninjured. In every case, the individual's Year of Birth (which appears elsewhere in the report) was checked against his or her CDS-reported AGE and the two values found to be consistent, i.e., the value of AGE plus the last two digits of Year of Birth equaled either 90 or 91. All such cases were treated as "consistent" with CDS.

## Sex

| PAR entry missing, CDS specific | | 1 | 1.3%* |
| CDS/PAR values: | Agree | 77 | 98.7% |
| | Disagree | 0 | 0.0%* |

* Overall discrepancy rate     1.3%

NOTE:    Unlike Age, Sex is reported for all occupants, injured or not.

## Injury Classification

| CDS/PAR values: | Agree or are consistent | 78 | 100.0% |
| | Disagree | 0 | 0.0% |

NOTE:    Injury Classification is another variable that appears in the section of the report that is normally not completed for uninjured drivers. In addition, since PSU 79B has no explicit code for "Not injured", the instructions state that this particular variable is to be omitted in the case of uninjured passengers, i.e., the absence of an entry is the only way to denote that a given passenger was uninjured. There were 42 such missing entries in all (21 involving drivers, 21 involving passengers); in each case, the corresponding CDS value was "No injury". All 42 cases were treated as consistent with CDS.

41

Transported to Medical Facility?

| | | |
|---|---|---|
| PAR specific, CDS unknown | 2 | 2.6% |
| CDS/PAR values: Agree or are consistent | 76 | 97.4% |
| Disagree | 0 | 0.0%* |

* Overall discrepancy rate  0.0%

NOTE:  As with Injury Classification, this variable is normally not reported for uninjured drivers or passengers. All persons for whom no entry was made were presumed to have been uninjured and therefore not in need of transportation to a medical facility. In each case, that interpretation was consistent with the CDS variable TREATMENT-MORTALITY.

Occupant Protection System Use

| | | |
|---|---|---|
| PAR entry missing, CDS specific | 1 | 1.3%* |
| PAR specific, CDS unknown | 32 | 41.0% |
| CDS/PAR values: Agree or are consistent | 37 | 47.4% |
| Disagree | 8 | 10.3%* |

* Overall discrepancy rate  11.6%

Air Bag Deployed?

As in PSU 13, the variable used to report air bag deployment is used to report on other safety equipment as well. One of its codes ("L") states that an air bag was deployed, another ("M") states that it was not. Code "M" never appeared in the sample for this PSU; Code "L" appeared only once and when it did, agreed with the CDS variable AIR BAG DEPLOYED. There were no cases in which CDS reported an air bag to have been deployed and the police did not.

Ejection

| | | |
|---|---|---|
| PAR entry missing, CDS unknown | 1 | 1.3% |
| PAR specific, CDS unknown | 1 | 1.3% |
| CDS/PAR values: Agree or are consistent | 76 | 97.4% |
| Disagree | 0 | 0.0%* |

* Overall discrepancy rate  0.0%

NOTE:  As with Age and Injury Classification, this variable is not required in the case of uninjured drivers. The 21 drivers for whom such occupant-specific information was omitted were assumed not to have been ejected. In each case, that interpretation coincided with the CDS variable EJECTION.

# 4. CONCLUSIONS AND RECOMMENDATIONS

This section summarizes, by variable, the findings presented in the preceding section. All of the target variables identified in the Statement of Work, with the exception of Damaged Area and Extent of Deformity, are treated on the pages that follow.[6,7]

Prior to reviewing this material, several comments seem in order:

a.  First, it should be noted that, as expected, not all variables display the same discrepancy rate. Some are more difficult to ferret out than others, some are more susceptible to misinterpretation. Vehicle Identification Number, for example, is a particularly troublesome variable. Those who rely on the hard copy PAR report as their source of information concerning VIN must recognize that one or more characters in the entered value may have been misread, transposed, or otherwise misentered by the police officer, or even if entered correctly, could easily be misread by the person extracting this information. A "Z" can easily be misread as a "2" (and vice versa); a "G" can easily be misread as a "6". The fact that there are ten such characters in all causes this particular variable to display a discrepancy rate, as herein defined, one full order of magnitude greater than most of the others.

b.  Second, some variables display significant variations in discrepancy rate across PSU/PAR versions. All variables were tested for such differences using standard chi-square measures. Where the differences among PSU/PAR versions were found to be significant at either the 5% ("significant") or 1% ("highly significant") level, the fact that such significance exists is highlighted in the material that follows and an explanation sought for the difference. In some cases, no explanation could be found; in others, the difference was considered to be attributable to one or more features inherent in either the design of the PAR form or the coding scheme and/or reporting instructions applicable to the variable in question. Observations of this nature are highlighted where relevant and

---

[6]  The basis for excluding the latter variables from the analysis that follows was previously explained on page 12. Both DAMAGED AREA and EXTENT OF DEFORMITY present certain complications that would hinder a direct comparison of the corresponding PAR and CDS values. Because of funding limitations, further pursuit of the precise statistical properties of these variables could not be justified.

[7]  To the right of each target variable, there appears in parentheses the corresponding CDS variable.

43

accompanied by recommendations where appropriate.

First Harmful Event          (CDS Variable:  OTHER VEHICLE NUMBER OR
                                             OBJECT CONTACTED)

The observed discrepancy rates for this variable ranged from 1.4 to
5.7 percent, as shown below:

            PSU 8         Variable not reported
            PSU 13        Variable not reported
            PSU 45             1.4%
            PSU 79A            5.7%
            PSU 79B            3.0%

The differences among PSU/PAR versions are not statistically
significant.

Manner of Collision          (CDS Variable:  MANNER OF COLLISION)

Observed discrepancy rates:

            PSU 8         Variable not reported
            PSU 13        Variable not reported
            PSU 45            27.5%
            PSU 79A           18.1%
            PSU 79B           12.1%

Although the differences among PSU/PAR versions are not
statistically significant, the values are sufficiently high and the
gap between PSUs 45 and 79B sufficiently pronounced that further
comment seems warranted.  In PSU 45, the vast majority of the
discrepancies noted involved single-car accidents for which the
police nonetheless entered a positive value ("Head-on", "Rear end",
etc.).  Forty of the 58 cases of disagreement in PSU 45 were of this
nature (see note, page 27).  In PSUs 79A and B, where single-vehicle
accidents were less common, the phenomenon was noted less often --
in nine of 18 cases in PSU 79A and in only one of four in PSU 79B.

One remedy to this problem, of course, would be to simply ignore the
value assigned by the police to this variable in single-vehicle
accidents.  If one were to subtract from the discrepancy rates noted
above the forty disagreements that were of this type in PSU 45, the
nine in PSU 79A, and the one in PSU 79B, the resulting discrepancy
rates would be strikingly similar:

            PSU 45             8.8%
            PSU 79A           10.7%
            PSU 79B            9.1%

Virtually all of these remaining discrepancies involved such matters
as one data source reporting the collision was "head-on" while the
other reported it was a "side-swipe".  To the extent that distinc-

44

tions of this nature are important in accident investigations, caution is advised with respect to the PAR version of this variable.

Time                                          (CDS Variable:   TIME)

Observed discrepancy rates:

    PSU 8          1.5%
    PSU 13         6.2%
    PSU 45         0.9%
    PSU 79A        2.5%
    PSU 79B        0.0%

The differences among PSU/PAR versions noted above are highly significant (p < .01), with PSU 13 the obvious outlier. While most jurisdictions experienced only a handful of disagreements (at most four) with respect to this variable, PSU 13 had a total of fourteen (see page 22). The fourteen disagreements noted were distributed as follows:

| Difference between CDS and PAR-reported Times | Number of Cases |
|---|---|
| 1 minute | 1 |
| 4 minutes | 1 |
| 5 minutes | 2 |
| 10 minutes | 1 |
| 20 minutes | 2 |
| 50 minutes | 1 |
| 2 hours | 1 |
| 10 hours | 3 |
| 12 hours | 2 |

No plausible explanation could be found for this phenomenon. There is no distinctive feature of the PSU 13 PAR form nor of the associated reporting instructions that would account for the materially higher discrepancy rate.

Month                                         (CDS Variable:   MONTH)

No discrepancies were observed with respect to this variable in any of the five jurisdictions.

Day of Week                                   (CDS Variable:   DAY OF WEEK)

Observed discrepancy rates:

    PSU 8               0.7%
    PSU 13             3.7%
    PSU 45             0.5%
    PSU 79A      Variable not reported
    PSU 79B            0.0%

45

The discrepancy rate for this variable is seen to be minuscule or non-existent for all PSUs but 13. The differences in rates among PSU/PAR versions are statistically significant ($p < .05$). Again, no plausible explanation could be found for this phenomenon.

It is worth noting that the source of these discrepancies may not lie in the variable Day of Week but rather in Accident Date, for if the latter variable were to be misreported, the CDS software system would automatically assign the wrong Day of Week to the accident. The computer-assigned value would then be at variance with the police-reported value. Further investigation of this issue may be warranted.

Number of Motor Vehicles        (CDS Variable:   NO. OF GENERAL VEHICLE
                                                 FORMS SUBMITTED)

This variable is reported only in PSUs 8, 13, and 45. In those PSUs, the only differences noted involved matters of definition (see comment, bottom of page 15). For all practical purposes, this data element may be regarded as reliably reported.

Occupant Fatalities             (CDS Variable:   TREATMENT-MORTALITY)

Observed discrepancy rates:

          PSU 8                     1.5%
          PSU 13          Variable not reported
          PSU 45                    1.4%
          PSU 79A                   0.8%
          PSU 79B                   3.0%

The differences among PSU/PAR versions noted above are not statistically significant. As noted earlier (see pages 16, 28, 33, and 38), any disagreement between the reported PAR and CDS values for this variable might simply be the result of a delayed fatality, one that took place after the PAR was completed.

Speed Limit                     (CDS Variable:   SPEED LIMIT)

Observed discrepancy rates:

          PSU 8                     8.6%
          PSU 13          Variable not reported
          PSU 45                    3.5%
          PSU 79A                   1.7%
          PSU 79B                   1.5%

The differences among PSU/PAR versions are in this case highly significant, with PSU 8 the obvious outlier. A possible contributing factor may be the manner in which speed limits are reported in PSU 8: instead of being vehicle-specific (as in the other PSUs as well as in CDS), they are roadway-specific, i.e., a

separate limit is reported for each roadway rather than for each vehicle. One must then access the accident diagram to determine the roadway on which each vehicle was traveling to determine the limit associated with that vehicle. This introduces a possibility for error that does not exist in the other jurisdictions. CDS researchers are specifically instructed <u>not</u> to use the police report as the basis for assigning a value to this variable.

Recommendation:      For consistency with other PSUs as well as with CDS, PSU 8 might wish to consider the possibility of converting to a vehicle-specific rather than roadway-specific format for reporting speed limit.

<u>Number of Occupants</u>      (CDS Variable:  NUMBER OF OCCUPANTS THIS VEHICLE)

Observed discrepancy rates:

| | |
|---|---|
| PSU 8 | Variable not reported |
| PSU 13 | 2.6% |
| PSU 45 | Variable not reported |
| PSU 79A | 6.9% |
| PSU 79B | Variable not reported |

Given the number of observations involved, the discrepancy rates for the two jurisdictions that report this variable do not differ significantly. Roughly half of the discrepancies -- seven out of twelve in PSU 13, six out of sixteen in PSU 79A -- involved PAR entries that were missing rather than explicitly in disagreement with CDS.

<u>Vehicle Identification Number</u>    (CDS Variable:  VEHICLE IDENTIFI-CATION NUMBER)

Observed discrepancy rates:

| | |
|---|---|
| PSU 8 | 17.8% |
| PSU 13 | 18.2% |
| PSU 45 | 30.3% |
| PSU 79A | Variable not reported |
| PSU 79B | Variable not reported |

The differences among PSU/PAR versions are in this case highly significant, with PSU 45 the obvious outlier. No reason could be found for the materially higher discrepancy rate in this jurisdiction.

The discrepancy rates for this variable are notably higher than they are for the others, primarily because of two factors:

    (a)   The greater number of opportunities for error. Effective

47

September 1980, all VINs must have seventeen characters. Although *only* the first ten are stored in the CDS General Vehicle File, an error in any of those ten locations constitutes a discrepancy.

(b)     The difficulty commonly encountered in distinguishing among handwritten alphanumeric characters.

To pursue the matter further, the COTR for this study, Dr. Carl Pierchala, investigated fifteen randomly selected cases of disagreement between the reported values for VIN in PSU 8, making use of software that identifies which characters in which positions are legal and which are not. He found that in all but one case, the PAR entry appeared to be in error, while in the fifteenth case, the source of the error could not be determined.[8] The fourteen errors confirmed by Dr. Pierchala to be PAR-related were as follows:

| Form of Error | Number of Cases |
|---|---|
| "5" substituted for "S" | 2 |
| "S" substituted for "5" | 2 |
| "6" substituted for "L" | 2 |
| "S" substituted for "2" | 1 |
| "3" substituted for "S" | 1 |
| "3" substituted for "B" | 1 |
| "B" substituted for "D" | 1 |
| "D" substituted for "O" | 1 |
| "O" substituted for "0" | 1 |
| "I" substituted for "1" | 1 |
| "GB" substituted for "BG" | 1 |

To reduce the uncertainty associated with this variable, police officers in PSU 13 are instructed to make the letters larger than the numbers. None of the other PSUs furnishes such an instruction.

Recommendation:     All PSUs should consider adopting the practice used in PSU 13, i.e., instructing the police officer to write the letters used to report VIN larger than the numbers.

Other measures for improving legibility -- for example, writing 7's and Z's with a horizontal slash, using a diagonal slash to distinguish the number "0" from the letter "O", etc. -- should be considered as well. Underscoring

---

[8]     The sole case of indeterminacy involved a foreign truck with a foreign vehicle identification number. Because the VIN was foreign, there was no way of knowing whether the character in a particular location should have been a "4" (as reported *by* the police) or a "Y" (as reported by CDS).

letters to distinguish them from numbers (or vice versa) is another possibility.

Make                                          (CDS Variable:   VEHICLE MAKE)

Observed discrepancy rates:

        PSU 8                       0.4%
        PSU 13                      1.4%
        PSU 45                      0.8%
        PSU 79A                     0.9%
        PSU 79B                     0.0%

These rates do not differ significantly by PSU/PAR version.

Model                                         (CDS Variable:   VEHICLE MODEL)

Observed discrepancy rates:

        PSU 8                         1.0%
        PSU 13              Variable not reported
        PSU 45                        3.5%
        PSU 79A                       3.9%
        PSU 79B                       1.5%

These rates, while higher than those for Make, also do not differ significantly. Typical discrepancies with respect to this variable were previously noted on pages 29 and 34.

Model Year                              (CDS Variable:   VEHICLE MODEL YEAR)

Observed discrepancy rates:

        PSU 8                         1.0%
        PSU 13                        1.1%
        PSU 45                        6.7%
        PSU 79A                       3.9%
        PSU 79B             Variable not reported

The differences among PSU/PAR versions are in this case highly significant. Whereas outright disagreements between the CDS and PAR-reported values for this variable were relatively rare in the other three jurisdictions -- only 13 disagreements in a total of 1,178 vehicles, a rate of slightly over one percent -- PSU 45 showed 22 disagreements in a total of only 373 vehicles, a rate of almost six percent. The PSU 45 disagreements were distributed as follows:

| Difference between CDS and PAR-Reported Value of Model Year | Number of Cases |
|---|---|
| 1 | 14 |
| 2 | 2 |

49

No explanation could be found for the materially higher rate of disagreement noted in PSU 45.

Body Type                              (CDS Variable:   VEHICLE BODY TYPE)

Observed discrepancy rates:

       PSU 8                    28.7%
       PSU 13                    6.4%
       PSU 45                    4.3%
       PSU 79A          Variable not reported
       PSU 79B                   6.2%

The differences among PSU/PAR versions are again highly significant. This time, PSU 8 is the obvious outlier.  The materially higher rate in that jurisdiction, however, is readily explained:  of the four PSU/PAR versions that report on Body Type, PSU 8 is the only one that provides a detailed breakdown for individual passenger cars (as does CDS).   Code 1, for example, denotes a convertible; Code 2 denotes a 2-door sedan, hardtop, or coupe; etc.  All of the other jurisdictions provide only a single code covering all passenger cars, regardless of body type.  Thus, there are opportunities for error in PSU 8 that do not exist in the other jurisdictions.

Taking this point a bit further:  of the 124 explicit disagreements between CDS and PAR involving Body Type in PSU 8, fully three-fourths (93) involved differences among passenger car types (calling a "hatchback" a "hardtop", calling a "two-door" a "three-door", etc.).  Disagreements of this nature would not be possible in the other jurisdictions.  If all such disagreements were to be removed from the PSU 8 discrepancy rate, the rate would drop to 9.2% and the differences among PSU/PAR versions would no longer be significant. The rate would nonetheless remain somewhat higher than some researchers might find acceptable.[9]

---

[9]    Two comments seem appropriate at this point.  First, it should be noted that no attempt was made in this report to define what is "acceptable" in the way of a discrepancy rate for a given variable; determinations of that nature would appear to lie within the province of the individual accident researcher and his or her stated intent in using police-reported data.

Second point:  the construct "discrepancy rate" is not always an acceptable proxy for "PAR data quality".  If, for example, the coding scheme for a given PSU/PAR version were to include only two codes for Body Type -- one defining "passenger car", the other defining "station wagon, van, or truck" -- the

Towed Due to Damage?        (CDS Variable:   POLICE-REPORTED VEHICLE
                                             DISPOSITION)

Observed discrepancy rates:

        PSU  8                3.3%
        PSU  13               4.4%
        PSU  45               1.8%
        PSU  79A              3.5%
        PSU  79B              4.6%

The differences among PSU/PAR versions are not statistically
significant.

Since the value reported for this variable by CDS is normally drawn
from the police report, one would not expect to encounter major
differences.  Of the 55 discrepancies noted in the total study
sample (i.e., across all jurisdictions combined), the majority (33)
involved situations in which there was no PAR entry at all.  The
remaining 22 involved situations in which the two sources explicitly
disagreed, with one source reporting the vehicle was towed while the
other said it was not.  Since the issue of whether or not a given
vehicle was towed may affect the composition of the CDS sample,[10]
the situation was studied further with the following results:

(a)     Of the 1,319 vehicles that CDS said were towed, the PAR agreed
        in 1,311 cases and disagreed in eight.  The conditional
        disagreement rate -- i.e., the percentage of cases in which the
        PAR said "No" when CDS said "Yes" -- was 0.6%.

(b)     Of the 297 vehicles that CDS said were not towed, the PAR
        agreed in 283 cases and disagreed in fourteen.  The conditional
        disagreement rate in this case was 4.7%, implying that
        disagreement is more likely when CDS indicates the vehicle had
        not been towed.

By and large, these 22 cases of disagreement had little impact on
the CDS sample.  In 21 of those cases, there was at least one other
vehicle that the PAR indicated had been towed, i.e., the accident
would have been qualified for the CDS sample in any event.  In only
one case, then, would differences between CDS and PAR with respect
to this variable have made a difference in the composition of the
CDS sample.  That one case involved a single-vehicle accident in
which the PAR indicated the vehicle had not been towed when CDS said

_____

        resulting discrepancy rate (explicit rate of disagreement with
        respect to CDS) might be remarkably low but little information
        of any value would be conveyed.

[10]    Only accidents in which at least one vehicle was towed qualify
        for inclusion in the sample.

it had.  The CDS researcher may in that case have used information
not contained in the original PAR report although there is always
the possibility, however remote, that the researcher may have been
in error and that this accident should not have been included.

Driver Alcohol/Drug Presence        (CDS Variables:   POLICE-REPORTED
                                                      ALCOHOL PRESENCE
                                                      and      POLICE-
                                                      REPORTED    OTHER
                                                      DRUG PRESENCE)

Observed discrepancy rates:

        PSU  8                 15.5%
        PSU 13                  3.8%
        PSU 45                  4.8%
        PSU 79A                 3.9%
        PSU 79B                 4.5%

The differences among PSU/PAR versions noted above are highly
significant, with PSU 8 the obvious outlier.  The higher discrepancy
rate in that jurisdiction may be related to the placement of this
particular data element on the PAR form.  In every other PSU/PAR
version, the data element used to report Driver Alcohol/Drug
Presence is co-located with a number of other data elements, all of
which are required to be completed.  In PSU 8, it is located by
itself in a remote section of the form that can easily be
overlooked.  Of the 74 discrepancies noted for this variable in PSU
8, 72 involved entries that were never made (i.e., were missing),
whereas the other four jurisdictions, with a substantially greater
total number of accidents, had a combined total of only 29 missing
entries.  If the number of missing entries in PSU 8 were to be
reduced to a level consistent with that of the others (roughly
2.5%), the resulting discrepancy rate for that jurisdiction would
drop to 2.9% and the observed differences among PSU/PAR versions
would no longer be significant.

Recommendation:        PSU 8 might wish to consider relocating the
                       data element on Driver Alcohol/Drug Presence to
                       a more prominent location.

A problem noted earlier in connection with PSU 13 warrants mention
at this point.  The PAR form for PSU 13 provides only a single data
element covering both alcohol and drugs but answerable only by a
simple "Yes" or "No".  There is no way, therefore, of distinguishing
between the driver's use of alcohol as opposed to drugs.  CDS
researchers are instructed to assume, in cases of this nature, that
it was alcohol that was present, rather than drugs.  To the extent
that the distinction between alcohol and drug use is important to
accident researchers, the following recommendation is made:

Recommendation:        PSU 13 should consider expanding the variable

52

on Alcohol/Drug Presence to permit alcohol and
drugs to be reported separately.

<u>Seating Position</u>                              (CDS Variable:   SEAT POSITION)

Observed discrepancy rates:[11]

        PSU  8                        1.8%
        PSU 13                        1.5%
        PSU 45                        2.2%
        PSU 79A                       1.3%
        PSU 79B                       1.3%

These differences are not statistically significant.  Apart from
occasional anomalies that complicate the task of linking PAR and
CDS-reported occupants (see table on page 11, also footnote 10 on
this page), Seating Position appears to be a generally reliably
reported data element.

<u>Age</u>                                          (CDS Variable:   AGE)

Observed discrepancy rates:

        PSU  8                        6.9%
        PSU 13                        9.1%
        PSU 45                        4.2%
        PSU 79A                       4.0%
        PSU 79B                       6.4%

The differences among PSU/PAR versions noted above are highly
significant, with PSU 13 the obvious outlier.   The higher
discrepancy rate in that jurisdiction is explained as follows:
unlike the others, PSU 13 does not require uninjured passengers to
be explicitly reported; the PAR instructions state that except for
Seat Position and Restraint, occupant-specific information on
uninjured passengers is "desirable" but not "mandatory".  There were
24 passengers in PSU 13 for whom no occupant-specific information
such as Age was reported; of these, 15 were persons who, according
to CDS, had not been injured.  The other nine cases were divided as
follows:  six were classified by CDS as having injuries of varying
severity, three were classified as "unknown".  If one were to delete
from the discrepancy rate the 15 passengers who sustained no injury
(and whose age therefore went unreported), the rate for PSU 13 would
decline to 6.9% and the differences among PSU/PAR versions would no
longer be significant.

---

[11]   The rates shown here are based on the table of seat position
       anomalies shown on page 11.  They were calculated by counting
       as discrepant the two rows in the table that are labeled
       "Differences in reported values" and "PAR missing".  All other
       anomalies (e.g., those involving multiple occupants in the same
       position or PAR-CDS differences in coding schemes) were not
       regarded as discrepancies.

Recommendation:     Since information on the characteristics of
                    persons who were uninjured in a given crash may
                    be of interest to at least some researchers,
                    PSU 13 might wish to reconsider its policy on
                    the reporting of uninjured passengers.

Sex                                                (CDS Variable:   SEX)

Observed discrepancy rates:

        PSU 8               2.0%
        PSU 13              5.1%
        PSU 45              2.8%
        PSU 79A             3.6%
        PSU 79B             1.3%

As in the case of Age and for the same reason, the differences among
PSU/PAR versions are statistically significant, with PSU 13 again
the obvious outlier.  If one were to delete from the discrepancy
rate the 15 uninjured and unreported passengers noted earlier, the
rate for PSU 13 would decline to 2.9% and the differences among
PSU/PAR versions would no longer be significant.

Recommendation:     Same as for Age.

Injury Classification          (CDS Variable:   INJURY SEVERITY)

Observed discrepancy rates:

        PSU 8               13.1%
        PSU 13              5.4%
        PSU 45              3.0%
        PSU 79A             4.0%
        PSU 79B             0.0%

The differences among PSU/PAR versions are highly significant.  The
higher discrepancy rate in PSU 8 results from the large number of
"unknowns" (44) reported by the police in that jurisdiction in cases
where CDS was quite specific; none of the other jurisdictions
reported even a single unknown.  Reason:  the PSU 8 classification
scheme includes an explicit code for "unknown"; the others do not.
The 44 occupants reported as having "unknown" injuries in PSU 8 were
divided as follows:  in one case, CDS reported the occupant had
sustained no injury; in the remaining 43, injuries of varying levels
were reported -- 32 were "possible", nine "non-incapacitating", and
two "incapacitating".  One is tempted to conclude that the "unknown"
code in PSU 8 was, in 1991, being overused.

Recommendation:     PSU 8 might wish to reconsider its use of an
                    "unknown" code for Injury.  Use of that code
                    may have inhibited, in at least some cases, use
                    of a more complete descriptor.

55

<u>Transported to Medical Facility?</u>          (CDS Variable:   TREATMENT-
                                                                 MORTALITY)

Observed discrepancy rates:

          PSU 8                 5.9%
          PSU 13               17.8%
          PSU 45                6.7%
          PSU 79A        Variable not reported
          PSU 79B               0.0%

The differences among PSU/PAR versions are highly significant, with
PSU 13 the obvious outlier.  PSU 13 had more missing entries for
this variable (111) than did all other jurisdictions combined (48).
Reason:  the PSU 13 PAR form does not provide, as the others do, a
separate entry for this variable as it pertains to each occupant;
instead, a single entry is provided for all occupants of the same
vehicle.  Perhaps because of this lack of specificity, the entry was
commonly omitted, leading to the high discrepancy rate noted above.

Recommendation:      PSU   13   should   consider   adopting   a   more
                     occupant-specific method for reporting this
                     variable.  Not all injured persons in the same
                     vehicle   are   necessarily   transported   to   a
                     medical   facility,   nor   are   those   that   are
                     transported   necessarily   taken   to   the   same
                     facility.

<u>Occupant Protection System Use</u>     (CDS Variable:   MANUAL   [ACTIVE]
                                                          BELT SYSTEM USE)

Observed discrepancy rates:

          PSU 8               37.4%
          PSU 13              19.3%
          PSU 45              19.6%
          PSU 79A             36.6%
          PSU 79B             11.6%

The  differences  among  PSU/PAR  versions  noted  above  are  highly
significant and the rates for all five versions surprisingly high.
While there were many different ways in which the CDS and PAR values
for each version could (and did) differ, a large percentage of the
disagreements in each jurisdiction were accounted for by only a
single combination of CDS and PAR values.  By PSU/PAR version, the
forms of disagreement that were most commonly observed were as
follows:

| PSU | Total Number of Disagreements | Form of Disagreement Most Commonly Observed | | Number of Times This Took Place |
| --- | --- | --- | --- | --- |
| | | PAR Value(s) | CDS Value | |
| 8 | 118 | ACTIVE RESTRAINT TYPE equals "Seat belt and Shoulder Harness"; ACTIVE RESTRAINT USAGE equals "Used". | MANUAL BELT SYSTEM USE ("MANUSE") equals 00.* | 67 (57%) |
| 13 | 87 | RESTRAINT BY OCCUPANT POSITION equals B ("Belt used"). | MANUSE = 00 | 81 (93%) |
| 45 | 86 | SEAT BELT - YES OR NO is "Yes". | MANUSE = 00 | 74 (86%) |
| 79A | 43 | RESTRAINT SYSTEM equals "Combination"; AVAILABILITY is "Yes"; USE is "Yes". | MANUSE = 00 | 20 (47%) |
| 79B | 8 | SAFETY EQUIPMENT equals "Lap/shoulder harness used". | MANUSE = 00 | 7 (88%) |

\* MANUSE = 00 is defined as "None used, not available, or removed/destroyed".

The fact that the disagreements involving this variable cluster so commonly in the manner described is a matter of concern; no explanation could be found for this phenomenon. CDS researchers are specifically instructed not to use the police report as their source of information for this variable.

Recommendation:     Given the importance of Occupant Protection System Use to automotive accident research, strenuous effort must be devoted to improving the accuracy with which this variable is reported by the police before police accident reports can be counted on as fully reliable in this regard.[12]

---

[12] This recommendation of course presupposes that in cases where the CDS researcher and the police accident report differ, it is the researcher who is correct. Given the greater number of sources available to the researcher (vehicle inspection is the primary source, supplemented by both an interview and examination of medical records), this would appear to be a reasonable supposition.

<u>Air Bag Deployed?</u>          (CDS Variable:   AIR BAG SYSTEM DEPLOYMENT)

Observed discrepancy rates:

          PSU 8                    19.1%
          PSU 13       Reported for only two occupants (see page 26).
          PSU 45          Variable not reported
          PSU 79A          Variable not reported
          PSU 79B     Reported for only one occupant (see page 42).

No useful comparison among PSU/PAR versions is possible.  For PSU 8, the high discrepancy rate resulted from the large percentage of cases (16.1%, see page 20) in which the police reported this variable to be "unknown".  The recommendation presented above for Occupant Protection System Use applies here as well.  In addition, the following recommendation is made with specific reference to Air Bag Deployment:

Recommendation:          Questions raised in the PAR form concerning the availability and deployment of air bags should be separated (as in PSU 8) from those pertaining to active restraints, to permit both sets of issues to be independently explored.


<u>Ejection</u>                              (CDS Variable:   EJECTION)

Observed discrepancy rates:

          PSU 8                     3.2%
          PSU 13       Variable not reported
          PSU 45                    5.0%
          PSU 79A      Variable not reported
          PSU 79B                   0.0%

The differences in discrepancy rates among PSU/PAR versions are not statistically significant.  A fundamental difference was noted, however, in the <u>types</u> of discrepancies observed in PSUs 8 and 45. The fourteen discrepancies on which the PSU 8 rate is based (see page 21) were mainly matters of omission, no entry having been made in thirteen of those cases.  (Indeed, the fourteenth case, although nominally a disagreement, was really a clerical error, the officer having placed the entry in the wrong box.)  In contrast, over half (15) of the twenty-five discrepancies in PSU 45 (see page 32) involved matters of outright disagreement.  In three of those cases, the PAR reported ejection while CDS reported none; in twelve cases, the PAR reported <u>no</u> ejection while CDS reported that either a partial or a complete ejection had taken place.  No plausible explanation could be found for these disagreements.

## Summary

By way of recapitulation, the following table shows, for each target variable, (a) the number of PSU/PAR versions for which that variable was reported in 1991, (b) the range of discrepancy rates observed, and (c) the median discrepancy rate across PSU/PAR versions.

| | Number of PSU/PAR Versions | Observed Range of Discrepancy Rates Min (%) | Max (%) | Median Discrepancy Rate (%) |
|---|---|---|---|---|
| First Harmful Event | 3 | 1.4 | 5.7 | 3.0 |
| Manner of Collision | 3 | 12.1 | 27.5 | 18.1 |
| Time | 5 | 0.0 | 6.2 | 1.5 |
| Month | 5 | 0.0 | 0.0 | 0.0 |
| Day of Week | 4 | 0.0 | 3.7 | 0.6 |
| No. of Motor Vehicles | 3 | 0.0 | 0.0 | 0.0 |
| Occupant Fatalities | 4 | 0.8 | 3.0 | 1.4 |
| Speed Limit | 4 | 1.5 | 8.6 | 2.6 |
| Number of Occupants | 2 | 2.6 | 6.9 | 4.7 |
| VIN | 3 | 17.8 | 30.3 | 18.2 |
| Make | 5 | 0.0 | 1.4 | 0.8 |
| Model | 4 | 1.0 | 3.9 | 2.5 |
| Model Year | 4 | 1.0 | 6.7 | 2.5 |
| Body Type | 4 | 4.3 | 28.7 | 6.3 |
| Towed Due to Damage? | 5 | 1.8 | 4.6 | 3.5 |
| Driver Alcohol/Drugs | 5 | 3.8 | 15.5 | 4.5 |
| Seating Position | 5 | 1.3 | 2.2 | 1.5 |
| Age | 5 | 4.0 | 9.1 | 6.4 |
| Sex | 5 | 1.3 | 5.1 | 2.8 |
| Injury Classification | 5 | 0.0 | 13.1 | 4.0 |
| Transp. to Med. Fac.? | 4 | 0.0 | 17.8 | 6.3 |
| Occupant Protection | 5 | 11.6 | 37.4 | 19.6 |
| Air Bag Deployed? | 1 | 19.1 | 19.1 | 19.1 |
| Ejection | 3 | 0.0 | 5.0 | 3.2 |

The most universally troublesome variables are seen to be those pertaining to Manner of Collision, VIN, Occupant Protection, and Air Bag Deployment. Strenuous efforts to improve the accuracy with which these variables are reported by the police are strongly recommended. Other variables -- Speed Limit, Body Type, Driver Alcohol/Drug Presence, Injury Classification, and Transportation to Medical Facility -- present problems in some jurisdictions but not in all.

APPENDIX A.

PAR VARIABLES, ASSOCIATED DATA ELEMENTS,
AND DATA ENTRY PROCEDURES, BY PSU

PAR VARIABLES, ASSOCIATED DATA ELEMENTS,
AND DATA ENTRY PROCEDURES, BY PSU

The statement of work for this project lists some two dozen target variables whose accuracy and completeness as reported on police accident reports (PARs) are to be studied.

As noted in the body of the report, however, not all police jurisdictions report on all target variables.  Some variables may not be reported at all; others may be reported in a form that differs somewhat not only from that of the other jurisdictions but also from that applicable to the same or similar variable in the National Highway Traffic Safety Administration's Crashworthiness Data System (CDS).

The pages that follow identify, for each of the five PSU/PAR versions included in this study (8, 13, 45, 79A, and 79B), which variables are reported in that particular jurisdiction in 1991 and which are not.  Also included is any special information that would help the reader understand and interpret both (i) the PAR reporting process and coding scheme employed, and (ii) the data entry process applied by Applied System Technologies, Inc. in creating a computerized file for comparison against the CDS file.*

Special terms and symbols that appear in this appendix are defined below:

| | |
|---|---|
| DEO | Acronym for Data Entry Operator. |
| -1 | Denotes a PAR entry that is missing, i.e., no entry was made at all. |
| -2 | Denotes a PAR entry that is ambiguous or otherwise difficult to interpret. |
| -3 | Denotes a PAR entry that is out-of-range or otherwise unanticipated. |

---

\* For information on the CDS counterparts to the variables described herein, the reader is referred to U.S. Department of Transportation/National Highway Traffic Safety Administration/ National Center for Statistics and Analysis publication, National Accident Sampling System:  1991 Crashworthiness Data System, Data Collection, Coding, and Editing Manual, January 1991.  A descriptive comparison of the two sets of variables is contained in the Applied System Technologies, Inc. publication, "Comparison of Coding Systems for all Target Variables", June 1994, originally cited on page 3.

A-1

## A. Accident-Level Variables:

### First Harmful Event

Not reported.

### Manner of Collision

Not reported.

### Time

Military (0000-2400). If reported as civilian, data entry operator (DEO) makes the necessary conversion.

### Month

May be alpha or numeric. If alpha, DEO converts to numeric.

### Day of Week

Alpha. DEO enters first three letters.

### Number of Motor Vehicles

DEO enters the value reported. Value includes both motorized and special vehicles. Does not include phantom vehicles, i.e., vehicles that may have contributed to the accident but were not involved in the collision.

### Occupant Fatalities

DEO enters the value reported. Value may include pedestrians. Fatalities are to be reported if they occur within 90 days of the accident.

## B. Vehicle-Level Variables:

### Speed Limit

Unlike other PSUs, this is an accident-, rather than vehicle-specific variable. Speed limits are reported for both the principal and intersecting roads. DEO is instructed to examine the accident diagram to determine which limit applies to which vehicle.

### Number of Occupants

Not reported.

## Vehicle Identification Number

DEO enters first ten characters. No special instructions to ensure clarity are provided in the state Police Accident Report Manual.

## Make and Model

Reported in alpha form. DEO does not enter. Will be separately encoded and added later.

## Year

DEO enters last 2 digits.

## Body Type

DEO enters the value reported. Codes 20-29 (motorcyles), 30-39 (buses), 70-79 (medium/heavy trucks), 80-89 (specialized motorized vehicles such as snowmobiles, farm equipment, etc.), 90-94 (non-motorized vehicles), and 95-96 (trains and trolleys) are non-CDS applicable. 98 is "other" and 99 is "unknown".

## Damaged Area

Not explicitly reported as such. Data element 64T (Initial Point of Contact), consisting of the twelve clock points plus a few others, is used instead as a surrogate. DEO enters the value reported.

## Extent of Deformity

Data element 13 (Vehicle Damage) conveys this information. Available values are None, Light, Moderate, and Severe. Some PAR forms code these entries 0 through 3; others code them 1 through 4. DEO enters the value reported, converting the former coding scheme, if used, to the latter.

## Towed Due to Damage?

Reported as Y or N. DEO enters the value reported.

## Driver Alcohol/Drug Presence

Available codes include 0-None, 1-Alcohol, 2-Controlled Substances, 3-Other Drugs, 4-Both Alcohol and Drugs, and 9-Unknown. DEO enters the value reported.

## C. Occupant-Level Variables (reported for all occupants including the driver):

<u>Seating Position</u>

Reported as 1 through 6, plus 8 ("other") and 9 ("unknown").
DEO enters the value reported.

<u>Age</u>

Values of zero are not allowed; all infants under the age of
2 are coded "1". Persons 98 or older are coded "98". "99" is
"unknown". DEO enters the value reported.

<u>Sex</u>

Reported as M, F, or U. DEO enters the value reported.

<u>Injury Classification</u>

Available codes include 0-None, 1-Death, 2-Major Injury, 3-
Moderate Injury, 4-Minor Injury, and 5-Unknown Injury. DEO
enters the value reported.

<u>Transported to Medical Facility?</u>

Can be inferred from data element 63M (Injury Transportation)
which identifies the <u>type</u> of vehicle used to transport injured
persons. Code 0 ("not applicable") denotes the person was not
transported. All other codes denote that transportation took
place. No recode is performed as part of the data entry
operation, however; DEO simply enters the value reported.

<u>Occupant Protection System Use</u>

Two data elements are needed to define this variable. 63E
identifies the type of active restraint with which the vehicle
is equipped (only one entry permitted per occupant); 63F
identifies whether the restraint in question was used or not.
DEO enters both values; they will later be recoded to produce
a single value for comparison against the CDS file.

<u>Air Bag Deployed?</u>

Implicit in data element 63G (Passive Restraint Type). Code
0 denotes that there was no airbag present; code 1 denotes PSU
that an airbag was present and deployed; code 2 denotes that
an airbag was present but not deployed. DEO enters the value
reported, whether it applies to airbags or not.

<u>Ejection/Extrication</u>

Relevant data element is 63L.  Codes 1 through 3 identify situations in which the person was ejected, either totally or partially.  Codes 3 (in which the person was partially ejected, then extricated) through 8 identify situations in which the person was extricated.  DEO enters the value reported.

## A. Accident-Level Variables:

First Harmful Event

Not reported.

Manner of Collision

Not reported.

Time

Civilian (AM/PM). If reported as military, DEO makes the necessary conversion.

Month

May be alpha or numeric. If alpha, DEO converts to numeric.

Day of Week

Reported by circling the correct initial letter (S, M, T, etc.). DEO converts to numeric (1-7).

Number of Motor Vehicles

PAR instructions provide no guidance as to the types of vehicles to be included or excluded, stating simply "Enter the total number of vehicles physically involved in the accident". DEO enters the value reported.

Occupant Fatalities

Not reported.

## B. Vehicle-Level Variables:

Speed Limit

Not reported.

Number of Occupants

Includes the driver and all passengers, both injured and uninjured. DEO enters the value reported.

Vehicle Identification Number

DEO enters first ten characters. PAR instructions are to make letters larger than numbers.

### Make and Model

Make is encoded (00-72); DEO enters the value reported. Model is not reported.

### Year

DEO enters last 2 digits.

### Body Type

Available codes are 1-Passenger car (including station wagon), 4-Van (passenger or utility), 5-Four wheel drive utility vehicle (Jeep, Blazer, etc.), 6-Pickup, 7-Truck, 8-Truck tractor, and 9-Other or Unknown. DEO enters the value reported.

### Damaged Area

Not explicitly reported as such. DEO enters the data element "Impact" instead, consisting of eight points around the perimeter of the vehicle plus a few others. The intent is to use this variable, as in PSU 8, as a surrogate.

### Extent of Deformity

The data element "Severity" conveys this information. Available codes are the 7-point Vehicle Damage Severity Scale developed by the National Safety Council. Code "0" denotes no damage. DEO enters the value reported.

### Towed Due to Damage?

DEO infers whether or not a given vehicle was towed based on information reported in the box "Vehicle Removed To/By". An entry recognizable as the name of a towing company or garage is coded Y; an entry such as a dash or "N/A" which indicates that towing was not required is coded N; the absence of an entry is coded -1.

### Driver Alcohol/Drug Presence

DEO enters Y or N, depending on which of these values was circled in the box labeled "HBD". "HBD" stands for "Had Been Drinking" but is defined to include controlled substances as well.

**C.  Occupant-Level Variables (reported for all occupants including the driver, with the exception that uninjured passengers need not be separately reported):**

Seating Position

Reported as 1 through 6, plus 7 ("other") and 9 ("unknown"). DEO enters the value reported.

Age

As of last birthday.  DEO enters the value reported.

Sex

M or F.  DEO enters the value reported.

Injury Classification

Available codes include K-Fatal Injury, A-Incapacitating Injury, B-Nonincapacitating Injury, C-Possible Injury (complaint of pain or momentary unconsciousness), and O-No Injury.  DEO enters the value reported.

Transported to Medical Facility?

Although this is in theory an occupant-specific variable, the PAR form permits only one entry per vehicle.  It is treated in this PSU, therefore, as vehicle-specific and its value is assigned to all occupants in the car, injured or not.  The value assigned is determined by the DEO based on information reported in the box "Injured Taken To/By".  An entry recognizable as the name of a medical facility is coded Y; an entry such as a dash or "N/A" which indicates that transportation was not required is coded N; the absence of an entry is coded -1.

Occupant Protection System Use

Available codes include A-No belt available, B-Belt used, C-Belt not used, D-Air bag, activated or non-activated, E-Child restraint device used and properly secured, F-Child restraint device not available, not used, or not properly secured, G-restraint failure (of any type), and H-Restraint use unknown. DEO enters the value reported.

Air Bag Deployed?

Information on the deployment of air bags is available, somewhat ambiguously and in a small percentage of cases, *from*

the preceding data element.   Code D states that an air bag was present but not whether it was deployed.   Code G could denote an airbag failure but there is no way of knowing for certain which device it was that failed.   No additional DEO entry is called for beyond that required to enter the preceding data element.

Ejection/Extrication

Not reported.

## A. Accident-Level Variables:

### First Harmful Event

PAR form defines ten possible roadway events (collision with other motor vehicle, collision with pedestrian, etc.) and two non-roadway events (ran off roadway and overturned or struck fixed object). "Other" is another possibility. DEO enters the value reported.

### Manner of Collision

Available codes include 1-Headon, 2-Rearend, 3-Angle, 4-Sideswipe, same direction, 5-Sideswipe, opposite direction, and 7-Other. DEO enters the value reported.

### Time

Civilian (AM/PM). If reported as military, DEO makes the necessary conversion.

### Month

Ordinarily entered as numeric. If alpha, DEO converts to numeric.

### Day of Week

PAR form has a separate box for each day of the week. Boxes are numbered 1 through 7. DEO enters the value reported.

### Number of Motor Vehicles

This number may include noncontact vehicles, defined as vehicles that "did not make contact with an object, another vehicle, and/or a pedestrian". DEO enters the value reported.

### Occupant Fatalities

DEO enters the value reported. Fatalities are to be reported if they occur within 30 days of the accident.

## B. Vehicle-Level Variables:

### Speed Limit

Posted speed limit is reported for each vehicle in the accident. DEO enters the value reported.

## Number of Occupants

Not reported.

## Vehicle Identification Number

DEO enters first ten characters. No special instructions to ensure clarity are provided.

## Make and Model

Reported in alpha form. DEO does not enter. Will be separately encoded and added later.

## Year

DEO enters last 2 digits.

## Body Type

DEO enters the value reported. Codes 30-48 (medium/heavy trucks), 50-58 (motorized cycles), 60-68 (buses), 70-78 (non-motorized vehicles such as bicycles), and 80-88 (special vehicles such as farm tractors, construction equipment, etc.) are non-CDS applicable. Code 99 is "unknown".

## Damaged Area

Not explicitly reported as such. The data element "Point of Initial Contact", consisting of eight points around the perimeter of the vehicle plus a few others, is used instead as a surrogate. DEO enters the value reported.

## Extent of Deformity

The data element "Officer's Estimated Amount of Damage" is used to convey this information. Most PAR forms have three checkoff boxes (Under $200, $200-$500, and Over $500) but some have only two. In the former case, DEO enters the value that was reported; in the latter case, he or she enters -2.

## Towed Due to Damage?

Reported as Y or N. DEO enters the value reported.

## Driver Alcohol/Drug Presence

The relevant PAR data element, "Condition of Driver or Pedestrian", addresses other issues besides alcohol and drugs. Of the eleven possible checkboxes, the only ones applicable to

alcohol or drugs are 2-Had not been drinking, 3-Had been drinking, 9-Unknown if drinking, and 10-Apparently Drugged. If only one box was checked, DEO enters that value, whether it pertains to alcohol or drugs or not. If more than one box was checked, DEO reports the box, if any, that pertains to alcohol or drugs; otherwise, he or she enters -1 (missing).

**C. Occupant-Level Variables (reported for all occupants including the driver):**

Seating Position

Reported as two digits. First digit distinguishes between front seat (1), second seat (2), and third seat (3). Second digit distinguishes between left (1), center (2), and right (3). Codes 41 through 49 identify other positions (e.g., sleeper section of cab or truck). Codes 51 through 82 are reserved for motorcycles, buses, and other non-CDS applicable vehicles. DEO enters the value reported.

Age

DEO enters the value reported.

Sex

Reported as M or F. DEO enters the value reported.

Injury Classification

Available codes are 0-No Injury, 1-Possible Injury, 2-Nonincapacitating Injury, 3-Incapacitating Injury, and 4-Fatal Injury. DEO enters the value reported.

Transported to Medical Facility?

Inferred by the DEO from the entry (or lack of entry) in the space labeled "Taken to" that is provided for all occupants. An entry recognizable as the name of a medical facility is coded Y; an entry such as a dash or "N/A" which indicates that transportation was not required is coded N; the absence of an entry is coded -1.

Occupant Protection System Use

For adults, the only reference to occupant protection is the data element "Seat Belt". DEO enters the value reported (Y or N). For children under the age of four, there is an additional data element, "Child Restraint Device". It includes three subelements, however -- Available, Used, and

Used Properly.  Rather than complicate the data entry, DEO is
instructed to ignore this additional data element, which
applies in only a handful of cases.  It will be separately
encoded and added later.

<u>Air Bag Deployed?</u>

Not reported.

<u>Ejection/Extrication</u>

Ejection is covered by the data element "Ejected" (Yes or No).
DEO enters the value reported.  There is no separate data
element for extrication.

## A.  Accident-Level Variables:

### First Harmful Event

PAR form defines ten possible events (non-collision, collision with other motor vehicle, collision with pedestrian, etc.) plus "other".  DEO enters the value reported.

### Manner of Collision

Available codes include A-Headon, B-Sideswipe, C-Rearend, D-Broadside, E-Hit object, F-Overturned, G-Struck pedestrian, and H-Other.  DEO enters the value reported.

### Time

Military (0000-2400).  If reported as civilian, DEO makes the necessary conversion.

### Month

If reported as alpha, DEO converts to numeric.

### Day of Week

Not reported.

### Number of Motor Vehicles

Not reported.

### Occupant Fatalities

DEO enters the value reported.

## B.  Vehicle-Level Variables:

### Speed Limit

Posted speed limit is reported for each vehicle in the accident.  DEO enters the value reported.

### Number of Occupants

Includes the driver and all passengers.  DEO enters the value reported.

### Vehicle Identification Number

Not reported.

## Make and Model

Reported in alpha form.  DEO does not enter.  Will be separately encoded and added later.

## Year

DEO enters last 2 digits.

## Body Type

Although the PAR form has a space labeled "Type", no special instruction or coding scheme is provided.  DEO is instructed not to enter this information.

## Damaged Area

Although there is a space for identifying the location of the damage, no special instruction or coding scheme is provided. DEO is instructed not to enter this information.

## Extent of Deformity

The data element "Vehicle Damage" is the relevant source of information.  Available codes are Minor, Moderate, Major, and Total.  DEO enters the value reported.  If nothing is checked, DEO enters the letter O.

## Towed Due to Damage?

DEO infers whether or not a given vehicle was towed based on information reported in the box "Disposition of Vehicle".  An entry recognizable as the name of a towing company or garage is coded Y; an entry such as a dash, "N/A", or other indication that towing was not required is coded N; the absence of an entry is coded -1.

## Driver Alcohol/Drug Presence

The relevant data element, "Sobriety-Drug-Physical", addresses other issues besides alcohol and drugs.  Of the eight possible checkboxes, five are applicable to alcohol or drugs:  A-Had not been drinking; B-HBD, under influence; C-HBD, not under influence; D-HBD, impairment unknown; and E-Under drug influence.  Since it is possible for more than one box to be checked, DEOs are instructed to proceed as follows:

-- If only one box was checked, DEO enters that value, even if it does not pertain to alcohol or drugs.

-- If more than one box was checked, DEO reports the box, if any, that pertains to alcohol or drugs. If none of the boxes that were checked pertains to alcohol or drugs, DEO enters -1.

## C. Occupant-Level Variables (reported for all occupants including the driver):

### Seating Position

Reported in alpha. DEO is instructed to code as follows: LF/CF/RF for front seat positions, LR/CR/RR for rear seat positions, O for other positions. If only one passenger was in the front seat, the applicable code is RF. If the left/center/right designation is omitted for rear seat passengers, the applicable code is XR.

### Age

As of last birthday. DEO enters the value reported. Exception: age followed by a question mark denotes that the person's age was estimated, in which case the DEO is instructed to enter -2 (ambiguous).

### Sex

M or F. DEO enters the value reported.

### Injury Classification

Available codes include K-Fatal Injury, A-Incapacitating Injury, B-Nonincapacitating Injury, and C-Possible Injury (complaint of pain or momentary unconsciousness). DEO enters the value reported, if any. Other situations are handled as follows: a zero, dash, slash, or other indication that the person was uninjured is entered as the letter O; the absence of any entry at all receives a -1 (missing).

### Transported to Medical Facility?

Not reported.

### Occupant Protection System Use

The relevant data element, "Restraint System", includes several subelements and is difficult both to interpret and apply. Rather than complicate the data entry, DEO is instructed to ignore this variable. It will be separately encoded and added later.

<u>Air Bag Deployed?</u>

Not reported.

<u>Ejection/Extrication</u>

Not reported.

## A. Accident-Level Variables:

### First Harmful Event

PAR form defines ten possible events (non-collision, collision with other motor vehicle, collision with pedestrian, etc.) plus "other". DEO enters the value reported.

### Manner of Collision

Available codes include A-Headon, B-Sideswipe, C-Rearend, D-Broadside, E-Hit object, F-Overturned, G-Struck pedestrian, and H-Other. DEO enters the value reported.

### Time

Military (0000-2400). If reported as civilian, DEO makes the necessary conversion.

### Month

If reported as alpha, DEO converts to numeric.

### Day of Week

Reported by circling the correct initial letter (S, M, T, etc.). DEO converts to numeric (1-7).

### Number of Motor Vehicles

Not reported.

### Occupant Fatalities

DEO enters the value reported.

## B. Vehicle-Level Variables:

### Speed Limit

Posted speed limit is reported for each vehicle in the accident. DEO enters the value reported.

### Number of Occupants

Not reported.

### Vehicle Identification Number

Not reported.

## Make and Model

Reported in alpha form. DEO does not enter. Will be separately encoded and added later.

## Year

DEO enters last 2 digits.

## Body Type

The relevant data element is "Vehicle Type". DEO enters the value reported in the left-hand box (the right-hand box applies only to towed vehicles). The only codes that pertain to CDS-applicable vehicles are 01 (passenger car, station wagon, or Jeep), 22 (pickups and panels), and 48 (police cars). Code 99 is "unknown". DEO enters the value reported, whether it is CDS-applicable or not.

## Damaged Area

Although there is a space on the form for shading in the damaged area, no special instruction or coding scheme is provided. DEO is instructed not to enter this information.

## Extent of Deformity

The data element "Vehicle Damage" is the relevant source of information on this subject. Available codes are Unknown, None, Minor, Moderate, Major, and Total. DEO enters the value reported.

## Towed Due to Damage?

DEO infers whether or not a given vehicle was towed based on information reported in the box "Disposition of Vehicle". An entry recognizable as the name of a towing company or garage is coded Y; an entry such as a dash, "N/A", or other indication that towing was not required is coded N; the absence of an entry is coded -1.

## Driver Alcohol/Drug Presence

The relevant data element, "Sobriety-Drug-Physical", addresses other issues besides alcohol and drugs. Of the nine possible checkboxes, five are applicable to alcohol or drugs: A-Had not been drinking; B-HBD, under influence; C-HBD, not under influence; D-HBD, impairment unknown; and E-Under drug influence. If only one box was checked, DEO enters that value, whether it pertains to alcohol or drugs or not. If more than

one box was checked, DEO reports the box, if any, that pertains to alcohol or drugs; otherwise, he or she enters -1 (missing).

C. **Occupant-Level Variables (reported for all occupants including the driver, with the exception that if the driver is uninjured, the only items reported are Sex and Safety Equipment):**

Seating Position

Reported as 1 through 6, plus 7 ("station wagon rear"), 8 ("truck or van rear"), 9 ("unknown"), and 10 ("other"). DEO enters the value reported.

Age

DEO enters the value reported. Exception: age followed by a question mark denotes that the person's age was estimated, in which case the DEO is instructed to enter -2.

Sex

M or F. DEO enters the value reported.

Injury Classification

Available values for this variable, reading from left to right, are Fatal Injury, Severe Injury, Other Visible Injury, and Complaint of Pain. DEO converts these entries to numeric code (Fatal = 1, ... , Complaint of Pain = 4). If nothing is checked, DEO enters -1.

Transported to Medical Facility?

Inferred by the DEO from the entry (or lack of entry) in the space labeled "(Injured Only) Transported By" on Page 3. An entry recognizable as the name of a medical facility is coded Y; an entry such as a dash or "N/A" which indicates that transportation was not required is coded N; the absence of an entry is coded -1.

Occupant Protection System Use

The relevant data element, "Safety Equipment", includes over twenty codes, not all of them mutually exclusive. In those cases where multiple codes apply (e.g., Lap Belt Used and Shoulder Harness Not Used), the officer is instructed to enter only a single code. No instruction is provided as to priorities, however, with the following exception: if an airbag was deployed, the corresponding code (L) receives precedence. In all cases, DEO enters the value reported.

## Air Bag Deployed?

Information on air bag deployment is available, in a small percentage of cases, from the preceding data element. Code L states that an air bag was deployed; code M states that it was not. No additional DEO entry is called for beyond that required to enter the preceding data element.

## Ejection/Extrication

Relevant data element is "Ejected". Available codes are 0-Not ejected, 1-Fully ejected, 2-Partially ejected, and 3-Unknown. DEO enters the value reported. There is no separate data element for extrication.

.

U.S. Department
of Transportation

**National Highway
Traffic Safety
Administration**

400 Seventh St., S.W.
Washington, D.C. 20590

Official Business
Penalty for Private Use $300