HSRI-SA-73-4

STATISTICAL INFERENCE FROM MULTIDISCIPLINARY ACCIDENT INVESTIGATION

JAMES O'DAY
WILLIAM L. CARLSON
RICHARD DOUGLASS
RICHARD J. KAPLAN

HIGHWAY SAFETY RESEARCH INSTITUTE
THE UNIVERSITY OF MICHIGAN
ANN ARBOR, MICHIGAN 48104

JUNE 1973
CONTRACT NO. DOT-HS-031-2-350
FINAL REPORT

PREPARED FOR:

DEPARTMENT OF TRANSPORTATION
NATIONAL HIGHWAY TRAFFIC SAFETY ADMINISTRATION
WASHINGTON, D.C. 20590

TECHNICAL REPORT STANDARD TITLE PAGE

1. Report No. DOT HS-801 111	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle STATISTICAL INFERENCE FROM MULTIDISCIPLINARY ACCIDENT INVESTIGATION.		5. Report Date August 1974 6. Performing Organization Code
7. Author(s) JAMES O'DAY, WILLIAM L. CARLSON, RICHARD DOUGLASS AND RICHARD J. KAPLAN		8. Performing Organization Report No. UM-HSRI-SA-73-4
9. Performing Organization Name and Address HIGHWAY SAFETY RESEARCH INSTITUTE UNIVERSITY OF MICHIGAN HURON PARKWAY & BAXTER ROAD ANN ARBOR, MICHIGAN 48105		10. Work Unit No 11. Contract or Grant No. DOT-HS-031-2-350
12. Sponsoring Agency Name and Address U.S. Department of Transportation NATIONAL HIGHWAY TRAFFIC SAFETY ADMINISTRATION WASHINGTON, D.C. 20590		13. Type of Report and Period Covered FINAL REPORT May 1972 - June 1973 14. Sponsoring Agency Code

15. Supplementary Notes

The opinions, findings and conclusions expressed in this publication are those of the author and not necessarily those of the National Highway Traffic Safety Administration.

16. Abstract

Multi-disciplinary accident investigation (MDAI) reports generated by more than twenty contractors have been compiled in digital form for analysis. Methods for drawing inferences regarding the relationship between crash severity and injury severity in the face of poorly defined samples are presented. The problems of determining accident and injury causation factors are discussed. Certain sub-sets of the MDAI data, particularly when considered with police-reported accident information from the same jurisdictions, can be used to estimate the frequency of occurrence of some phenomena of interest. Finally a plan is presented for a modification of the present MDAI program which should achieve both a statistically representative sample of accidents for a large portion of the United States and a capability to conduct special short term studies to answer specific questions.

17. Key Words MOTOR VEHICLE ACCIDENTS ACCIDENT INVESTIGATION METHODS STATISTICAL ANALYSIS REGRESSION MODELING		18. Distribution Statement Document is available to the public through the National Technical Information Service, Springfield, Virginia 22151	
19. Security Classif. (of this report) UNCLASSIFIED	20. Security Classif. (of this page) UNCLASSIFIED	21. No. of Pages 237	22. Price

Form DOT F 1700.7 (8-69)

STATISTICAL INFERENCE
FROM
MULTIDISCIPLINARY ACCIDENT INVESTIGATIONS

List of Tables

List of Figures

List of Figures Continued

CHAPTER I

INTRODUCTION

The set of all crash events constitutes a population which extends over both time and geographic dimensions. From this population a sample is obtained and analyzed in order to learn about all crashes in the population. This sample should represent the population. That is, relationships resulting from analyses of the sample should also apply to the entire population. A random sample is the best way of insuring representativeness. Unfortunately no random sample of United States crashes exists. Instead there are groups of case studies (e.g., CPIR file) or data which purport to be a census of crash events in a particular geographic area (e.g., Washtenaw County police investigation reports). The latter type of sample can be used to study crashes in the particular geographic area (e.g., Washtenaw County). Unfortunately the police reported data generally do not contain enough detail to answer the sophisticated questions presently being asked by the highway safety community. An adequate evaluation of energy absorbing steering columns requires precise measurement of impact angles, interior components struck, and location of body injuries. Thus the user of crash data faces a dilemma. Either he must analyze data from a large group of crashes with little or no detail or he must analyze a small detailed set of data from a poorly defined sample of crashes.

This report presents an approach to handling the CPIR data which contains precise measurements. Methods for using the present file are presented. In addition a plan for collecting future data, which considers both the problems of complete accurate measurement <u>and</u> representativeness, is presented. This plan provides a basis for establishing a national crash information program which can provide the basis for decisions regarding safety standards and countermeasures. However, the major portion of the report discusses uses of the CPIR data which are presently available. These data are available at no cost. They can and should be used to gain greater understanding of the crash process. The alternative is to wait until a national crash information program is established before doing additional analyses.

PURPOSE OF THIS STUDY

The combination of in-depth investigation reports, the digital files resulting from them (and the associated police-report and population data from tri-level MDAI teams) constitutes a resource which has been used sporadically but remained largely untapped because of a lack of understanding of its capabilities and limitations. This information can be viewed as one part of the broader compilation of many sources of accident data, but the MDAI data collection system, being under the direct purview of the federal government, is thereby susceptible to change by direction of the NHTSA.

Data about accidents occurring in the United States may be looked upon as a sort of window on the highway traffic system--the visible result of our efforts toward improved traffic safety. The data resulting from the MDAI program should not be considered as an end in itself. Rather it may be viewed as a part of the information necessary to the solution of some stated problem. One may generate a problem statement such as "how many fatal accidents are 'caused by' cars parked on the shoulder of high speed roads, and could these be inhibited by some new warning signal to be carried by each vehicle and properly employed by each driver?"

The mere statement of this problem implies that some data has already been obtained (although a question could be asked independent of any data). This could range from a single observation reported by letter to a compilation of accident data from a large jurisdiction. The data which led to the problem statement may not have indicated that there were any fatalities (simply that these were accidents of this type), or it may have suggested that there were a substantial number of such fatal accidents (but without a very precise estimate of the national total). Our observation of the present highway traffic system is almost always in terms of some sample of the entire population, and our ability to infer to the national totals depends in large measure on how well that sample can be defined.

Chapter III describes the data file and defines the rules for its collection. In particular a methodology for using the present data is introduced. Each crash investigated is viewed as an experiment--defined by crash classification variables--which results in a measurable outcome such as personal injury measured using the AIS scale. The measurable outcomes are viewed as random variables conditional on the crash classification variables. Following this approach it is possible to obtain results concerning the relationship between classification variables, and for example, occupant injury.

Chapter IV presents data describing the entire CPIR file. This can be useful for gaining some initial impressions of the file. Of particular interest are tables which indicate the frequency and injury severity for particular parts of the body striking various vehicle interior components. These tables provide a first level screening for the contribution of various interior components to injury production.

Chapter V presents an application of the retroactive experiment approach which uses a multiple regression model to control for the effect of crash severity on injury. The next chapter (VI) indicates how estimates of the frequency of various crash phenomena can be obtained from the present CPIR data. As indicated these estimates are somewhat tenuous; but estimates made using these procedures may be the best results available.

The remaining two chapters describe a proposal for a data collection system which overcomes the problems of lack of precise measurement and non-representativeness which plague the present data files. This system provides both a continuous sample capability which can be used to monitor the crash population, to conduct studies of various relationships between crash variables, and to estimate the frequency of occurrence for various crash variables. In addition it provides a rapid response capability to acquire data for special studies.

Supporting material is contained in appendices A through I. While these are not necessary to a general understanding of the concepts presented, they should be of value to readers who will be making direct use of the MDAI files.

CHAPTER II

SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

The present collection of MDAI cases is a sample of an undefined and relatively undefinable population, thus limiting severely the capability to draw inferences to the national accident picture. The cases in themselves are often of considerable anecdotal value, and indeed may lead to further data analyses which can lead to statistical inferences. Certain subsets of the data, most particularly the census of injury accidents resulting from the CALSPAN jointly funded program and two MVMA funded programs, and the set of accidents investigated in Washtenaw County (under separately funded NHTSA and MVMA support) can be used to infer frequencies of occurrence of phenomena of interest; but these are limited by their provinciality and only guarded inferences to the national population can be made.

Methodology has been presented in this report for defining what is called a "retroactive experiment", i.e., selecting experimental data points from the existing reports to fill in the cells of an experimental design. While this technique can illuminate some problem areas, it is again difficult to extrapolate these findings to a national picture because of some known built-in biases in the selection of MDAI cases. In the early instruction to MDAI teams it was noted that cases should be selected which demonstrated the greatest disparity between vehicle damage and injury severity--suggesting that the best cases for inclusion would be fatal accidents with no damage to the vehicle or the opposite. It is clear from analysis of the data that teams concentrated on the former, and that the present set of data represents a worse than average distribution of injuries relative to the vehicle damage extent.

Some insight into injury causation factors can be gained by analysis of the existing data, generally by accounting for some of the above mentioned biases by statistical manipulation. In particular the method of regression analysis is discussed and applied to some problems.

The stated purposes of the MDAI programs, contained in most of the MDAI team contracts, are to:

(1) Determine causation of accidents

(2) Determine causation of injuries

(3) Identify functional problems of the highway transportation system

(4) Identify the need for countermeasures

(5) Indicate appropriate countermeasure methods

(6) Assess the effectiveness of new safety features

(7) Evaluate the performance of vehicle safety standards

(8) Evaluate the performance of highway safety standards

As a guide to individual investigations, and to the writing of reports on individual investigations, this is a succinct statement of the expectations of the NHTSA for the accident investigation program. Indeed, in writing up individual cases the investigators keep all of these points in mind and address them.

A major purpose of this study performed by HSRI was to determine whether the combined information from the many individual investigations can be used in a statistical sense to address these same purposes. To a limited extent causative factors involved in accidents can be identified--however without exposure information the existence of potential causative factors is difficult to interpret.* The causes of injuries (in the sense of which parts of the vehicle interior were involved) can be studied with respect to, say accident configuration (rollover vs. non-rollover). Functional problems in the highway transportation system might be identified by noting the occasional occurrence of some factor (like injury due to hood penetration of the windshield), but the national frequency of such events can only be estimated by reference to one of the MVMA subsets of the data--and there with an unknown provincial bias.

---

* In addition the methodology for reporting causative factors was not well defined and information from various teams cannot be aggregated in a consistant fashion. A scheme similar to that employed by the Indiana MDAI program if consistently applied by all teams, would permit some more useful analysis of accident causes.

The need for countermeasures, and the identification of the appropriate countermeasure may be suggested by data analysis, but must in the end rest with experts in countermeasure design--for example with highway or vehicle engineers who can invent a solution to a problem identified in the data.

The effectiveness of new safety features, and the evaluation of the performance of vehicle safety standards can only be solved in a limited fashion in the present data--in part because of the sampling biases (e.g., early crash investigations concentrated on the most severe crashes) and in part because of the lack of data on older vehicles and on non-injury accidents which are necessary for comparison. Finally the performance of highway safety standards, which may often be noted in individual reports, is exceedingly difficult to determine in the collected data because of the lack of an exposure measure--for example we may know in some detail whether the drivers in the accident data set have had a particular kind of driver education course, but we have no idea of the extent to which this course has been taken by drivers not involved in crashes. The evaluation of some of these safety standards is exceedingly difficult even in well designed experiments (as discussed in Appendix C) and the MDAI program should hardly be expected to solve all these problems.

This tabulation may seem rather negative, but the reader might keep in mind that negative results also come from highly controlled experiments in these same areas. This set of data has many limitations which are discussed in some detail in the text of this report, but it is also better in many ways than anything else which is available.

It is clear that we as analysts would be better off if we had planned at the outset carefully to collect a set of data more appropriate to our questions, but we note that the work to-date has been of value in a number of ways--learning what data could be obtained and developing reporting methods for consistent and precise recording; learning more about ways of analyzing accident data and thence defining sampling plans more carefully; and developing a cadre of investigators who have the capability to provide information more attuned to the needs of statistical inference.

The problems noted above are still with us, and we might ask the question "can we make better use of our present capability to build a more useful data acquisition system?" We note that accident data reported by operating agencies (primarily police departments) has the advantage of being more of a census of* uncertain measurement precision. The percentage of all injuries identified as "A" (relatively serious), in police reported data varies from 65% (Virginia), 42% (Florida), 28% (North Carolina), 23% (Connecticut), 15% (New York), to 12% (Oklahoma). Compilations of police injury data which include information from several parts of the United States are of doubtful value. Vehicle damage data information is similarly confounded, reporting systems being different--and even where the same interpreted in a different manner.

It is suggested therefore, that a principal requirement for a set of data which can lead to national inferences about injury and vehicle damage is that the data be representative of the nation and that it be consistently reported. The present MDAI program is viewed as a starting point from which to launch such a data collection activity.

In Appendix A of this report a geographical sampling design is presented which would include data from approximately 30 Standard Metropolitan Areas (SMSA). These areas are chosen because adequate census data is currently available to describe them, because they represent all SMSA's and thus more than half the population of the United States, and because the facilities and personnel of the census bureau and other surveying organizations can easily be used

---

* Local police agencies have different selection rules because of official reporting criteria,daily workloads, and local policies. Such practices lead to undefinable errors when police data are aggregated over many agencies.

in further exposure studies of such areas. In addition results of other studies of population characteristics can be compared to the crash involvement studies. A stratified sampling plan is discussed based on both regions of the country and SMSA population size. Among the present or past MDAI teams some 15 have operated within acceptable SMSA's and modification of their accident sampling plans would lead to an immediate capability. From a longer term point of view it will be of value to add new data taking teams, and it is expected that some of these could be direct federal employees operating in conjunction with, say, an NHTSA regional office.

THE TEAM ACTIVITIES:

Three separate but complementary activities would be performed by these investigation teams:

(1) Collection of full MDAI cases (SCRUB - Selected Collection of Reports with Unknown Biases)

(2) A continuing sample of all collisions (of a defined accident severity) occurring within the SMSA, (VACUUM - Valid Accident Data Collection for Universal Utility Measurements)

(3) Periodic special studies, responsive to direction from NHTSA, to collect very specific data for a small number of identified collisions (CARPETSWEEP - Computer Assisted Rapid-response Procedure for Evaluating Technology in Specific Wide area Environments of Emerging Problems)

These three activites are amplified below:

(1) The MDAI cases, performed much as in the present operation, would be limited to about 10 or 20 per year per team. They would serve several purposes, the principal one being to provide anecdotal information about a small number of interesting crashes to the sponsor , accompanied by the reasoned interpretations of the local investigation experts. It is assumed that particular teams would concentrate on areas of interest to them--one looking at cases involving vehicle defects, for example, and another concerned with broken bones. Another purpose of the MDAI investigation is to pro-

vide a continuing rapport with local police agencies, hospitals, wrecking yards, motor vehicle departments, and experts in various fields related to accident investigation so that, when they are necessary to some other directed study, the channels will be open. Finally, the MDAI cases will serve to provide some degree of continuous training to the investigators so that they will have the knowledge and skills necessary to reporting of information in the directed studies.

(2) The continuing sample of collisions should be based on some vehicle damage criterion (such as towaway), and should provide a limited amount of (primarily) injury and damage information on approximately 500 vehicles per year per team. The present CPIR form could serve as a starting point for such investigations, but in the future it is anticipated that a somewhat shorter form tailored to the expected needs of the NHTSA would be developed. The purpose of this sample would be to provide a national sample which could be summarized monthly and annually in terms of injury and vehicle damage distributions by make, model, and size of vehicle; to identify marked changes in injury rates to, for example, motorcyclists; to identify a change in injury patterns occasioned by some new feature. With each of 30 teams providing 500 cases per year, some 15,000 such reports will be acquired per year. These reports would, in general, be entirely digitized, involve no photography, and would cost approximately $150 per case.

(3) The special investigations would be conducted by the same teams using the CARPET SWEEP concept detailed later in this report. Each of the teams would allocate approximately one third of its effort to collecting a small amount of data for special studies identified by some central selection process. For example, if the controlling agency wished to determine the frequency of use of various kinds of child seats, and the injuries associated with accidents for each, it might be desirable to investigate 1000 accidents involving 1 to 4 year old children as vehicle occupants. Using the 30 SMSA teams, each accident (or some defined sample of accidents)

fitting this criterion could be investigated in order to determine the seated location of the child, the model number of the child seat, the injuries to the child (using the AIS codes), the damage to the vehicle (using the VDI code), and other information specific to this study. It is anticipated that the information would be limited enough (in extent) so as to require a minimum of investigation time, and that it would be communicated to the study headquarters by teletype so that reporting and processing times would be minimized. In this way data collection for this subject might be completed in 30 days, conclusions drawn, a report written, and a subsequent study begun.

TIME PHASING

A continuing sample of all collisions could be defined for each of several teams presently operating, and this process (item number 2 above) begun without delay. New teams should be implemented on some priority basis, probably choosing more complete regional representation first (there are currently no teams in the northwest area of the United States), and then implementing activities in the major cities identified in the sampling plan.

COST

A preliminary cost estimate has been based primarily on the costs of existing MDAI teams. It has been estimated that each team performing the three tasks defined above might operate for approximately $150,000 per year, although the total would depend on local salary and expense levels, and on whether the full MDAI cases were accomplished.

Each of the teams might be considered a sort of Tri-level operation, although the tri-levels are defined somewhat differently from the present operations. Access to computerized police data for each team would be useful--perhaps necessary to the design of the sampling plan and to the design of CARPETSWEEP operations. It is anticipated that most teams would not need to maintain their own banks of police level data for this purpose, but could draw on local compilations.

DATA NEEDS/OBJECTIVES INTERFACE

The objectives of a federal accident investigation program can be stated in summary form as three items:

(1) development and publication of several national indices of the traffic accident situation.

(2) information on the relationship between injury degree and type and crash configuration variables (such as impact vector, size of car, vehicle damage, presence of safety equipment, etc.)

(3) identification of factors leading to the occurrence of crashes (i.e., crash causation).

Most of the analysis in this report deals with the second objective stated above, and examples are given with the use of the present data in this regard. Objectives (1) and (3) have been considered and discussed, and in general the present data is not adequate to answer these questions because of both the lack of exposure information and the lack of a definable sampling plan.

Five factors interact in the definition of a data collection program. These are:

(1) the objectives of the study (for example, what percentage of the fatal involvements have injuries resulting from hood penetration of the windshield?)

(2) the type and extent of information to be collected (make and model of windshield, type of damage, description of injuries, etc.)

(3) the number of dollars available for the study

(4) the number of cases to be collected

(5) the precision desired in the result

In this report we have recommended two complementary types of data collection activities--one is continuous and the other is problem oriented and exhibits rapid response. For the continuous data collection program we have specified in this report recommended levels for factors 2 through 5 based on our conversations with many data users in government, industry, and research institutions. In summary the continuous data collection program (VACUUM) would be embedded in a national program costing approximately $5,000,000 per year, collecting about 15,000 cases per year, the information being essentially equivalent in detail to what is presently contained in the CPIR form, and providing the ability to see changes (from year to year) of the order of 1% in occurrence of crash variables of interest (e.g., if lap belt usage increased from 24% to 25% a sample of this size could detect that fact).

As a supplement to this the special data collection program (CARPETSWEEP) would require a specific design for each problem. As discussed in Chapter VII the investigator would ordinarily try to answer his question using existing data, but would proceed to design a specific program when that did not suffice. The choice of type of information, number of cases, and desired precision of the result in the CARPETSWEEP operations will depend on the users needs and the number of dollars available.

CHAPTER III
THE MDAI FILE

HISTORY

The Multi-disciplinary Accident Investigation program of the National Highway Traffic Safety Administration was begun when a small number of dedicated professionals undertook a series of investigations of (mostly) serious accidents, and forwarded the reports to the (then) NHSB. Initially, these reports were relatively unstructured, but in about 1969--mainly as a result of discussions held at a symposium for accident investigators--the teams then operating began to report a portion of their observations using the Collision Performance and Injury Report, Revision III (sometimes known as the General Motors Long Form).

In the same period several teams supported by the Motor Vehicle Manufacturer's Association also investigated a series of accidents, and provided their reports using the same forms. In general the MVMA reports are less detailed with regard to the driver and environmental factors in the accident, but were equivalent in regard to injuries and vehicle damage information.

Since 1969 there have been a total of 22 investigation teams operating at one time or another under the sponsorship of NHTSA. At the time of writing of this report a total of 1133 vehicle reports have been coded into a digital file, with information on 1936 occupants.

Chronologically speaking the first few teams in the MDAI program came into being either because they were already investigating accidents for other reasons (Cornell and UCLA) or because their principal investigators showed an interest in the activity and received funding to begin such an effort. Subsequently teams were selected on a geographical basis so as to have some coverage of most of the continental United States. Within the geographic constraints, however, it was necessary to have a group of interested persons propose to conduct such studies; and the teams which came into existence tended to be one of three general types--(1) those associated primarily with a medical examiner's office, (2) those associated with a University teaching or research facility, and (3) those associated with a research institute. With few exceptions, then, the MDAI teams tend to be located in or near large cities (where there are Research Institutes, Medical examiners, or Universities), and it might be expected that the complete set of data resulting from these studies would not generally be representative of accidents in the United States.

The 1133 case vehicle reports from MDAI teams have been combined into a digital file with 1500 reports from MVMA sponsored teams. This study contract has been concerned primarily with the problems of drawing inferences from the MDAI data, but it will touch on the utility of the MVMA-provided data where appropriate.

Several of the MDAI teams have expanded to become tri-level or multi-level operations with respect to the kinds of data collected and analyzed. The tri-level concept has developed over the past several years with the intent that in-depth investigations could be complemented by police reported accident information, or in some cases, data collected to some intermediate level of detail, in order to permit better estimation of national accident characteristics.

CASE SELECTION BY THE TEAMS

The complete MDAI digital file at the present time is made up of data submitted by NHTSA and MVMA teams over the period since 1969. In part because of case selection, and in part because of longer processing time required for the more severe collisions, the severity of accidents in the file varies with time...being the most severe in the earlier period, and least severe in the most recent period (1972). The processing time problem will be discussed first.

The time between the occurrence of the accident and the submission of the case to the sponsor is somewhat variable, depending to a large extent on the difficulty of getting all of the information required to complete the case. As a general rule, the more severe cases (e.g., those involving multiple fatalities and multiple car accidents) require more investigation and have the most delays in obtaining information. When the next-of-kin must be interviewed, or when the investigator must obtain information through the attorney for a driver charged with negligent homicide, much time is consumed. At the other extreme, an accident involving a minor or moderate injury with a single vehicle may be reported quickly.

These differential delays seem to be increased as the case proceeds to the computer; many cases involving little or no injury get into the computer within 30 to 90 days, but some serious accident cases currently being coded for input are more than 18 months old.

The time delay problem, then, must be taken into account in some analyses in which this bias would be important. But ultimately it can be solved by waiting, or corrected for by limiting the analysis to time periods for which the data are complete.

The more serious problem is that of variation in the case selection or sampling rules used by the several teams. In 1969, at the Airlie House Symposium, Dr. Flamboe* noted that "The present criteria (for MDAI teams) determining which crashes are to be investigated are those involving late model cars, 1968-69, and where there is a great degree of disparity between vehicle damage and injury severity."

Using this criterion one should expect to find a collection of MDAI cases which are more severe (in the injury measure) than cases selected randomly from collisions of specified (vehicular) damage severity. This is in fact the case, and the effect upon the ability to do certain analyses with the data is somewhat devastating. More

---

* "A National Collision Data Collection System" by Eugene E. Flamboe, proceedings of conference at Airlie House Symposium 1969, Warrenton, Virginia.

recently most MDAI teams have been instructed to investigate a "representative" sample of collisions within their jurisdictions, but no precise definition of representativeness has been given.

Figures 1, 2, 3, and 4 show the distribution of injury severities (using the AIS scale) for several groups of accident severity (using the Vehicle Damage Index Scale). The vertical bars indicate the percentage of cases (in the indicated subset) for which the most severe injury in the vehicle was as shown. To provide a standard for comparison, the subset identified as "Washtenaw County" is given. This includes 772 vehicle reports of American-manufactured cars occurring over a period of about 20 months, and is intended to be a census of towaway involvements.*

Figure 1 compares the complete set of NHTSA sponsored cases with the Washtenaw County data, taking the complete range of Vehicle Damage Indices. Note that the percentage of fatal involvements in the NHTSA cases is more than four times that in the "towaway" census; the percentage of non-injury involvements in the NHTSA collection is about one-third that of the comparison set. It is clear that, on the average, cases involving injuries more severe than would occur in a sample of all towaways were selected.

Figures 2, 3, and 4 show the same comparison divided into three sub-groups by an accident severity measure. In Figure 2 all cases with a vehicle damage index of 1 or 2 are compared. Note that in the towaway set there were no cases of injury at the levels 4 (severe), 5 (critical), or 6 through 9 (fatal), although about 5% of the injuries in the NHTSA set are in this range. This is certainly consistent with the instruction to "find those cases in which there is a disparity between the impact severity and the injury severity".

---

* It represents, in fact, about 50% of the eligible towaways and is probably biased slightly toward more severe collisions.

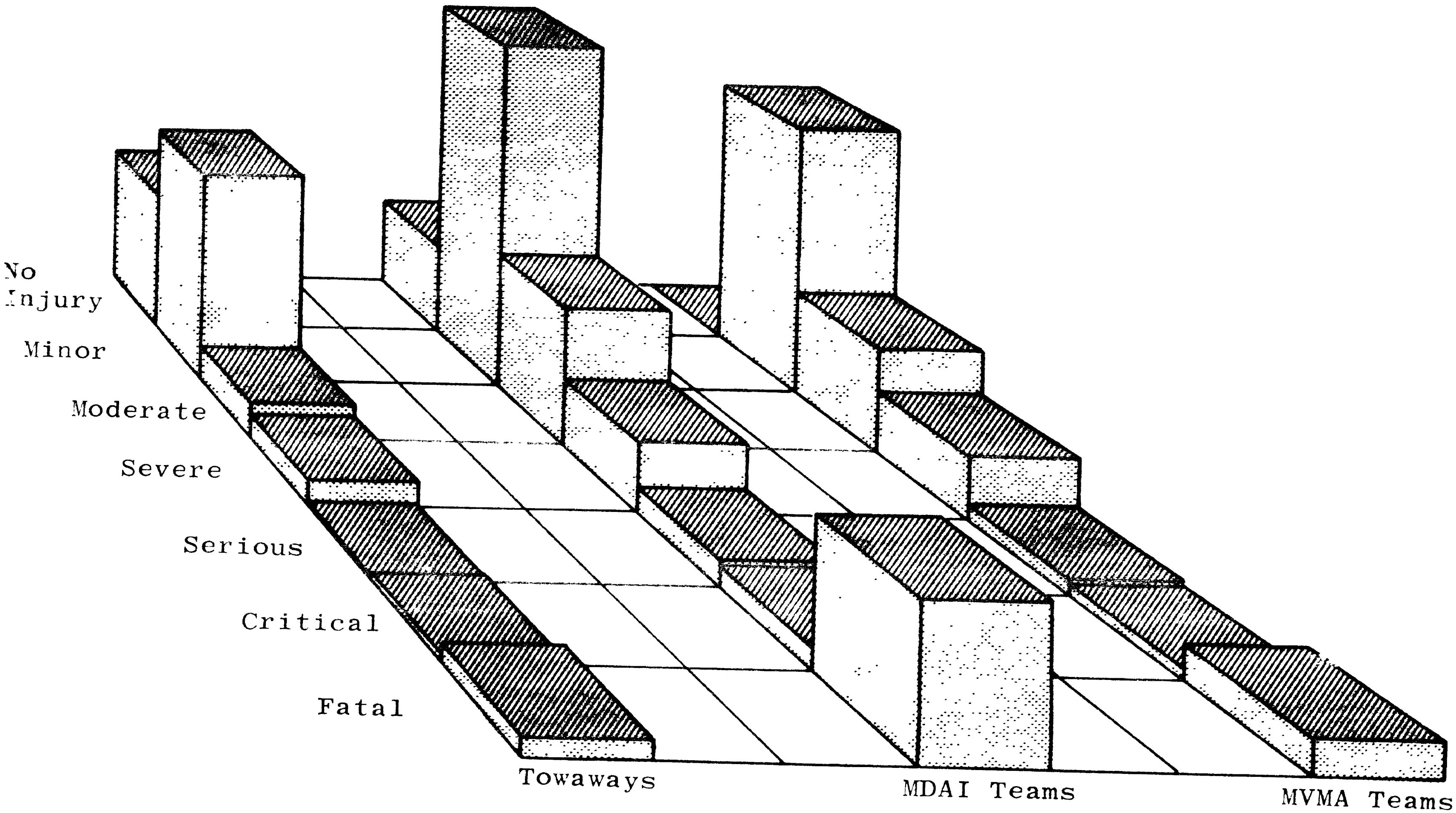


Figure 1. Injury severity distributions for three sub-groups of data in the MDAI file (all impact severities combined).

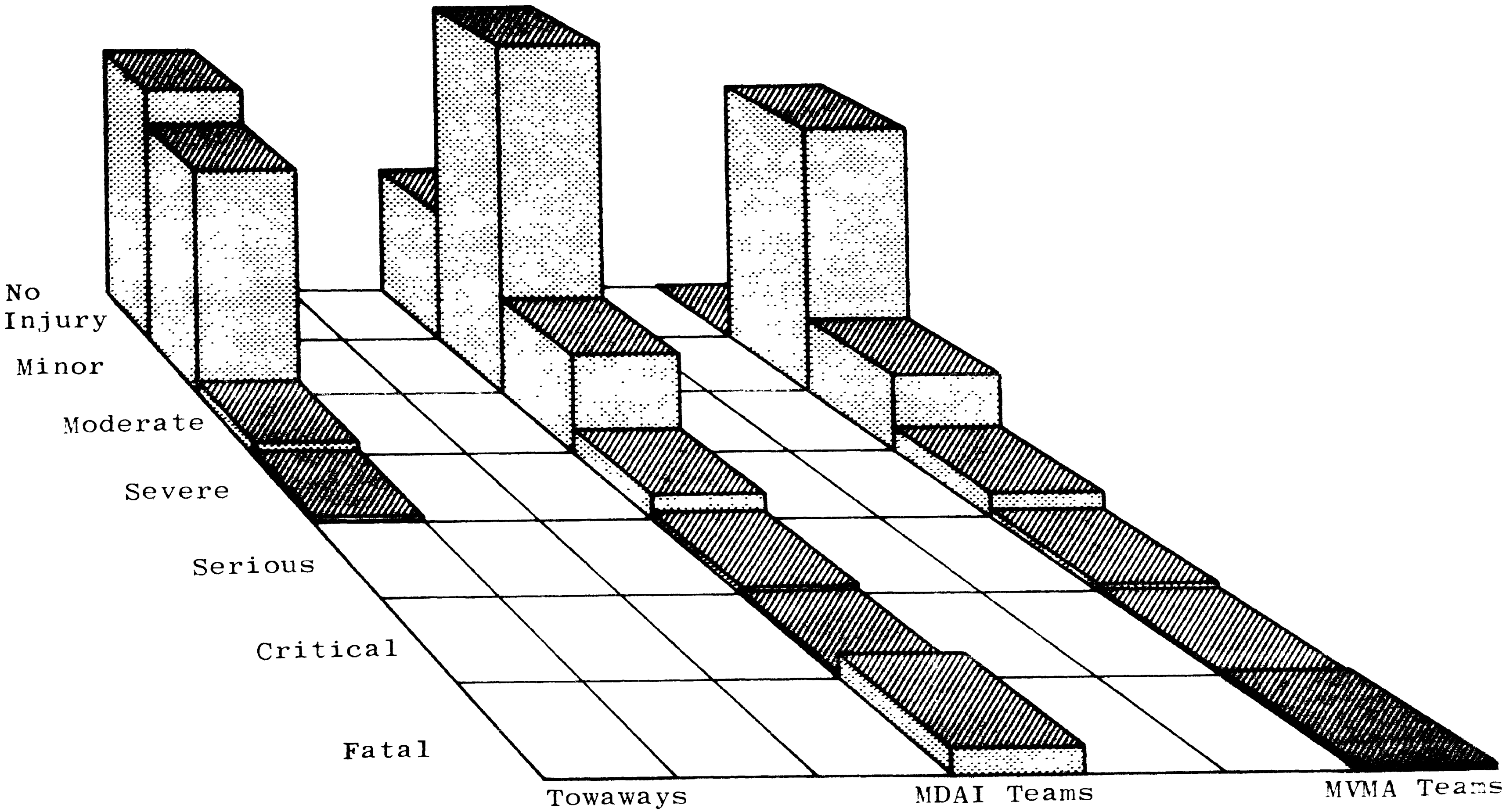


Figure 2. Injury severity distributions for three sub-groups of data in the MDAI file (low severity impacts, VDI=1-2).

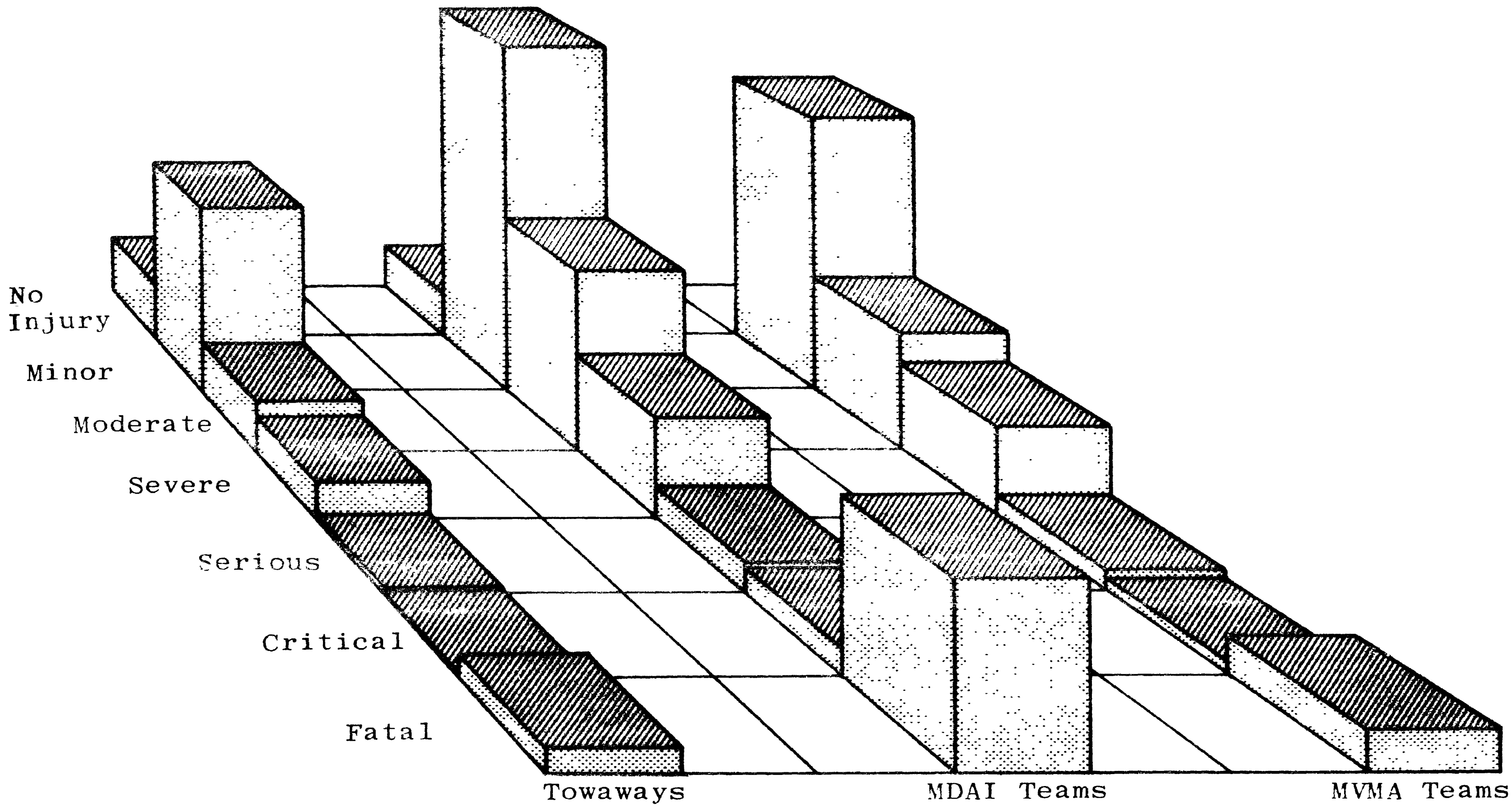


Figure 3. Injury severity distributions for three sub-groups of data in the MDAI file (moderate severity impacts, VDI=3-4).

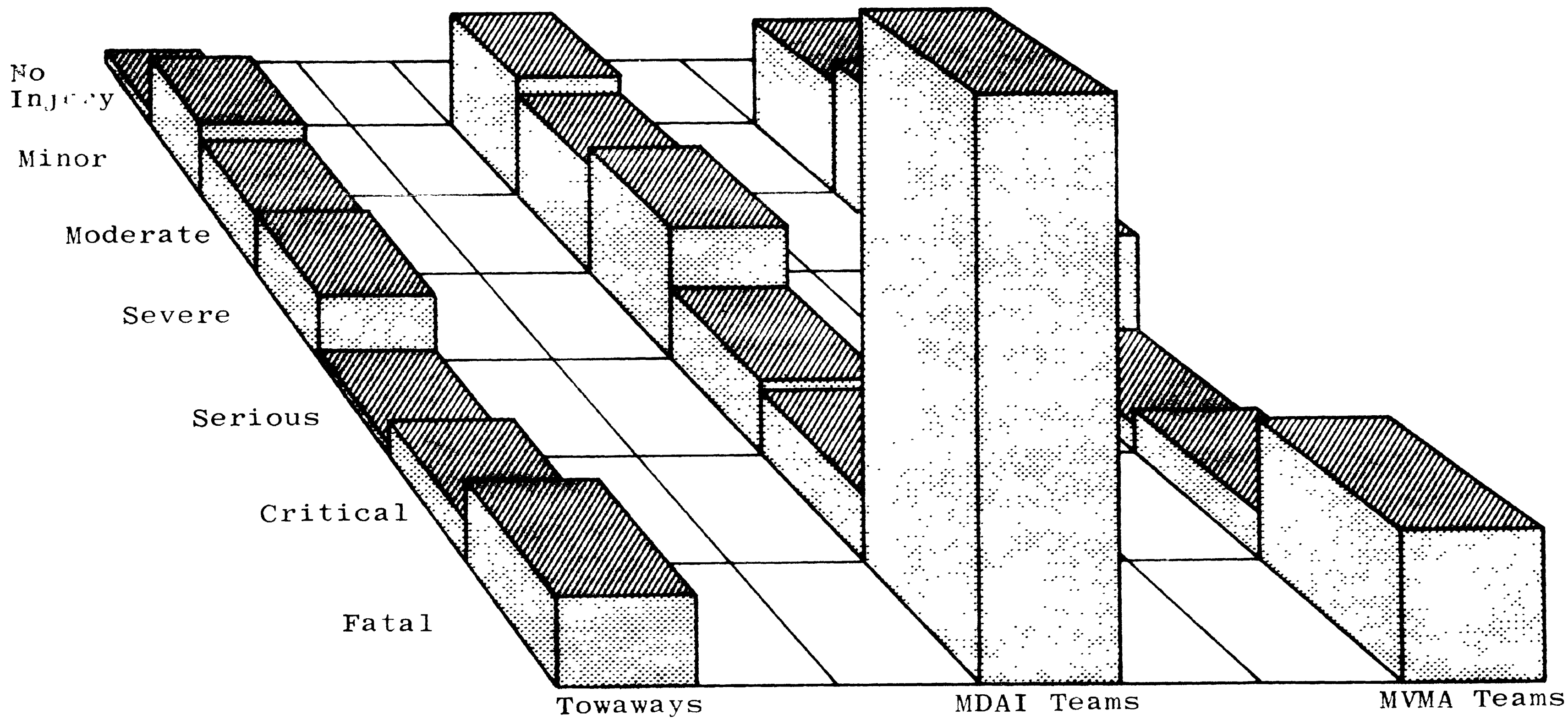


Figure 4. Injury severity distributions for three sub-groups of data in the MDAI file (high severity impacts, VDI=5-9).

The same effect is clear in Figure 3 (for vehicle damage indices of 3 and 4), and in Figure 4 (for vehicle damage indices of 5 through 9). The instruction to find those cases with a great disparity between vehicle damage and personal injury levels could have been interpreted in either of two ways--i.e., to find cases in which no one was injured even though the accident was severe, and vice versa. The data indicate that the latter was the general rule followed.

If this instruction had been taken to the extreme, one might expect to find only two kinds of cases reported--fatal accidents with little or no damage to the vehicle, and non-injury accidents with the vehicles demolished. In fact there are many accidents reported which are more normal, but in looking at the entire set of NHTSA data it is not generally possible to segregate these from the "outliers".

The MDAI data itself represents a relatively undefined sample of accidents, and this suggests that its principal utility may be in finding new problems and not in defining their frequency of occurrence. For example, a particular MDAI case report involving a school bus which rolled over during a crash notes that the driver had no training in the handling of the bus in emergencies, and suggests that initial and periodic training in emergency maneuvers while perhaps not useful in inhibiting the crash, could have prevented the rollover. But the entire set of MDAI cases does not give a picture of the number of school bus rollovers in the country (i.e. is this really a problem), nor does it tell what percentage of the school bus drivers currently have adequate emergency maneuver training. Yet the immediate finding seems important, and deserves to be augmented by additional frequency information.

Parts of the MDAI data set do constitute a more precisely defined sample, being an attempted census of certain accidents occurring within given jurisdictions. With respect to such subsets of the data some inferences about frequency of occurrence may be drawn. These will be defined later in this report, and examples of their use given.

The MDAI data might be viewed as a set of information acquired in considerable detail so that future questions (those not thought of in detail at the time of data taking) would have some high probability of being answered. With respect to injury and vehicle damage problems, this is largely the case. To some extent, the police reported data (as in the Tri-level programs) serves the same purpose. But new questions will be asked in the future for which no existing data is appropriate, and it will be necessary to generate new data collection programs to provide the answers. In this report we consider each of these aspects--the present use of MDAI data, the use of police reported data, in conjunction with it, and methods of using the MDAI organization to acquire new information pertinent to new and specific questions.

A major criticism that has been made of the MDAI program is that the data were not acquired in response to a carefully designed sampling plan. The purist will claim that no statistical inferences can be made from this data, and that it thus represents merely a series of anecdotes which may generate much discussion but few conclusions.

It is our intent to suggest that in spite of the lack of a sampling design, the data are in fact useful in that they represent a broad variety of highway accidents, contain very detailed information, and may be assembled into groups from which certain useful inferences can be drawn. Each crash might be viewed as a random sample of a subset of crash classification variables. Examples of these classification variables include vehicle damage index extent code, impact velocity, crash configuration, etc. Much of the remainder of this report is concerned with the methods for making use of this data, and making logical extensions based on the methodology of the data collection process currently in use.

CONTENTS OF THE MDAI FILE

Before attacking the problem of what kinds of statistical inferences can be made from MDAI-type data and what methodologies might be involved in these endeavors, we must take a critical look at what these data represent. In some sense this involves a correction of some common misconceptions about this file of data. After developing a basic definition of the file we will present a justification for its existence in terms of newly established principles for its use.

First of all, the MDAI file as a whole is not a sample--of anything. The cases in the file do not accurately represent any definable subset of the total crash population, even though in specific instances the marginal distributions between the file and some population can be shown to match quite closely. Cases do not get into the file as a result of random sampling, and therefore, results obtained from analyses on these data are not generalizable to the specific events in any population, defined geographically or otherwise. It in no way can be considered a microcosm of the national accident population which we can use to draw conclusions about the characteristics of that population.

If the MDAI file is not a sample, what is it, and of what use is it? We consider it to be a collection of experimental data, each case in the file representing a single experiment or a data point in a large experimental series designed to answer questions yet to be defined. The methodology employed in its construction is that of the "functional" experiment, so named because its goal is to determine the functional form of the variables in relation to one another. The selection of cases to be included in the file is purposive, the attempt being to include in the data all levels of the independent variables which are within the range of interest. The criticism that the file is replete with selection bias, while justified if the file is thought to be a random sample, becomes irrelevant when it is looked at in this light. The inclusion criticism is that a particular case adds to our knowledge about one or more of the independent variables of interest, by extending the range, filling a gap, or providing a combination not available before.

The use to which this data file can be put derives from the large number of variables included for each case as well as the meticulous detail with which accuracy is insured. The methodology for exploration of the file contents involves formulating the solution to a question or problem in terms of the relationship between or among variables. The functional form of these relationships can be expected to provide, if not a basic truth, at least some insight into what happens when an accident occurs. Inferences can be made about the degree to which changes in one variable affect levels of other variables. For example, one might wish to compare the injuries received by occupants wearing lap restraints to those of unrestrained occupants. For crashes of similar severity the file could be interrogated and this comparison made (See Chapter V in which such an analysis is performed). The generalization from the relationship holds for the accident population even though the proportions of accidents in the file representing different levels of variables may be vastly different from those in the general population. Inferences are made to these interactions among variables by this procedure, not necessarily to facets of accident effects in the whole population.

The distribution of data points for any particular variable over the range of interest is an indication of our state of knowledge about that variable and hence our ability to predict from it. The continual analysis of these data, therefore provide a direction for research in the area of accident data collection. The gaps in our understanding of functional relationships represent a need for additional research, and these gaps will become apparent as we undertake to answer more and more questions about the interactions among variables.

It should be mentioned that all the data are conditional on the occurrence of accidents. The effect of exposure is considered neither in the building of the data nor in its primary analytical use, since what we are concerned with is the inclusion of the ranges of variables of interest no matter what the prevalence of that variable might be in the total accident population or the population in general. The results of analyses of these data attempt to explain the underlying cause and effect relationships among variables when accidents do occur, rather than describe or predict the relative frequencies of different kinds of accidents.

HOW IT CAN BE USED

We will use a concept here defined as "retroactive experimentation"--i.e., a process of designing an experiment for which all or part of the data has already been collected and resides in the MDAI file. The approach to using the data, then, is to formulate a specific question which requires experimentation, lay out an experiment using the classical principles of experimental design, and then look to see how much of the data is already available.

Let us consider a specific example of this process. The energy absorbing steering column has come about through the following series of steps:

1. Accident studies indicated that driver injuries were predominantly associated with steering columns.
2. Design engineers formulated a countermeasure in the form of an energy absorbing steering column. Initially a theoretical (physical) model of how the device would behave was calculated.
3. Prototype energy absorbing steering columns were developed and tested under laboratory conditions.
4. Experimental cars were constructed with these new devices, and tested using instrumented dummies.
5. Production automobiles were introduced in which the newly developed steering column was installed.

The obvious question at this point is to determine whether the device performs as well as expected from steps (1) through (4). A way to answer such a question is to conduct an experiment.

The first tasks in experimental design include the specification of:

1. The dependent variable or the measure of interest, in this case the degree of injury measured by the overall Abbreviated Injury Scale (AIS)* - incurred by the vehicle driver striking the steering column.
2. The independent variables, or the parameters whose relationships to injury we are interested in determining. In this case we might be interested in the weight of the car, the speed of impact, and the presence or absence of an energy absorbing steering column. By identifying the effects of car weight and impact velocity we propose to determine the effect of energy absorbing steering columns with greater precision.
3. The control variables, or those which we wish to hold constant for the conditions of the experiment. In this case we might wish to hold the age and weight of the driver within specified ranges.
4. The sample size, a complex decision to be made on the basis of cost, convenience, and the size of the effect that is considered to be important.

The designed experiment might look as shown in Table 1.

	W_1	W_2	W_3
10 mph	5	5	5
30 mph	5	5	5
50 mph	5	5	5

Cars with EA Steering Columns

	W_1	W_2	W_3
10 mph	5	5	5
30 mph	5	5	5
50 mph	5	5	5

Cars without EA Steering Columns

Table 1

Experimental Design for Energy Absorbing Steering Column Evaluation

----

* See Figure 17 on page 131 for fuller description of this scale.

W_1, W_2, and W_3 represent three different weight classes, and the row headings indicate three different impact speeds. The tests will control on driver age (all 20 to 30 year olds) and weight (all between 160 and 180 lbs.), and 5 runs will be conducted at each speed-weight intersection. The assignment of drivers to cells will be done randomly. The number of observations in each cell is chosen at the outset as an estimate of the number needed to give enough stability to the data to permit comparison. Since the variance of the dependent variable is not known at the outset, we might view the design as a sequential experiment in which we will try the number of runs shown, evaluate the results, stop if there is enough data, and run all or part of the experiment again if there is not.

The results of this experiment can be expected to answer the question about the effectiveness of the energy absorbing steering column. But one can hardly conceive of this planned experiment being run.

The data, however, for many of the cells has been generated by chance and is stored in the MDAI accident files. In a sense the experiment has been run, the data collected, and made ready for analysis. We will probably find that some of the cells have a larger number of cases than required by the design, and that a few of the cells will have a smaller number of cases than required. The excess cases can be accepted, and will in fact help to strengthen the results. The missing cases, however, will have to be obtained.

The MDAI file was reviewed in order to select data for this retroactive experiment. Vehicles have been restricted to those having impact vectors of 11, 12, and 01 o'clock, and the Vehicle Damage Index Extent (VDI) has been used as a surrogate for impact speed. Five levels have been tabulated rather than three, since five were available. The weight classifications were (1) less than

2500#, (2) 2501-3500#, and (3) more than 3500#. It was required that the steering wheel or spokes were contacted by the occupant. Drivers in this set of data are restricted to 150 to 190 pounds, and between 20 and 40 years of age.

There are some obvious difficulties in using VDI-extent as a surrogate for speed; yet, for purposes of this study it is the best single variable available on which to control for accident severity. For frontal collisions VDI-extent is defined largely as a measure of the percentage of the vehicle in front of the windshield which has been crushed. For conventional cars, a VDI extent of 3 means that there is penetration about halfway to the windshield. The case of the microbus is an obvious outlier,* but even for conventional passenger cars there is some variation in frontal stiffness, and further, the crush is quite dependent on the width of the object struck and the precise location of the object in the horizontal plane. It would thus seem preferable if we could know the striking speed of the vehicle into a solid barrier, and to have a deceleration record of the impact so as to assure that the data from the two groups is comparable. Alternative measures for crash severity will be discussed later in this report, but for the present we will continue to depend on the VDI (more recently called the Collision Deformation Index) for this study.

Table 2 presents the results taken from the MDAI file. Since data are tabulated in the file for VDI-extents of 1, 2, 3, 4, and 5 all five levels are presented. This shows the number of cases in the file which meet each cell requirement, and the average value of overall injury to the driver (using the AIS scale values ranging from 0 (uninjured) to 9 (fatality with 3 or more fatal injuries)). The third number presented shows the standard deviation (of injury level) in each cell.

Note that three of the cells in the "With EA Column" table, and 13 of the 15 cells in the "Without EA Column" table did not reach the level of 5 cases hoped for at the outset. The columns

---

* Such obvious anomalies would be excluded from the analysis.

labeled "All weights combined" have at least five cases in each cell, but comparing these assures that the effect of weight cannot be studied.

The combined weight data is plotted in Figures 5 and 6. Figure 7 shows two least squares regression lines (derived from the original data) and these suggest that vehicles equipped with EA devices produce a lessened overall injury.

These regression lines are estimates of the "true" relationship between Accident Injury Severity (AIS) and Vehicle Damage Index (VDI). An important question is: How closely do these estimated relationships correspond to the "true" relationships? To answer this question 95% confidence intervals have been constructed for each of the estimated relationships. These are indicated by the dotted, parabolic, curves that are shown about the estimated lines. The interpretation of these confidence intervals is that 95% of the time experiments of this type are conducted the "true" relationship will be included within the computed confidence intervals. However only one experiment has been conducted.

A reasonable operational strategy is to conclude that the "true" relationship is somewhere within these confidence intervals. We do this knowing that on the average one out of twenty times we will be wrong! Specifically from Figure 7 it is concluded that the "true" relationship between AIS and VDI has not been shown to be different for cars equipped and not equipped with energy absorbing steering columns.

This analysis leads to one of two conclusions: either there is little difference between the (average) injury level to drivers of cars with and without EA columns, or there is not sufficient data in the present sample to make this determination. An alternative hypothesis, of course, is that biases in the data destroy the validity so that it is impossible to determine this relationship from the present information. There is some evidence that this

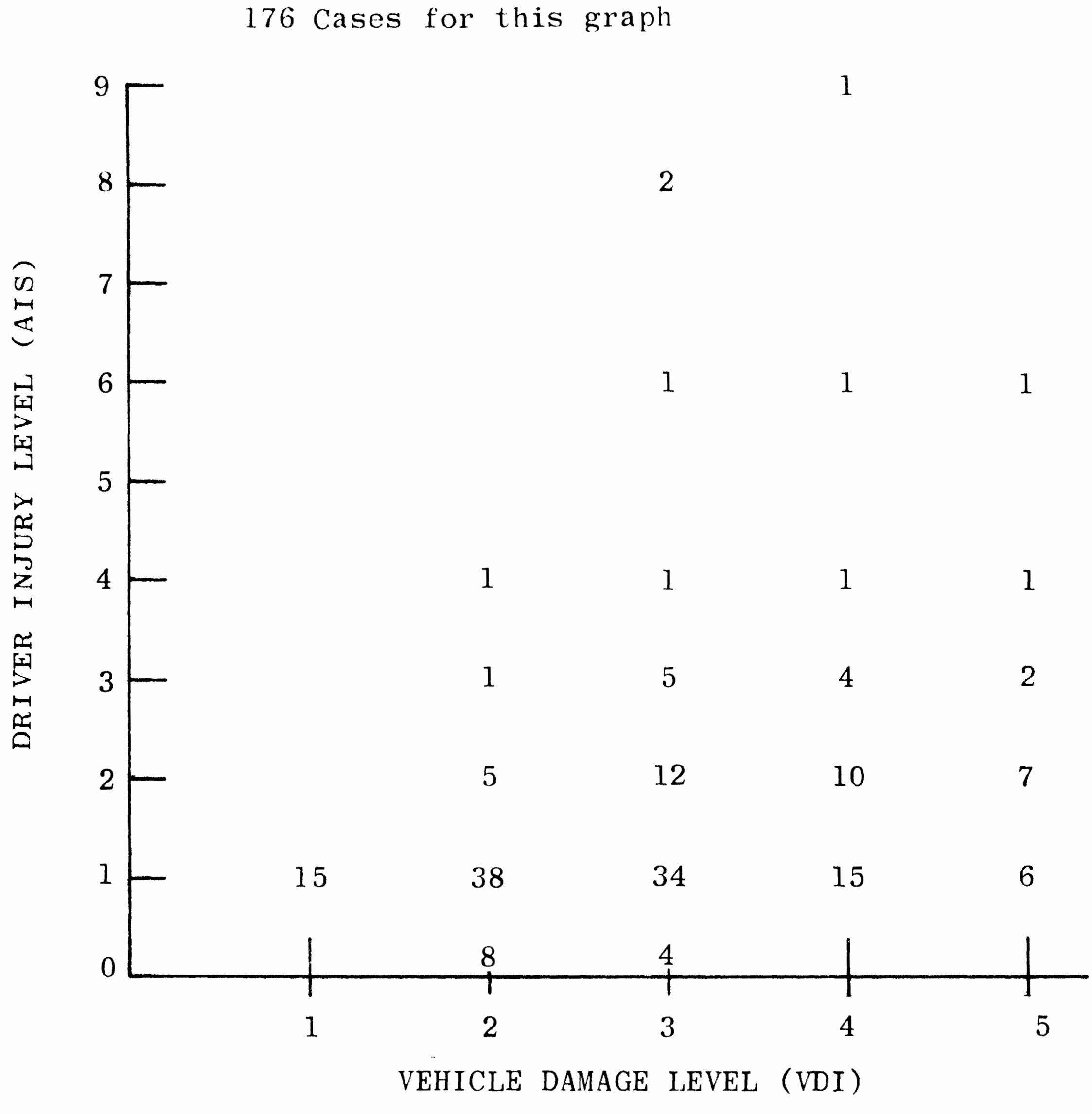


Figure 5. Scattergram, driver injury vs. VDI extent, cars with EA.

50 Cases for this graph

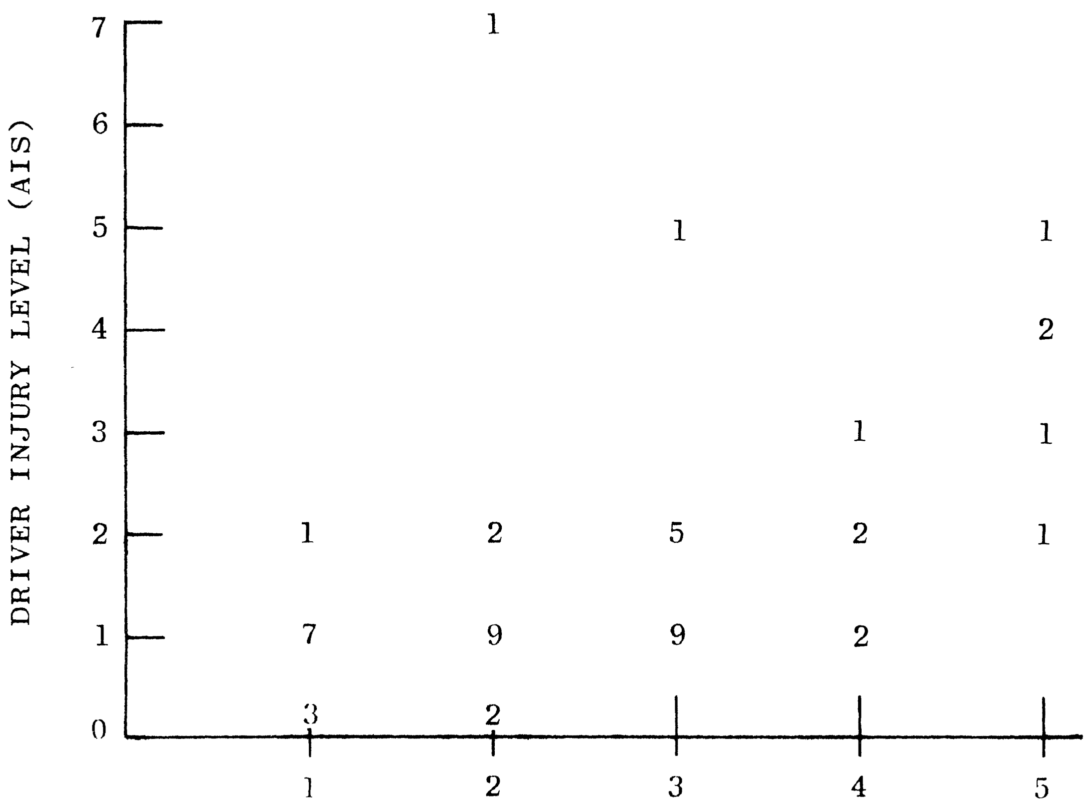


Figure 6. Scattergram, driver injury vs. VDI extent, cars without EA.

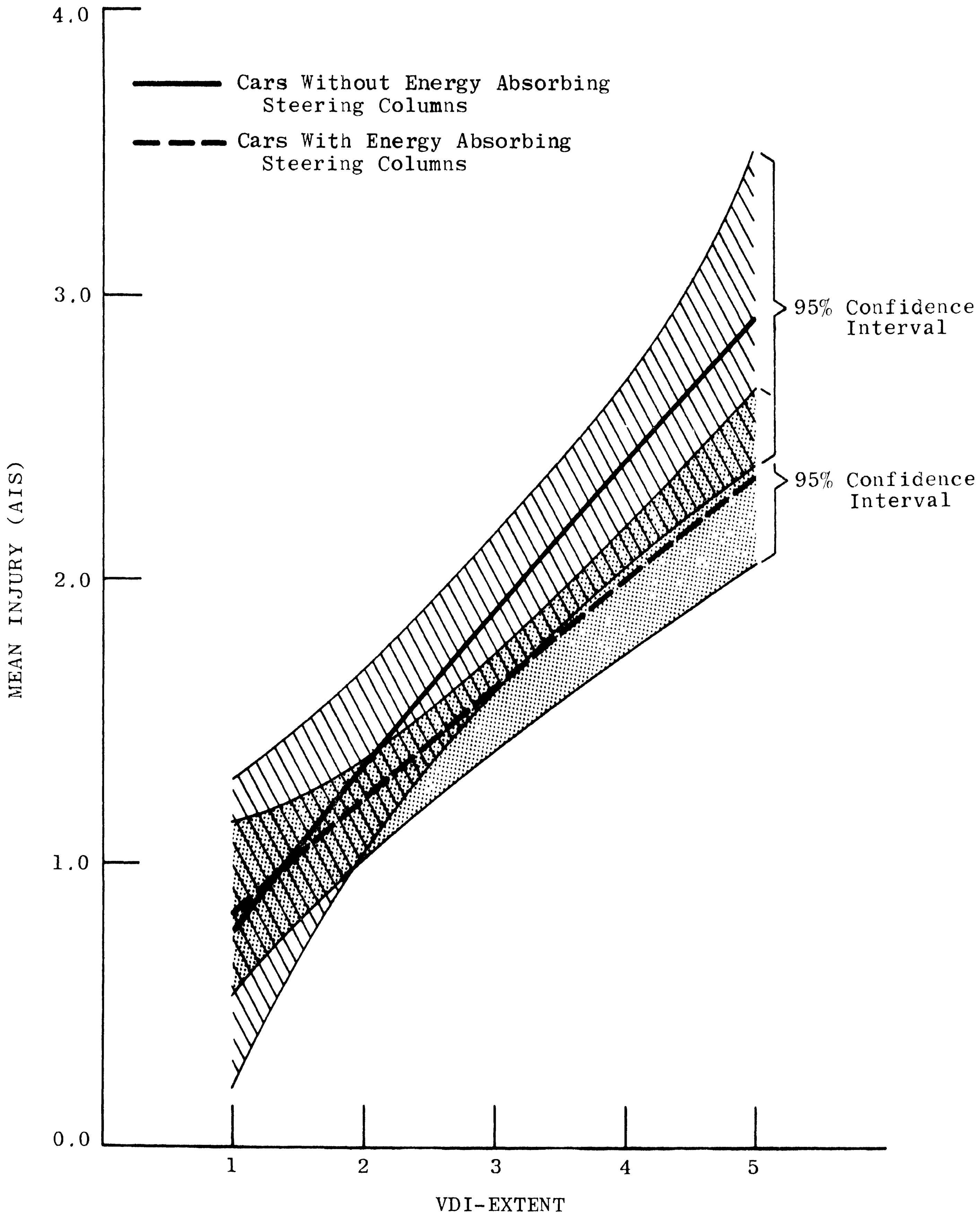


Figure 7. Average ıɲjury vs. VDI for cars with and without EA columns.

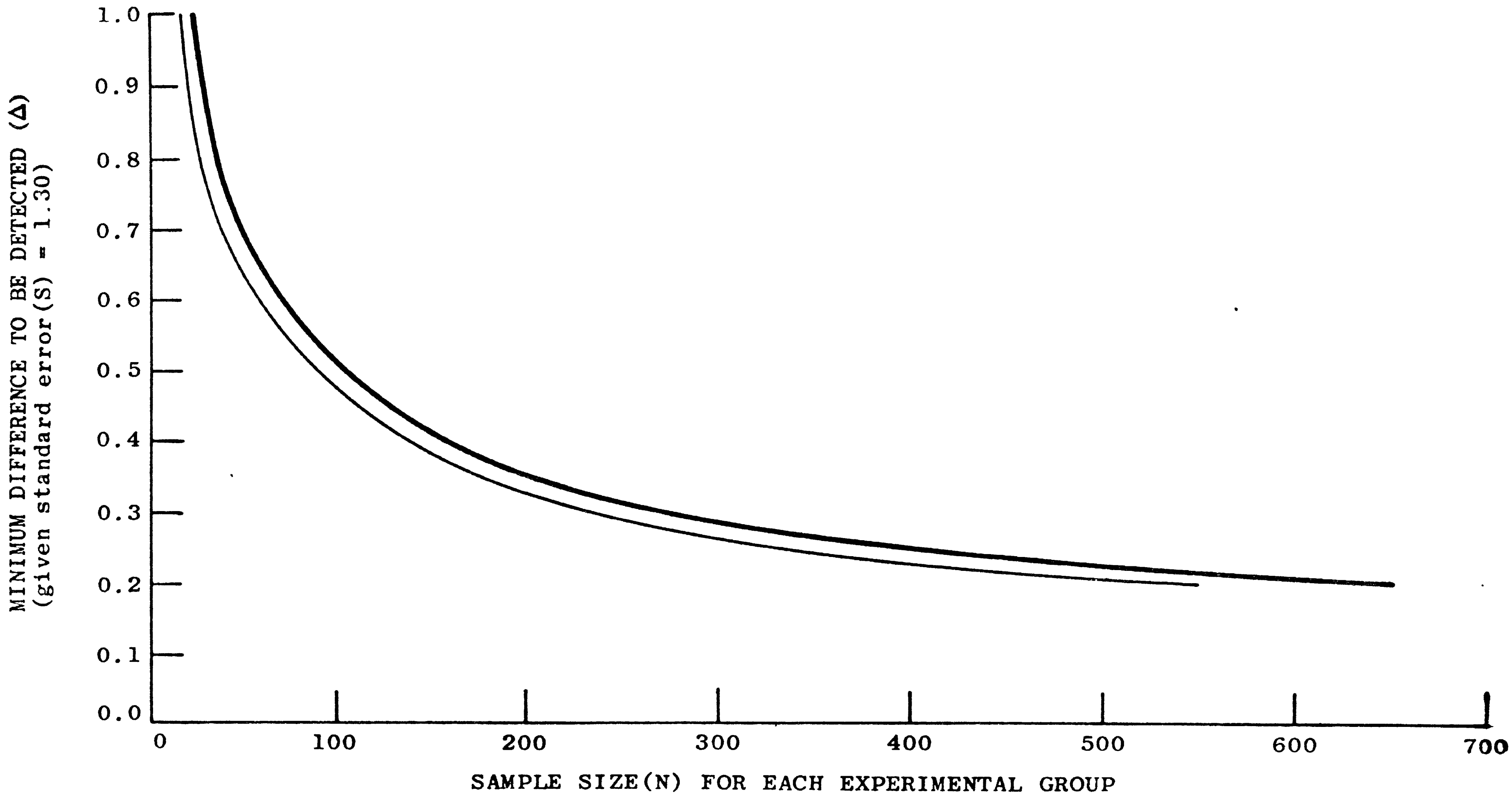


Figure 8. Sample size required as a function of the minimum expected difference; = probability of detecting a difference if it does not exist; = probability of failing to detect a difference that does exist;

$\Delta = \frac{S^2(Z_\alpha + Z_\beta)^2}{N}$; Z_α, Z_β - Normalized deviation for significance level α or β.

is true; but for purposes of discussing the method here it will be neglected. A set of data acquired without the kinds of biases detailed in Appendix B, "Bias in Crash Injury Severity by Investigation Year", would certainly be preferred. The observed difference between the two estimated regression lines at the average VDI (3.0) is about 0.25 AIS units. While the meaning of a change in the mean AIS is not easily defined in absolute terms, it can be considered in relative terms. Elsewhere in this report it is shown that the mean difference in injury level for occupants with and without full restraint systems is approximately 0.3 on this same scale.

But our confidence in the 0.25 value for EA columns (which if the data were adequate might be considered about as effective as restraint systems) is not great. One may ask the question "How much data (i.e., how many cases) are needed to detect a change of 0.25, 0.5, or another value". Figure 8 plots the number of cases needed as a function of the true difference in AIS to be discerned. If the true difference in injury level ascribed to this factor is as low as 0.1 AIS units, approximately 1600 cases would be necessary to detect this fact.

It is possible to test the relationships between impact severity, column type and driver injury severity. The last two columns of Table 2a contain the results of combining the weight categories for all speed and column groups. This data was analyzed using an analysis of variance program that does not require balanced cells. That analysis, summarized in Table 2b, indicates that injury severity is a function of impact severity and the interaction between steering column and impact severity--but not a function of Energy absorbing Steering Columns.

Inspection of Table 2a reveals that the interaction effect occurs for crashes at 40 and at 50 miles per hour impact speed. Specifically at 50 mph cars with EA columns showed lower average injury, while at 40 mph cars with EA columns showed higher average injury. This result is counter-intuitive, and may well be the

Table 2a

Tabulation of data from present MDAI file for EA column experiment

VDI (Impact Speed)	With EA Column			Without EA Column			All Weights Combined	
	Weight <2500#	Weight 2501-3500#	Weight >3501#	Weight <2500#	Weight 2501-3500#	Weight >3501#	With	Without
1 10 mph								
No. of Cases	4	5	6	4	2	4	15	10
Average Injury	1.000	1.000	1.000	1.000	.500	.750	1.000	.800
Std. Dev.	0	0	0	0	.707	.957	0	.632
2 20 mph								
No. of Cases	12	24	17	0	1	10	53	11
Average Injury	1.333	.917	1.000	---	1.000	1.600	1.038	1.545
Std. Dev.	.888	.408	.866		0	2.011	.706	1.916
3 30 mph								
No. of Cases	7	31	18	1	3	7	56	11
Average Injury	2.000	1.677	1.722	2.000	2.333	1.427	1.732	1.727
Std. Dev.	2.646	1.620	.958	0	2.309	.535	1.578	1.191
4 40 mph								
No. of Cases	5	15	4	0	2	3	24	5
Average Injury	2.200	2.467	1.250	---	1.000	2.333	2.208	1.800
Std. Dev.	1.304	2.200	.500		0	.577	1.865	.837
5 50 mph								
No. of Cases	2	5	8	1	3	1	15	5
Average Injury	2.000	2.600	2.125	2.000	4.333	3.000	2.133	3.600
Std. Dev.	0	1.949	1.126	0	.572	0	1.407	1.140

result of small sample size and/or biased selection of cases for inclusion in this data set.

Table 2b

Analysis of Variance of the Effect of Speed and EA Columns on Driver Injury for All Vehicle Weights

Source	Sum of Squares	Degress of Freedom	Mean Square	F
Impact Severity	1.83	4	0.46	3.53*
Steering Column	0.01	1	0.01	.11
Interaction	4.29	4	1.07	8.28*
Error	25.16	195	0.13	
Total	31.29			

* Significant @ $\alpha \leq 0.01$

This example (the energy absorbing steering column) has been presented primarily to indicate a method of using MDAI-like data. There are clearly some deficiencies in the study in addition to the sparsity of data on older vehicles. There may be biases in the selection of cases which are different for older cars (those without EA columns) and for newer cars (those with EA columns). Given the present data set it is possible that the older cars will tend to be associated with teams which produce more severe accidents (or more severe injuries given an accident). While all of these factors have not been discussed in detail here, one analyzing these data should question his findings with regard to the effect of such biases, and invite further questioning from his colleagues before announcing a conclusion. But it is concluded that the desired measure (i.e., the efficacy of the EA column) cannot be obtained from the present set of data.

PRESENT TRI-LEVEL MDAI OPERATIONS:

There are presently three MDAI teams operating under the banner of a "tri-level" program. These are located at CALSPAN, Indiana University, and The University of Michigan. All operate in a somewhat different fashion, and their programs will be reviewed here briefly.

CALSPAN's tri-level program has been designated to obtain measures of injury causation. It has been jointly sponsored by NHTSA and MVMA for several years, and in the earlier years was funded only to acquire data to make it available to the sponsors. For the past two years the program has been augmented with separate activities to perform certain analyses on the data.

The three levels of data are as follows: For all accidents occurring in an eight county region in western New York State police accident reports are compiled in digital form, along with driver records and vehicle registration records from the same region. These then constitute a census of accidents in the region, but with a level of detail consistent with the police reporting. These are the level one data.

For all accidents in the same region involving 0 to 2 year old cars a"level two" file is created and placed in digital form. For as many cases as possible in this data set vehicle damage is coded in terms of the Vehicle Damage Index (VDI) and injuries are coded using the Abbreviated Injury Scale (AIS). These indices are more complete for some cases than others, the difference depending primarily on availability of supplementary information from the police departments. Although the level two data nominally represents a census of new car involvements in the region, that portion of it with VDI and AIS fully coded does not represent a randomly selected subset.

For new American manufactured cars (0 to 2 years of age) and certain trucks in which there was (relatively) serious injury to the vehicle occupants, a Collision Performance and Injury Report is completed. There are approximately 350 of these per year, and they too are intended to be a census of this category of accident in the region. At the time of this writing the completeness of that census is not known, but this matter will be discussed under the University of Michigan program below.

Finally there are produced about 50 "Level Three" or full MDAI cases per year. These are generally selected from new car involvements, but represent a somewhat older car population than those above. Both the 350 above and these 50 are ultimately encoded into the MDAI file.

INDIANA UNIVERSITY

For nearly three years Indiana University has operated a tri-level program designated as a pre-crash or accident causation study. Level one data includes the police accident reports and associated vehicle registration and licensing data, and special surveys such as one of motor vehicle inspection in the county. While the data are not specifically identified as "Level Two" or "Level Three" after that, the data most closely identified as level two comes from approximately 500 on-scene accident investigations per year, and the level three data comes from approximately 100 cases per year in which vehicles are brought into the laboratory for analysis.

Although an attempt is made to make the level three cases a random subset of level two the final selection for level three depends in large measure on the willingness of owners to have their cars analyzed in-depth. This same condition exists to a large extent in all of the MDAI programs in which a case can only be completed if certain interviews and vehicle examinations can be accomplished--thus it has been difficult to achieve any sort of random selection at this level.

Analysis of the data acquired at Indiana is an integral part of the program, and has been concerned primarily with pre-crash factors.

UNIVERSITY OF MICHIGAN

The tri-level program at the University of Michigan has been funded in part by NHTSA and in part by MVMA for almost two years. Level I data is similar to that in other areas, being a compilation of all police accident reports in one county, a sample of registered vehicles and licensed drivers, and data from ASAP roadside surveys and exposure studies conducted in the same area.

The Level II data is composed of full CPIR reports made on about 600 vehicles per year--this being an intended census of new American manufactured cars towed away from accidents in this county. In fact this "census" turns out to constitute something like 50% of the eligible towaways, as revealed by comparison with the police level data. It is of interest to consider whether the 50% that are acquired are a random subset of the whole, and it is possible to employ the police level data to study this.

A study has been conducted as to why cases are not included. Approximately 30% of the total are missed because the staff of investigators is not able to get to the vehicles before they disappear from wrecking yards or repair facilities. About 5% disappear quickly, often being driven out of state. Much of the remainder were not really towaways, although so listed on the police reports. They may have involved a drunk driver whose car was towed because he could not drive it; or they may have been "marginal" towaways--say with a fender bent against the wheel which could have been adjusted on scene. It seems likely that the 50% which are acquired are indeed biased toward more severe collisions, thus limiting any inferences drawn from this data when matters of severity are involved. This is discussed in Chapter VI with regard to rollover involvements.

Finally, there are approximately 50 full MDAI cases produced per year by a team at the University of Michigan. These are selected on the basis of their interest and in no way can they be considered a statistically representative sample of the accidents in the county.

GENERAL COMMENTS

The tri-level operations have all been engaged in compiling data at several levels within closed jurisdictions, making analyses of these data, and furnishing reports of these analyses to the sponsor. Specific studies are not reported in this document, but may be read in reports from the appropriate teams. While certain conclusions with regard to accident and injury causation have resulted from these studies, and while they may provide considerable insight into these processes, we can presently not infer much about the national population from these. The data acquired are both provincial (in that they come from only a few and rather unique regions of the country), and non-representative (simply because of a lack of definition or application of random selection methods at the local level).

CHAPTER IV

DESCRIPTIVE STATISTICS

The data in the MDAI file as of October 1972, was described in some detail in a final report on the contract for automating that data. Figures 9 and 10 , reproduced from the report, display the number of cases by month and by team of occurrence as of that time. Since then approximately 1000 additional cases have been processed, although relatively few of these are from the federally sponsored teams. The entire file as of May, 1973, contains 3500 case vehicles and a total of 5500 occupants.

Appendix A presents a brief discussion of the characteristics of each team which has contributed data, as well as some descriptive statistics derived from the data file. Generally one can conclude that, while any one team is likely to be biased (relative to the national population) in many ways--e.g., too many urban accidents, too many female drivers, too many limited access highways, etc.--the sum of the data is much closer to the national averages except with respect to severity (MDAI accidents tend to be severe) and age of vehicle (by definition the teams have been asked to look at new cars). The mean age of cars in this set of data is 1.22 years, calculated simply by subtracting the model year from the accident year. The few older cars which do enter the data result principally from being the second car in a two-car collision involving a newer car; a very small number of older cars in single vehicle collisions came into the file through one of the special purpose teams studying, for example, vehicle defects.

Table 3 presents some summary statistics for the NHTSA-sponsored MDAI teams as of the time of writing this report. For selected variables a comparison is made with an available statistic from a larger or different sample. No significance computations are made; comparison with the several million accidents in the national accident summary would yield a high level of statistical

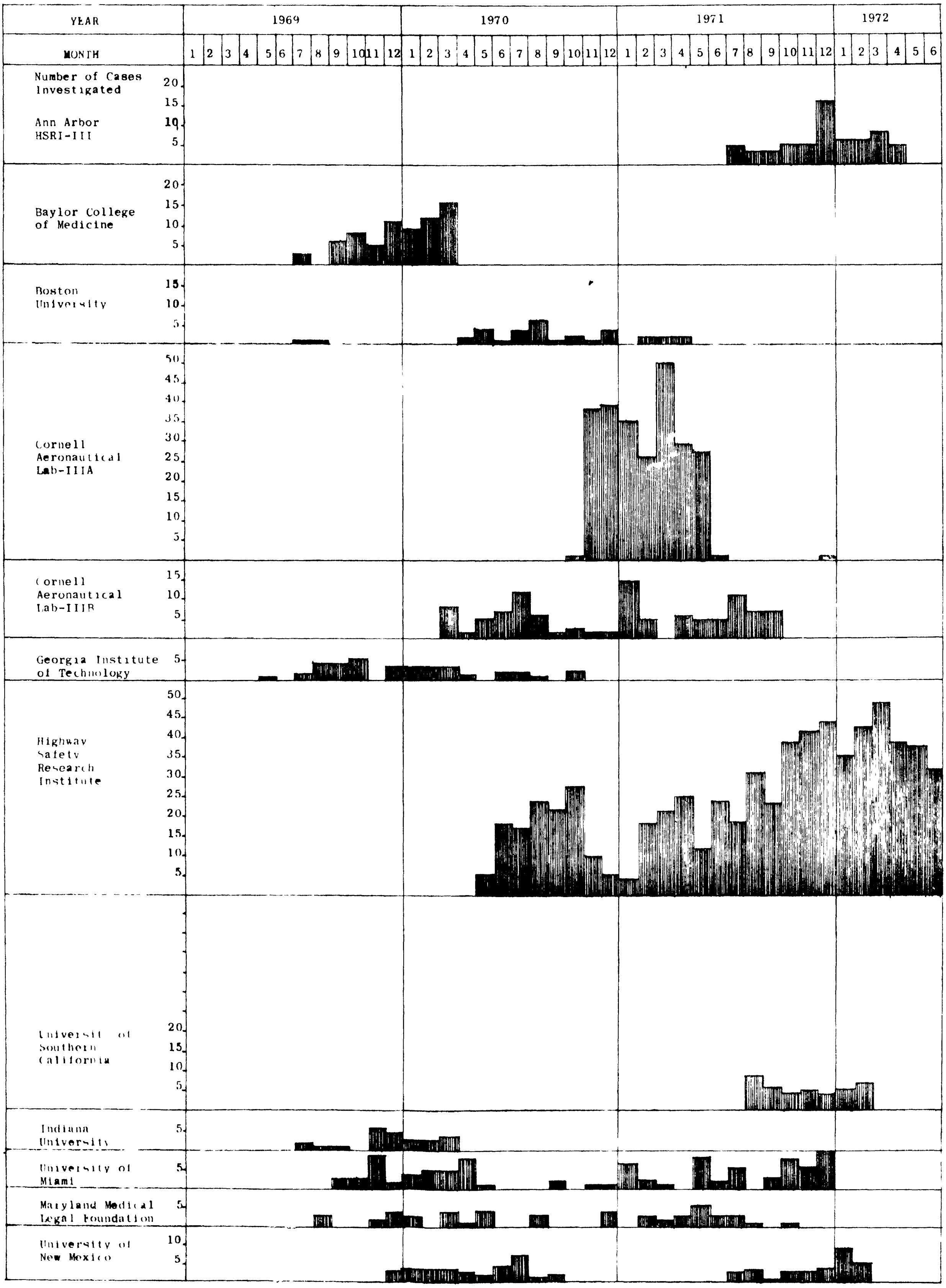


FIGURE 9
DATE OF ACCIDENT

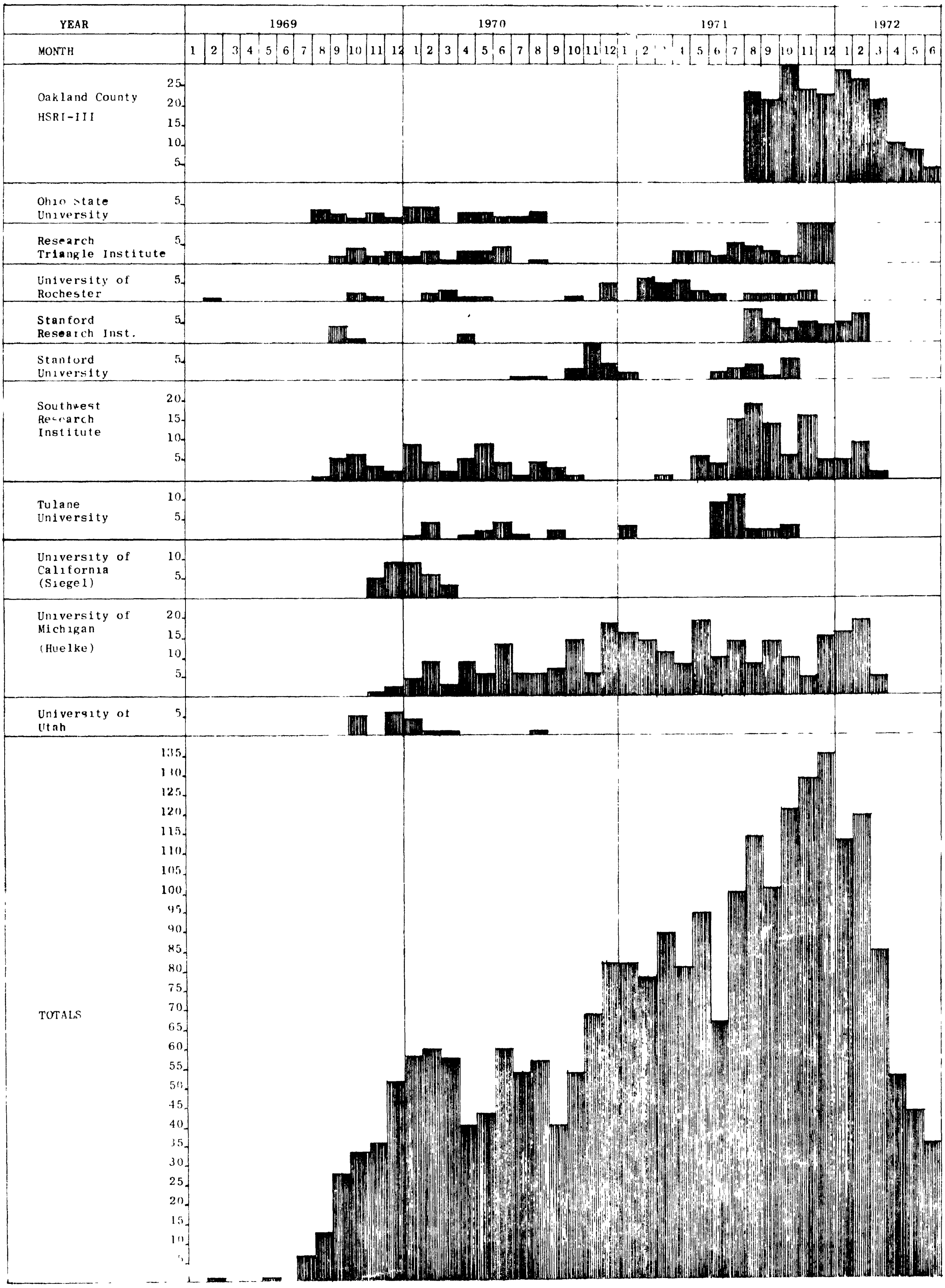


FIGURE 10
DATE OF ACCIDENT

TABLE 3. SOME SUMMARY STATISTICS FOR THE MDAI TEAMS

% Occup. wearing lap belt	18.5%
% drivers wearing lap belt	21.8%
% female occupants	35.8%
% female drivers	25.7%---(26.7% of passenger car drivers in national accident summary are females)
Average occ. inj. severity*	2.078 (AIS)
Average driver inj. severity	2.124 (AIS)
Average occupant age	30.52 years
Average driver age	34.144 years (35.7 years approximated from national accident summary)
Average occupant height	66.641 inches
Average driver height	68.334 inches
% rural accidents	33.4% (37.0% rural in the national accident summary)
% lim. access hwy. accidents	18.8%
% accidents on curves	20.0%
% accidents on slippery surf.	17.3%
% vehicles rolling over	11.9%
% accidents at night, dawn, or dusk	46.2% (35% of passenger cars in national accident summary are at night, dawn, or dusk)

*Vs. 0.89 for all occupants in the Washtenaw County subset (i.e., all towaways involving new cars whether there was an injury or not).

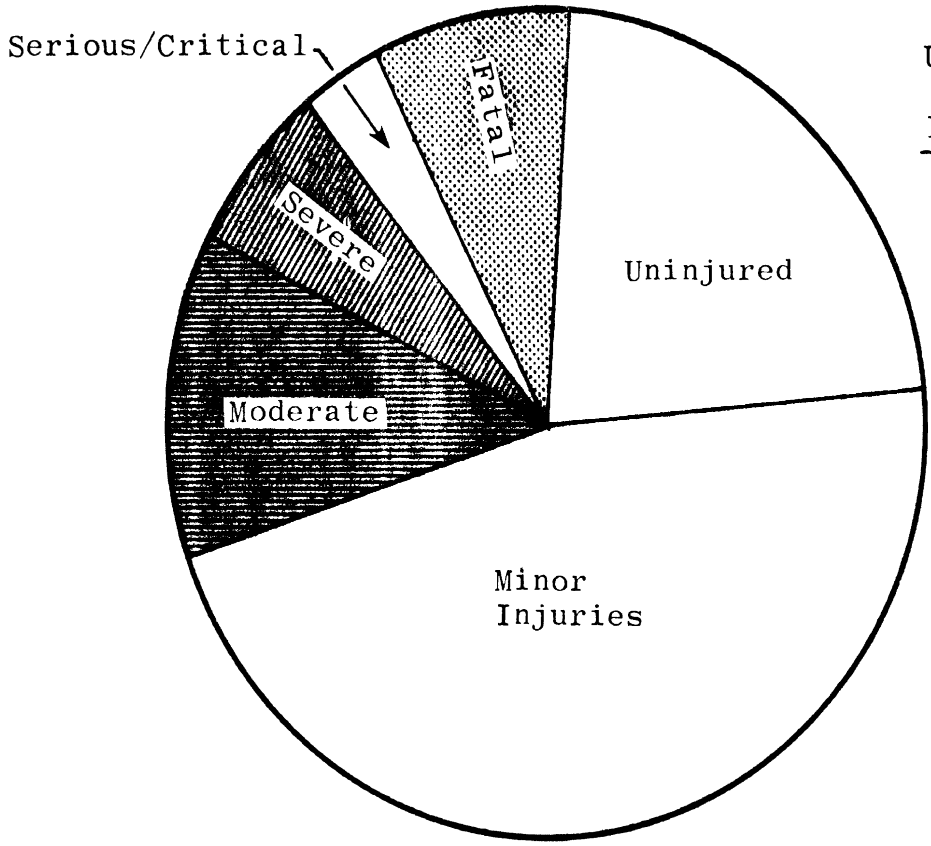
significance, but there is little practical difference between the items noted except for the percent of nighttime accidents and severity.

INJURY CAUSATION:

Injury reports for the MDAI cases are usually provided by physicians, and the reporting codes are those of the abbreviated injury scale (AIS) developed by an American Medical Association committee. The AIS is an ordered scale ranging from a value of 0 (no injury) to 9 (dead with three or more fatal injuries).

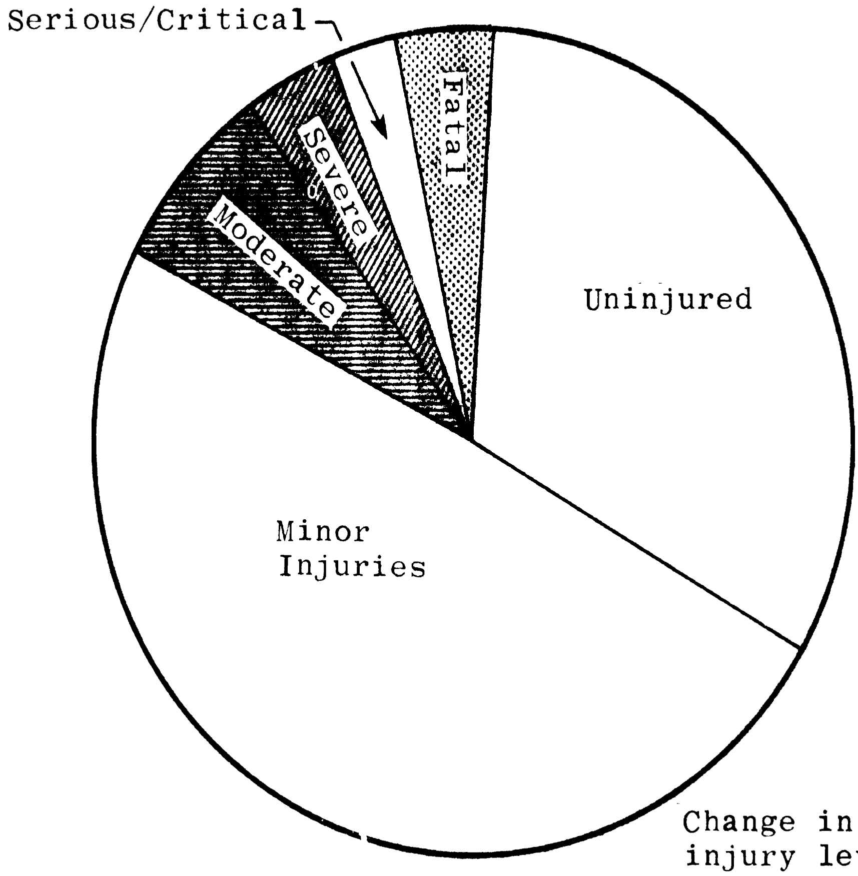
Numerical analysis with such a scale has some obvious deficiencies. Although the scale is ordered, it is not obvious that the difference between 1 (minor) and 2 (moderate) is equivalent to the difference between 4 (serious) and 5 (critical). The difference between 8 (two fatal injuries) and 9 (three or more fatal injuries) is even more confusing. Nevertheless, the average value of occupant injury calculated over some subset of the data (e.g., team) is intuitively useful. A more comprehensive comparison between two groups might be made by inspecting the full distribution of injury levels, and indeed there will be cases in which this is preferred.

Figure 11 shows the distribution of injuries using the AIS scale for occupants in this set of data who are using restraint systems as opposed to those who are not. The mean value of injury (calculated on the same scale) is also shown, once for only injury cases and once for all cases. The difference in the average AIS value is 0.310 (for all cases), and by inspection it can be seen that the shift in the distribution was not in any single cell. This data has not been adjusted for any differences in accident severity experienced by the belted and unbelted groups, and is intended here primarily as an example of the relationship between the full AIS distribution and the mean value of this ordinal variable. A comparison including such an adjustment is given in Chapter V.



Unbelted Mean Injury

Injury Only	All
1.92	1.48



Belted Mean Injury

	Injury Only	All
	1.72	1.17
Change in injury level	-0.200	-0.310

Figure 11. AIS injury distributions for restrained and unrestrained occupants.

The utility of the injury scale would be greatly expanded if one could assign some dollar benefit to an observed change. For example, if a change in a different AIS scale of 0.3 were worth $200 one might calculate a cost/benefit ratio to be associated with a countermeasure of that effectiveness.

One can imagine countermeasures which would change the form of the injury distribution but would not change the mean value. For example, some modification to the vehicle might save lives at the expense of increasing some minor injuries to the moderate level. And in this case, the mean value would not contribute much to the understanding of the pheonomenon. If the scale were linear with value and the mean did not change, one could conclude that the change produced no benefit.

The AIS scale was evidently conceived without extensive thinking about its use as an analytical variable. Rather it was a convenient form of noting something about the severity of injury in an ordered form. Its use as an analytical variable is viewed tentatively as a convenient way of comparing subsets of the data with the application of the possibility of misinterpretation, and the option of further comparisons treating it as a categorical variable.

If the injury distribution is well behaved with respect to countermeasures, i.e., if the countermeasures produce a shift in the same direction throughout the AIS scale, one must still be concerned with the weighting of the scale. This matter is discussed more completely in Appendix E, and a test of the data has been conducted by transforming the scale in several ways (making it logarithmic, exponential, etc.) and studying the effect of such change on observed differences across a number of independent variables. While the statistical firmness of conclusions varies somewhat with the different scales, there were no changes of direction induced by this method.

For individual parts of the human body, injuries are coded with the same scale; but the scale is limited to values 0 to 6. Overall injury to each occupant is coded as the maximum value to any of the parts of the body except in the case of fatalities where the scheme mentioned above is employed.

INJURIES TO OCCUPANTS OF LARGE AND SMALL CARS

Viewing the entire file of data in the light of its origin--i.e., a broad mixture of case selection rules varying over time and place--one must judge that the reported injuries are representative of a set of severe collisions (more severe than either a "towaway sample" or even "a sample of all injury accidents"). Injury information is reported in some detail with respect to the location of the injury on the body, the likely contact points within the vehicle, the type of injury (fracture, contusion, etc.), and the severity of the injury (using the Abbreviated Injury Scale) which provides six levels ranging from minor to fatal.

Intuitively one might expect to see some variation in the severity of injury in a random selection of cases between occupants of small and large cars; police level data have indicated that, given a collision, it is more likely that there will be an injured occupant in a small car than in a large one, and further that, on the average, more persons will be injured per collision in the small car. Yet if the sampling method used to put data into this file is to find those cases involving the more severe injuries it is possible that this bias will minimize any real differences between cars of different sizes. With this in mind, data on injuries in small and large cars are presented with the understanding that they are descriptive statistics from the present file. The reader is cautioned to consider the reported injury variations as descriptive statistics which deserve further confirmation before action is taken.

It is noted that there is some confusion in the reporting of injury contact points. When there are several injuries to a given part of the body, and several contact points, the method of data reporting does not permit one to differentiate between these. For example, if the head has a reported laceration and a contusion, and contact with the windshield and with the gearshift lever, it is not possible to determine which injury was occasioned by which contact. Further, investigators often report several likely contact points, and they are not able to report (in the present form) their judgement as to whether these are certain, probable, or possible.* In spite of these deficiencies in both the sampling and the reporting methodology, the general distribution of injuries provides some insight into what happens to the occupants of vehicles involved in relatively serious accidents. The tradeoff to be considered is that a complete census of police injury reports identifying injuries as "A", "B", or "C" tells us very little about what parts of the body are injured and how. The present MDAI data, with its unfortunate biases, does provide sufficient detail to see what kinds of injuries occur in this set of real accidents.

Injury severity has been used as a numeric dependent variable in these tables. The minimum value is 1, representing a minor injury), and the maximum value is 6, representing a fatal injury to that part of the body. Table 4 displays the number of injuries by part of the body and the mean value of the injury level for each noted contact point within the car (or exterior to the car in a few cases). Heavy cars have been defined as those with a curb weight of more than 3100 pounds; light cars are those weighing 3100 pounds or less. The table is presented without comment except to note that there is a total of approximately 12,000 injuries tabulated, and the relative frequency of combinations of struck components and parts of the body may be computed from this.

---

* This matter is more fully discussed in "Existing Accident Injury Causation Data Recording and Proposed Improvements" Marsh, J.C., proceedings of 16th annual AAAM conference, held at University of North Carolina October 18-21, 1972.

Table 4A

Part of Body Injured vs. Contact Point

	Internal Organs				Brain				Face			
	N-133		N-194		N-187		N-322		N-790		N-1271	
	Light Car		Heavy Car		Light Car		Heavy Car		Light Car		Heavy Car	
	$\bar{Y}$	N	$\bar{Y}$	N	$\bar{Y}$	N	$\bar{Y}$	N	$\bar{Y}$	N	$\bar{Y}$	N
Front												
05 - Instrument Panel	5.08	26	5.47	19	3.27	15	2.21	19	1.28	79	1.43	139
09 - Steering Assembly	5.18	49	5.12	89	3.00	8	2.91	23	1.32	137	1.30	364
12 - Windshield					2.66	32	3.14	64	1.57	238	1.32	320
02 - Glove Compartment	6.00	1	4.20	5	4.33	3	2.00	3	1.33	6	1.00	22
03 - Hardware Items					5.00	1			2.00	2	1.00	3
04 - Heater or AC Ducts			6.00	1	5.00	2			1.67	3		
01 - AC or Vent Outlets			6.00	2			6.00	3	1.00	3		
06 - Mirrors					2.50	6	1.75	12	1.20	60	1.16	50
07 - Parking Brake												
08 - Radio												
10 - Sunvisors & Fittings	4.00	1			2.73	15	2.00	21	1.68	65	1.49	65
11 - Trans. Selector Lever	4.00	2									3.00	2
53 - Parcel Tray											1.00	1
Sides												
20 - Surface of Side Interiors	4.78	18	4.03	34	4.44	16	5.13	8	1.15	27	1.42	19
19 - Hardware											1.00	1
13 - Armrests			3.25	4								
22 - Window Glass					1.50	2	2.28	18	1.53	59	1.87	62
21 - Window Frames			6.00	2	2.00	3	4.20	10	1.33	6	1.58	19
14 - A Pillar			5.00	1	2.23	13	4.59	44	1.36	33	1.77	73
15 - B Pillar							5.00	3	1.00	1		
16 - C Pillar									1.00	1		
17 - D Pillar												
18 - Courtesy Lights												
Interior												
29 - Front Seatbacks	3.60	5	5.50	2	1.00	2	5.00	2			1.00	2
33 - Restraint System Hard.					2.00	1					2.00	1
34 - Restraint System Web.	4.00	4	4.00	1								
30 - Head Restraints							2.00	1			1.00	1
32 - Other Occupants	4.00	1	4.67	3	5.00	2	6.00	1	1.00	3	1.50	4
31 - Interior Loose Object							4.00	1			1.00	3
50 - Rear Seat					6.00	2						
51 - Front Seat Cushion											1.00	1
52 - Internal Flying Glass									1.00	6	1.13	8
Roof												
26 - Roof Side Rails			3.00	1	4.15	13	3.47	17	1.38	13	1.56	16
10 - Sunvisors & Fittings	4.00	1										
25 - Roof or Conv. Top	5.00	6	5.00	1	2.33	6	3.58	12	1.50	8	2.13	16
24 - Coat Hooks												
18 - Courtesy Lights												
Floor												
11 - Trans. Selector Lever	4.00	2										
40 - Floor							3.67	3	1.00	1		
28 - Foot Controls											1.00	1
27 - Console											2.00	1
Rear												
23 - Backlight (rear window)					2.00	1						
39 - Backlight Header												
Exterior												
35 - Hood							5.80	5			2.89	9
36 - Objects Exterior to Car	4.00	10	5.79	19	4.12	32	5.12	39	1.92	26	2.57	35
37 - Outside Surface of Car	3.83	6	4.00	3	6.00	6	6.00	3	3.71	7	1.00	2
38 - Other	5.50	4	5.86	7	6.00	5	4.38	8	1.60	5	1.85	26
98 - Impact Force, "Whiplash", Hyperextension/compression					2.50	2	3.50	2	1.00	1	1.00	5

Table 4B

Part of Body Injured vs. Contact Point

	Head				Neck (Cervical Region)				Shoulder Girdle				Right Upper Limb			
	N=532		N=794		N=210		N=372		N=170		N=248		N=283		N=434	
	Light Car		Heavy Car		Light Car		Heavy Car		Light Car		Heavy Car		Light Car		Heavy Car	
	$\bar{Y}$	N	$\bar{Y}$	N	$\bar{Y}$	N	$\bar{Y}$	N	$\bar{Y}$	N	$\bar{Y}$	N	$\bar{Y}$	N	$\bar{Y}$	N
Front																
05 - Instrument Panel	2.77	22	1.56	52	1.50	4	2.43	14	1.50	22	1.28	29	1.35	83	1.33	120
09 - Steering Assembly	1.73	15	1.60	42	2.83	24	1.89	27	1.44	16	1.20	30	1.00	46	1.34	102
12 - Windshield	1.24	160	1.60	202	2.42	12	2.60	15			1.00	4	1.16	19	1.06	16
02 - Glove Compartment	4.60	5	1.75	4					1.00	2	1.00	3	1.00	2	1.00	8
03 - Hardware Items	2.00	2									1.50	2	1.00	6	2.00	4
04 - Heater or AC Ducts	2.00	1	1.00	1												
01 - AC or Vent Outlets	2.00	1	2.67	3												
06 - Mirrors	1.05	38	1.23	57	1.00	3	1.67	3					1.29	7	1.00	5
07 - Parking Brake																
08 - Radio													2.00	1	1.00	3
10 - Sunvisors and Fittings	1.58	40	1.28	78	6.00	3	2.40	5			1.80	5				
11 - Trans Selector Lever					1.00	2							1.00	4	1.29	7
53 - Parcel Tray																
Sides																
20 - Surface of Side Interior	1.59	27	1.86	29	1.75	8	1.53	15	1.27	59	1.41	81	1.37	38	1.44	61
19 - Hardware			1.00	1							1.00	1			1.20	5
13 - Armrests													1.00	1		
22 - Window Glass	1.73	56	1.59	56	1.00	4	1.50	6	1.00	5	1:00	6	1.14	7	1.00	15
21 - Window Frames	1.53	15	1.64	11	1.00	4			1.00	8	1.00	2	1.00	4		
14 - A Pillar	1.75	24	2.15	59	1.57	7	4.25	8			1.29	7	2.80	5	2.25	4
15 - B Pillar	1.00	2	2.67	3					1.00	2						
16 - C Pillar																
17 - D Pillar																
18 - Courtesy Lights																
Interior																
29 - Front Seatbacks	1.00	3			1.85	13	1.32	22	1.00	3	1.18	11			1.67	3
33 - Restraint System Hard.	1.00	1	1.33	3												
34 - Restraint System Web.					1.00	6			1.00	3	1.00	2				
30 - Head Restraints	1.00	7	1.00	2	1.00	8	1.77	22			1.00	3				
32 - Other Occupants	1.00	5	2.00	5			1.50	4	1.50	4	1.50	8	1.00	1	1.14	7
31 - Interior Loose Object			1.00	8			1.00	3			1.00	3			1.00	1
50 - Rear Seat	6.00	1														
51 - Front Seat Cushion																
52 - Internal Flying Glass	1.00	4	1.00	6	1.00	1	1.00	1					1.00	7	1.00	12
Roof																
11 - Trans. Selector Lever																
40 - Floor	1.00	1	1.00	2			1.00	1			1.00	1	1.00	1	4.00	1
28 - Foot Controls																
27 - Console							1.00	1	1.00	1	1.00	1	1.00	1	1.00	3
Rear																
23 - Backlight (rear window)	1.00	3	1.00	1												
39 - Backlight Header			5.00	2												
Exterior																
35 - Hood			4.73	11												
36 - Objects Exterior to Car	4.00	32	4.30	50	4.60	5	4.06	17	1.53	17	2.05	20	),28	29	1.93	28
37 - Outside Surface of Car	6.00	2	3.00	10	2.00	3	6.00	1	2.00	3			2.00	3	1.00	1
38 - Other	4.88	8	3.56	18	1.30	10	1.35	26	1.33	6	1.20	5	1.00	8	1.73	11
98 - Impact Force, "Whiplash", Hyperextension/Compression	1.00	3	1.00	4	1.22	90	1.20	171	1.09	11	1.22	18	1.14	7	1.00	1

Table 4C

Part of Body Injured vs. Contact Point

	Left Upper Limb N-320 Light Car		Left Upper Limb N-457 Heavy Car		Chest & Upper Back (Thorax) N-447 Light Car		Chest & Upper Back (Thorax) N-790 Heavy Car		Lower Back (Lumbar Region) N-90 Light Car		Lower Back (Lumbar Region) N-138 Heavy Car		Abdomen N-125 Light Car		Abdomen N-187 Heavy Car	
	Ȳ	N	Ȳ	N	Ȳ	N	Ȳ	N	Ȳ	N	Ȳ	N	Ȳ	N	Ȳ	N
Front																
05 - Instrument Panel	1.19	64	1.23	95	2.00	64	2.55	98			1.00	8	3.11	18	2.77	13
09 - Steering Assembly	1.42	36	1.16	89	2.20	231	2.16	394	2.50	4	1.00	3	2.31	49	1.51	74
12 - Windshield	1.00	10	1.00	11	1.00	3	2.50	2								
02 - Glove Compartment	2.00	2	1.57	7	1.80	10	2.40	10	1.00	1			1.00	3		
03 - Hardware Items	1.00	6	1.11	9			2.00	1								
04 - Heater or AC Ducts																
01 - AC or Vent Outlets							2.33	3								
06 - Mirrors	1.00	3	1.00	3			4.00	2					5.00	1		
07 - Parking Brake																
08 - Radio																
10 - Sunvisors & Fittings																
11 - Trans. Selector Lever	1.00	5	1.00	1	1.67	6	3.00	1	1.50	2	3.00	1				
53 - Parcel Tray																
Sides																
20 - Surface of Side Interior	1.52	86	1.26	117	2.94	53	3.09	139	1.00	5	1.19	16	3.67	6	3.46	13
19 - Hardware	1.00	2	1.29	7	4.00	1	1.33	3			2.00	2				
13 - Armrests	1.00	7	1.20	10	2.00	3	2.14	7			1.67	3	1.00	4	2.00	7
22 - Window Glass	1.00	16	1.29	35	1.00	3	1.00	1	1.00	1			1.00	2		
21 - Window Frames	1.86	14	1.83	6			3.00	2	2.00	3			1.00	1	1.50	2
14 - A Pillar	1.00	5	1.45	11	2.33	3										
15 - B Pillar	1.67	3	2.00	4			2.50	2								
16 - C Pillar																
17 - D Pillar																
18 - Courtesy Lights																
Interior																
29 - Front Seatbacks			2.50	2	1.13	8	1.80	10	1.18	11	1.25	16				
33 - Restraint System Hard									1.00	3			1.38	8	1.50	12
34 - Restraint System Web.					1.00	4	3.25	8	1.50	6	2.83	6	1.80	20	1.59	32
30 - Head Restraints					2.00	1										
32 - Other Occupants	1.86	7	1.50	2	2.29	7	1.88	17			1.00	1	1.00	1	2.67	3
31 - Interior Loose Object					1.00	1	1.00	1							1.00	1
50 - Rear Seat																
51 - Front Seat Cushion							1.00	1			1.00	1				
52 - Internal Flying Glass	1.00	8	1.00	6	1.00	1	1.00	1								
Roof																
26 - Roof Side Rails	1.00	1			1.00	1	2.00	2			3.00	1				
10 - Sunvisors & Fittings																
25 - Roof or Conv. Top	1.75	4	1.25	4	2.33	3	4.00	7	3.33	6						
24 - Coat Hooks																
18 - Courtesy Lights																
Floor																
11 - Trans. Selector Lever																
40 - Floor							1.00	1					6.00	2		
28 - Foot Controls																
27 - Console							2.00	1			1.00	2				
Rear																
23 - Backlight (Rear Window)									1.00	2						
39 - Backlight Header																
Exterior																
35 - Hood			2.00	1												
36 - Objects Exterior to Car	1.41	22	1.50	16	1.67	12	4.50	38	1.09	11	1.63	8	2.83	6	3.78	23
37 - Outside Surface of Car	1.88	8			5.00	5	3.78	9							6.00	1
38 - Other	1.00	7	2.79	14	2.86	7	1.92	12	1.00	5	1.56	16	2.50	2	2.25	4
98 - Impact Force, "Whiplash" Hyperextension/Compression	1.00	4	1.00	6	1.50	10	1.12	17	1.30	30	1.24	54	6.00	2	1.50	2

Table 4D

Part of Body Injured vs. Contact Point

	Pelvic Girdle				Right Lower Limb				Left Lower Limb				Whole Body			
	N = 143		N = 235		N = 572		N = 848		N = 564		N = 820		N = 46		N = 64	
	Light Car		Heavy Car		Light Car		Heavy Car		Light Car		Heavy Car		Light Car		Heavy Car	
	$\bar{Y}$	N	$\bar{Y}$	N	$\bar{Y}$	N	$\bar{Y}$	N	$\bar{Y}$	N	$\bar{Y}$	N	$\bar{Y}$	N	$\bar{Y}$	N
Front																
05 = Instrument Panel	3.00	19	2.15	34	1.29	269	1.40	478	1.30	278	1.38	421	1.50'	6	1 00	7
09 = Steering Assembly	2.05	22	1.58	33	1.33	43	1.28	43	1 18	39			1.00	2	1.00	5
12 = Windshield	1.00	1					2.00	2			2.00	3	1.00	2	1.00	1
02 = Glove Compartment	1.40	5	4.00	1	1.11	28	1.32	47	1.33	15	1.00	28			1.00	1
03 = Hardware Items																
04 = Heater or AC Ducts			3.00	1	1.60	15	2.41	17	1.00	10	2.33	12				
01 = AC or Vent Outlets			3.00	1	1.67	6	1.50	8	2.00	3	1.14	7			2.00	2
06 = Mirrors					1.00	3			1.00	2						
07 = Parking Brake					1.00	1			1.45	15	1.19	21				
08 = Radio					1.00	5	1.33	6	1.50	4	1.00	3				
10 = Sunvisors and Fittings							1.00	1								
11 = Trans. Selector Lever	1.00	4	1.50	6	1.00	9	1.29	14	1.00	10	1.14	7				
53 = Parcel Tray					1.00	4			1.60	5						
Sides																
20 = Surface of Side Interior	2.09	44	1.86	59			1.91	43	1.87	46	1.52	65	1.00	6	1.60	5
19 = Hardware					2.00	3	1.00	2	1.00	10	1.18	11				
13 = Armrests	1.57	7	1.46	24	1.00	1	1.00	5	1.00	3	1.22	9				
22 = Window Glass			1.00	2			1.50	2	1.00	3	1.00	3	1.00	1	1.00	2
21 = Window Frames	2.33	3	1.00	1	1.00	2	2.00	2	1.00	3						
14 = A Pillar					2.00	4			1.75	4						
15 = B Pillar																
16 = C Pillar																
17 = D Pillar																
18 = Courtesy Lights																
Interior																
29 = Front Seatbacks	2.00	1	2.50	2			1.00	2	1.00	2	1.00	1	1.00	1		
33 = Restraint System Hard.	1.00	2	1.00	6												
34 = Restraint System Web.	1.26	19	1.57	23	1.33	3	2.29	7	1.00	1	2.50	6			2.00	1
30 = Head Restraints																
32 = Other Occupants	2.00	3	1.00	5	1.00	4	1.00	6	2.20	5	1.33	3	1.00	2		
31 = Interior Loose Object			1.00	1	1.00	1	1.00	1	1.00	1						
50 = Rear Seat																
51 = Front Seat Cushion			1.00	2					1.00	1					1.00	1
52 = Internal Flying Glass							1.00	2							1.00	2
Roof																
26 = Roof Side Rails																
10 = Sunvisors & Fittings					1.00	1										
25 = Roof or Conv. Top			3.00	1			1.00	3	1.00	1	1.00	2			1.00	3
24 = Coat Hooks							1.00	1								
18 = Courtesy Lights											1.00	1				
Floor																
11 = Trans. Selector Lever																
40 = Floor	3.00	2	2.00	2	2.26	27	2.03	33	1.95	22	2.17	42				
28 = Foot Controls					1.69	36	1.53	34	1.59	17	1.78	23				
27 = Console	1.00	2	1.00	2	1.00	2	1.00	4	1.00	3	1.40	5				
Rear																
23 = Backlight (Rear Window)			1.00	1												
39 = Backlight Header																
Exterior																
35 = Hood																
36 = Objects Exterior to Car	1.43	7	2.47	15	1.80	20	1.10	20	2.36	22	2.10	31	1.00	9		
37 = Outside Surface of Car	3.00	1	4.33	3			1.00	2	5.00	1	2.33	3				
38 = Other	1.00	1	1.00	3	1.10	30	1.80	15	1.25	24	1.54	26	3.00	5	3.56	9
98 = Impact Force, "Whiplash", Hyperextension/Compression			1.60	5	1.00	3	1.20	5	1.00	2	2.20	5	1.00	12	1.00	22

Occupants of the heavy cars in the file sustained 7397 recorded injuries; occupants of light cars sustained 4601. Table 5 displays the percentage of these injuries (at all points of the body) attributable to each of the major contributors, along with the average injury level computed using the Abbreviated Injury Scale.

Table 5

Average injury level from contact with various parts of the vehicle, and percent of contacts with that part. Light cars <3100#, heavy cars >3100#.

	Heavy Cars	7397 Injuries	Light Cars	4601 Injuries
FRONT				
Instrument panel	25.0	1.30	25.5	1.58
Steering assembly	17.7	1.87	15.8	1.98
Windshield	8.6	1.72	10.2	1.52
Glove compartment	1.9	1.42	1.8	1.66
Hardware items	1.3	1.14	1.2	1.18
Mirrors	1.8	1.29	2.6	1.24
Sunvisors and fittings	2.5	1.44	2.6	1.90
Trans. selector lever	0.5	1.46	1.4	1.25
SIDES				
Surf. of side inter.	9.5	2.02	9.6	2.07
Window glass	2.8	1.76	3.5	1.47
Window frames	0.6	1.48	1.4	1.42
A-pillar	2.8	2.57	2.1	1.73
INTERIOR				
Front seatbacks	1.0	1.61	1.0	1.57
Rest. System webbing	1.2	2.34	1.4	1.56
Other occupants	0.9	1.74	1.0	1.82
ROOF				
Roof side rails	0.9	2.39	1.0	2.78
Roof (or conv. top)	1.5	2.54	1.7	1.85
FLOOR				
Floor	1.2	2.12	1.2	2.21
Foot controls	0.6	1.62	1.1	1.66
EXTERIOR & MISCELLANEOUS				
Objects exterior to car	4.8	3.34	5.6	2.31
Outside surface of car	0.5	3.45	1.0	2.55
Impact force	4.2	1.22	3.8	1.28
OTHER	8.2	1.66	3.5	1.53
AVERAGE INJURY LEVEL		1.784		1.740

Heavy cars	Light cars
1.40 occupants/veh	1.36 occupants/veh
1589 vehicles	1095 vehicles
2217 occupants	1484 occupants
1706 occupants with inj.	987 occupants with inj.
1.07 inj. persons/veh.	.9 inj. persons/veh.

ACCIDENT CAUSATION:

The MDAI cases generally contain much detail about the accident circumstances, and include an opinion of the investigator about causes or contributing factors. While these causes are noted in the discussion they have not been systematically coded into the digital file, principally because of a lack of consistency in reporting. A certain amount of factual information which may be related to the causation problem is encoded, partly in the original CPIR form and partly in supplementary coding from the text of the reports.

The CPIR file includes reporting of vehicle malfunctions, weather conditions, visibility obstructions, and driver impairment for example. The supplementary codes report such items as trip origin (home, work, bar, etc.), marital status, type of driver education, occupation class, and hazardous road conditions.

To arrive at any conclusions regarding accident causation from an accident data file it is necessary to have some idea of the exposure information associated with the same group of drivers. For example, if 15% of the drivers in this set of accident data were drinking, and if only 5% of the drivers in an exposure set sampled from the same population were drinking, we might conclude that drinkers were overrepresented in the accident population. With the lack of precise definition of the sampling involved in creating this set of data it does not seem appropriate to try to construct an exposure data set. It is possible to look at the distributions of those factors which may be considered "causative" within this data set, simply to characterize it; but one should be wary of concluding that, for example, males are more accident prone than females.

There are some accident causation factors noted which are identified as direct and sole causes--a wheel fell off the vehicle, or the driver had a fatal heart attack. These may be taken as anecdotal information about accident causation, but, again, because of the lack of an exposure data set, should not be viewed as representative of the frequency of such phenomena.

With all of these cautionary remarks, we now present some descriptive statistics of the MDAI file with regard to causative factors.

INDUCED EXPOSURE:

A method which has been employed to account for exposure in the past is to use information within the accident file itself. This has generally been called an "induced exposure" method, and has been discussed by Hall[1], Thorpe[2], and Carr[3]. In principal one uses information within the accident data set itself to determine culpability of the drivers, and then compares other driver or vehicle characteristics as a function of culpability. In previous studies culpability was determined by such devices as identifying those drivers who were given citations by the police, and then comparing their characteristics (such as age) against those drivers who were not cited in connection with the accident.

In editing of the MDAI reports two separate editors are employed. The first does the initial detailed editing of the various forms, essentially preparing the information in a consistent form for

----

[1] "An Empirical Analysis of Accident Data Using Induced Exposure", W.K. Hall, Hit-Lab Reports, September 1970.

[2] "Calculating Relative Involvement Rates in Accidents Without Determining Exposure", by J.D. Thorpe, Australian Road Research, September 1964.

[3] "A Statistical Analysis of Rural Ontario Traffic Accidents Using Induced Exposure Data" by B.R. Carr, Symposium of the Use of Statistical Methods in the Analysis of Road Accidents, Road Research Laboratory, April 1969.

placing it in the computer. The second editor, a check editor or supervisor, reviews the entire case and encodes a supplementary code sheet. One of the items on this supplementary sheet is the assignment of responsibility for the accident. The editor looks at the whole accident and assigns (to each vehicle involved), the order of responsibility. In a three vehicle accident, for example, one driver will be noted as the first most responsible, another as the second, and another as the third. The rule by which this assignment is made is that the last driver who could have done something to avoid the accident is listed first, etc.

On rare occasions the driver in a single vehicle accident may not be considered the first most responsible, for example in a "phantom" vehicle accident.

An adaptation of the induced exposure method has then been applied to this set of data. A new binary variable was created with values (1) when the case vehicle driver was the most responsible for this accident and (2) when the case vehicle driver was not the most responsible for this accident. Looking then at any subset of the file, one can observe the degree of culpability associated with that group of cars or drivers.

The Automatic Interaction Detection program (AID) was used for this analysis. Culpability or responsibility is the dependent variable, and a number of vehicle and driver factors were chosen as independent variables. These included trip origin, marital status, occupation category, worst injury in vehicle, driver impairment, odometer reading, type of driver education, familiarity with route, driver age, weight and height, and whether or not a restraint system was worn by the driver.

The AID diagram is presented in Figure 12 , and shows that the first split was between drivers who were noted as impaired (in some way) and those who were not. The impaired group splits a second time on the same variable. Please note that the editor's assignment

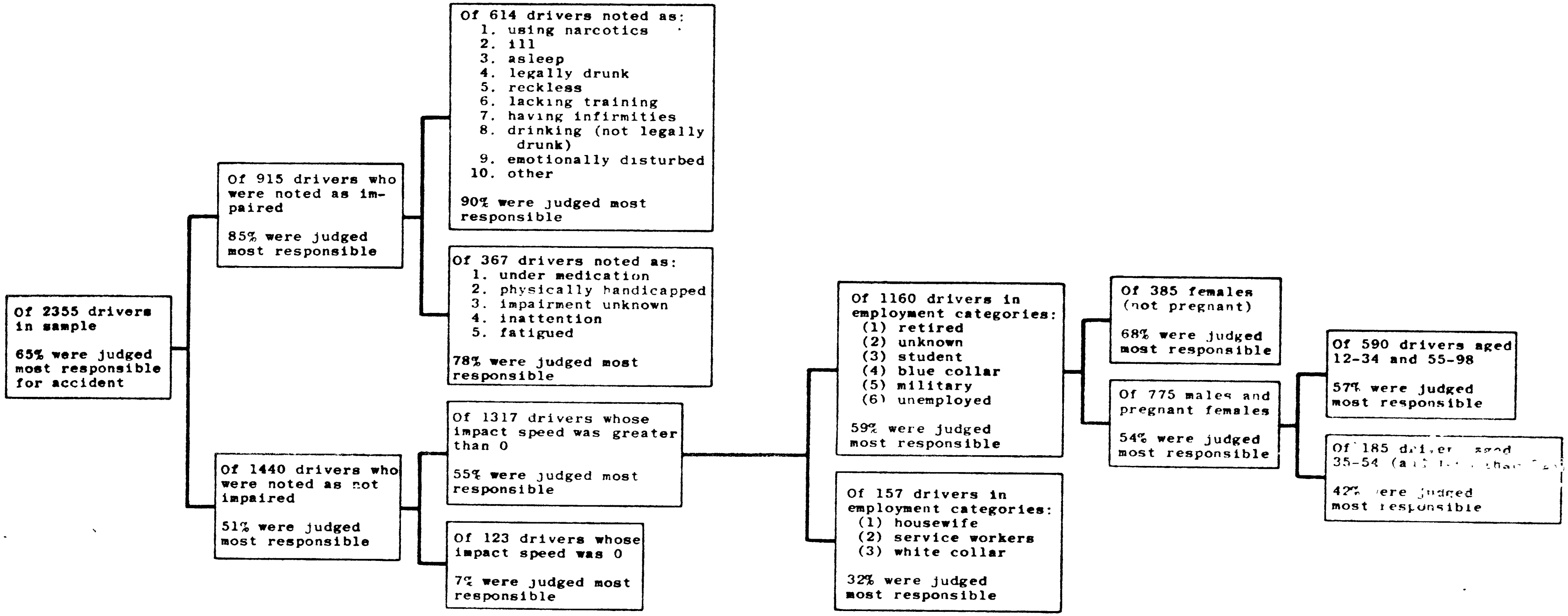


Figure 12. AID diagram for driver responsibility in accident.

of responsibility is not expected to be based on his knowledge of impairment, but simply the assessment of the last vehicle which could have avoided the collision.

Tables 6 through 17 present the one way relationships between proportion most responsible and the various accident "causation" variables. These are presented as descriptions of the accident cases rather than as evidence of causation. F statistics are shown for each table together with the associated significance level. It should be emphasized that the significance levels apply only on an individual table basis. Thus for example one can conclude that responsibility is significantly different at the 0.01 level for, Origin of Trip, or for, Marital State, taken individually. However the joint significance level of both variables is not 0.01.

SUMMARY OF THE INDUCED EXPOSURE APPROACH

The lack of sampling methodology in the file leads to constraints in inferring from this kind of causation analysis to a national population. The method employed to study causation, i.e., the method of induced exposure, was decided upon in retrospect after looking at the data. If there had been an initial plan to employ this method there might have been a more precise specification of "responsibility" or "impairment" than presently exists.

It is noted that the relatively hard data reported in the MDAI cases is likely to be correct and precisely reported. The number of cylinders in the engine, the driver age, whether or not the windshield was broken or the tire deflated must be viewed as generally accurate information. Instructions for reporting these harder facts are more complete. Many of the softer data elements seem to have been added without clear definitions--for example how do different reporters differentiate between an emotional state, lack of training, recklessness, and inattention?

In spite of these difficulties it is informative to study the collection of investigations using this induced exposure technique. But the reader must keep in mind that the conclusions drawn are with respect to whatever population this set of data is a sample of, and that that population is still (and perhaps forever) imprecisely known.

Therefore conclusions based upon these kinds of analyses are somewhat anecdotal and are best used to develop hypotheses that can be tested by collecting data in a more rigorous manner.

Table 6

Driver Education Levels vs. Percent Responsible Drivers

DRIVER EDUCATION	NUMBER	% RESPONSIBLE DRIVERS
Commercial	40	68
Unknown	1607	66
High School	224	65
Yes, but type unknown	21	62
None	356	61
Informal	65	60
Professional	15	60
Other	5	60
Military	22	32

Not significant @ 10% level

Table 7

Lap Belt Worn vs. Percent Responsible Drivers

LAP BELT WORN	NUMBER	% RESPONSIBLE DRIVERS
No	1701	63
Yes	523	62
Not applicable (i.e. no belts installed or available)	69	59
Unknown if worn	62	53

F-ratio (neglecting unknowns) = 1.871 Not sig. @ 10% level

Table 8

Driver Impairment vs. Percent Responsible Drivers

DRIVER IMPAIRMENT	NUMBER	% RESPONSIBLE DRIVERS
Drugs (narcotic)	4	100
Illness	13	100
Asleep	29	97
Drunk (by local legal standards)	120	94
Recklessness	28	89
Lack of training	18	89
Infirmities	8	87
Drinking	351	87
Emotional state	36	86
Other	7	87
Medication	6	83
Physically handicapped	5	80
Unknown if impaired	113	80
Inattention	156	77
Fatigue	21	71
Not impaired	1440	51

F-ratio (neglecting unknown and not impaired) = 2.172 Sig. @ $<$1%

Table 9

Sex vs. Percent Responsible Drivers

SEX	NUMBER	% RESPONSIBLE DRIVERS
Female	673	66
Male	1670	64
Pregnant female	8	38

F-ratio = 1.45 Not. sig. @ 10% level

Table 10

Trip Plan-Route Usage vs. Percent Responsible Drivers

TRIP PLAN-ROUTE USAGE	NUMBER	% RESPONSIBLE DRIVERS
Annually (1-3 times)	21	90
Never before	35	83
Less than annually	16	75
Unknown	1803	66
Quarterly (1-2 times)	8	63
Weekly (1-4 times)	201	61
Monthly (1-3 times)	60	57
Daily	211	53

F-ratio (neglecting unknown) = 3.740 Sig. @ $<$ 1% level

Table 11

Driver Age vs. Percent Responsible Drivers

DRIVER AGE	NUMBER	% RESPONSIBLE DRIVERS
12-16 years	64	84
17	82	74
64-98	109	72
1 month to 12 yrs.	20	70
22-24	337	68
20-21	247	68
18-19	226	66
55-63	140	64
30-34	205	62
25-29	334	61
35-44	315	59
45-54	276	59

F-ratio = 2.892   Sig. @ $<$1% level

Table 12

Speed at Impact vs. Percent Responsible Drivers

SPEED AT IMPACT	NUMBER	% RESPONSIBLE DRIVERS
81-90 mph	3	100
91-100	1	100
71-80	15	93
unknown	234	82
51-60	84	80
61-70	49	80
41-50	171	73
1-10	457	68
31-40	274	66
11-20	549	63
21-30	383	59
stopped	135	11

F-ratio = 23.8 (neglecting the unknown category)   sig. @ $<$1%

Table 13

High-Performance Vehicle vs. Percent Responsible Drivers

HIGH-PERFORMANCE VEHICLE	NUMBER	% RESPONSIBLE DRIVERS
Yes	234	71
No	2091	64

F-ratio = 5.227 Sig. @ 1% level

Table 14

Body Style vs. Percent Responsible Drivers

BODY STYLE	NUMBER	% RESPONSIBLE DRIVERS
Convertible	80	74
Other (bus, jeep)	10	70
2-door sedan or coupe	662	67
2-door hardtop	892	66
Truck (small)	88	64
4-door hardtop	135	63
station wagon	158	59
4-door sedan	294	59
Van	36	43

F-ratio = 1.966 Sig. @ 5% level

Table 15

Origin of Trip vs. Percent Responsible Drivers

ORIGIN OF TRIP	NUMBER	% RESPONSIBLE DRIVERS
Cocktail lounge/ bar/wet party	74	91
Friends/relatives	145	75
Church	4	75
Recreation	61	74
Other	34	68
Unknown	1450	64
Shopping	76	61
School	37	59
Home	264	58
Work	210	56

F-ratio (neglecting unknown) = 5.53 Sig. @ $<$ 1%

Table 16

Marital State vs. Percent Responsible Drivers

MARITAL STATE	NUMBER	% RESPONSIBLE DRIVERS
Common law	3	100
Separated	29	83
Widowed	23	83
Divorced	47	74
Single	340	71
Unknown	1356	65
Married	557	57

F-ratio (neglecting unknown) = 6.88 Sig. @ $<$ 1%

Table 17

Occupation vs. Percent Responsible Drivers

OCCUPATION	NUMBER	% RESPONSIBLE DRIVERS
Military	19	79
Retired	25	76
Farm Workers	3	75
Unemployed	20	70
Student	97	68
Blue collar	136	66
Unknown	1710	66
White collar	234	56
Housewife	48	54
Service workers	62	53

F-ratio (neglecting unknown) = 1.88 Sig. @ $<$ 10%

FREQUENCY DISTRIBUTIONS OF CRASH CAUSATION VARIABLES

Tables 18 through 26 contain the frequency distributions of variables from the CPIR file which might be loosely defined as crash causation variables. Such information as type of driver impairment and type of vehicle defect are examples. This information is generally reported by the investigator in a relatively unstructured way, and has generally been encoded into the file from comments in the text of the accident report. They can be used to identify a given set of crashes for further study--for example if one were interested in studying those crashes in which driver alcohol impairment was indicated, he could select those by letting the computer identify all cases at that level of the "impairment" variable.

Due to the non-random nature of the file it is not possible to infer that the amount of alcohol involvement in crashes in this data is representative of any other definable population. One might get some insight into the accident process by comparing, say the extent of belt wearing among drinking and non-drinking drivers. This same data recorded for a proper sample (and consistently reported) might lead to some understanding of the relative frequency of various causative factors.

But for the present we tabulate here some further descriptive statistics of this file which are intended only to give the reader some knowledge of the several kinds of accident causation information in the file.

It should be noted that there are a number of unreported or unknown responses for all of these "causation" variables. This adds another possibility for bias which must be kept in mind.

Table 18

Crash Causation Analysis
Frequency Distribution of Case Vehicle* Driver's Ability to Drive Impaired for Crash Involved Vehicles in CPIR File

Category	Frequency	Percentage	Percentage of Reported Responses
Unknown	290	5.1	---
None	4148	73.5	---
Drinking Involved	444	7.9	36.8
Drunk by Legal Stds.	137	2.4	11.3
Asleep	61	1.1	5.1
Fatigue	63	1.1	5.2
Recklessness	84	1.5	7.0
Inattention	214	3.8	17.4
Lack of Training	36	0.6	3.0
Emotional State	70	1.2	5.8
Medication	11	0.2	0.9
Drugs (Narcotic)	14	0.2	1.2
Illness	21	0.4	1.7
Infirmities	17	0.3	1.4
Physically Handicap	9	0.2	.8
Other	25	0.4	2.1

* This is a multiple response variable, so that a particular driver could be identified with up to four of the categories--e.g., untrained, emotionally upset, drunk, and reckless at the same time. The counts and percentages are based on the sum of all reports.

Table 19

Crash Causation Analysis
Frequency Distribution of Driver Stress That Day
For Crash Involved Vehicles in CPIR File

Category	Frequency	Percentage	Percentage of Reported Responses
Unknown	2038	72.2	---
Argument with Relations or Friends	28	1.0	24.4
Argument with Boss or Co-worker	6	0.2	5.2
Loss of Friend or Relative	2	0.1	1.7
Financial Difficulty	6	0.2	5.2
School Problems/Work Problems	25	0.9	21.7
Legal/Police Problems	5	0.2	4.4
Social Agency/Counselor Problems	2	0.1	1.7
Other	41	1.5	35.7
None	669	23.7	---

Table 20

Crash Causation Analysis
Frequency Distribution of Permanent Physiological
Condition for Crash Involved Vehicles in CPIR Files

Category	Frequency	Percentage	Percentage of Reported Responses
Unknown	1229	43.6	---
Infirmities (Arthritis, Senility, etc.)	13	0.5	9.3
Diabetes	13	0.5	9.3
Brain (Epilepsy, Stroke)	9	0.3	6.4
Cardio-Vascular (Heart failure, Angina, Infection)	17	0.6	12.1
Vision/Hearing Restricted	55	1.9	39.3
Respiratory Condition	1	0.0	0.7
Paraplegic, Amputee	1	0.0	0.7
Other	31	1.1	22.1
None	1453	51.5	---

Table 21

Crash Causation Analysis
Frequency Distribution of Transient Physiological
Condition for Crash Involved Vehicles in CPIR Files

Category	Frequency	Percentage	Percentage of Reported Responses
Unknown	573	20.3	---
None	1640	58.1	---
Blackouts	15	0.5	2.5
Dozing	28	1.0	4.6
Fatigue	41	1.5	6.7
Drunk	141	5.0	23.2
Drinking Involved	336	11.9	55.2
Drug or Medication	18	0.6	3.0
Flu, Head cold, etc.	13	0.5	2.1
Pregnancy	10	0.4	1.6
Hangover	1	0.0	0.2
Not Wearing Corrective Lenses	1	0.0	0.2
Other	5	0.2	0.8

Table 22

Crash Causation Analysis
Frequency Distribution of Non-Impact Medical Condition
For Crash Involved Vehicles in CPIR Files

Category	Frequency	Percentage	Percentage of Reported Responses
None	2697	95.6	---
Yes - time and type unknown	2	0.1	4.8
Pre-crash fatal (Clinical Death at Wheel)	2	0.1	4.8
Pre-Crash Non-Fatal (Prior injury, stroke)	21	0.7	50.0
Post-Crash Fatal (Drowning)	10	0.4	23.8
Post Crash Non-Fatal	7	0.2	16.7
Unknown	83	2.9	---

Table 23

Crash Causation Analysis
Frequency Distribution of Pharmacological Agents
Noted for Crash Involved Vehicles in CPIR Files

Category	Frequency	Percentage	Percentage of Reported Responses
Unknown	788	27.9	---
Yes, Unknown or Other	30	1.1	6.8
None Noted	1592	56.4	---
Stimulants, Perscriptive/Narcotics (Amphetamines, Cocaine, Bennies)	0	0.0	0.0
Stimulants, over-the-counter (Caffine,'no-doz')	0	0.0	0.0
Depressants, Prescriptive/ Narcotics (Barbituates, Opiates, Tranquilizers)	14	0.5	3.2
Depressants, Over-the Counter (Alcohol, sleeping compounds)	388	13.7	87.8
Antihistamines	2	0.1	0.5
Hallucinogens (LSD, DMT, Mescaline, Psilocybin)	1	0.0	0.2
Marijuana	7	0.2	1.6

Table 24

Crash Causation Analysis
Frequency Distribution of the Primary Factor Responsible for Accident Noted for Crash Involved Vehicles in CPIR Files

Category	Frequency	Percentage	Percentage of Reported Responses
Driver Ommission or Unaware Error	1191	42.2	52.4
Driver Commission or Aware Error	797	28.2	35.1
Vehicle Defect	59	2.1	2.6
Road Defect	32	1.1	1.4
Ambience	193	6.8	8.5
Unknown	550	19.5	---

Table 25

Crash Causation Analysis
Frequency Distribution of the Primary Error of
Most Responsible Vehicle Noted for Crash Involved
Vehicles in CPIR Files

Category	Frequency	Percentage	Percentage of Reported Responses
No Error	71	2.5	3.1
Underestimation	108	3.8	4.7
Falling Asleep, Blackout, Death-at-Wheel	82	2.9	3.5
Diverted Attention	165	5.8	7.1
Inexperienced Driving or Erratic Driving	121	4.3	5.2
Drunk Driving, Drinking Involved, or Narcotics or Medication	362	12.8	15.6
Right-of-Way	386	13.7	16.7
Turning Error	93	3.3	4.0
Speeding	328	11.6	14.1
Overtaking	53	1.9	2.3
Following too Closely	91	3.2	3.9
Signs, Signals Disobeyed	250	8.9	10.8
Wrong Way into Oncoming Traffic	72	2.6	3.1
Lack of Lights	7	0.2	0.3
Lack of Brakes	31	1.1	1.3
Other	49	1.7	2.1
Avoidance Maneuver	50	1.8	2.2
Unknown	503	17.8	---

Table 26

Crash Causation Analysis
Frequency Distribution of Hazardous Road Conditions for
Most Responsible Vehicle Noted for Crash Involved Vehicles in CPIR Files

Category	Frequency	Percentage	Percentage of Reported Response
None	1592	56.4	67.6
Surface Under Water	4	0.1	0.2
Surface Slippery (oil, ice, water, etc.)	501	17.8	21.3
Shoulders Slippery	8	0.3	0.3
Weather Obstructions (snow, fog, etc.)	67	2.4	2.9
Light (sun, Headlight, etc.)	24	0.9	1.0
Obstacle in Road	70	2.5	3.0
Road Construction, Repair or Disrepair	71	2.5	3.0
Other	18	0.6	1.0
Unknown	467	16.5	---

CHAPTER V

REGRESSION MODEL APPROACH

The importance of comparing injury reduction countermeasures over similar crashes has been indicated. For the previous evaluation of energy absorbing steering columns, crash subsets were defined by VDI groupings. This provided discrete subsets of crashes classified by a single variable--V.D.I. However crash severity tends to occur over a continuous range which can be defined by combinations of several classification variables. Thus separating crashes into discrete subsets requires the creation of arbitrary division points.

An alternative to discrete subsetting is a multiple regression model which uses a number of discrete and continuous crash classification variables to predict a measure of crash severity. Such a model has been developed. Details of the development are presented in Appendix E. . This model predicts an expected AIS value which can be used as a standard of severity for each crash. Deviations from this standard can be used to compare vehicle safety features.

The investigation leading to the development of the multiple regression injury prediction model provides a method for equating crashes to a common basis. From that work it was found that 53% of the variation in injury from crash to crash could be explained by using the following variables as predictors; vehicle damage extent code, crash impact velocity squared, indication of windshield bond separation and indication of whether or not the occupant was ejected. A common adjustment between multiple and single vehicle crashes was also included in the model. However that improvement was marginal. Explanation of 53% of the variability occurred when the model was fitted to Washtenaw County crash investigations, which include injury and non-injury crashes on a basis more closely approximating their occurrence in the total set of crashes. The same model provides similar predictions when applied to all in-depth crash investigations. However the percentage of explained variability is lower for the entire set of crash investigations due to the fact that more "unusual" crashes have been selected in the entire set.

This prediction model provides a basis for equating crashes to a similar standard because it explains a large portion of the variability using a few predictors that are familiar to crash investigators and which can be measured on a reasonably common basis. Basically the model provides an estimate of expected injury resulting from crash dynamics. Therefore an expected baseline injury level is established for crashes over a wide range of crash types and intensities. It is then possible to compute the difference or residual between predicted and observed injury (Residual = Observed Injury Minus predicted Injury) and use that as a "controlled" measure of crash injury. The residual can then be used directly as a statistic for analyzing the effectiveness of changes in vehicle or environment components. With respect to the original problem an analysis could be conducted by merely comparing the residual injury statistics for vehicles with and without the new component. Depending upon the problem one might wish to further detail the analysis by considering only those crashes in which the component being evaluated could reasonably be expected to have an effect. Thus an analysis of side door beams would probably be restricted to only side impact crashes.

Two basic approaches can be used to obtain the residuals for crash injury analysis. First, one might use a specific regression model that has been derived using a representative set of crash data, for example the regression model fitted to all crashes in the file. Since this data represents the range of crashes there is a good argument for using it as a global estimate of expected injury. Alternatively one might use the predictor variables and fit a unique regression model for the actual set of data being used to perform the analysis. Residual injury statistics could then be computed using that model and the analysis would proceed in the same manner. A disadvantage of the second procedure is that the model would tend to reduce any unusual characteristics of the particular data set being used. This might remove some of the true differences attributable to the component being analyzed. From a conceptual viewpoint the former approach seems best. However the evidence establishing its superiority is very limited and unclear.

Tables 27 and 28 and Figure 13 present an analysis of the effect of restraint systems - Table 27 considering all occupants and Table 28 considering only persons in the front seat. Since upper torso restraints are only installed in the front seat Table 28 provides a comparison which is not partially masked by the non-availability of upper torso restraints. In both occupant populations those persons not using any restraint system have worse injuries. Examination of the residual injury variables indicates that persons without restraints are injured worse than expected while persons with restraints are injured less than expected. Further those occupants wearing both seat belts and upper torso restraints have even lower residaul injuries. These differences lead to an F test which indicates statistical significance at the $\alpha < 0.01$ level. Examination of the predicted injury from either model indicates that the accident severities were different for the restrained and unrestrained occupants. Those occupants who were not wearing any restraint device have higher expected injuries, i.e., their crashes were on the average more severe. Predicted crash severity does not depend upon whether or not the upper torso restraint was added to the seat belt. But the observed and residual injuries are lower when the upper torso restraint is added.

Table 27
Analysis of Injury Severity with
Respect to all Vehicle Occupants

Levels		N	Overall AIS	Predicted AIS from Complete Model	Residual AIS from complete Model
1	Neither	2029	1.39	1.35	.043
2	Seat Belt only	485	1.06	1.17	-.11
3	Seat & Upper Torso Belt	57	.89	1.13	-.24
4	Not Indicated	66	.09		-.42
Mean Sq. Within			2.45		1.52
Mean Sq. Between			18.56		8.03
F-statistic			7.87		5.27
Signif.			.00		.00

Table 28
Analysis of Injury Severity with
Respect to Front Seat Occupants

Levels		N	Overall AIS	Predicted AIS from Complete Model	Residual AIS from complete Model
1	Neither	1666	1.42	1.31	.11
2	Seat Belt Only	447	1.08	1.14	-.062
3	Seat & Upper Torso Belt	57	.89	1.13	-.24
4	Not Indicated	9	1.33		.095
Mean Sq. Within			2.44		1.46
Mean Sq. Between			17.69		5.40
F-statistic			7.26		3.70
Signif.			.00		.01

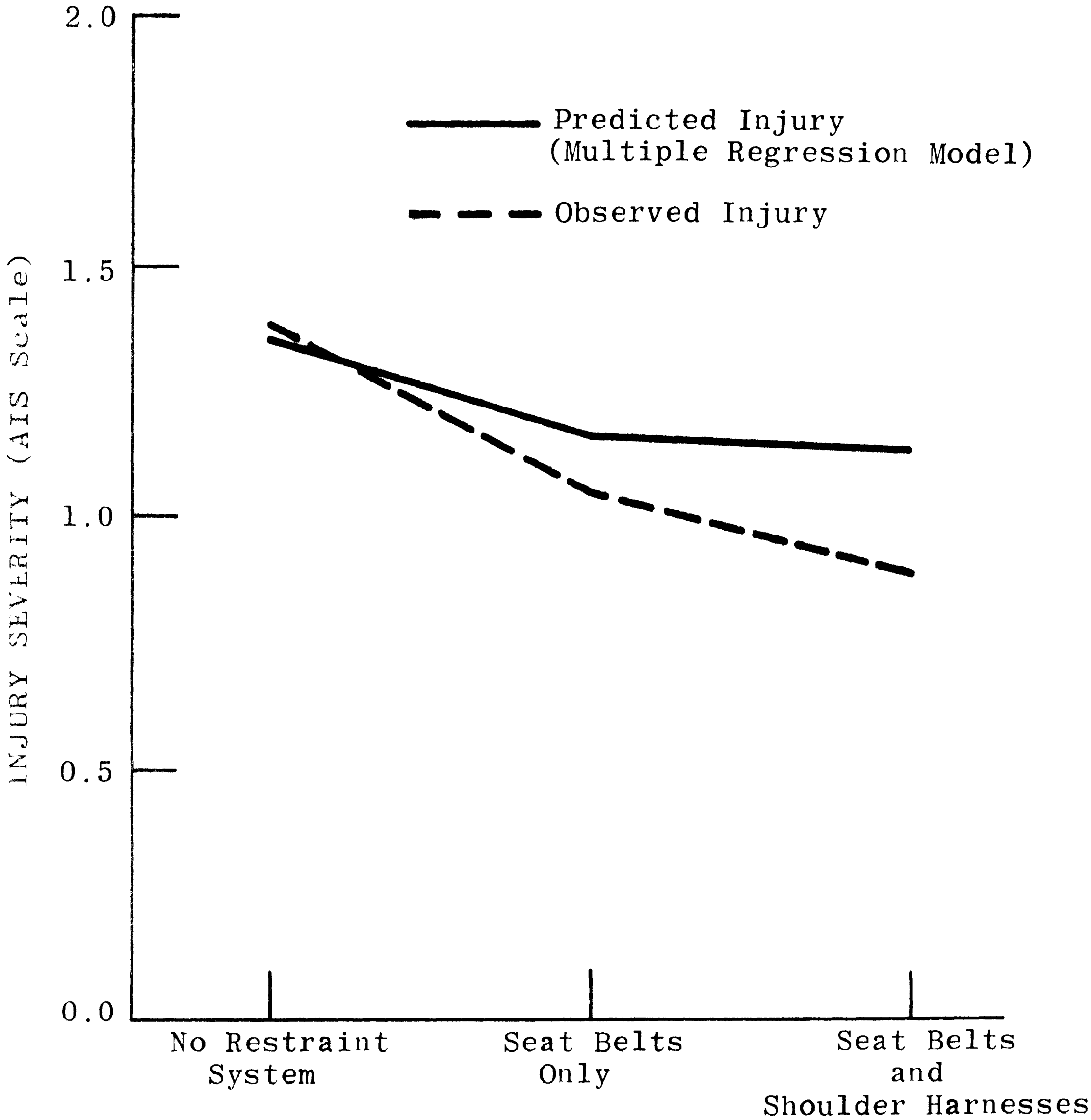


Figure 13. Relationship between restraint system usage and occupant injury severity.

The conclusions of this analysis are:

1. Occupants of vehicles who were not wearing any restraint device were involved in crashes which had a slightly higher expected injury severity.

2. Seat belts and upper torso restraints each contribute an increment of occupant injury severity reduction, even after the differences in crash severity have been controlled.

Tables 29 and 30 show an analysis of observed and residual injury for various crash configurations. The objective in presenting these results is a sensitivity analysis of the prediction model. As shown there are differences in residual injury that are identifiable by crash configuration. Having made that observation the next action is dependent upon the objective of the specific study for which the model is to be used. For example by using the results presented in Tables 29 and 30 one might argue that certain crash types - e.g., T-intersection - are inherently more severe than others. In developing the regression model a different method was used to compute impact velocity for the different crash configurations. Clearly that procedure did not remove all of the differences in injury prediction between crash types. Therefore either impact velocity has not been properly represented over crash configuration or the process leading to injuries is different. Such a conclusion might be useful by itself. However if a more precise comparison of the effect of particular components - e.g., restraint systems - is required, it may be desirable to include crash configuration as a variable in the prediction model.

Tables 31 and 32 indicate some of the inherent biases which exist in the data collection system and also indicate how the model compensates for these biases. Specifically the mean AIS is lower for the later years compared to early years. One reason for this decrease is that a greater emphasis - particularly by HSRI - has

Table 29
Analysis of Injury Severity with Respect to Crash Configuration for all Crashes in the File

Levels		N	Overall AIS	Residual AIS from Complete Model
1	Unknown	2	1.00	-1.60
2	Single Vehicle	723	1.64	..00
3	Head-on	326	1.91	.16
4	Intersection Type L	796	1.03	-.037
5	Side-swipe	115	.90	-.13
6	Rear-impact	436	.84	-.17
7	Other	7	.43	-1.37
8	Intersection Type T	300	1.45	.25
Mean Sq. Within			2.33	1.51
Mean Sq. Between			55.67	8.75
F-statistic			23.85	5.79
Signif.			.00	.00

Table 30
Analysis of Injury Severity with Respect to Crash Configuration for Washtenaw County Crashes Only

Levels		N	Overall AIS	Residual AIS from Complete Model
1	Unknown	2	1.00	-1.60
2	Single Vehicle	587	1.71	.11
3	Head-on	261	1.93	.25
4	Intersection Type L	666	1.04	.012
5	Side-swipe	87	1.00	-.027
6	Rear impact	364	.87	-.12
7	Other	6	.50	-1.11
8	Intersection Type T	280	1.49	.32
Mean Sq. Within			2.33	1.45
Mean Sq. Between			47.76	7.83
F-statistic			20.54	5.40
Signif.			.00	.00

Table 31
Analysis of Injury Severity with
Respect to Year of Investigation

Levels		N	Overall AIS	Residual AIS from Complete Model
70	Collision Year	524	.78	.052
71	Collision Year	1403	1.34	.021
72	Collision Year	778	.94	-.073
Mean Sq. Within			2.39	1.53
Mean Sq. Between			111.42	3.09
F-statistic			46.59	2.02
Signif.			.00	.13

Table 32
Analysis of Injury Severity with
Respect to Year of Investigation
for Washtenaw County Crashes Only

Levels		N	Overall AIS	Residual AIS from Complete Model
70	Collision Year	149	1.75	-.037
71	Collision Year	592	1.13	-.096
72	Collision Year	494	.62	-.26
Mean Sq. Within			1.87	.97
Mean Sq. Between			82.77	4.90
F-statistic			44.35	5.06
Signif.			.00	.01

been placed on obtaining data from non-injury crashes. This has contributed to an even greater deviation in 1972 crash investigations because more of the non-injury crashes from 1972 have been added to the file. Examination of the residuals from the complete model provides some additional confidence in the capability of the model to adjust for differences in crash severity. Since the residuals from the total model (Table 5) are not significantly different it is concluded that the model adequately explains the differences in severity.

Tables 33 and 34 present an analysis of injury levels from crashes defined by type of road. Examination of overall AIS indicates that crashes occurring on limited access expressways result in higher injury severity. However examination of the residuals indicates that after correcting for crash severity residual injuries are actually lower for crashes occurring on limited access expressways. This may be the result of expressway crashes being more likely to be single vehicle crashes - whose differences are adjusted for in the injury prediction model. Alternatively expressway crashes may not involve as many hostile environmental factors as do crashes on for example undivided four lane roads.

Tables 35 and 36 indicate that observed and residual injuries do not differ significantly as a function of vehicle manufacturer. It should be noted that model development considered only data from these four U.S. manufacturers.

Table 37 compares injuries by automobile body type. Observed injuries are not significantly different, however residual injuries from the complete model are. A somewhat surprising result is that four door hardtops and sedans have the highest residual injury. Since four door vehicles are generally larger one would suspect lower residual injury. The fact that residual injury is higher implies either a model prediction bias - e.g., a lower VDI for a given "true" crash severity would give this result - or possibly there is some other problem contributing to excess injury in four-door vehicles. We tend to favor the former.

Table 33
Analysis of Injury Severity with Respect to Road Type

Levels		N	Overall AIS	Residual AIS from complete Model
1	1 lane	13	1.62	.84
2	2 lane	1279	1.38	-.012
3	3 lane	73	.85	-.29
4	4 or more lane	785	1.08	.068
5	4 or more lane divided	525	1.56	-.038
6	Parking lot driveway	25	.95	-.27
7	Other (e.g., RR Track)	8	.87	-.16
Mean Sq. Within			2.44	1.53
Mean Sq. Between			17.13	3.65
F-statistic			7.02	2.39
Signif.			.00	.03

Table 34
Analysis of Injury Severity with Respect to Road Type for Washtenaw County Crashes

Levels		N	Overall AIS	Residual AIS from Complete Model
1	1 lane	10	1.70	.84
2	2 lane	560	1.09	-.12
3	3 lane	47	.66	-.33
4	4 or more lanes	379	.68	-.15
5	4 or more lanes divided	231	1.36	-.23
6	Parking lot, Driveway	3	0.	-.95
7	Other (e.g, RR tracks)	5	.60	-.34
Mean Sq. Within			1.94	.97
Mean Sq. Between			14.47	2.55
F-statistic			7.47	2.64
Signif.			.00	.02

Table 35
Analysis of Injury Severity with Respect to Manufacturer

Levels		N	Overall AIS	Residual AIS from Complete Model
1	General Motors	1114	1.35	.008
2	Ford	1008	1.25	-.041
3	Chrysler	476	1.31	.018
4	American Motors	107	1.42	.23
Mean Sq. Within			2.47	1.53
Mean Sq. Between			2.45	2.48
F-statistic			.99	1.62
Signif.			.40	.18

Table 36
Analysis of Injury Severity with Respect to Manufacturer for Washtenaw County Crashes

Levels		N	Overall AIS	Residual AIS from Complete Model
1	General Motors	435	1.11	-.14
2	Ford	510	.94	-.21
3	Chrysler	225	.97	-.092
4	American Motors	65	.89	-.052
Mean Sq. Within			2.00	.97
Mean Sq. Between			2.57	1.03
F-statistic			1.29	1.06
Signif.			.28	.36

Table 37
Analysis of Injury Severity with
Respect to Automobile Body Type

Levels		N	Overall AIS	Residual AIS from Complete Model
1	2-door hardtop	1129	1.28	-.074
2	2-door sedan	810	1.25	-.014
3	4-door hardtop	170	1.24	.12
4	4-door sedan	345	1.53	.19
5	Station wagon	197	1.37	-.066
6	Convertible	54	1.22	-.033
Mean Sq. Within			2.47	1.53
Mean Sq. Between			4.44	4.50
F-statistic			1.80	2.95
Signif.			.11	.01

Table 38 compares injuries and residual injuries over the various teams. This provides an indication of the biases which exist when individual teams are treated uniquely. In particular it should be noted that a number of MDAI teams have higher absolute AIS codes and higher residual injuries. Highly deviant examples are the team from Boston University (residual injury equals 1.95) and the Maryland Medical-Legal Team (residual injury equals 1.05). The objective of these teams is to obtain high injury crashes. As indicated this objective resulted in a large injury bias.

Table 39 and Figure 14 compares injuries and residual injuries by occupant age group. Younger persons have less severe injuries than do older persons. In addition injuries of younger persons have negative residuals--indicating less severe injuries than expected--while occupants over 35 have positive residuals--indicating more severe injuries than expected.

Table 38
Analysis of Injury Severity with Respect to Accident Investigation Team

Levels		N	Overall AIS	Residual AIS from Complete Model
1	Ann Arbor, HSRI-III	81	2.38	.47
2	Baylor Coll. of Med.	25	.56	-.54
3	Boston Univ.	23	4.74	1.95
4	Cornell Aero Lab IIIA	345	1.57	.18
5	Cornell Aero Lab IIIB	111	1.06	-.033
7	Georgia Inst. of Tech.	40	1.43	-.037
8	H.S.R.I.	811	.44	-.35
9	Indiana Univ.	8	3.25	1.06
10	McGill Univ., Mont.	5	1.00	.47
11	Univ. of Miami	77	2.26	.49
12	Md. Med/legal Found.	31	3.35	1.05
15	Univ. of New Mexico	56	1.63	-.30
16	Oakland Co., HSRI-III	431	1.39	.18
17	Ohio State Univ.	14	3.00	.93
18	Research Triangle Inst.	75	1.27	-.26
19	Univ. of Rochester	36	2.06	.64
20	Univ. of S. Cal.	26	.96	.037
21	Stanford Research Inst.	0		
23	Stanford Univ.	12	1.75	-.33
24	Southwest Research Inst.	93	1.29	-.27
26	Tulane Univ.	39	1.23	-.49
27	Univ. of Cal. (Siegel)	20	1.85	-.55
30	Univ. of Mich. (Huelke)	348	2.01	.17
Mean Sq. Within			1.93	1.40
Mean Sq. Between			71.44	17.87
F-statistic			36.94	12.72
Signif.			0.	.00

Table 39
Analysis of Injury Severity with Respect to Age Groups

Levels		N	Overall AIS	Residual AIS from Complete Model
1	Age Group (1-5)	61	.72	-.49
2	Age Group (6-10)	65	.68	-.44
3	Age Group (11-15)	146	.96	-.44
4	Age Group (16-20)	598	1.24	-.16
5	Age Group (21-25)	578	1.21	-.057
6	Age Group (26-33)	434	1.32	.008
7	Age Group (34-45)	258	1.45	.21
8	Age Group (46-55)	254	1.59	.34
9	Age Group (56-65)	141	1.65	.43
10	Age Group (66-80)	78	2.13	.88
Mean Sq. Within			2.33	1.45
Mean Sq. Between			18.61	22.04
F-Statistic			8.00	15.15
Signif.			.00	.00

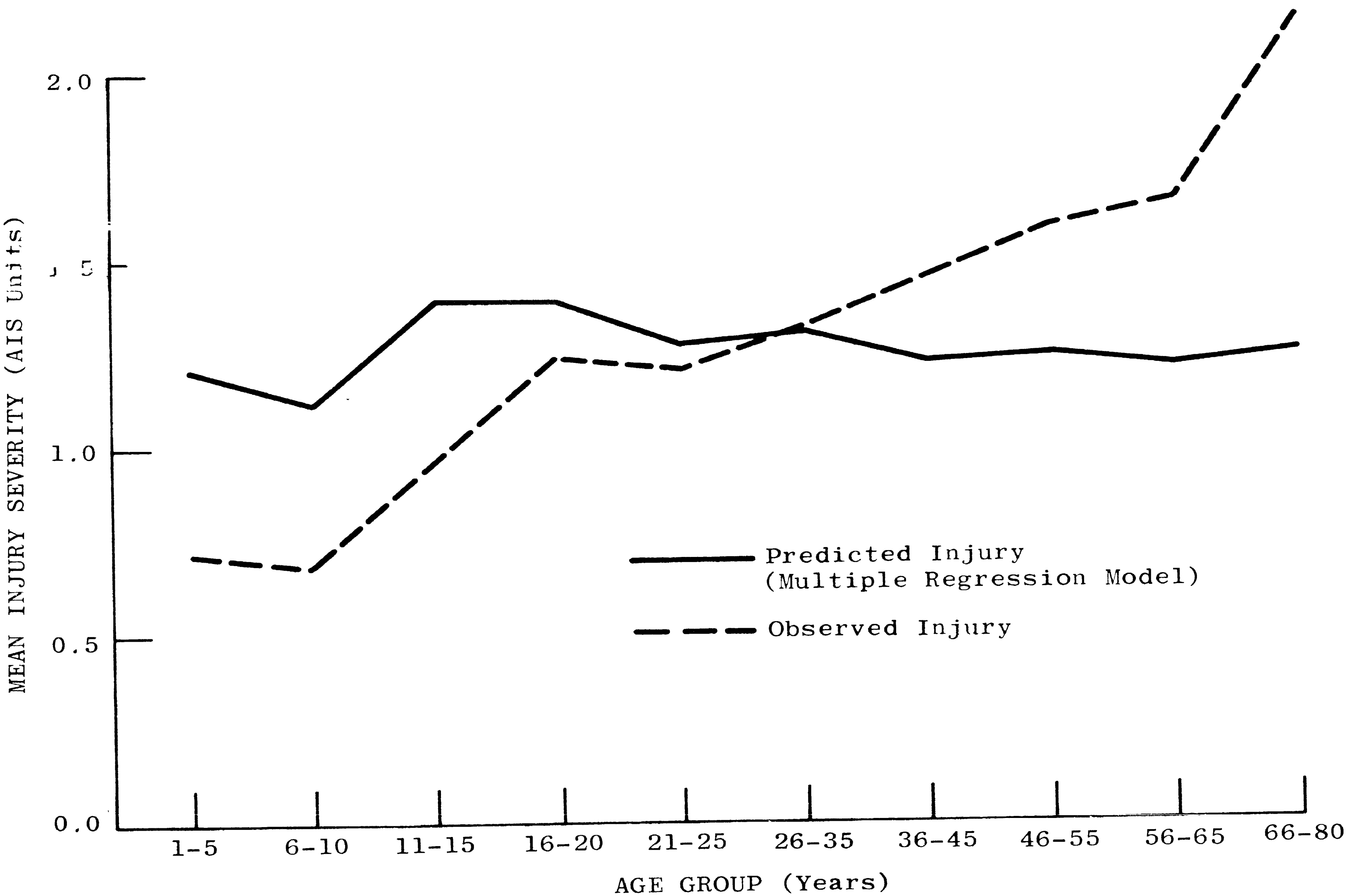


Figure 14. Relationship Between Age and Occupant Injury Severity

CHAPTER VI

USING SUBSETS OF THE DATA TO ESTIMATE THE FREQUENCY OF OCCURRENCE OF CERTAIN PHENOMENA:

As discussed above, several large subsets of the present data file were acquired with selection rules which are more easily determined or defined than the bulk of the MDAI reports. One grouping of interest is the three teams operating primarily in response to an MVMA specification requiring the investigation of (serious) new car injury accidents within their jurisdiction. This includes the cases known as Cornell III-A (supported jointly by MVMA and NHTSA), Oakland County cases (supported wholly by MVMA), and those cases investigated by Dr. Huelke (supported wholly by MVMA but included in the files of the tri-level operation at the University of Michigan which are supported by NHTSA.) All three of these teams have been asked to obtain a census of serious injury accidents involving American manufactured passenger cars (and certain foreign cars manufactured by subsidiaries of American manufacturers). For a variety of reasons these cases are somewhat less than a census, and the criterion for selection on the basis of injury is not quite as precise as one might wish. Nevertheless, there is evidence in the data that the three teams do acquire comparable data with regard to the relationship between impact and injury severity and that they vary in an exepcted fashion because of geographic differences among the covered areas. Specifically, Dr. Huelke's cases are 61% rural, Cornell's are 53% rural, and Oakland County's are 9% rural (the Oakland County investigations being limited to six large urban jurisdictions within that county). For purposes of synthesizing a sample of "all" injury accidents, we have combined the data from these three teams and call them the MVMA injury group.

The second subset of interest is that data obtained in Washtenaw County from Dr. Huelke's investigations, a complementary set of non- or little-injury cases (both of these sponsored by the MVMA), and a small number of cases investigated by the MDAI team operating at the University of Michigan. This combination is intended to represent a census of towaway involvements of 0 to 2 year old American manufactured cars occurring within this county. In fact, it is not a complete census--for a variety of reasons it includes only about 50% of the vehicles identified on police reports as "towaways", and it is of interest to consider biases introduced by this shortcoming. The cases that "get away" do so because (1) they occur on week-ends and since the investigators work a normal five-day week the cars have occasionally departed the county before Monday, (2) a very heavy run of accidents prevents a case from being done before the car is moved to a repair facility and repairs have been initiated, (3) some cars are listed as "towed away" on the police report but are in fact towed only because the driver was incapacitated by injury or drunkenness. From an analysis of the police data it is known that the percentage of injury and non-injury involvements in the lost data is approximately the same as in the acquired data, but it seems likely that the lost cases are generally less severe collisions.

Extrapolation from either of these sets of data to a national picture has obvious difficulties. There may be overriding regional biases which make such estimation invalid. The reader may wish to regard any conclusions drawn from such subsets as examples of what could be done with a more defensible sampling plan. Yet it will be argued that for many purposes these groups of data are useful in putting bounds on some problem areas.

Table 40 compares these two groups of teams with the remainder of the CPIR file (i.e., the MDAI teams) and with the entire file. Note that these are not mutually exclusive, since Dr. Huelke's data is in both of the first two columns. These data were developed in a study of the frequency of hood/windshield involvement. If the same values held up for a more adequate sample, one would infer that hood penetration occurs in something like 2% of the involvements in a towaway set, 4.6% of a set collected to the MVMA injury specification, and that this phenomenon may be overrepresented in the MDAI cases.

Table 40

Several factors compared for 3 subsets of the CPIR file, passenger cars only

	877 cases from Washtenaw Cty. towaway set	691 cases from MVMA injury set	847 cases from MDAI teams	2253 cases from entire CPIR file
% lap-belted drivers	21.5	13.7	20.0	18.8
% cases in which rear edge of hood penetrated the windshield	1.8	4.6	6.2	3.8
% cases in which the rear edge of hood contacted the windshield	8.7	18.5	16.4	13.5
% cases in which the rear edge of the hood was elevated	35.5	55.5	56.7	47.5
% vehicles which rolled over in the crash	9.7	13.7	10.4	10.1

Rollovers seem to hover around the 10% level in all except the MVMA injury data, and further analysis might find this to be associated with a higher incidence of rural crashes in that data. The bias in collection of the Washtenaw County towaway sample discussed above is estimated to lead to a reduced percentage of rollovers in a total population of towaways. Assume, for example, that none of the 50% of the towaways that got away rolled over--then the rollover frequency in towaway accidents might be on the order of 5% rather than 10%.* Finally, the percentage of towaway involvements in the Washtenaw County police level data is about 40%. Multiplying by this factor, one might estimate that only 2% of the vehicles reported in accidents in this county rolled over.

While it is clear that one cannot state statistical confidence levels in these values (with regard to national representation), it is interesting to compare them with what other data is presently available. Rollovers are listed in mass (police) accident data usually only if they represent the first event in an accident sequence. If a car is struck at an intersection and it subsequently rolls over, the rollover information is generally lost. The national accident summary, compiled by combining police reported information from many states, lists rollover frequency as 0.5% of all reported involvements.

The hood windshield contact and penetration information is just not available in police level data. Further, it is relatively difficult phenomenon to observe. The process of inspecting a car using the CPIR form, and noting the relationship between a raised rear edge of a hood and a tear in the windshield is time consuming, and requires a certain amount of expertise on the part of the investigator. The MDAI file then provides some idea of the frequency of this event under different (e.g. towaway, all injury) conditions. It seems likely that an adjustment similar to the one performed for the rollover data would be in order for the hood windshield

---

* An alternative hypothesis might be that all of the lost 50% rolled over. Although the likelihood of this can be studied by reference to police level data in this case, in other cases it cannot.

information, assuming that the towaways not in the sample were not generally as severely damaged. So the estimated frequency of hood windshield penetration in "all" reported towaways may be of the order of 1%, and of all accidents something like half of that.

For some purposes, then, it may be useful to use these identified subsets of the CPIR file to get an estimate of the frequency of certain phenomena in these particular populations. It is clear, however, that a broader sample geographically representative of the nation and more strictly representative of some definable accident population would be of more value in arguing the costs/benefits of a proposed countermeasure.

CHAPTER VII

Methodology for Optimum Procurement of Data for Underlying Processes MOPUP

A NEW TRI-LEVEL SURVEILLANCE CONCEPT

While the general goals of the MDAI team activities have been stated as: Determine accident and injury causation, identify functional problems in the highway transportation system, identify the need for countermeasures, etc., it should be clear that accident investigation is only one tool or method in approaching a solution to such problems.

These general goals lead to a number of specific operational questions faced by the NHTSA staff, such as: What is the anticipated benefit of the proposed addition of air bags to automobiles? Which vehicle interior components are still contributing excessively to injury? Is there a need for a federal standard to require that each automobile be equipped with a portable stopped vehicle warning device? Solutions to such problems may occasionally come from analysis of some existing set of accident data, but more often than not they will involve most of the steps shown in Figure 15 -- analysis of existing literature, running controlled tests to investigate the phenomena of interest, analyzing available sources of data, specifying a field data collection program, and hopefully arriving at an acceptable answer to the problem. While we are concerned here primarily with the acquisition and analysis of data from the real world of accidents, it should be clear that this accident investigation process is not an end in itself but rather fits into a general scientific approach to solving problems. Four types of data collection and analysis systems are important for NHTSA operations. The first of these is represented by the present and past MDAI efforts in which expert investigators collect very detailed information about a modest number of accidents, consider the data in

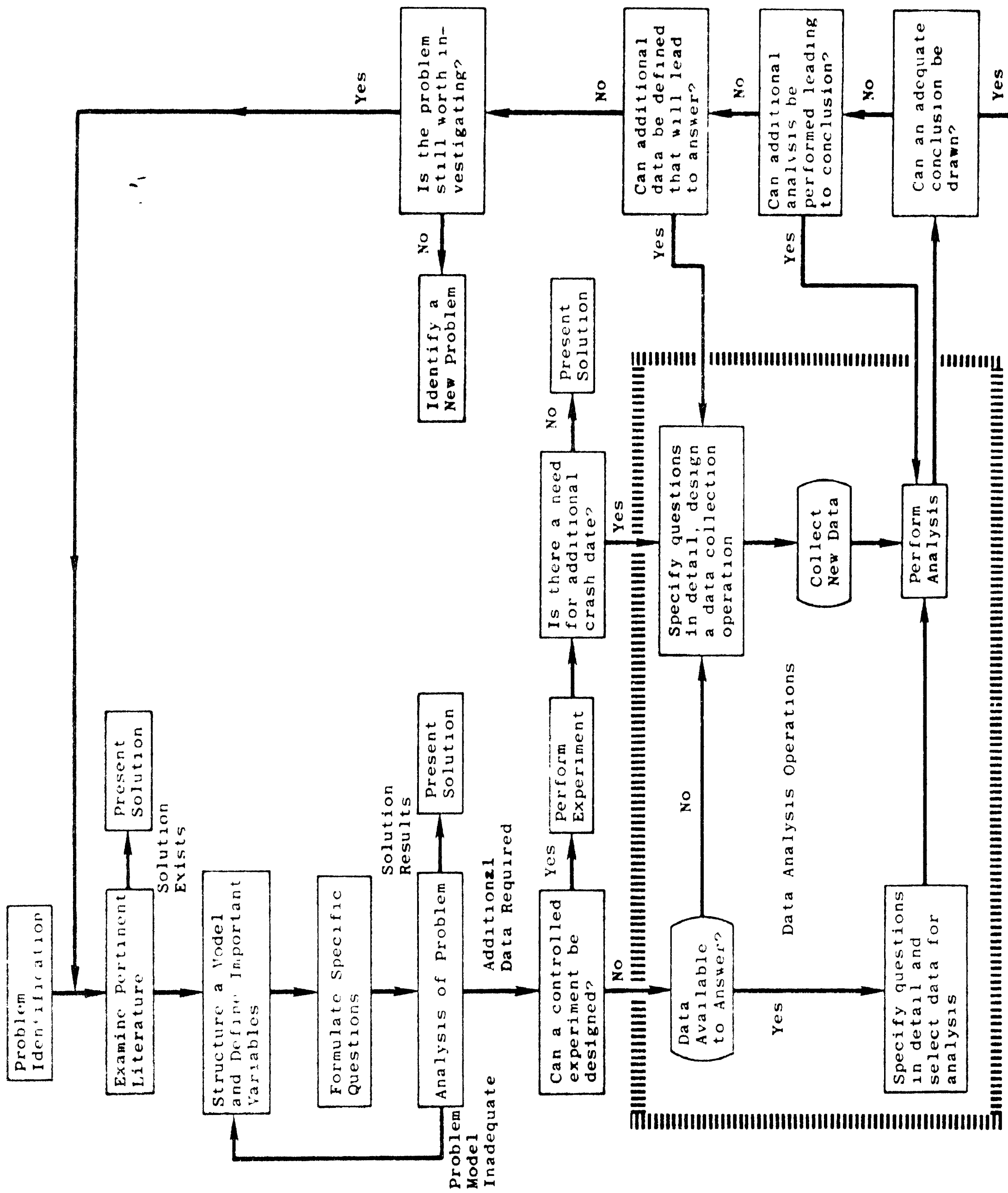


Figure 15. NHTSA problem solving system.

the light of their own expertise, and make recommendations for change or further study. The second is a need to routinely monitor the crashes on the highway to determine whether or not losses are increasing or decreasing, and to measure the characteristics of the losses. The third is a need for occasional sampling of the accident population to study a problem of current interest for which existing data will not suffice.

A final data input concerns the identification and analysis of system defects, e.g., the accelerator that sticks in a unique set of circumstances, Such problems are not likely to be routinely discovered by in-depth crash investigation. This comes about because such defects are "rare"events--i.e., they might occur in less than 1% of the crashes. Detection of system defects such as these are better left to the routine police investigation reports--or preferably to various inputs from individual drivers,public service groups, news reporters, or automotive companies. Government sponsored crash investigation teams should be included in this set of reporting agencies, but only as one of many information sources and not as a primary source.

Other government agencies have systems which yield answers to specific questions continuously over time. Such systems ordinarily proceed from a carefully defined problem and should include some pre-defined analytic approaches. A number of excellent examples exist in the Federal Government including:

1. The Current Population Survey (CPS)

This survey consists of a probability sample of 50,000 U. S. Households conducted each month. Results are made public within approximately two weeks of the time that the interviews are conducted. The best known item of information developed is the national unemployment rate. In addition statistics on such varied topics as smoking and voting have been collected at various times This system which has been operating for 25 years requires a well designed sampling strategy, having specific objectives. In addition a well developed organization is required to provide the rapid response and to check on the accuracy and completeness of the data.

2. The Consumer Price Index (CPI)

The CPI is constructed by first assuming a "market basket" of items defined as common purchases for a family. The large number of items are separated into five major groups; food, housing, apparel and upkeep, transportation, and health and recreation. These groups are each divided so that in total 52 categories of expenditures are included. Statistical sampling is then used to determine the price to consumers of items in each category for 56 cities. These categories are then combined, according to purchasing patterns, for the various cities to obtain an index for each city. These indices are then combined to develop the CPI. Finally, an index number, which shows the ratio of present prices to a base period, is computed for comparison purposes.

3. The National Crime Survey

The Census Bureau, in cooperation with the Law Enforcement Assistance Administration, has field tested a survey of criminal activity in several cities, in the United States, and will begin shortly to interview 60,000 families and 10,000 businesses each month to measure the country's crime rate. The development of this method was begun in 1966 with a survey of 10,000 households, and was subsequently more fully tested in San Diego and Dayton. While it is currently believed that police departments under report criminal activity (although they have been increasing the reporting frequency in recent years), this survey is expected to yield a monthly national crime statistic which may then lead to allocation of resources, identification of problem areas, etc.

In all three of these examples several characteristics come through. First, the surveys have well defined objectives which are met by the development of a comprehensive data collection process.

The development of this process reflects a compromise between theory and practice. The data is collected by a large well trained staff which rigorously adheres to a set procedure. This provides a set of numbers that are comparable from one period to the next. More importantly the surveys provide numbers that are used by decision makers. Thus, while the average man represented by these surveys may not exist, the numbers developed have proven to be useful measures, i.e., they provide a measure of how the nation's economy or crime rate is moving.

Development of such a system for highway safety would require specifically defined objectives that are seen as important by management in DOT. There will be a need to devote significant resources continuously over time to develop and maintain such a system. However, without such a system it will be difficult to obtain system performance data that is comparable over time. Generally, any extensive data collection system will include certain known and unknown biases. Some of these are recognized but their removal would be too expensive. Other biases are unknown or only recognized as possibilities. Clearly a balance between cost and system performance must be maintained.

Three parallel data collection efforts can be operated to provide the in-depth, the continuing surveillance, and the special investigations needed for NHTSA's decision making. These are described here under three acronyms:

S C R U B - Selective Collection of Reports with Unknown Bias

V A C U U M - Valid Accident Data Collection for Universal Utility Measurements

C A R P E T S W E E P - Computer Assisted Rapid Response Procedure for Evaluating Technology in Specific Wide Area Environments of Emerging Problems

Each of these concepts will be described in more detail below.

Selective Collection of Reports with Unknown Biases (SCRUB)

This effort will be essentially a continuation of the present in-depth reporting, with relatively complete information being developed about each accident investigated. It is expected that numerous photographs would be taken of the vehicle and the scene, and in some cases of the injuries.

Conventions presently in force for MDAI cases would continue with complete interviews with drivers and witnesses, detailed measurement of vehicle damage, and full medical reports for each occupant.

The major contribution of these cases, however, should be the expert opinion of the investigating team members regarding causative factors, and particular problems observed in the vehicle, human, or environmental aspects of the accident. Suggestions for countermeasures should be welcomed.

The set of information developed from these reports may be placed in computer form if the quantity is large enough (we have estimated a total input of perhaps 300 cases per year from a number of teams), but the computer searching would be used principally for finding cases pertinent to a particular subject, and then the investigator would have to review the cases in detail to gain a better understanding of his subject.

No attempt would be made to make the SCRUB cases statistically representative of the national or even local population, although there should be some emphasis placed on getting cases from all regions of the country. Lacking a further reason, teams might be instructed to obtain SCRUB cases which were interesting - in this sense they might look for the unusual, or for a wide range of accident types. Thus the resulting collection of SCRUB cases should ultimately have one of almost any conceivable type of accident in it, but would in fact be statistically representative of nothing.

The usefulness of this SCRUB information would be almost the same as for the present MDAI cases. Recommendations from individual investigators may lead to further data analysis or experiments (for example when a severe injury induced by some new type of roadside furniture is reported). Viewing the entire collection of cases as a reference library would permit an analyst to read about several cases in which there was severe injury at low speed, to understand the range of mechanisms for fires starting in connection with crashes, to see pictures of broken glass in two different kinds of windshields, etc.

The usefulness of the SCRUB program from the investigators point of view, however, is somewhat different. This SCRUB activity would insure that investigating teams maintained open communication with all agencies concerned with traffic accidents in their jurisdictions--hospitals, police agencies, traffic engineers, wrecking yards, motor vehicle departments, insurance companies and others. It would also furnish an opportunity for training field investigators over a broad range of investigative techniques, so that they would be able to investigate and report for special studies (to be described below) without further training. Finally, the remainder of the tasks to be assigned to these new tri-level teams are perhaps more mundane, and the SCRUB investigations should serve to maintain interest of professionals in this work.

Valid Accident data Collection for Universal Utility Measurements (VACUUM)

This effort will be a very strict representative sampling of accidents occurring with a well defined geographic region, in order to obtain a set of accident reports which can be combined with those of other regions for national inference. The relative frequency of a large number of crash subsets could be estimated from this sample. It should be viewed as a collection of data which provides a continuous pulse measurement of the traffic accident situation.

(The recommended regions, and the total sample size will be discussed in section VIII below).

In concept the data acquired in this program is intended to furnish a continuing assessment of the state of traffic safety in the country (in terms of the number and type of accidents and injuries; it is assumed that fatality counts will continue to come from a census taken under the fatality analysis file). In addition, this set of data should provide information for identification of new problems, and assessment of the efficacy of some solutions. Generally it will provide information to answer questions which were not asked specifically in advance, but for which there was a general idea of their type.

For example it has recently been of some interest to determine the number and severity of injuries occasioned by hood penetration of passenger car windshields in the United States. While the present MDAI file can be interrogated with respect to this, and, in fact some guarded inferences drawn with respect to frequency, there are obvious shortcomings in the data. This particular question of hood-windshield involvement was in part anticipated by the developers of the Collision Performance and Injury Report in that hood penetration of the windshield, hood contact with the windshield, and elevation of the rear edge of the hood are all recorded. Further, injuries to various parts of the body are attributed to contact with particular vehicle components--such as "head struck windshield" or "face struck hood". It is difficult to suggest the next specific question which may be asked--it could be "what are the points of contact for the more severe injuries incurred by drivers when their cars are struck on the left side by other vehicles", or "of all collisions in which roadside furniture is struck, what percentage (this year) involve breakaway posts?".

No attempt is made in the present study to detail the content of the reports made under the VACUUM program. In the interests of an early start it is suggested that the present CPIR form would be an appropriate starting point, but that during a trial period (of a year or so), a more careful determination of NHTSA's needs would likely delete some items from the form and add others. An advantage in beginning with the present form is that the data handling mechanism is set up to accept it, and that the kinds of data reported (while primarily concerned with vehicle damage and injury) have been generally used in early analyses.

Present MDAI cases (and their MVMA counterparts) supplement the CPIR form with numerous photographs. While these have been of value to design engineers in the past, the logistics of handling a large number of these is awesome; and it is clear that there would be a tradeoff (for the same cost) in photographs and the total number of cases. For the VACUUM program photographs are not recommended. One shortcoming of this choice is that there is little redundancy left in the reporting scheme, and it will not be easy for editors to check reports or resolve conflicts in these. This then would indicate a strong need for good quality control throughout the entire data collection system.

An alternative to this relatively expensive and expansive data collection activity would seem to be to accept police level data, and to try to insure its completeness and sampling adequacy. This is not recommended for several reasons. The type of information to be reported in this program, and the level of detail and precision needed to NHTSA, is such that the investigators must be highly trained. The sampling process must be precise and relatively complete--i.e., if every tenth towaway accident within a particular jurisdiction is to be included in the sample considerable effort must be made to insure its inclusion. In order to achieve both the technical capability and the sampling precision needed the investigation activity should be dedicated to this task rather than accepting it as a part time operation. The part time statisticians within police departments are the reason for the new national crime survey.

Determining unemployment rates by asking cities or states each to report their unemployment statistics is not a workable scheme. And there is evidence aplenty in the traffic safety literature of the variability of reporting among police departments.*

There should be a strong central control of the VACUUM operation, and if it is difficult to get research oriented teams like the present MDAI operations to provide the input precision required, NHTSA should consider staffing the teams with federal employees. At this juncture there seems to be no reason why there could not be a mixture of both contract and federal teams providing information for VACUUM.

Questions appropriate for the data to be acquired under VACUUM are many. With respect to the vehicle one might wish (at some later date) to determine how many wheels fell off cars as a function of age, what percentage of the crashed vehicles had the windshield penetrated by the hood; whether the energy absorbing steering columns are effective; what percentage of a particular recalled vehicle (in accidents) has been properly repaired. With respect to occupants one might wish to determine injury patterns as a function of driver age and accident configuration, or keep tabs on restraint system usage as a function of region of the country, time of day, purpose of trip, or season. With respect to the environment one might wish to determine the extent of icy pavement as a causative factor in accidents (and understand well the severity of the resulting accidents), or to study the relationship between some driver factor (e.g., belt wearing) and weather.

It is expected that the data elements to be collected in the VACUUM program would change, but only slowly so that there would be consistent data for many elements over a long period of time.

---

* Garrett, J.W.; Braisted, R.C.; and Morris, D.F., "Tri-Level accident research study. Second annual report. Final report" Cornell Aeronautical Laboratory, Inc., Buffalo, N.Y., May 1972. 87 p. Report No. CAL VJ-2893-V-2/ DOT-HS-800 679., NHTSA.

Wuerdemann, H., "Estimating the quality of routine traffic accident reporting by police personnel", Center for the Environment and Man, Inc., Hartford, Conn. 1970. 7 pages., National Safety Congress Transactions, Vol. 24, 1970.

Scott, R.E., and Carroll, P.S., "Acquisition of Information of Information on Exposure and on Non-fatal Crashes", Vol. II, Final Report, May 12, 1971. Highway Safety Research Institute, Ann Arbor, Michigan, Contract No. FH-11-7293.

The total amount of data will depend on choice of sample size, the amount of funding available for the program, and the precision of results desired. But what must be striven for is a set of data of accuracy and size sufficient to stand up under a variety of criticism.

Computer Assisted Rapid response Procedures for Evaluating Technology in Specific Wide Area Environments of Emerging Problems (CARPETSWEEP)

When a problem arises, either the answer will be available or it will not. Availability means that the data exist in a report or in a computer file and can be retrieved, together with the necessary statistical analysis, to allow a decision to be made about the matter at hand. If the data are not available they must be obtained The procedure to do this is currently a time consuming and costly one, involving the formulation of the problem in the framework of a request for proposals sent to qualified research entities, the evaluation of the proposals which are submitted in response to this, the negotiation of contractual arrangements, the monitoring of research progress, and finally the analysis of the results reported by the contractor. These two ways of answering questions are extremes on continuums of both time and effort. While each can lead to satisfactory results in specific instances, we will suggest here than an intermediate approach may be advisable. The procedure we are defining stems from two considerations. First, there is a need for an intermediate level, rapid response data collecting technique to fill the gap in NHTSA's ability to respond to the requests of its clients, and secondly, a great part of the mechanism for accomplishing the tasks set out already exists because of the Administrations previous efforts.

The name we have given to this new procedure is Computer Assisted Rapid-response Procedure for Evaluating Technology in Specific Wide-area Environments of Emerging Problems: CARPET SWEEP. The acronym is intended to convey the idea that the technique is to be used in situations where immediate attention must be given to a specific area. A handy simple mechanism is needed which is kept at a high state of readiness and can be put into action as soon as a problem is identified. The process is turned on for as long a time as necessary to get the job done and it then returns to its ready state awaiting the next requirement.

In several ways , the existing MDAI teams meet the needs for the CARPET SWEEP process. They consist of highly trained personnel capable of collecting accident data accurately and efficiently. Their relationships with the community, i.e., the police and fire departments, ambulance dispatching services, hospital emergency room organizations, and the judicial system, are such that access to a continually changing set of data can be readily assured. Familiarity with the data input and its subsequent analytical use, furthermore, makes them sensitive to the issues that are critical to the success of any data collection program.

The institution of the procedure we have in mind would necessitate a somewhat different contractural relationship between NHTSA and the MDAI teams. The SCRUB and VACUUM portions of the team's work would continue as before under fixed contracts, and this part would provide for the collection of data on an on-going basis. Another portion of the team's work would be left open and would be dedicated to the performance of ad hoc studies under the CARPET SWEEP concept.

A study under this procedure starts with the identification of a problem, and the source of this activity will most often be at NHTSA itself. Along with a statement of the problem must come a complete description of the data which are to be collected and an understanding of how these data are to be processed. It is strongly suggested, however, that the teams themselves have a considerable degree of participation in this stage of the process, both because the perspective of field teams is a valuable asset and because of the sense of identification and contribution which is important to the morale and efficiency of those working in areas remote from the directing agency.

Communication between the central office and the MDAI teams is to be handled almost completely by a remote computer terminal link. The instructions for collecting the data are stored in a computer file and the field team is given instructions to access that file. This will tell them the criteria to be used in selecting cases for this study, the variables on which data are to be collected and the format to be used in the reporting of data. Reporting is to be done on a case-by-case basis directly from the terminal at the MDAI site to the computer in the central office. With each study design an interactive program[1] will be written to prompt the user in his reporting and will accumulate the data from the many sources in a single file for analysis. The data can be analyzed at any point in the CARPET SWEEP operation so that an indication may be gotten as to when enough information has been collected to answer the original question.

Let us take as an example of a problem which could be attacked by this procedure, a question or set of questions which the standards writing division at NHTSA might have about child restraint systems. The goal of this inquiry might be to determine whether to require that such a device be provided with every new car if the purchaser desires it. The questions which arise in connection with this investigation can be answered from many sources. The total population of children is available from standard census data, and will give an idea of the size of the group the decision will effect. The number of families that would use such a device if it were provided to them, free of charge or at some defined cost, might be attacked by something resembling a market survey. The more serious questions from the Administration's point of view, however, have to do with the public acceptance and use of the device and its performance in reducing or eliminating injuries and fatalities. It is toward this later set of questions that an <u>ad hoc</u> study will be directed.

---

[1] This interactive program will be composed of sub-program modules that can be quickly put together according to the needs of a specific study.

The solution will be framed around a collection of variables, relatively small in number, which bear directly on the problem. The data collection procedure will be designed and the analysis and reporting process defined, both based on computer codes previously checked out and used, but modified for this specific case. The MDAI sites at which data will be gathered will be selected so that a sufficient quantity of data will be collected in the time required. This set up process* should be completed in no more than one week. At that time each of the designated MDAI teams will receive a message on its remote terminal similar to the following:

CARPET SWEEP STUDY NO. 8**

Beginning November 1, 1974 and continuing until December 30, 1974, unless otherwise directed, you will investigate all motor vehicle accidents in which 1) a VDI of 1 or greater is established and 2) a child between the ages of 6 months and 4 years is an occupant of any vehicle involved. A separate report will be made for each such occupant.

To enter a report signon and issue the command "SOURCE CPS8: ENTER" you will then be prompted at the terminal to enter data with regard to the following variables for each case:

1 - Make and model of vehicle
2 - Model year
3 - Driver age
4 - Driver sex
5 - Accident Type
6 - VDI
7 - Impact Speed
8 - Road conditions
9 - Date/Time

---

* i.e., the mechanics of the set up process; the initial formulation of the problem may take any amount of time.

** Appendix H shows interactive computer conversation as it would appear to the reporting crash investigator.

10 - Case occupant
   a) age
   b) sex
   c) weight
   d) position
   e) injury severity
   f) restraint system data (V461-471, CPIR)*

11 - Free form test comments

The ENTRY routine has an error checking capability and will halt at each step until a valid input is received, repeating a step should an error occur. The command "???" instead of an entry value will display a complete description of the variable awaiting values. The command "CHANGE X", where X is a variable number in this report, will delete previously entered values and allow new values to be entered. The command "END ENTER" stores this report in the central file and prints a copy for your records.

To allow you to monitor the progress of this program, two status reports are available which summarize the data held in the central file at the time of the request. The command "SOURCE CPS8: STATUS-ME" will present the status of your own contributions to the file, and the command "SOURCE CPS8: STATUS-ALL will provide a summary of the entire project.

For a training, educational, or demonstration purposes, the command "SOURCE CPS8:DEMO" may be used. This will cause the terminal to behave exactly as it does when in the ENTRY mode, except that no data are entered into the central files.

Mr. A.C. Dent at Area Code 202-724-0000 will be able to answer any questions you might have regarding CPS8.

Please signon and issue the command "SOURCE CPS8:OK" to acknowledge receipt of this message.

---End of Message---

---

* i.e., In the same manner as you would record data for variables 461 through 471 of the Collision Performance and Injury Report.

The recipients of the message would have about two weeks in which to prepare for the new effort; adjusting the scheduling of their on-going MDAI cases, preparing whatever forms were thought necessary to the data collection, advising police, hospital, etc. of their informational needs, and indoctrinating personnel as to the goals and methods of the program. The total time for the completion of the entire project, therefore, would be less than three months.

Implementation of the CARPETSWEEP concept would provide NHTSA with the means of acquiring essential data on a timely basis to provide solutions to specific problems. Its power is derived from the combination of using the expertise already in existence at the MDAI team sites and employing a remote access computer as both the communication network and a rapid data processor.

We list below a dozen problems which might be attacked by the CARPETSWEEP technique, together with an indication of the goals of each study and the variables relevant to answering our questions. These are intended only to reflect the scope of such problems and not to suggest to NHTSA any ordering of importance or priority. Naturally, in each case the specific of the sample and the data to be collected would have to be worked out in complete detail before the study were initiated.

1) Steering Column: The goal here is to find out if energy absorbing steering columns are effective in reducing injury severity in frontal collisions. Each collision identified in the time period for a new model car containing an EA column would be matched* on the

---

* Here and elsewhere in this section reference is made to the collection of matched samples. It should be noted that the investigation teams themselves would not be given the opportunity to do the matching. They might be asked for example, to obtain information about crashes with and without airbags. with 20 crashes in each group. Within each set there might be a requirement to obtain a particular number at each of several Vehicle Damage Extent levels. But the actual selection of cases for inclusion would ultimately result from a random process employed by the team. Ultimate matching for crash severity would be performed analytically using something like the regression model discussed in this report.

basis of collision severity (impact speed or VDI) with an older car not equipped with an EA column. The main data item sought would be extent of injury to the driver, and the sample would most likely be subdivided further on some control variables such as the use of restraints.

2) Air Bag: When the air bag is in common enough use as a passive restraint, we will be interested in evaluating its effectiveness in preventing or reducing injury. For each crash in which an air bag is deployed, a similar crash will be taken in which the vehicle was not equipped with the device. A comparison of these two populations will then be made to evaluate differences in injury production.

3) Bumper systems: Low speed rear-end collisions will be used in this study, and the sample will be restricted to cases where one of the vehicles has a new bumper system and the other does not. The "new" cars will equally often, in the sample, be the striking and the struck vehicle. The cost of repairs for each car will be collected and compared to see if the new bumper system is achieving its intended goals.

4) Roadside Obstacles: This study will be restricted to the study of single vehicle accidents in which the crash involves hitting some permanent feature of the highway. The object of this study will be to develop a distribution of accidents over the kinds of obstacles which are struck and to define the severity of the injury related to each. Knowledge of the type of roadways and other environmental conditions will help in decisions as to the possibility or necessity of taking corrective actions.

5) Unlicensed Drivers: All crashes in which the driver did not possess a valid driver's license will constitute this sample, and the goal of the study is to discover the motivating factors underlying this kind of behavior. Did the driver know he did not have a valid license? How long has he been driving with a revoked license? What risks does he perceive as a result of his behavior?

A concurrent study on exposure aimed at unlicensed drivers might be useful in establishing the magnitude of the problem.

6) Recreational Vehicles: The aim of this study would be to discover the special circumstance surrounding recreational vehicle accidents. The nature and duration of the trip components, the number and ages of vehicle occupants, familiarity with the area travelled, length of time the vehicle has been owned, and many other variables would help to distinguish this type of accidents from the general population. Injuries which are attributable to specialized vehicle configuration will also serve to indicate whether countermeasures should be directed at this class.

7) Shoulder Harnesses: There are many indications that upper torso restraints are an effective means of reducing injury severity, and the goal of this study will be to test the truth of these observations. Crashes in which shoulder harnesses are worn by the vehicle occupants will be matched against similar crashes in which the restraints were not used, both samples being controlled for the use of lap belt restraints. The between group comparison of injury production will tell us something about the effectiveness of the respective restraint systems.

8) Warning Buzzers: The effectiveness of the requirement for a warning buzzer and light to be activated when seat belts are not properly fastened is the object of this study. We may either sample from the crash population or perform an exposure survey to determine if occupants of cars with the warning system are using their lap belts and shoulder harnesses. Our major question has to do with the percentage increase in restraint usage for vehicles which have the warning system over those which do not. We don't expect usage in cars with the warning devices to be 100% and we also wish to know why and how the owners have defeated the system.

9) School Buses: Since school buses are provided for a part of our population which requires more than average protection, we are concerned with whether they are receiving it. Any accident involving a school bus will be studied to find out the reasons for the crash, including such things as the condition of the vehicle and the experience of its driver. In addition vehicle structures associated with passenger injury would be reported in a uniform manner.

10) Trucks: Trucks and other special heavy-load vehicles present special problems on the highway because of their size, weight, and operating characteristics which make them different from the main portion of the traffic stream. One particular problem is that passenger cars in a rear-end collision with a truck might "underride" the truck chassis. This study would be directed at finding out the prevalence of this problem in the general population of accidents, and by investigating the specific instances of it, determine if injury production is augmented.

11) Stalling: One of the side effects of the introduction of emission control devices on newer model cars, is that engine stalling occurs more frequently upon start-up or during acceleration. The purpose of this study would be to determine how frequently the situation arises, and when it does, is it a contributory cause to an accident. All such cases will be studied in detail to discover the relevant facts.

SUMMARY (OF NEW TRI-LEVEL APPROACH)

The SCRUB, VACUUM, and CARPETSWEEP programs are intended to provide a nationally representative set of data which can be used for drawing inferences about the state of traffic safety in the United States and for investigation of specific technical problems as they arise. It does not duplicate the compilation of state-provided information; rather it supplants it with a carefully controlled sampling procedure.

Individual teams operating in this program will be expected to have access to local police level data, and this may be the mechanism through which their sampling is accomplished. Further, it will be useful to develop exposure information about the several regions contributing data to the program, and one or more of the special (CARPETSWEEP) studies may be directed to the collection of exposure data.

There is no recommendation here for the compilation of the police level data for all of these regions. It is considered to be a local resource, of value for day to day direction of the team operations. But the variability of collection rules among police agencies in different parts of the country would make any complete compilation suspect.

Cost of this program has been estimated using knowledge of existing MDAI costs. In the next section of this report we will develop an argument for approximately 30 teams each of which provide about 500 VACUUM cases per year each. This number of cases (i.e, 15,000 total) will permit analysts to ask questions several levels down into the data, and still have a sufficient number of events to make statistical judgments. For example, with 15,000 vehicles, one might wish to look at the distribution of injuries to male and female drivers (adjusted for impact severity) for seven different collision configurations (rear-end, head-on, side-swipe, intersection-L, intersection-T, rollovers, and other single vehicle crashes). From other information we expect head-on involvements to be the smallest group (about 5%). This would yield about 750 cases with perhaps two-thirds male drivers and one-third females. It might be possible to go one further step - say for various times of the day.

With 30 teams and a total of 15,000 cases, costs of the total field program are estimated at $4.5 million. Program administration and data handling will add $500,000 to $1,000,000 to this figure.

Costs

System	Cases/Yr.	$/Case	Total
SCRUB	20	2500	50,000
VACUUM	500	150	75,000
CARPETSWEEP	500	40	20,000
			145,000

TRAFFIC SAFETY INDICES:

With the exception of the annual fatality count, and several derived statistics based on this, there are no really solid national indices of the state of traffic safety. Injury statistics provided by the National Safety Council are based on compilations of reports from many states, and the biases in the original reports are little known but generally quite varied.

The VACUUM program described herein has the characteristics required to develop a continuing index of traffic safety--a well defined sampling procedure, a consistent geographical sampling, and data in sufficient detail to provide several potentially useful statistics.

One important question is to determine whether the proportion of crashes involving a certain degree of injury is rising or falling. If, for example, the true proportion of crashes in the finite area involving injury is 50% in one month, and it declines to 45% in the succeeding month, with a sample size of 1200 this can be detected with a high degree of certainty.* In comparing this same proportion on an annual basis it is possible to detect a change from 50% to 48.5% with the same confidence.

Other measures (based on monthly comparisons) would have similar precision. For example, one could compare the AIS injury distribution for all occupants, the average and residual injury severities (as in the regression model developed in this report), the number of accidents and injuries per 100,000 population, or the observed percentage of seat belt wearers in the accident sample. Seasonal variations could be observed, and the time series patterns of a large number of parameters of potential interest to highway safety managers could be developed and followed.

Any data collection system of this type is likely to have unknown biases imposed by the particular geographical site selection and the survey team investigation procedures. This will be true even though every effort is made to keep the teams consistent by training and monitoring of operations. But the process of using the same locations and teams each month will control for biases of these types if they occur, and the true differences from one period to the next will not be affected by the unknown fixed biases. As opposed to fixed biases, the random or variable biases will contribute to the unexplained error. For example, if seat belt wearing is over-reported in summer months (because of a better opportunity to make observations) and underreported in winter months (and this phenomenon is not detected) this will contribute to the unexplained variability and hence reduce the precision of the statistics being estimated. This emphasizes again the need for consistent training and strong monitoring of the entire operation.

---

* The alpha and beta errors are both .05. This example is worked out in Appendix F.

Chapter VIII

THE GEOGRAPHICAL SAMPLE DESIGN FOR MØPUP

THE RESEARCH QUESTION

Depending upon the questions for which information is needed a survey approach to data collection may or may not be appropriate. For instance, if the inter-relationship of certain variables, given a crash experience, is under study, a controlled experiment is the only appropriate design. Assuming that the interrelationship between severity of human injury and a specific VDI cannot be studied because of ethical consideration, the current MDAI data provide us with surrogates of such an "experiment". This has been previously discussed under the use of MDAI data as a retroactive experiment. Although such data lack controls which would be ideal, the variable interrelationships of drivers, occupants and vehicle components can be studied, given a crash experience.

But what about the question regarding frequencies of occurrence? Current MDAI data files are inappropriate to determine how many people are killed annually in crashes with any given set of characteristics. MDAI data, because of personal judgment in team location and case selection cannot be used to infer the frequency of any crash or injury category relative to any large population than the MDAI data file itself. It is to this classification of research question that the following recommendations are addressed.

A SAMPLING PLAN FOR THE MDAI SYSTEM

MDAI teams are highly trained and closely coordinated with the various accident reporting systems in their locale. Training and jurisdictional system development are time consuming and expensive which restricts the number of such teams available and the capacity of the current system to be altered. It is unlikely that a purely statistical sampling methodology could be employed to select locations throughout the U.S. which will be representative of the nation. First, there is no clearly appropriate sampling frame of accidents from which to select team locations. Secondly, a random sample of locations which would ideally satisfy statistical sampling requirements would call for more teams than would be practical.

If the nation is stratified into geographic regions, East, MidWest, South, West, Pacific, each region except Pacific is now represented by at least one MDAI team. Although it is not possible to justify current team locations as "representative" of its particular region, the fact of each region's representation is an asset which can be utilized. The present geographic arrangement of the MDAI team is analogous to a cluster sampling design. Clustering is usually a necessary compromise to simple random sampling because the method reduces the cost of each data set collected. A properly conducted cluster sample produces no biases; however the MDAI situation was not created according to a sampling plan. In order to employ the present team location as if selection were based on a probabilility design, it is necessary to support the assumption of representativeness of the present MDAI locations. With some modifications it can be shown that as a sample, 15 of the MDAI teams operate in areas with characteristics much like the national norms when several matching characteristics are examined. This point will be elaborated below and provides the basis for the first proposed data collection modification of the MDAI systems.

MODIFICATION 1

If we assume that the teams operate in jurisdictions which represent their region, then, in order to apply inferential analysis, the case selection procedure must be altered to satisfy the requirements of probability sampling. Essentially these requirements are: (1) the probability of each case selection should be known and (2) case selection should not be biased by judgmental decision-making. Cases should be selected in a mechanical, random fashion.

Cases can be selected according to sampling procedures in a variety of ways. For instance, each k-th case of a given type can be selected by a team for investigation. Only each k-th case shall be investigated. When an adequately large number of cases are studied, then inferences can be made about the entire population of such cases for the team location. As previously mentioned if a specific crash or injury type is rare, all or most such cases can be studied; it is only essential that the size of the population N be known in order to calculate the probability of case selection.

It is important to be cognizant of sampling problems within selected team locations which can influence the sampling procedure. In any SMSA or other team location, accidents are reported from a variety of sources. Police agencies, wrecking companies, street departments and hospitals among others compile and report accident data. It is essential that such sources of accident reports use consistent definitions and data from such varied sources be consistent for the entire MDAI system. The importance of rigorous operational procedures in these areas cannot be over-emphasized because the sampling frames will be only as good as the data from reporting sources. It follows that no matter how rigouously random case selection procedures are adhered to, the sample is only as good as the sampling frame.

If the sampling frame consists, for instance, of all towaway accidents for a specified time period, then "towaway" must be consistently defined across all jurisdictions within team locations and across all team locations. We consider it to be primarily a problem of management to identify sources of definitional ambiguities and to provide the means for achieving consistency throughout the system.

Another operational problem which needs to be addressed is the jurisdictional complexity of several team locations (present and future). Clearly an investigating team cannot construct daily sampling frames of all accidents, in, for instance, New York City, if all accidents for the entire area are to be included. Sub-sampling within the team location is an operational necessity. But, what precincts or other jurisdictions can provide "representative" accidents on a routine basis with which to construct sampling frames? It is necessary to initiate in these cases, a multistage sampling design such that jurisdictions are selected on a probability basis within the team location. Then, cases are routinely sampled from the accident reports of the sampled jurisdiction(s).

In a complex team location a representative sampling frame of accident reports can be constructed on a daily basis with rapid response team capabilities if n reporting jurisdictions are randomly selected and then all accidents are compiled into a list from these n reporting sources. Random case selection procedures can become operational in a complex area with many reporting jurisdictions as well as in a simpler area once the daily sampling frame construction becomes systematic, consistent and routine.

This design can be schematically presented in Figure 16.

Figure 16

ACHIEVEMENT OF NATIONAL REPRESENTATIVENESS OF THE MDAI SYSTEM

CURRENT MDAI SYSTEM

INITIATE RANDOM
CASE SELECTION PROCEDURES

REGIONAL AND POPULATION SIZE STRATIFICATION
OF STANDARD METROPOLITAN STATISTICAL AREAS

INITIATE PACIFIC TEAM TO COMPLETE
REGIONAL REPRESENTATION

STAGEWISE EXPANSION OF SYSTEM BY SIZE AND
REGIONAL STRATA TO ACHIEVE NATIONAL REPRESENTATIVENESS

At the least, cases selected in this fashion could be representative of all such cases in the jurisdiction in which teams are located. If representativeness of team location to the U.S. generally can be demonstrated, then inferences can be suggested about the larger population.

MDAI LOCATIONS WITHIN SMSA's

Fifteen of the current MDAI teams are situated in Standard Metropolitan Statistical Areas. Data was taken on 19 matching variables for these SMSA's from Statistical Abstracts.* The 15 SMSA's were analyzed as a sample in relation to the total population of SMSA's and to the U.S. general population. Population rates for 13 of these variables are presented in tabular form in Table 39. It can be seen that for the variables presented, the jurisdictions now represented by MDAI team are not greatly different than either the average rates for the U.S. general population or for the total population of SMSA's. Given the restricted number of teams (n=15), the sample is fairly representative of the U.S. except for specific characteristics such as children under 5 years old, physicians per capita, old age public assistance recipients and ADC recipients. The variables that differ from the SMSA and U.S. populations are expected to differ because of the regional judgmental MDAI team location selection procedure. From Table 41 the similarity of the MDAI-SMSA's to the U.S. general population and the SMSA population can be noted.

In order to determine, from available data, which matching variables are most consistently associated with concerns of highway safety, motor vehicle fatality rates were used as the dependent variable in a stepwise multiple regression analysis using all the variables as predictors. Not surprisingly, exposure related variables 15 and 16,** predicted most significantly. The results of this regression analysis are shown in Figure 17.

One consideration of cluster sampling is the expectation that the within cluster variance is more homogeneous than the variance of the larger population. Therefore, if a representative sample

----

* Statistical Abstracts provides matching data only for SMSA's over 200,000 population.

** Var. 15 - workers using public transportation to work
Var. 16 - workers employed outside country of residence

Table 41

POPULATION RATES FOR SELECTED VARIABLES*

Population Rates	(n=15) MDAI-SMSA's	U.S. Population	All SMSA's
Children under 5 years old	.12263	.07599	.08283
Adults over 65 years old	.10740	.12000	.10800
Physicians	2.1846	1.5377	1.8916
Hospital Beds	3.9415	4.1492	4.1459
Families	.24981	.25186	.23123
Unemployment	.02888	.02458	missing
Total working force	.42589	.39402	.40615
Female working force	.16540	.15013	.15714
Workers taking public transportation to work**	.03128	.03501	.05196
Workers commuting out of county of residence to work**	.056867	.07014	missing
Old age public assistance recipients	.01016	.00981	.00739
Aid to dependent children recipients	.07068	.05182	.05962
Motor vehicle fatalities (12 mos)	.00014525	.00025595	missing

* All data taken from Statistical Abstracts, 1971 except Motor Vehicle Fatalities which were individually compiled from published sources by state.

** Exposure Variables (analytic no. 15 and 16)

Figure 17

Stepwise Regression Analysis of Two Exposure Variables and Other 1971 Census Statistics to Predict Motor Vehicle Fatalities by SMSA for N-15 MDAI Represented SMSA's*

Prediction of Motor Vehicle Fatalities "MVFATAL"

Selection of Regression

Step 1. Variable 15, "Workers Using Public Transportation to Work"
Analysis of Variance MVFATAL N=15 EQN=1

Source	DF	Sum of Sqrs	Mean Square	F-Statistic	Sig.
Regression	1	$.29284 \times 10^6$	$.29284 \times 10^6$	30.602	.0001
Error	13	$.12440 \times 10^6$	9569.2		
Total	14	$.41724 \times 10^6$			

Multiple R = .83776 R-SQR= .70185 SE= 97.822

Variable	Partial	Coefficient	Std. Error	T-Statistic	Signif.
Constant		99.432	35.623	2.7912	.0153
Var 15	.83776	2.7072	.48937	5.5319	.0001

Remaining	Partial	Signif.	
Var 2	.11661	.6914	median age
Var 3	.17304	.5541	total population
Var 4	-.19062	.5139	central city population
Var 5	-.05374	.8552	white population
Var 6	.46383	.0948	population density
Var 7	.02714	.9266	central city population under 5 years old
Var 8	.26920	.3520	central city population over 65 years old
Var 9	.13690	.6407	number of physicians
Var 10	.25062	.3875	number of hospital beds
Var 11	.15905	.5871	number of families
Var 12	.19996	.4931	total 1971 unemployment
Var 13	.19086	.5134	size of total labor force
Var 14	.15200	.6039	size of female labor force
Var 16	.80071	.0006	workers employed outside county of resid.
Var 17	-.22716	.4348	recipients of old age assistance
Var 18	-.06904	.8146	recipients of juvenile dependence assistance

* Caution should be exercised in drawing conclusions from analysis based on a sample of cases as small as n=15.

Figure 17 continued

Step 2. Variable 16, "Workers Employed Outside County of Residence"

Analysis of Variance of MVFATAL N=15 EQN=2

Source	DF	Sum of Sqrs	Mean Square	F-Statistic	Sig.
Regression	2	$.37259 \times 10^6$	$.18630 \times 10^6$	50.078	.0000
Error	12	44642	3720.2		
Total	14	$.41724 \times 10^6$			

Multiple R=.94499 R-SQR=.89301 SE=60.993

Variable	Partial	Coefficient	Std. Error	T-Statistic	Signif.
Constant		68.941	23.167	2.9758	.0116
Var 15	.84966	1.9387	.34734	5.5815	.0001
Var 16	.80071	.74934	.16184	4.6302	.0006

Remaining	Partial	Signif	
Var 2	.31154	.3001	median age
Var 3	.02025	.9476	total population
Var 4	-.14708	.6316	central city population
Var 5	-.16913	.5807	white population
Var 6	.04101	.8942	population density
Var 7	.01320	.9659	central city population under 5 years old
Var 8	.06440	.8344	central city population over 65 years old
Var 9	-.02969	.9233	number of physicians
Var 10	.00951	.9754	number of hospital beds
Var 11	.02632	.9320	number of families
Var 12	.08101	.7925	total 1971 unemployment
Var 13	-.00068	.9982	total labor force
Var 14	-.02414	.9376	female labor force
Var 17	-.13239	.6664	recipients of old age assistance
Var 18	-.10355	.7364	recipients of juvenile dependence assistance

Regression of MVFATAL Summary Forward Selection

EQN	R-SWR	STD ERR	#VAR		VARIABLE	PARTIAL	T-STAT	SIGNIF
1	.70185	97.822	1	in	Var 15	.83776	5.5319	.0001
2	.89301	60.993	2	in	Var 16	.80071	4.6302	.0006

of clusters (MDAI-locations) is expected, the clusters should represent the full-range of variablity of the larger population of all possible clusters (all SMSA's).

Table 42 presents the present 15 SMSA's represented by an MDAI team and the range of variable measures for SMSA's greater than 200,000 people. Population rates for variables consistently associated with motor vehicle fatalities are presented along with the full range of such rates for all SMSA's. It can be seen that the range of variability for these matching variables is somewhat covered by the current set of SMSA's which provides some evidence of the representativeness of present MDAI locations as adequate "cluster sites" for an inferential analysis to the population of SMSA's. It is also apparent that the variability of these exposure variables is different for the current MDAI/SMSA's than for the full SMSA population, indicating the need, eventually for a larger sample. With these considerations in mind, the first modification suggested here, Indicating Random Case Selection Procedures, will permit guarded inference to metropolitan areas in the U.S.

FURTHER MDAI MODIFICATIONS

More reliable and satisfying inferences of accident data can be made only if (1) the sample of team locations is enlarged and (2) case selection procedures are both systematic and consistent among the teams.

CRITERIA FOR THE DESIGN

Standard survey sampling procedures* have been referenced and the proposed MDAI modifications are based on these sources. In addition national sampling designs of the Survey Research Center, of the Institute for Social Research, University of Michigan have been examined in order to define operational considerations as well as statistical requirements. One operational consideration identified

----

* Kish, Leslie. Survey Sampling. John Wiley & Sons, Inc. New York 1967.

Lansing, John B., and Morgan, James N. Economic Survey Methods. University of Michigan, Ann Arbor, 1971.

Moser, C.A.. Survey Methods in Social Investigation. Heinemann, London, 1969.

TABLE 42

MINIMUM AND MAXIMUM VALUES OF TWO EXPOSURE VARIABLES AMONG
ALL SMSA'S AND AMONG MDAI-REPRESENTED SMSA'S
(1970 Census Data)

Variable 15, Workers Using Public Transportation to Work

	All SMSA's	MDAI SMSA's
Minimum	1.1% (Bakersfield, California)	2.3% (Salt Lake City, Utah)
Maximum	48.0% (New York City)	20.4% (New Orleans, La.)

Variable 16, Workers Employed Outside County of Residence

	All SMSA's	MDAI SMSA's
Minimum	1.5% (Phoenix, Arizona)	2.8% (Albuquerque, N.M.)
Maximum	44.0% (Richmond, Virginia)	36.0% (Atlanta, Georgia)

by Moser, Kish and ISR is the "apparent" representativeness of cluster sites on a national design when most clusters are defined as cities or SMSA's.

Determination of cluster sites for a national inference design for the MDAI system is influenced by both statistical and "political" criteria. On the one hand a "pure" random sample of cities, counties or SMSA's, properly stratified by size, would produce an efficient representative sample for inferential purposes. However, data on which to test representativeness for the population of highway accidents, accident-related injuries, etc., is unavailable. The factors (human and environmental) characterizing these study populations can be expected to vary in undeterminable ways from one political jurisdiction to the next and therefore a pure random sample might well be "undefensable" in the political sense. Therefore, in order to be above criticism, and because the modifications proposed are purposively optimal, a combination of statistical and political criteria have been employed.

For instance, it might be true that the highway accident exposure factors in Philadelphia, Baltimore, and Cleveland are virtually identical in most respects, and accident data from one might be generalizable to the others, but critics in these cities might not accept arguments to that effect which are based on the minimal matching data at hand. All of these three cities and the SMSA's represented are part of this national inference design, despite the fact that a purely statistically defined sample might not require them all.

The sample is in fact a probability sample where the probability of inclusion of any study site is known. It is also true that the probability of inclusion of several sites is unity. Example of the latter case are most SMSA's currently represented by an MDAI team and the largest U.S. cities.

A STRATIFIED PROBABILITY SAMPLE OF SMSA's

SMSA's have been routinely used by both university and private survey research institutes as well as the Census Bureau as primary sampling units for representative national inference. A host of supportive arguments for utilizing SMSA's can be presented including:

(1) Data availability

(2) Resource availability

(3) Potential for comparison with other surveys based on SMSA's as primary sampling units.

(4) Operational considerations

(5) SMSA's comprise a sampling frame which includes ½ the U.S. Population and all metropolitan areas with populations greater than 50,000

It is a relatively safe assumption that as well as being ideal primary sampling units (PSU's) for the U.S. human population, SMSA's are aggregate PSU's for the automobiles and related auto accident populations.

SAMPLE DESIGN

The modification of the MDAI system presented here culminates in a probability sample representative of metropolitan areas in the U.S. The design, as discussed above, is defensable both on statistical and operational grounds as long as systematic random case selection procedures are employed by all MDAI teams.

THE POPULATION

The sample is based on the metropolitan human populations of the U.S. where "metropolitan" is defined as a population cluster with 50,000 or more people. These clusters are documented as SMSA's by the Bureau of the Census. Although several SMSA's include rural components (counties, town, parishes) reliable inferences should only be made to metropolitan areas.

It is necessary to base the design on human urban or semi-urban populations because adequate data for accident populations is unavailable. We assume a high association between human and automobile populations. Therefore the design described here is probably not representative of either rural human or accident populations and inferences from the sample to rural populations must be regarded as inappropriate.*

SAMPLING FRAME

The most suitable and available sampling frame for a national probability sample of the U.S. population clusters is the list of SMSA's provided by the Bureau of the Census.

U.S. SMSA's have been stratified into four levels according to population size and into five geographic regions. Table 41 presents levels 1 and 2 and the SMSA's included. The number of SMSA's in levels 3 and 4 is much larger than is levels 1 and 2 because of the similarity of size of most U.S. metropolitan areas.

Also in Table 41 the current and desired sample size of MDAI team locations are indicated. Three of the seven largest SMSA's are now or have been represented by in-depth teams. Similarily, three of the 37 level 2 SMSA's are or have been represented by MDAI teams.

It is proposed that all eight level 1 SMSA's eventually be included in the sample and maintain an MDAI team ($p=1.0$). This constitutes sampling stage I. The 8 largest SMSA's represent 21% of the total U.S. population and 34% of the SMSA populations.

Level 2, Level 3 and Level 4 SMSA's have been stratified into five geographic regions. It is proposed that 10 Level 2 SMSA's be included in the final sample. Level 3 and Level 4 represent the greatest number of SMSA's and the largest variance in population size. Seventy-five SMSA's between 250,000 and 599,999 are included in Level 3 and the remaining 106 SMSA's in Level 4. Two MDAI teams

---

* For continuous surveillance of the traffic safety problem it is important to have a consistently defined population. Month-to-month or year-to-year changes in the traffic accident picture can be expected to be relatively small, and a well defined sample and adequate exposure information will be prerequisite to the measurement of time varying changes.

have operated in Level 3 and 2 in Level 4. For economy these teams should be included (p=1.0) and the remaining sample sites be randomly selected with proportionate considerations to geographical region. It is suggested that 10 to 13 additional sites be selected which provides that between 29 and 31 teams should be operating in the proposed optimal MDAI system.

Clearly it is not possible to initiate such a large system at one time or in one fiscal period. In fact, immediate establishment would not be wise for a number of reasons:

(1) Economy

(2) Sampling statistics for accident sample size such as S.E. and variance are not known for the U.S.

(3) Team development is time consuming and expensive.

With these considerations in mind a stagewise implementation schedule is suggested which eventually results in the optimal national sample design. The implementation schedule is presented in Table 43.

Table 43

Proposed MDAI National Sampling Frame of Standard Metropolitan Statistical Areas - Stratified by Region and Size - With Indication of Current MDAI Sites and Optimal Sample Size Within Strata

* Indicates Current or Former MDAI Location
\# Detroit SMSA is in part covered by the Oakland County MVMA team

SMSA		Region
Level 1. Eight largest metropolitan areas, minimum size = 2,730,000		
Boston, Massachusetts	*	East
Chicago, Illinois		Midwest
Detroit, Michigan	* #	Midwest
Los Angeles - Long Beach	*	West
New York, New York		East
Philadelphia, Pennsylvania		East
San Francisco - Oakland, Calif.		West
Washington, D.C.		East

Optimal National Sample Includes All Level 1. SMSA's
Probability of Inclusion = 1.0

Level 2. Metropolitan Areas between 700,000 and 2,729,999 population.

Albany-Schenectady-Troy, New York		East
Baltimore, Maryland	*	East
Newark, New Jersey		
Patterson-Clifton-Passaic, New Jersey		
Providence-Pawtucket-Warwick		

Optimal National Sample Includes One Level 2, Eastern Region SMSA

Buffalo, New York		Midwest
Cincinnati, Ohio		
Cleveland, Ohio		
Columbus, Ohio	*	
Dayton, Ohio		
Indianapolis, Indiana		
Louisville, Kentucky		
Milwaukee, Wisconsin		
Minneapolis-St. Paul, Minnesota		
Pittsburgh, Pennsylvania		
Rochester, New York	*	
St. Louis, Missouri		

Optimal National Sample Includes Three Level 2. Mid Western SMSA's

Table 43 continued

Atlanta, Georgia		South
Birmingham, Alabama		
Memphis, Tennessee		
Miami, Florida	*	
New Orleans, Louisiana	*	
Tampa-St. Petersburg, Florida		

Optimal National Sample Includes Two Level 2. Southern SMSA's

Anaheim-Santa Ana-Garden Grove, California		West
Dallas, Texas		
Denver, Colorado		
Fort Worth, Texas		
Kansas City, Missouri		
Phoenix, Arizona		
Sacramento, California		
San Antonio, Texas	*	
San Bernadino-Riverside-Ontario, California		
San Diego, California		
San Jose, California	*	

Optimal National Sample Includes Three Level 2. Western SMSA's

Portland, Oregon	Pacific
Seattle, Washington	

Optimal National Sample Includes One Level 2. Pacific SMSA

Of the 37 Level 2. SMSA's the Optimal National Sample should include 10. Of the 10 SMSA's included, 7 should be existing MDAI-SMSA locations for reasons of economy. Priority should be given to inclusion of one Pacific Region SMSA to provide Regional representation.

Level 3. Metropolitan areas between 250,000 and 5,999,999 population.

Of the 75 level 3. SMSA's, only two, Albuquerque, New Mexico and Salt Lake City, Utah now maintain an MDAI team. Both of these existing locations are in the Western Region. The Southern Region is over represented in Level 3 as it is under represented in Level 1. It is proposed that for the Optimal National Sample of SMSA's for a national MDAI system, two of the Level 3. locations be from the Southern Region, both of the existing Western locations be maintained, and two additional locations be randomly selected from the combined remaining SMSA's. Six Level 3. SMSA's should be included.

Level 4. Metropolitan areas between 50,000 and 249,000 population.

Of the 106 remaining SMSA's which constitute Level 4. only two now maintain an MDAI team, Ann Arbor, Michigan and Raleigh, North Carolina. It is suggested for the Optimal National Sample, that both of the existing MDAI locations in Level 4. be maintained and one additional location be selected for each SMSA Regional Strata. Seven Level 4. SMSA's should be included.

SUMMARY

Level 1.	n=8	N=8
Level 2.	n-10	N=37
Level 3.	n=6	N=75
Level 4.	n=7	N=106
Total	n=31	N=226

TABLE 44

Stagewise MDAI System Modification

(1) Initiate Random Case Selection Procedures (RCSP) in SMSA locations currently represents and MDAI team

Purposes:

1.1. Permit statistical analysis not possible before (for FY74 - data).
1.2. Test feasibility of procedures (RCSP)
1.3. Permit computation of variances of accidents populations necessary to determine optimal sample size for national inference

(2) Establish a new team in one of the un-staffed seven major cities (Chicago).

Purposes:

2.1. Test feasibility of (RCSP) in a new setting as a pilot for other system expansion in future
2.2. Increase SMSA representativeness of existent files
2.3. Test feasibility of using Federal teams operating from Regional offices
2.4. Determine cost estimates for new team establishment

(3) Establish team in Seattle or Portland

Purposes:

3.1. Permit SMSA coverage for Pacific N.W. not yet attained
3.2. Test (RCSP) feasibility and cost in a self-contained level 2 SMSA

(4) Establish teams in reamining Strata 1 SMSA's

Purposes:

4.1. Permit inference to all Strata 1 populations - nationwide
4.2. Refine logistical considerations of team development and case selection methodologies

(5) Stepwise implementation of teams in Strata 2 - 4 beginning with level 2

Purposes:

5.1. Stepwise increments in inferential capabilities - ultimately to all SMSA populations - nationwide
5.2. Final Sample Fraction represents approximately 40.4% of total U.S. population and 64.3% of total U.S./SMSA population

SUMMARY

The present MDAI data collection system might be called a non-probability judgment cluster sample of locations combined with unsystematic non-probabilistic case selection. As such, analytic statistics are inappropriate and population inferences from the file so collected cannot be made beyond the limits of the file itself.

The present MDAI Sample: In terms of sampling, the current data collection system is within the framework of quota or judgment samples. Team locations were selected by judgment, not as a probability sample from a population of potential investigation sites. Moreover, case selection is non-probabilistic because here, as with team location, the cases investigated are largely identified by judgmental decisions of accident reporters or the investigator's personal interests. Elements of personal choice enter the process of selection at both the team location and the case selection levels. One potential argument supporting the present procedure is that it appears to be efficient in terms of information gained per dollar spent on data collection. However, this argument is misleading because as a result of the elements of personal choice in the selection procedures, there is an inherent tendency to ambiguity and uncertainty as to exactly what is done. There is uncertainty as to the merits of the claim of more information per dollar when the information cannot be utilized to represent anything except itself.

There are three methodological difficulties with procedures of this type. First, the method may be biased. This is clearly the case of the current MDAI files which focus on specific crash types, extent of injury severity, and exceptional cases. Second the method may lead to an underestimate of the variability of the population. Again, this appears to be true of the MDAI files which are heavily skewed toward extreme values of property damage, extent of injury, and age of vehicle involved. Third, it may be difficult or impossible to estimate the variance of the sample estimates. This difficulty is only relevant if statistical presentation is planned which is intended to infer to a larger population. No such presentation (such as the frequency of accident-involved injuries among children who were not restrained in any way) is now possible with the MDAI data. This third difficulty can only be overcome if the basic sampling design is altered in such a way as to permit inferential analysis.

The current data collection system is appropriate if it can be argued that there is no feasible method of sampling for the questions such as those regarding accident causation. In this case a judgment sample may at least permit a preliminary assessment of inter-variable relationships, which can in fact be done with the current file. It is necessary, however, to alter the current system by introducing some kind of probability sampling if larger inferences are expected to be derived. Note that probability sampling is not the same as simple random sampling. All that is required is that the probability of case selection be known, the probability can be unity or any lesser value.

Appendix A

Comparison of the Characteristics of the Several Teams

TEAM CHARACTERISTICS

The present set of digital data contains information from 22 separate MDAI teams, four teams which have been supported by the MVMA, one jointly supported team, and (although the number is presently rather small) four teams supported by the Canadian ministry of transport. This report is primarily concerned with the characteristics and utility of the data emanating from the NHTSA-sponsored teams, and most comparisons will be made of that data alone. Nevertheless, approximately half the data at this time stems from the other sources, and they will be briefly described here.

MVMA TEAMS

The University of California at Los Angeles provided accident reports under MVMA sponsorship in 1969 and 1970, and 67 cases are included in this data set. These were all injury accidents involving recent model cars, and in general they occurred in Los Angeles, California. Average injury severity* in the vehicles (on the AIS scale) for these cases is 2.12.

Dr. Donald Huelke, at The University of Michigan, has investigated injury accidents involving new cars of American manufacturer continuously since 1969, and currently has 296 cases in this file. These are intended to represent a census of such vehicles in accidents in Washtenaw County, Michigan. In addition to the passenger cars recorded in the digital file a number of reports of truck accidents have been made. The average injury severity for these cases is 1.98.

---

* I.e., the mean value of the worst injury in each vehicle for which a CPIR form was completed.

Mr. Robert Scott, also at The University of Michigan, has investigated injury producing accidents in six communities in Oakland County, Michigan, since July of 1971. There are currently 277 cases in the file, again representing injury accidents with relatively new American-manufacturer passenger cars. The average injury severity for these cases is 1.39.

Mr. Ralph Darby, at The University of Michigan, reports on non- (or very minor) injury accidents involving new American-made passenger cars in Washtenaw County, Michigan. These are intended to complement Dr. Huelke's reports, and in combination they aim toward a census of the new-car towaway involvements in that county. In practice, they represent about 50% of the vehicles which qualify for investigation. The average injury severity for these cases is .415 on the AIS scale (0 = no injury, 1 = minor injury).

The Cornell Aeronautical Laboratory (now CALSPAN) provides reports on investigations conducted on injury-producing accidents involving new American made cars under a program sponsored jointly by NHTSA and MVMA. Their cases are identified in two groups in the digital file--as CAL III-A and CAL III-B. The former are done essentially to the MVMA specification, and may be compared with the UCLA, Huelke, or Oakland County cases. The latter are done to the NHTSA MDAI specification, and will be discussed later in the comparison of MDAI teams. Average severity for these (III-A) cases is 1.59.

It is noted that the reports furnished under MVMA sponsorship consist essentially of a Collision performance and Injury Report supplemented by a brief description of the accident and injury circumstances and a set of 35 mm. slides primarily relating to the vehicle damage and injuries. Since the full digital file contains much information available only in the NHTSA-sponsored cases, missing data codes are entered for many variables for the MVMA cases.

While the MVMA data is somewhat less complete, it has an advantage (from a statistical point of view) in that it is easier to describe the population from which these cases come, and to define them as a subset of that population. Two groups of data may be considered for specific analyses--one, identified as the Washtenaw County group and consisting of cases from teams 8 and 30 (Huelke and Darby) represent approximately half of the new car towaways in that county. Police level data for the same area permits some identification of the "lost" cases, and the analyst may judge the effect of this on a particular problem.* The second group consists of the Cornell III-A cases, the Oakland County cases, and the Huelke cases; these together are intended to represent the census of American-made new-car towaway injury accidents in their communities. The "new-car" definition is essentially for a car two years old or less, and thus a sliding group of new cars is presented for analysis of new vehicle features. The average injury level for these three teams (Cornell, Huelke, and Oakland) is seen to be different, and the reason for this is not entirely clear. Oakland cases are primarily urban, Huelke's primarily rural, and the CAL III-A cases approximately evenly split between urban and rural. Mean injuries (on the AIS scale) for these three teams is shown in table 45 . The F ratios shown are significant at better than the 1% level.

	N_{occ} N_{driv}	Mean Occup. Injury	Mean Driver Injury
Cornell III-A	438(276)	1.587	1.717
Oakland	451(275)	1.388	1.458
Huelke	505(296)	1.976	2.041
Average		1.663	1.746
F(2,1390),F(2,844)		21.103	11.622

TABLE 45. MEAN INJURIES FOR DRIVERS AND OCCUPANTS IN THREE MVMA-SPONSORED TEAMS

----

* See discussion of this matter in Section V of this report.

NHTSA TEAMS

The principal characteristics of the NHTSA-sponsored teams are discussed here in terms of (1) any special assignments given to the team by the sponsor, (2) the interests of the principal investigators, and (3) the environmental characteristics of the regions in which these teams operate. Following this discussion certain comparisons will be made of the data in the file. Data in the digital file at this writing is unfortunately not as complete as the number of cases submitted (because of a variety of delays in the reporting and encoding system), so that these comparisons are necessarily limited to the available data. Teams will be discussed in the order in which they appear in the file structure--i.e., alphabetical by the code name used in the file. ANN ARBOR--This is the MDAI team operating at The University of Michigan. Its basic assignment is considered the "standard MDAI" assignment--passenger cars of the last three model years in a "representative" set of accidents. Generally these will involve injury, although a few accidents may be included which do not. Trucks may be reported on (and are), although they do not appear in the present digital file except as the "other" vehicle.

The team members have shifted in emphasis over the period for which data is in the file from a relatively heavy emphasis on human factors to a more even emphasis on vehicle, human, and environmental factors. Accidents have intentionally been chosen to be of different types. The principal investigator's background is in engineering, and the team members who attend the accidents are supplemented by medical and engineering specialists as consultants.

Since this team operates in the same area as two of the MVMA sponsored teams, an attempt is made not to duplicate cases. As a result there will be fewer cases of new American cars and more cases of older cars, foreign cars, and pedestrians than might otherwise be reported on. Essentially all reported cases have on scene investigations.

BAYLOR

Initially the Baylor teams was led by a psychiatrist, and their investigations were oriented toward motivation and behavioral problems. For their earlier cases, case selection was essentially at their option. In about 1969 they were asked by the sponsor to concentrate on problems associated with vehicle inspection, and for the next year they investigated primarily cases involving vehicle defects. The geographical area covered was greater Houston. A particular feature of the team was the ability to impound vehicles for further study, and this along with a strong mechanical engineering talent should make for rather complete reports on vehicle-related factors. This team operation was discontinued in about 1970, and there are currently 68 cases in the present data file.

BOSTON

The Boston team is directed by the Boston medical examiner, and their on-scene investigator works in the medical examiner's office. It is understood that there's is an on-scene investigation in about one-third of the cases. Essentially all of the cases involve a fatality. Out of 60 cases reported, five or six are diagnosed as heart attacks, and three or four as probable suicides.

Boston reports are quite detailed in the medical and human factors areas, but competent in the vehicle area. A professional traffic engineer participates in analyzing and writing up some, but not all, of the cases.

Geographically the team operates in the "greater Boston" area, but most of the deaths reported occur within the city limits of Boston.

GEORGIA TECH

The principal investigator of the Georgia Tech team is a civil engineer, with interests in traffic and transportation planning engineering. The geographical area covered is metropolitan Atlanta, including DeKalb County which is a residential are adjacent to the city. Most of the accidents reported are in urban areas, and a substantial fraction are on freeways.

The basic selection rules call for recent model vehicles in accidents with at least one vehicle towed from the scene. In their earlier cases emphasis was placed on severe accidents. Sampling of the accident population was done with the intent of representing all kinds and times of day of accidents; a comparison made by the team showed that, in their first hundred cases, the time distribution was not significantly different from the general population of accidents.

Medical information was generated by two consulting surgeons (one a general surgeon, and the other a neuro-surgeon). Members of the Georgia Tech engineering and psychology departments provided the automotive and human factors expertise.

INDIANA UNIVERSITY

For the cases presently in the digital file the team was managed by two people with a primary background in law (although both of them also had degrees in the physical sciences). They were supported by a pathologist, several consulting physicians, and a team of on-scene investigators which included persons with a police investigative background. Essentially all accidents involved an on-scene investigation.

The geographical area covered was Monroe County, Indiana, and contact for on-scene investigation was made by monitoring the radio nets of the Bloomington (city) police, the county sheriff, and the Indiana State police. Many of the accidents investigated occurred in rural areas, mainly on two-lane roads; this county contained no limited access highways at the time of these investigations.

Selection rules were those referred to as "Standard MDAI"-- i.e., recent model year vehicles in (primarily) injury producing accidents. This team, as were several others, was occasionally called upon (by the sponsor) to investigate an accident which occurred outside the county, but such remote investigations do not constitute any substantial proportion of the data.

MIAMI UNIVERSITY

The Miami University team has been directed by a civil engineer, and generally provides good detail on the environmental factors. Human factors are covered by a team psychologist, and vehicle factors by member of the faculty in mechanical engineering.

The area covered is Dade County, Florida, and their normal selection rules include injury and fatal accidents involving passenger cars of the past three model years plus trucks of various ages. The Miami team normally fields five people to cover the scene of the accident, and provides good photography and note taking about the volatile information about an accident scene.

Injury detail is good, and the team is noted for its careful evaluation of the accidents regarding application of federal standards.

MARYLAND

The principal investigator of this team is the state medical examiner, and he is supported by a mechanical engineer and a traffic engineer, who are called in when the case warrants their attention. Retired police officers do the vehicle reporting. Human factors are covered intensively by a psychiatrist and others.

The geographical area covered is nominally the greater Baltimore area, although most of the cases are from the city of Baltimore. The team rarely investigates the accidents on scene.

Nearly all of the accidents investigated involve a fatality. There are occasional truck accidents and a few pedestrians.

NEW MEXICO

The principal investigator at the time cases now in the file were prepared was a physician, with a co-principal investigator who was a professor of mechanical engineering.

The geographic area covered by this team was Albuquerque, New Mexico, their notification of an accident coming primarily from the Albuquerque police. It can be expected that they will provide mostly urban accidents.

OHIO STATE UNIVERSITY

The principal investigator for this team was a mechanical engineer. Their geographic coverage was in and near Columbus, Ohio. Primarily urban accidents were investigated.

RESEARCH TRIANGLE INSTITUTE

There have been two principal investigators concerned with this team's operations--the first a mechanical engineer (until about mid-1971) and then a civil engineer for the remainder of the cases in the file.

Most of the accident sites investigated by this team can be defined as rural, their coverage including Durham, Orange, and Wake Counties in North Carolina. There are few freeways in this area.

The team attended most accidents on scene. Their reporting of environmental factors is felt to be particularly strong. They have good coverage of two-lane rural road accidents.

ROCHESTER

The principal investigator has been an orthopedic surgeon, and he is supported by the city traffic engineer and other medical consultants.

The geographic region covered by this team is the city of Rochester, New York and the surrounding rural area. The earlier cases were mostly identified by the fact that a patient with an orthopedic problem was seen by one of the physicians, and then the accident was followed up. Beginning in about the summer of 1971 the team began to investigate a number of accidents on scene, and cases with broader injury problems should appear in their data after that time.

In operation there are usually two people on scene at the accident, a vehicle technician and an interviewer. Their reports generally emphasize the vehicle crash phase and the medical details with minimal information on pre-crash details.

SOUTHERN CALIFORNIA

The team's principal investigator is a mechanical engineer, and he is supported by other engineers, a physician, and an interviewer.

Their geographic coverage is primarily within the city limits of Los Angeles, and they cover numerous freeway accidents.

Notification of accidents is obtained from the fire rescue service, and team personnel (two or three in a car) get to the scene of most accidents quickly. They choose cases, within the limits of the standard vehicle criteria, on the basis of severity.

STANFORD RESEARCH INSTITUTE (1 and 2)

SRI cases in the digital file are identified in two separate groups as SRI-1 and SRI-2. There have been two discrete MDAI programs at the Stanford Research Institute with a time period between the programs.

Those cases identified as SRI-1 were investigated under a special program aimed at studying the association between vehicle defects and accident causation. The principal investigator had an engineering background.

The team operated broadly in the Palo Alto area, with notification being provided by various police agencies when they observed a vehicle in which a defect was presumed to have contributed to the accident.

Those cases identified as SRI-2 have been acquired under a program which began about May of 1972. They have followed the usual selection rules (cars three years old or less) and have generally operated in the immediate vicinity of Menlo Park, California (although there are no specific jurisdictional limits on their investigations). Notification is provided by police agencies, and the team ordinarily investigates on scene. In addition to the principal investigator (who is an instrumentation engineer), a mechanical engineer, an orthopedic surgeon, and a civil (traffic) engineer participate in the program.

STANFORD UNIVERSITY

The principal investigator of this team was a physicist, with support from an orthopedic surgeon. This team operated for about one year (the time interval between the two SRI operations discussed above), and to some extent made use of the same consultants.

Their geographic area of coverage was also the Palo Alto-Menlo Park vicinity, and they concentrated on vehicles of three years or less in age.

Particular strengths in their reporting were in injury details and in reconstruction of the occupant dynamics.

UCLA-TRAUMA

UCLA had two investigation teams simultaneously, one under NHTSA sponsorship and the other under the sponsorship of the (then) Automobile Manufacturer's Association. Cases identified in the digital file as "UCLA-Trauma Research Group" constitute the output of the NHTSA-sponsored team.

The geographic area covered was generally the greater Los Angeles area. They covered relatively few rural accidents, and many freeway accidents.

The principal investigator was a mechanical engineer and he was supported by a physician, a traffic engineer, and others. Particular strengths of these reports are on vehicle damage and injury production factors. Case vehicles were selected from the last three model years. This team provided a larger number of truck and bus cases than most other teams, although these vehicles are not included in the present CPIR file because of its orientation to passenger cars.

SOUTHWEST RESEARCH INSTITUTE

The principal investigator's background is in mechanical engineering, and in the bulk of reports currently in the digital file he participated directly in the investigation and writing of the case. Human factors information was provided by an interviewer with a background in social work. Autopsy information was furnished by the county medical examiner, and other injury information was generally furnished by consulting physicians in area hospitals. Environmental reporting was performed by engineers and technicians associated with the team.

The geographical area covered was Bexar County, Texas. Members of the team monitored the city and county police radio nets, responded to what police termed "major" accidents. This resulted in a large number of dry runs, and selection of cases for full investigation was based on observations made on scene. The general rule for selection was that one of the vehicles in an accident had to be of the last three model years, and a vehicle had to be towed from the scene. Most accidents involve some injury.

Accidents were selected to represent a broad cross-section of types of events in the area. The percentage of nighttime accidents was roughly in proportion to those of the general population. The human factors information is considered to be emphasized more strongly than other details, and the written reports should reflect this. Other areas are adequately reported in terms of the material in the digital file.

TULANE UNIVERSITY

The principal investigator of this team was a physician in the University School of Medicine, but his direct participation in the operations were limited. A professor of civil engineering was the person most directly concerned with team operations.

The geographic area covered was metropolitan New Orleans, with the majority of the accidents being investigated in urban areas.

These cases are felt to be somewhat weak in interpretation of the highway safety standards, and minimal regarding examination of the vehicle. Their selection rules were similar to the other "standard" MDAI teams--i.e., vehicles three years of age or less, etc.

UNIVERSITY OF UTAH

The principal investigator is a physician and a medical examiner. The co-principal investigator is a member of the mechanical engineering department at the University of Utah, and has a background in vehicles and vehicle dynamics. A professor of civil engineering was associated with the team for the earlier reported cases, and in the more recent cases the environmental workups have been provided by contract with the highway department of the State of Utah. Strong support in the field of psychology has been provided.

The area covered is defined as the greater Salt Lake City region, with a large percentage of rural accidents. The on-scene investigator monitors radio channels of city, county, and state police and responds to accidents of interest. Most accidents are investigated on scene by at least one person.

Vehicles are investigated in detail most vehicles being brought into the laboratory for examination. Particular strengths are in the environmental reporting and in the interviewing and physchological followup.

OTHER TEAMS

Several teams have been in operation for such a short time that little data is presently available in the digital file. These will be discussed in a single group here.

A team operating at the University of Houston concentrates on pedestrian accidents, and completes a set of data forms relevant to that subject which are not compatible with the general MDAI files. Thus no cases from that team are included here.

A team has been established at the University of Kentucky; at the present time only three cases from this source are included in the digital file. And a team has been established at the University of Oklahoma, and two cases have been added to the file from there.

CONCLUSIONS

From observation of the present and past MDAI team operations, one may infer that the entire set of data will be biased strongly in two directions--first, there should be preponderance of relatively new cars, second, the reported accidents should be generally more severe than, say, the complete set of police-reported collisions in the same areas.

Older cars will be included in the data set since the present practice is to complete all of the reporting forms on each vehicle in an accident which qualifies for investigation. Thus if a 1971 passenger car is in a two vehicle collision with a 1950 pickup truck, the latter will also be reported in detail. If an analyst desires to study the entire set of data with respect to some factor which depends on vehicle age (for example the effectiveness of energy absorbing steering columns which were introduced in 1967 and 1968 models), he should recognize that the most of the pre-1967 vehicles in the data will have been in two vehicle collisions, whereas a substantial number of the newer vehicles will have been in single vehicle collisions--the former implying an excess of side damage, and the latter an excess of frontal damage .

The injury severity bias comes about for two reasons--several teams produce reports only on fatal accidents--specifically Boston and Baltimore. Most of the other teams exhibit somewhat of a mixture of injury severities, but nearly all of them investigate a greater proportion of fatal accidents then occur in the population of police reported accidents.

An exception to the severity bias occurs in the set of data defined by the combination of two MVMA teams and the MDAI team operating in Washtenaw County, since there an attempt is made to get a census of the new car towaway (injury or not) accidents. In Chapter V of this report the use of this subset of the data is discussed.

TEAM CHARACTERISTICS AS REFLECTED IN THE ACCIDENT DATA

As observed above, one might expect to see some variation in the team reports because of their geographic environment, their team selection rules, and the personal characteristics of the investigators. A search of the digital files comparing the several teams across a number of characteristics reveals that there are few measures in the file for which the teams are similar.

On the other hand, the entire collection of data may be compared with a broader source for some variables. Table 3 (in Chapter 3) presents some summary statistics for the combined data of all NHTSA-sponsored teams, and indicates that the percent female drivers is about the same as (for passenger cars) in the National Accident Summary; the average driver age is about the same; the percent rural accidents is slightly lower than in the NAS, but not greatly dissimilar, and the percent of accidents occurring in hours of darkness is somewhat greater than the NAS.

Also compared in this same table is the mean injury severity (on the AIS scale) for all occupants of the MDAI case vehicles and all occupants of vehicles in the Washtenaw County "towaway" set. The latter includes an attempted census of recent model year vehicles involved in accidents in Washtenaw County in which the damage was sufficient to require towing of the vehicle (nearly half of these cases involve no injury). From this comparison it can be seen that the severity of the MDAI cases is relatively high.

Individual team comparisons are presented in tables 46 through 48. These were tabulated using a one-way analysis of variance program with MDAI team as the independent variable. The number of cases (drivers, occupants, vehicles) varies slightly from column to column depending upon the missing data, but the values shown are approximately correct. For each tabulated variable an F-ratio is given, and it can be seen that there is significant variation in nearly all of these. One may conclude that no single team is very representative of the whole, and that the entire set of data, while biased toward the more serious collisions, does not vary widely in those characteristics compared from the National Accident Summary.

Table 46
Comparison of Accident/Environmental Characteristics
MDAI Teams

Team #	Team Name	Number of Vehicles	% Rural	% on Curve	% Not on Limited Access Hwy	% Not on Slippery Surface	% Not Rollover
1	Ann Arbor	61	55.7	21.3	80.3	75.4	85.2
2	Baylor	67	1.5	5.9	76.5	80.6	92.6
3	Boston	31	28.1	66.7	68.7	67.7	84.4
5	Cornell-B	119	31.9	14.3	96.6	81.5	94.1
7	Georgia	63	14.3	23.8	74.6	85.7	88.9
9	Indiana	25	72.0	72.7	100.0	60.0	68.0
11	Miami	96	20.6	14.4	92.8	90.6	88.7
12	Maryland	56	42.9	33.9	76.8	78.6	80.4
15	New Mexico	70	14.3	7.1	67.1	94.3	85.7
17	Ohio	30	83.8	23.3	73.3	76.7	90.0
18	Res Tri	76	90.8	31.6	93.4	80.3	84.2
19	Rochester	52	36.5	11.5	90.4	65.4	98.1
20	Southern Cal.	36	100.0	5.6	80.6	94.4	97.2
21	SRI-2	1	100.0	100.0	100.0	100.0	100.0
22	SRI-1	7	26.6	14.3	100.0	100.0	100.0
23	Stanford	36	22.2	11.1	55.6	75.0	97.2
24	Southwest R	168	37.5	14.3	83.3	85.7	86.9
25	UCLA-Trauma	68	29.0	33.3	59.4	89.7	82.6
26	Tulane	46	13.0	10.9	80.4	84.8	93.5
29	Kentucky	3	100.0	100.0	100.0	100.0	100.0
31	Oklahoma	2	100.0	100.0	100.0	100.0	100.0
33	Utah	19	84.2	26.3	68.4	84.2	63.2
	Totals/Average	1131	33.4	20.0	81.2	82.7	88.1
	F-Ratio		17.028	8.104	5.523	2.634	2.453
	Significant		$<1\%$	$<1\%$	$<1\%$	$<1\%$	$<1\%$

Table 47
Comparison of Occupant Characteristics, MDAI Teams

Team #	Team Name	Number of Occupants	Average Occupant Age	Average Injury Severity (AIS)	% Not Belted	% Female	Average Height (Inches)
1	Ann Arbor	112	33.724	2.383	80.8	48.2	65.773
2	Baylor	94	29.007	1.073	77.8	35.1	68.547
3	Boston	62	30.837	4.694	98.1	35.5	68.172
5	Cornell-B	177	32.891	1.083	78.0	40.7	66.549
7	Georgia	103	30.437	2.029	82.8	27.2	67.641
9	Indiana	51	27.440	2.471	85.4	31.4	67.800
11	Miami	197	28.859	2.450	81.8	31.5	65.175
12	Maryland	95	30.284	3.796	87.8	20.0	69.078
15	New Mexico	124	30.528	1.879	82.4	45.2	66.273
17	Ohio	49	30.809	2.271	79.2	38.8	65.682
18	Res Tri	143	28.486	1.535	84.0	39.2	67.569
19	Rochester	84	35.711	2.141	76.3	25.0	67.845
20	Southern Cal	59	39.661	0.934	70.4	42.4	66.286
21	SRI-2	4	15.250	2.000	100.0	50.0	67.333
22	SRI-1	7	33.143	2.000	20.0	42.9	69.143
23	Stanford	48	29.042	2.320	73.0	33.3	67.813
24	Southwest R	267	28.530	1.647	80.3	37.8	65.191
25	UCLA-Trauma	134	30.830	2.850	82.2	33.6	68.405
26	Tulane	79	30.557	1.743	91.2	30.4	65.788
29	Kentucky	7	28.286	1.000	80.0	57.1	63.286
31	Oklahoma	5	13.400	0.800	60.0	40.0	65.667
33	Utah	35	25.371	2.829	93.7	40.0	67.000
	Total/Average	1936	30.520	2.078	81.5	35.8	66.641
	F-Ratio		2.709	13.674	3.035	1.927	2.535
	Significant		<1%	<1%	<1%	<1%	<1%

Table 48

Comparison of Driver Characteristics, MDAI Teams

Team #	Team Name	Number of Drivers	Average Age	Average Injury Severity (AIS)	% Not Belted	% Female	Average Height (Inches)
1	Ann Arbor	60	35.393	2.180	77.6	31.7	67.712
2	Baylor	68	31.868	0.941	76.7	27.9	68.455
3	Boston	32	38.969	4.937	96.4	9.4	68.389
5	Cornell-B	119	36.252	1.110	77.5	31.9	68.044
7	Georgia	63	33.429	1.905	82.5	20.6	68.170
9	Indiana	25	27.800	2.240	77.3	12.0	70.292
11	Miami	97	34.274	2.726	79.1	20.6	67.697
12	Maryland	56	33.464	4.073	83.0	12.5	69.547
15	New Mexico	70	34.771	1.971	72.7	35.7	67.547
17	Ohio	30	34.033	2.833	86.2	30.0	68.630
18	Res Tri	76	33.434	1.560	76.7	34.2	68.212
19	Rochester	52	37.981	2.000	72.0	11.5	68.682
20	Southern Cal	36	39.944	0.972	71.4	30.6	67.941
21	SRI-2	1	17.000	2.000	100.0	100.0	68.000
22	SRI-1	7	33.143	2.000	20.0	42.9	69.143
23	Stanford	35	29.771	2.400	65.4	31.4	68.629
24	Southwest R	167	32.497	1.833	76.9	27.5	68.435
25	UCLA-Trauma	70	34.493	3.282	79.7	18.6	68.677
26	Tulane	46	35.543	1.791	90.2	26.1	68.364
29	Kentucky	3	38.667	1.000	100.0	100.0	71.333
31	Oklahoma	2	19.500	0.500	50.0	50.0	68.000
33	Utah	19	27.421	2.737	88.2	31.6	67.667
	Totals/Average	1133	34.144	2.124	78.2	25.7	68.334
	F-Ratio		1.568	9.875	1.529	1.852	1.181
	Significant		2.4%	1%	5%	1%	Not Sig.

Appendix B

Bias in Crash Injury Severity by Investigation Year

One of the stated objectives of the in-depth crash investigation program is to:

> "Determining accident and injury causation as well as the effectiveness of new safety features"

It has been shown in Chapter V and Appendix E that a multiple regression model can be used to compute a predicted AIS for each crash. The residual (observed AIS minus predicted AIS) from this regression model is a measure of injury severity after the contributions due to crash severity have been removed. For example, this measure was previously used to indicate the injury reduction from restraint system usage. The fundamental assumption underlying the use of residual AIS is that a random sample of occupant injuries are obtained from the crash population subsets defined by the regression model predictor variables.

Evaluation of vehicle standards is performed by comparing the mean residual injury from crashes in which the vehicle had the required equipment versus those in which the vehicle did not have the required equipment. This strategy will provide the required evaluation if there are no other confounding variables which are correlated with the change in equipment. For example new vehicle standards have been implemented with changes in model year. Therefore if more severe crashes were selected for investigation from different model years, then differences resulting from the vehicle standards could be either reduced or increased by the crash selection procedure.

Table 49 indicates that such a bias did occur. Specifically it can be seen that both the mean AIS and residual AIS decreased over the investigation years 1969 through 1972. In addition crash

TABLE 49. ANALYSIS OF INJURY BIASES IN CPIR FILE (Model Year/Accident Year Comparison)

Investigation Year	Vehicle Model Year	Number of Occupants	Mean AIS	Residual AIS
1969	1967	26	2.54	.41
	1968	53	2.85	.38
	1969	91	2.64	-.01
	1970	16	2.63	.38
1970	1967	26	2.81	1.76
	1968	70	2.29	.27
	1969	132	2.17	.02
	1970	373	1.66	-.01
	1971	19	1.26	.29
1971	1967	23	2.17	.48
	1968	48	1.58	-.04
	1969	190	1.95	.05
	1970	616	1.22	-.03
	1971	547	1.28	-.01
	1972	50	1.12	-.14
1972	1967	1	1.00	.16
	1968	5	1.00	-.09
	1969	16	2.88	.91
	1970	231	.84	-.15
	1971	346	.94	-.08
	1972	185	.89	-.22
Total	3064*			
F Statistic			17.15	3.99
Significance			0.00	0.00

*Includes all occupant injuries in the January 1973 that had sufficient data for including in the multiple regression model.

investigations were generally conducted for late model vehicles during any given investigation year. For example most of the 1968 model year vehicles in the file were investigated during 1969 and 1970. Thus comparison of 1968 model year vehicles with 1970 model year vehicles is confounded with differences between investigation years. This confounding could be removed by restricting any analyses to only those crashes investigated during 1971 and 1972. However such a strategy severely reduces the number of 1968 model year cases that are included in the analysis.

Tables 50 and 51 which compare injuries by domestic manufacturer within investigation year, indicate a similar problem. Specifically the proportion of General Motors vehicles in the sample, by investigation year, have decreased while the proportion of Ford vehicles have increased. This resulted from the large number of Washtenaw County crashes included in the file (e.g., Washtenaw County has a large proportion of Ford cars). Any direct comparison between Ford and General Motors would be biased in favor of Ford, since a larger proportion of Ford crashes were obtained in later investigation years.

Analysis of vehicle safety standards are severely restricted by the crash investigation bias. During 1971 and 1972 this bias has been largely removed. However the exclusion of older vehicles during these years has also severely handicapped the capability for evaluating vehicle safety standards.

TABLE 50. ANALYSIS OF INJURY BIASES IN CPIR FILE
(Accident Year/Manufacturer Comparison)

Investigation Year	Manufacturer (Domestic Only)	Number of Occupants	% of Investigation Year	Mean AIS	Residual AIS
1969	GM	121	50.4%	2.26	.22
	Ford	76	31.7%	2.41	.05
	Chrysler	38	15.8%	3.26	.14
	AMC	5	2.1%	2.20	.01
		240			
1970	GM	307	45.4%	2.06	.06
	Ford	221	32.6%	1.74	-.04
	Chrysler	122	18.0%	1.96	.17
	AMC	27	4.0%	2.00	.19
		677			
1971	GM	697	43.8%	1.40	-.02
	Ford	537	33.8%	1.31	-.06
	Chrysler	292	18.4%	1.47	-.03
	AMC	64	4.0%	1.44	.13
		1590			
1972	GM	315	38.8%	1.06	.03
	Ford	333	41.1%	.87	-.27
	Chrysler	134	16.5%	.90	-.03
	AMC	29	3.6%	1.00	-.15
		811			
Total		3318			
F Statistic				15.1	1.39
Significance				0.0	.14

TABLE 51. ANALYSIS OF INJURY BIASES IN CPIR FILE (FRONTAL IMPACT CRASHES ONLY)

Investigation Year	Manufacturer (Domestic)	Number of Occupants	% of Investigation Year	Mean AIS	Residual AIS
1969	GM	38	52.1%	2.32	.37
	Ford	20	27.4%	1.90	.14
	Chrysler	13	17.8%	3.00	.72
	AMC	2	2.7%	3.00	.25
		73			
1970	GM	119	47.0%	2.07	.05
	Ford	81	32.0%	1.99	.30
	Chrysler	43	17.0%	1.35	-.34
	AMC	10	4.0%	2.30	.45
		253			
1971	GM	246	43.1%	1.37	.11
	Ford	195	34.2%	1.11	.00
	Chrysler	105	18.4%	1.32	.23
	AMC	25	4.4%	1.12	.08
		571			
1972	GM	120	41.1%	.96	.03
	Ford	118	40.4%	.90	-.12
	Chrysler	44	15.1%	.93	.15
	AMC	10	3.4%	.90	-.24
		292			
Total		1189			
F Statistic				6.46	1.37
Significance				0.0	.15

THE USE OF THE MDAI DATA IN EVALUATING THE EFFICACY OF THE HIGHWAY SAFETY PROGRAM STANDARDS:

The Highway Safety Program Standards express the importance of each state's supporting activity toward compliance with the standards. Methods of complying with the standards are suggested in detail by a series of documents prepared by professionals in each of the several fields, and with respect to any particular accident report it is possible to judge whether the lack of compliance (as when a defective vehicle is implicated) might have been a causative factor, or whether compliance with the standard (as in the installation of breakaway highway signs) was a factor in reducing the severity of the accident or injury. Such factors related to the state standards are routinely reported in the MDAI cases, and some of these are encoded into the digital file. It is of interest to consider whether such information can be used in a statistical analysis sense to assess the value or shortcomings of the highway safety standards.

One means of approaching this problem would be to measure the incidence of some factor (such as whether or not a driver had taken a formal course in driver education, or whether or not the vehicle had been inspected) in the accident data and to compare it with the incidence in some control or exposure population. But as has been noted elsewhere in this report the exposure population has not been measured in most cases. However there are some notable exceptions in the case of random samples of night drivers conducted as part of the many Alcohol Safety Action Programs (ASAP).*

Even without exposure information it would be possible to analyze the data using the method of induced exposure (discussed in Chapter IV), but this would be of limited value unless the crash sampling method could be specified.

---

* Carlson, W.L., Chapman, M.M., Clark, C.D., Filkins, L.D., and Wolfe, A.C. "Washtenaw County BAC Roadside Survey Report", Ann Arbor, Michigan, Highway Safety Research Institute, The University of Michigan 1971.

With this in mind, possible uses of a better defined data set will be discussed with respect to each of the highway safety program standards. The comments are summarized briefly in Table 52 which displays the likely value of the present data and of data from a future controlled sample in the evaluation or continuing assessment of the highway safety standards activities. The assessments are based on a reasoning of the adequacy of the data and methods of analysis.

Table 52

State Standards Evaluation Summary

STANDARD	Can the efficacy of this standard be evaluated using the present MDAI data?	Would a future controlled sample be useful in the evaluation of of this standard?
Vehicle inspection	No	Doubtful
Vehicle Registration	No	No
Motorcycle safety	No	Possibly
Driver education	No	Possibly
Driver licensing	No	No
Codes and laws	No	Possibly
Traffic Courts	No	No
Alcohol	No	Possibly
Identification of accident locations	No	No
Traffic records	Partially	Possibly
Emergency medical services	No	Possibly
Highway design	No	Possibly
Traffic control devices	No	Possibly
Pedestrian safety	No	Possibly
Police traffic services	No	No
Debris removal	No	Possibly
Pupil transportation systems	No	Doubtful
Accident investigation	No	No

1. MOTOR VEHICLE INSPECTION

If periodic motor vehicle inspection is effective in preventing accidents the population of inspected vehicles should be less likely (other things being equal) to become involved in accidents that those which were not inspected. It is usually not possible to have both inspected and uninspected vehicles within the same jurisdiction, so one might try to associate accident rate with time since inspection--judging that a lower involvement rate for vehicles inspected more recently would be evidence supporting vehicle inspection. Such measures have been attempted (for example in a study by B.J. Campbell in North Carolina) using police reported accident data, and in an unpublished study by NHTSA personnel in Washington, D.C.

Within the MDAI file there is no exposure information on the parent population, i.e., there is really no way of determining the number of inspected vehicles in the parent population. It is possible to tabulate the number of vehicle defects which contributed to accidents in the MDAI file, and to consider whether such defects might have been eliminated by an inspection system. In part because of a particular bias in the file (two teams submitted only cases involving vehicle defects), and in part because of some misunderstandings in the reporting system (many defects are listed which are not in fact identified as causative factors), it is not possible to infer any national frequency of defects which are associated with crashes--from the present set of information. A better defined sample might do this, and thus might lead to useful ideas about inspection methodology, but it is concluded that the present compilation of data is of little value in this regard.

2. MOTOR VEHICLE REGISTRATION

This standard requires that each state provide for recording and making available (for accident research) such information as make, model, license plate number, name of owner, etc., for each vehicle licensed to operated by that state. The relationship to highway safety is circuitous in that it is expected that such recorded information will permit data analyses which identify accident causation factors, or would allow vehicles to be identified for such purposes as recall campaigns. If, for example, vehicles under recall but not repaired were overrepresented in the accident population (relative to those in the non-accident population) one might infer that the registration functions could be improved by developing a system to routinely notify owners of recalled vehicles. With the lack of exposure in the present file such inference would not be possible; but even in a larger more defined set of data such an analysis seems unlikely to be productive.

3. MOTORCYCLE SAFETY

While there are a small number of motorcycle accidents reported by MDAI teams, the present digital file is tailored specifically to passenger cars and small trucks and does not contain enough information about motorcycles or motorcyclists to be of any statistical value.

4. DRIVER EDUCATION

The safety value of formal driver education courses has been the subject of numerous studies both before and since the creation of the National Highway Safety Administration, most of which are inconclusive. This is not to say that such educational programs have no value, but simply that the experiments which have been conducted have generally been so imprecise or confounded as to make any firm conclusion impossible. Again because of the lack of exposure information it is not possible to use the MDAI data to evaluate driver education directly.

In section IV of this report a discussion of the induced exposure method is given, and some evidence of the association of driver education with accident prevention may be gained by conducting such an analysis if the MDAI data were sufficiently representative of a larger population. With the present data there was no clear indication that formally educated drivers were less likely to "cause" accidents than those without formal education, but for the present this must be viewed as a conclusion relative to the existing data set and not generalizable.

5. DRIVER LICENSING

One could speculate that unlicensed drivers are more likely to be involved in accidents than those with valid licenses. An alternative hypothesis is that unlicensed drivers drive very carefully and are less likely to be involved in accidents. Neither can be confirmed without some knowledge of the control population--i.e., what number of unlicensed drivers are there operating on the roads.

Again it is possible that individual cases will report licensing system improvements which might have prevented a specific crash. In a particular case a drinking driver who caused two successive accidents within a span of about a minute was reported (following the MDAI team investigation) to have been without a driving license for 15 years, although he had driven regularly throughout that period. It was noted by the investigator that the licensing system had some loopholes which might well be bottled up, and this anecdotal information may be useful in some way. But no statistical analysis of the present data is seen to be of value.

6. CODES AND LAWS

This standard suggests that all states should adopt motor vehicle codes consistent with their neighbors, and strive for consistency of laws within the state. Anecdotal information about misunderstanding of traffic laws in connection with an accident (e.g., someone from California making a right turn on red in a state where such an action is prohibited) may strengthen the arguments for standardization. The present data file does not code this item; but if it did it would be of interest only as a collection of such anecdotes and without statistical value because of the lack of sample definition. If the sample were more precisely defined, and this item of information more consistently reported, some estimate of the frequency of law misinterpretation as an accident causation factor would be possible This would seem to be of some use even without an exposure measure.

7. TRAFFIC COURTS

The traffic court standard requires, among other things, that drivers charged with all moving hazardous violations appear in court, that the court system be operated under uniform procedures and be financially independent of fees.

MDAI reports include information on the number of violations and accidents assigned to drivers of accident involved vehicles, but no information about the type of court system in which such violations were tried. It is concluded that there is not sufficient detail in the present MDAI investigations to make any statistical analysis relative to the efficacy of the court standard.

8. ALCOHOL IN RELATION TO HIGHWAY SAFETY

The place of alcohol as a contributing factor in highway accidents is as apparent in the MDAI data as in many other sets of data. But the finding that drunk drivers "cause" accidents can hardly be considered as a worthy output of this program. The alcohol standard, and its sequelae such as the ASAP programs, are intended to decrease this problem.

MDAI cases can continue to provide anecdotal information in this regard, and it is possible that a program operated in conjunction with an ASAP area might provide some statistical information appropriate to that area. But the entire compilation of MDAI information is presently of little value in assessing the effectiveness of this standard because of sampling problems.

9. IDENTIFICATION AND SURVEILLANCE OF ACCIDENT LOCATIONS

Individual investigation have often led to the identification of hazardous locations through publicity or contact between the investigators and highway or traffic engineering departments. The general intent of this standard, however, is for the identification of such locations within each jurisdiction by continuous surveillance of the roadway network and the analysis of mass accident information.

No method is seen at this time which would make the collected MDAI data of value in evaluating this standard.

10. TRAFFIC RECORDS

The traffic records standard requires that states maintain certain kinds of data in such a form as to permit a continuing assessment of their traffic accident problems. One particular value of the MDAI program is to furnish a sort of standard for comparison with the police reported information.

Certain analyses have been performed within tri-level jurisdictions comparing, for example, police injury reporting scales with the abbreviated injury scales used in MDAI cases. Such (police) scales vary widely from one jurisdiction to another, and it seems unlikely at this time that there would be much value in a national comparison, but the analyses conducted to date have served to identify the differences in reporting.

Such analyses should ultimately lead to a better understanding of the police level data, and one can surmise, to better results from the application of the traffic records standard. In addition, the investigation methodology developed by in-depth teams could be used to update procedures used by police. But it seems unlikely that any direct measure of the effectiveness of this standard can be made by analysis of the present MDAI files.

11. EMERGENCY MEDICAL SERVICES

Because of the lack of sample design it is not possible to perform more than a cursory survey of the role or effect of emergency medical capabilities in the MDAI cases. While this item is usually discussed in the case reports, information about it is not currently coded into the digital file--e.g., there is no digital record of the time spent waiting for an ambulance, the quality of the emergency medical treatment, the distance to the hospital, etc. Even if there were there would be little value in it.

If the data had been acquired in response to a sampling plan, and information relating to the emergency medical standard were recorded, it would be possible to get a measure of the characteristics of individual (sampled) emergency medical systems, and to describe a combined system represented by all of the sampled regions. If an appropriate sampling plan is set up it would be useful to record emergency medical information and to place it in a digital record for analysis.

12. HIGHWAY DESIGN, CONSTRUCTION, AND MAINTENANCE

The highway design standard requires that the states and local jurisdictions maintain existing highways properly, and that new highways conform to the best safety practices. This implies, for example breakaway light poles, installation of median barriers on narrow medians, etc.

The present MDAI file contains only sporadic information about highway features--in part because of the lack of any formal report form for this information. But again, because of the lack of sampling, one could take the compiled information only as an indication of the types of problems which exist and not of their frequency. There are, for example, some 250 accidents in this file involving guard rails. Information may be reported by the investigator regarding the adequacy of the guard rails (too high, too rigid, etc.), but this is not presently coded into the digital file.

If the sampling were adequate it would be possible to draw some inference about changes in the highway system by reporting and analyzing such roadway information. At the present time it is not possible to make any judgment about the adequacy of implementation of the highway design standard.

13. TRAFFIC CONTROL DEVICES

This standard is concerned with the state implementation of traffic control device improvements that bear directly on reducing accidents. This includes signs, signals, markings, and a variety of electronic controls.

The incidence of substandard traffic signals, or of the contribution of a malfunctioning signal to an accident is occasionally noted by investigators. There is at present no digital information maintained about this item.

If the sampling were appropriate, and the proper data recorded, it would be possible to keep a pulse on the changes in response to this standard. But at the present time there is little that can be done with the compiled data here.

14. PEDESTRIAN SAFETY

The pedestrian safety standard is concerned both with pedestrian education and with the use of traffic controls and pavement markings to protect pedestrian movements.

Although pedestrian accidents are occasionally reported by teams in the present MDAI program, the information developed is generally not appropriate to the digital files which are concerned mostly with vehicle damage and occupant injury. Both the frequency of cases, and the lack of sampling lead to an inability to assess the national pedestrian accident situation from this data.

15. POLICE TRAFFIC SERVICES

This standard requires increased training of police officers, and that their traffic enforcement practices be consistently applied. But there is presently no data in the compilation of MDAI cases which is appropriate to the evaluation of this standard.

Individual case reports, of course, may make note of the lack of police patrols in a dangerous area. It is conceivable that, given a proper sampling plan, information appropriate to this matter could be recorded and analyzed. But this is not the case now.

16. DEBRIS REMOVAL

This standard is concerned with appropriate capabilities and practices to remove the after effects of accidents on the highway. There are numerous mentions of information pertinent to this standard in the MDAI reports, including descriptions of the debris removal operations and recording of the times involved. This information has not been consistently recorded in the digital file--probably because of the lack of a form for reporting the information. But even if it were it would be of little value in estimating a national picture. If this item were reported in a proper sample of accidents, it could serve as a measure of change in regard to debris removal capabilities in the country. But this cannot presently be done.

17. PUPIL TRANSPORTATION STANDARD

While a few school bus accidents have been reported, and the information from each individual accident may be of value in assessing problems in pupil transportation, there is not enough data in the combined file to be of much statistical value. Whether there would be enough information in a larger more controlled sample is doubtful. A sample of 15,000 involvements per year might contain reports of 50 school bus accidents (based on this ratio in Oakland County, Michigan).

18. ACCIDENT INVESTIGATION STANDARD

This standard requires the states to modify and improve their accident investigation and reporting procedures. With the exception of the comparisons noted in discussion of the traffic records standard above, there is no obvious way of evaluating performance with respect to this standard.

THE USE OF MDAI DATA IN EVALUATING THE EFFICACY OF THE MOTOR VEHICLE SAFETY STANDARDS

There are presently 43 motor vehicle safety standards which have been enacted, and all but a few of these are currently in force. Accident investigation teams are generally familiar with these standards, and to some extent with the other standards generation activities in earlier phases (there are another 25 titles which may lead to motor vehicle standards in the future). Most investigators consider in connection with each investigated accident whether the standard played some part in the accident--for example if a head restraint system could be given credit for minimizing or inhibiting a neck injury, or if the lack of a rear view mirror contributed to the causation of the accident.

The encoding of information relative to the standards in the present data file has been sporadic. The principal digital information in the file comes from the CPIR form, which was developed largely within the automotive industry. While there is considerable overlap of that form and vehicle standards information, it is clear that the structured data taking was not designed specifically to answer the question "was this standard effective". Further, the case selection rules which concentrate on outliers or unusual cases would seem to inhibit any statistical inference about the frequency with which a particular standard has been effective or not. Individual reports may identify a shortcoming, and this may be further studied by means external to the multi-disciplinary investigation itself.

In spite of this bleak picture, there are certain modifications to the vehicle which may shine through. By careful analysis of the MDAI data--careful in the sense that known biases are accounted for, or subsets chosen for analysis in which biases relating to a particular problem are minimal--one can study the efficacy of changes. For example it is clear from the regression model analysis of the MDAI data (which attempts to account for accident severity) that occupants wearing restraint systems are noticeably better off.

It is possible to search the MDAI files for cases in which occupants were injured as the result of hood-windshield contact or penetration. It is even possible to get some guarded estimate of the frequency with which hood-windshield involvement occurs. But it is not possible to estimate the national frequency of most events, nor to get a national estimate of the efficacy (in the sense necessary to a cost/benefit computation) from the present data. The fault is not in the precision and detail of the information available in the MDAI file; it is with the sampling procedures.

A very particular shortcoming of the present MDAI file with regard to evaluation of the motor vehicle standards is that there is by choice little or no data on older vehicles--those vehicles with which the newer models (with the standards implemented) could be compared. As shown in Section II of this report, the sample is inadequate for evaluating the energy absorbing steering column because of this. Nearly all of the cars in the file have head restraints of some sort. And essentially all of the cars in the file have the HPR windshield. Most cars are equipped with upper torso restraints, and essentially all with lap restraints--and it is only the fact that the restraint system is not passive that permits some evaluation to be made of that.

Several examples of the use of the present data in gaining insight into the vehicle related standards are given in the text of this report. In addition the reader is referred to other publications using this data as a base for study. These include a study of restraint system effectiveness by Shortridge and Preston,* a study of the incidence of hood-windshield involvement in crashes by Scott* among others.

Table 53 lists in brief form the present motor vehicle safety standards and provides four columns for noting whether particular sets of data may be useful in either identifying problems or in drawing nationally useful inferences about activities relevant to that standard. Entries in the table are the result of reasoned judgment, and represent a concensus of experienced analysts and investigators who were asked to consider how they might approach or solve these problems.

---

* To be published.

Table 53: Federal Motor Vehicle Safety Standard

Number	Short Title	Individual Case Studies	Present MDAI Data	VACUUM	CARPET SWEEP
101	Controls/displays/ergonomics	IIIIII	--IIIW	--IWWW	-IWWWS
102	Shift Lever Sequence	----II	----W	---IWW	-IWWWS
103	Defrosting/defogging	---III	----IW	---WWS	-IWSSS
104	Wiping/washing	IIIIII	-IIIII	--IIWS	-IWWSS
105	Hydraulic brake systems	-IIIII	IIIIIW	-IWWWW	-IWSSS
106	Brake hoses and assemblies	-IIIII	--IIIW	--IWWW	IISSSS
107	Glare from reflecting surfaces	IIIIII	--IIII	--IIWW	-IIWSS
108	New lighting systems	---III	---IIW	-IIWWS	IWWSSS
109	Pneumatic tires	--IIII	--III	--IIWW	-WWWWS
110	Tire selection and rims	-IIIII	---IIW	-IIWWW	-IWWSS
111	Mirror systems	IIIIII	--IIII	IWWWWW	WWWWWS
112	Headlamp concealment devices	--III	---II	-IIIW	-IWWS
113	Hood Latch System	IIIIII	---IIW	-IIWSS	-WWWWS
114	Key locking system (theft)	---II	-----	-IWWS	-WWSS
115	Vehicle Identification No.	-----S	-----S	----IS	----WS
116	Brake fluids	-IIIII	--IIIW	-IIWWS	IWWWSS
117	Retreaded tires	--IIII	--IIIW	IWWWWW	IWWSSS
118	Power operated windows	---III	-----I	---III	--IIWS
119	NOT CURRENTLY IN FORCE				
120	NOT CURRENTLY IN FORCE				
121	Air brake system	--IIII	--IIWW	-IWWWW	-WWWSS
122	Motorcycle brake systems	--III	----IW	--IIWW	-WWWWS
123	Motorcycle controls and displays	--III	------	----IW	---IWS
124	Accelerator control system	-IIIII	---IIW	IIWWWW	WWWWSS
125	Warning devices (for stopped vehicle)	-----I	------	----IS	IIWWWS
126	Truck/camper loading	--IIII	----II	-IIIWW	WWWSSS
201	Occupant protection from interior impact	-IIIII	--IWWS	WWSSSS	WSSSSS
202	Head restraints	-IIIII	-IIIWW	IIWWSS	WWSSSS
203	Impact protection from steering control system	IIIIII	-IWWWW	IISSSS	WSSSSS
204	Steering control rearward displacement	-IIIII	-WWWWW	IWSSSS	SSSSSS
205	Glazing Materials	-IIII	---IW	-WWSS	-WWWS
206	Door locks and retention components	-IIIII	--IWWW	-IIWSS	--WWWS
207	Seating system	IIIII	IIWWW	IISSS	WSSSS
208	Occupant crash protection	-IIIII	-IIWWS	IIISSS	WWSSSS
209	Seat belt assemblies	IIIIII	-IIIWW	IWWWWS	IWSSSS
210	Seat belt assembly anchorages	-IIIII	--IIWW	IWWWWS	IWSSSS
211	Wheel nuts, discs, hub caps	--IIII	------	---IIW	--IIWW
212	Windshield mounting	IIIIII	-IIIWW	IIWWWS	IWSSSS
213	Child seating systems	IIIIII	--IIII	IWWWSS	IWSSSS
214	Side door strength	IIIII	IIIII	IIWWW	IWWWS
215	Exterior protection	IIIIII	-IIIWW	IWWWSS	IWSSSS
216	Roof crush resistance	IIIIII	IIIIIW	IWWWWW	IWWSSS
217	Bus window retention/release	--IIII	----IW	--WWWS	--WSSS
301	Fuel system integrity	-IIIII	--IIWW	-IWWWS	WWWSSS
302	Flammability of interior materials	-IIISS	--IIWS	--IWSS	IWWSSS

- Undetermined
I Insight
W Weak Inference
S Strong Inference

Appendix E

Prediction of Automobile Crash Injury Using Multiple Regression

The study of the relationship between automobile crash parameters and injury is an important problem in the field of accident research. Improved understanding of the underlying mechanisms leading to injury can provide important insight for the design and evaluation of countermeasures. The study of this problem can be classified under three major categories. First are the many bio-medical studies which include dynamic testing on crash sleds and crash testing of instrumented vehicles. A second approach involves intensive case study investigation by expert medical-engineering teams. Third there are statistical analysis of accident reports prepared routinely by police traffic investigators. These categories span the range from controlled precisely measured results which provide considerable knowledge of crash physics, given particular hypothesized crash events, to imprecise measurements of a complete representation of all crash events.

This study represents a combination of categories 2 and 3. Specifically the objective is to use the results of a population of case study crash investigations to obtain increased precision in the measurement of crash characteristics which represent a broad spectrum of the total population of crash types. This study does not represent the final answer to this problem but is seen as the beginning of a developing methodology.

Crash injury results from a complex dynamic process which is influenced by a large number of variables. In fact one might argue that all crashes are unique events. However this uniqueness is merely a matter of degree. A fundamental principle of this study is that injury can be predicted using a model of the following form:

$$Y = f(X_1, X_2, \ldots, X_k) \pm E \tag{1}$$

where:

Y = a measure of occupant crash injury severity which is common to all crashes.

X_j = (j=1,..,k) Variables common to all crashes which can be measured independently of Y.

E = a measure of that portion of crash injury which is not "explained" or predicted by the X variables.

It is assumed that E is a random variable which is independent of the X's. That is:

$$Pr(E \leq E^*) = \int_{-\infty}^{E^*} f(E, \theta_1, \theta_2)\,dE \tag{2}$$

and

$$Pr(E \leq E^* \mid X_1, \ldots, X_k) = Pr(E \leq E^*) \tag{3}$$

The data used for developing the injury prediction model consisted of Occupant Injury reports for 2705 occupants that were obtained from intensive crash investigation reports coded on the Collision Performance and Injury Report Form (CPIR), popularly known as the "GM Long Form". These data have been collected by a number of crash investigation teams located throughout the United States. A list of these teams and their locations is presented in Table 54. The investigation reports prepared by these teams have been coded and built into several computerized data files that are readily accessible through the University of Michigan MTS system computer. Access can be made directly in batch mode or through remote batch or terminal devices using the telephone lines,* The specific version of the data used was a file which contains one observation per occupant in the vehicles investigated. Generally these are late model U.S. vehicles. Crashes selected for investigation

---

* "Multidisciplinary Accident Investigation Report Automation" final report, Five volumes, by Joseph C. Marsh, prepared for Department of Transportation, National Highway Traffic Safety Admin., October 1972.

MDAI INVESTIGATOR CODES

Report Prefix V2	Report Number V3	Team Number V5	Team* Sponsor V6	Team Case Number V16	Team
AA-	00105	1	1	AA-105	Ann Arbor, HSRI-III
BA-	00012	2	1	MVD-12,4ME31	Baylor College of Medicine
BU-	70017	3	1	BU-70-17	Boston University
CA-	71063	4	3	CAL-71-63A	Cornell Aeronautical Lab-IIIA
CB-	71016	5	3	CAL-71-16B	Cornell Aeronautical Lab-IIIB
DT-	71027	6	4	DTS-027-71	Ministry of Transport, Ottawa, Canada
GI-	00071	7	1	GIT-71	Georgia Institute of Technology
HS-	00131	8	2	HS-131	Highway Safety Research Institute
IU-	69013	9	1	MCR-69-13	Indiana University
MG-	71021	10	4	MAC1P021-71, MGU-02S-71	McGill University, Montreal
MI-	00121	11	1	MI-697002	University of Miami
ML-	70008	12	1	MMF-70-8	Maryland Medical/Legal Foundation
MU-	71016	13	4	EPM-016-71	University of Montreal, Ecole Poly-technique
NB-	71009	14	4	UNB-009-71	University of New Brunswick
NM-	00039	15	1	UNM-39	University of New Mexico
OK-	00010	16	2	OK-10	Oakland County, HSRI-III
OS-	00012	17	1	OSU-12	Ohio State University
RT-	00032	18	1	RTI-32	Research Triangle Institute
RU-	00099	19	1	RAI-99	University of Rochester
SC-	00006	20	1	USC-71-6	University of Southern California
SI-	00002	21	1	SRI-2-002	Stanford Research Institute (2)
SR-	00081	22	1	SRI-0081	Stanford Research Institute (1)
SU-	00019	23	1	SU-019	Stanford University
SW-	69003	24	1	SWRI-6903	Southwest Research Institute
TR-	01143	25	1	UC-1143D	Trauma Research Group, UCLA
TU-	00013 71005	26	1	TU-13B2970 TU-71-5	Tulane University
UC-	00450	27	2	UC-450	University of California (Siegel)
UH-	00002	28	1	HOU-2	University of Houston
UK-		29	1		University of Kentucky
UM-	00513	30	2	UM-513-71	University of Michigan (Huelke)
UO-	·	31	1		University of Oklahoma
UT-	71023	32	4	TOR23-08-71	University of Toronto
UU-	70013	33	1	Utah-013-70	University of Utah

* Sponsor Number:
1. NHTSA/DOT
2. AMA
3. Joint AMA, NHTSA
4. Ministry of Transport, Canada

Table 54

Teams and Locations in the MDAI File

are generally not chosen by a random sampling process. Usually they are chosen because of some special characteristics that are of interest to the investigation team. Crash investigations are conducted on a case study basis and thus no attempt is generally made to obtain a "random" sample. Therefore conclusions based upon the analysis of this data file are conditional on the occurrence of a crash.

Basically the study consists of using points in the space defined by all automobile crashes. At each of these points an "experiment" (e.g., a crash) is run. This experiment results in injuries of varying degrees of severity. In many cases several occupants are tested in the same experiment. These may be viewed as nested experiments or as several replications of a single experiment. The difference results from the variables chosen to define the experimental space. For example this study has shown that older persons (above 40) and persons sitting in the front seat tend to have more severe injuries when the differences between crashes are controlled for. Thus if one choses to use age and seating position as definers of the experimental space a nested experiment results. If one does not then the several occupants are replications and the differences in injuries due to age and seating position become part of the unexplained variability.

Returning to our basic model (equation 1) we are assuming that the injuries incurred by a specific occupant are a random variable conditional on the characteristics, X's which define the crash. Since multiple regression assumes that the independent or predictor variables are fixed and that all of the variability is in the dependent (injury) variable the application of multiple regression is valid. As indicated previously important methodological questions deal with the choice of the variable measuring injury.

In choosing a set of crashes for fitting a model there was one important selection bias problem that had to be overcome. Because of the case study nature of the data collection most of the investigations consisted of more severe crashes. The investigator wanted to know how a serious injury occurred or how a particular vehicle component held up under very high stress. This is counter to the objective of representing the entire space of crashes, which is essential for fitting an injury prediction model. To overcome this problem the Highway Safety Research Institute (HSRI) began a program of also investigating non-injury crashes involving late model cars.* In addition HSRI has two other teams which investigate the more severe crashes. All of these investigations are concentrated in Washtenaw County, Michigan, which results in some geographic and cultural biases. However this investigation strategy has resulted in a set of crashes whose identifying variables span the range of crash type and crash dynamics. To illustrate this spanning of the range see Table 55, which indicates the distribution of occupants by impact velocity and vehicle damage extent code.

Measurement of Injury

The selection of a criterion variable, Y, for this approach is an important problem. This results from the fact that a common generally accepted measure of injury severity with an interval scale does not exist. Such a measure would require precisely defined severity levels for specific injuries. (e.g., chest trauma). Further there is a need for a weighting function to combine different categories of injury such as head, chest, or leg injury. Development of such a scale will require considerable time and effort. Therefore it was decided to use the Abbreviated Injury Scale (AIS). This scale is presented in Figure 18.

The use of the AIS code in this manner raises some important methodological questions. Specifically AIS is generally thought of as an ordinal variable rather than an interval variable. That is, unit differences between levels of the variable do not imply the same magnitude of injury severity difference at all points on the scale. For example a 4 injury level is not necessarily twice as severe as a 2 injury level. However we assume that there is a scale - unknown to this writer - that does have the necessary interval variable property. Further, when such a variable is defined it

Figure 18: The Abbreviated Injury Scale as used in the CPIR.

AMERICAN MEDICAL ASSOCIATION *

Abbreviated Injury Scale

SEVERITY CODE	SEVERITY CATEGORY/INJURY DESCRIPTION	POLICE CODE
0 (Zero)	NO INJURY	0 or D
1	MINOR	C
	GENERAL ---Aches all over. ---Minor lacerations, contusions, and abrasions (first aid--simple closure). ---All 1° or small 2° or small 3° burns. HEAD AND NECK ---Cerebral injury with headache, dizziness, no loss of consciousness. ---"Whiplash" complaint with no anatomical or radiological evidence. ---Abrasions and contusions of ocular apparatus (lids, conjunctiva, cornea, uveal injuries), vitreous or retinal hemorrhage. ---Fracture and/or dislocations of teeth. CHEST ---Muscle ache or chest wall stiffness ABDOMINAL ---Muscle ache, seat belt abrasion, etc. EXTREMITIES ---Minor sprains and fractures and/or dislocation of digits.	
2	MODERATE	B
	GENERAL ---Extensive contusions, abrasions, large lacerations, avulsions (less than 3" wide). ---10-20% body surface 2° or 3° burns. HEAD AND NECK ---Cerebral injury with or without skull fracture, less than 15 minutes unconsciousness, no post-traumatic amnesia. ---Undisplaced skull or facial bone fractures or compound fracture of nose. ---Lacerations of the eye and appendages, retinal detachment. ---Disfiguring lacerations. ---"Whiplash" - severe complaints with anatomical or radiological evidence. CHEST ---Simple rib or sternal fractures. ---Major contusions of chest wall without hemothorax or pneumothorax or respiratory embarrassment. ABDOMINAL ---Major contusion of abdominal wall EXTREMITIES AND/OR PELVIC GIRDLE ---Compound fractures of digits. ---Undisplaced long bone or pelvic fractures. ---Major sprains of major joints.	
3	SEVERE (Not Life-Threatening)	B
	GENERAL ---Extensive contusions, abrasions, large lacerations involving more than two extremities, or large avulsions (greater than 3" wide). ---20-30% body surface 2° or 3° burns. HEAD AND NECK ---Cerebral injury with or without skull fracture, with unconsciousness more than 15 minutes, without severe neurological signs, brief post-traumatic amnesia (less than 3 hours). ---Displaced closed skull fractures without unconsciousness or other signs of intracranial injury ---Loss of eye, or avulsion of optic nerve. ---Displaced facial bone fractures or those with antral or orbital involvement. ---Cervical spine fractures without cord damage CHEST ---Multiple rib fractures without respiratory embarrassment. ---Hemothorax or pneumothorax ---Rupture of diaphragm ---Lung contusion. ABDOMINAL ---Contusion of abdominal organs ---Extraperitoneal bladder rupture ---Retroperitoneal hemorrhage ---Avulsion of ureter ---Laceration of urethra ---Thoracic or lumbar spine fractures without neurological involvement EXTREMITIES AND/OR PELVIC GIRDLE ---Displaced simple long-bone fractures, and/or multiple hand and foot fractures ---Single open long-bone fractures ---Pelvic fracture with displacement ---Dislocation of major joints ---Multiple amputations of digits ---Lacerations of the major nerves or vessels of extremities	
4	SERIOUS (Life-Threatening, Survival Probable)	B
	GENERAL ---Severe lacerations and/or avulsions with dangerous hemorrhage ---30-50% surface 2° or 3° burns HEAD AND NECK ---Cerebral injury with or without skull fracture, with unconsciousness of more than 15 minutes, with definite abnormal neurological signs, post-traumatic amnesia 3-12 hours ---Compound skull fracture CHEST ---Open chest wounds, flail chest, pneumomediastinum, myocardial contusion without circulatory embarrassment, pericardial injuries ABDOMINAL ---Minor laceration of intra-abdominal contents (to include ruptured spleen, kidney, and injuries to tail of pancreas) ---Intraperitoneal bladder rupture ---Avulsion of the genitals ---Thoracic and/or lumbar spine fractures with paraplegia EXTREMITIES ---Multiple closed long-bone fractures ---Amputation of limbs	
5	CRITICAL (Survival Uncertain)	A
	GENERAL ---Over 50% body surface 2° or 3° burns HEAD AND NECK ---Cerebral injury with or without skull fracture with unconsciousness of more than 24 hours, post-traumatic amnesia more than 12 hours, intracranial hemorrhage, signs of increased intracranical pressure (decreasing state of consciousness, brady-cardia under 60, progressive rise in blood pressure or progressive pupil inequality). ---Cervical spine injury with quadriplegia. ---Major airway obstruction. CHEST ---Chest injuries with major respiratory embarrassment (laceration of trachea, hemomediastinum, etc.) ---Aortic laceration. ---Myocardial rupture or contusion with circulatory embarrassment ABDOMINAL ---Rupture, avulsion or severe laceration of intra-abdominal vessels or organs, except kidney, spleen or ureter EXTREMITIES ---Multiple open limb fractures	
6	FATAL (Within 24 Hours)	K
	---Fatal lesions of single region of body, plus injuries of other body regions of Severity Code 3 or less. ---Fatal from burns regardless of degree.	
7	FATAL (Within 24 Hours)	K
	---Fatal lesions of single region of body, plus injuries of other body regions of Severity Code 4 or 5	
8	FATAL	K
	---2 fatal lesions in 2 regions of body	
9	FATAL	K
	---3 or more fatal injuries ---Incineration by fire	
99	X SEVERITY UNKNOWN	
	---Injured, but severity not known	
98	Z PRESENCE UNKNOWN	
	---Presence of injury not known	

* Developed by the American Medical Association Committee on Medical Aspects of Automotive Safety, in cooperation with physicians representing medical specialties most involved in the diagnosis, care and treatment of crash injuries, and General Motors Corporation

will be possible to transform the present AIS code uniquely to the new injury severity scale. That is each AIS level will be equivalent to a unique range of values for the appropriate interval scale. Given these assumptions the problem becomes one of determining an appropriate transformation for the AIS variable. At this point in the analysis several different transformations have been used on the AIS code without changing the conclusions of the analysis. Examples of these are given in Appendix I.

Given these methodological problems two choices exist. First one could wait until a more suitable injury scale is developed or one can proceed cautiously with the overall approach using the best available injury scale. The latter approach has been chosen for several reasons. Most important is the need to develop a methodology for equating crashes to a common base so that injuries can be compared from one crash to another. Second this study should provide guidance for the development of an interval injury variable. Having chosen this approach it is important that the results be tested and evaluated at every stop of the process. In the end the final judgement of the approach used concerns whether or not it contributes to an understanding of the injury production process.

As indicated previously there are some important questions concerning the interpretation of the AIS code. Since the approach used in fitting a regression model requires the computation of means it is useful to determine what exactly is implied by a mean AIS code. Table 56 presents the distribution of AIS codes within subsets of the crash population defined by VDI extent code and impact velocity. In addition the mean AIS has been computed for each subset. Examination of the distributions within subsets indicates that the mean defines a point of central tendency about which are grouped - in

----

* The criterion for selecting these crashes was that the vehicle had to be towed from the scene of the crash.

most cases closely grouped - the observed AIS codes. Thus for example in the subset, defined by VDI equal to 1 and impact velocity less than 24.5 mph, there were 214 occupants with a "0" AIS code and 240 with a "1" AIS code and a mean AIS of 0.53. In contrast the adjacent subset with VDI code 2 has 235 occupants with an "0" AIS code, 356 with a "1", 28 with a "2" AIS code and 12 with higher codes for a mean AIS of 0.72. The implication is that the second subset represents a "worse" injury condition since there were more persons with injuries. Specifically the difference in mean results from a shift in the central portion of the distribution rather than merely a change of a few occupants from a "2" to a "6" AIS code. The same general observations applies to other subsets shown. Naturally conclusions are not as clear for those subsets with smaller sample sizes. However the underlying assumption is that the mean will represent the point about which most of the AIS codes are clustered given a large enough sample within each subset.

Table 55

Distribution of Occupants by Impact Velocity and Vehicle Damage Extent Code

Vehicle Damage Index Extent Code (cases without Ejection)*

Impact Velocity	1	2	3	4	5	≥6	Total
< 24.5 mph	455	631	303	69	11	13	1482
< 37.4 mph	63	224	196	131	30	44	688
< 51 mph	21	49	65	87	33	38	293
< 65.6 mph	12	21	40	32	30	32	167
> 65.6 mph	2	10	31	1	4	26	74
Total	553	935	635	320	108	153	2704

* Note: There were only 15 cases in which occupant ejection occurred.

Table 56

Distribution of Injury Severity Codes Within Sub-Groups Defined by Vehicle Damage Index and Impact Velocity

Two-way Cross Tabulation   Stratum = 1

Impact Velocity <24.5 mph

AIS CODE	Total	VDI Extent Code 1	2	3	4	5	6
Mean AIS		.53	.72	1.15	1.26	1.82	2.31
0	198	214	235	60	9	0	0
1	191	240	356	178	43	6	4
2	17	1	28	37	11	2	4
3	6	0	7	21	4	2	4
4&5	1	0	3	4	1	1	0
6 or more	1	0	2	3	1	0	1
Total	414	455	631	303	69	11	13

Two-Way Cross Tabulation   Stratum = 2

24.5 mph ≤ Impact Velocity <37.4 mph

AIS CODE	Total	VDI Extent Code 1	2	3	4	5	6
Mean AIS		.48	.83	1.40	1.73	2.27	2.55
0	39	33	58	24	15	0	5
1	102	30	150	114	62	11	13
2	20	0	15	34	22	10	10
3	16	0	0	12	22	5	4
4&5	1	0	0	8	4	1	4
6	1	0	1	4	5	3	7
9	1	0	0	0	1	0	1
Total	180	63	224	196	131	30	44

Table 56 Continued

Two-Way Cross Tabulation   Stratum = 3

37.4 mph ≤ Impact Velocity 51 mph

AIS CODE	Total	VDI Extent Code 1	2	3	4	5	6
Mean AIS		.71	1.06	1.78	2.37	2.64	4.22
0	7	11	13	10	3	1	1
1	21	9	30	31	37	9	3
2	13	0	3	10	17	9	8
3	11	0	1	4	12	7	8
4	5	0	0	4	5	0	3
5	9	0	1	1	3	3	5
6&7	4	1	0	4	6	4	4
8&9	1	0	1	1	4	0	6
Total	71	21	49	65	87	33	36

Two-Way Cross Tabulation   Stratum = 4

51 mph < Impact Velocity

AIS CODE	Total	VDI Extent Code 1	2	3	4	5	6
Mean AIS		1.29	.94	2.17	3.21	3.03	4.53
0	6	4	9	12	0	1	5
1	20	8	16	29	13	11	9
2	4	0	1	13	4	5	2
3	10	1	3	4	5	7	8
4	1	0	0	1	2	2	4
5	3	0	0	1	1	1	3
6	7	0	2	4	3	3	10
7	1	1	0	2	1	2	1
8&9	9	0	0	5	4	2	16
Total	61	14	31	71	33	34	58

Fitting of the Multiple Regression Model to Predict Injury

Our analysis has used two basic data sets from the ICPR file dated February 5, 1973, one is the occupant injuries from Washtenaw County crash investigations conducted during 1970 and 1972. The other consists of all occupant injuries in the file coming from crashes occurring during 1969 through 1972. Clear, the first is a subset of the second. The rationale for using the Washtenaw County subset has been discussed previously.

Table 57 presents a comparison of the A.I.S. code for the samples from the two populations studied. The distributions of injuries are different as would be expected given the previously discussed crash selection rules. Specifically, the Washtenaw County group has a larger proportion of 0 and 1 injury codes. In addition the Washtenaw County group has 2.9% fatalities while the other group has 4.0% fatalities.

Multiple regression models predicting A.I.S. were fitted to the data in each of these sets. These models are as follows:

Model I Washtenaw County Crashes

$$Y = -.21 + \underset{(.42)}{.34X_1} + \underset{(.30)}{.29X_2} + \underset{(.31)}{3.00X_3} + \underset{(.19)}{.74X_4} + \underset{(.0)}{-.01X_5}$$

$R^2 = 0.53$ N-1235

$S = 0.97$

Model II All Crashes

$$Y = 0.0 + \underset{(.33)}{.35X_1} + \underset{(.20)}{.22X_2} + \underset{(.21)}{1.71X_3} + \underset{(.21)}{.79X_4} - \underset{(0.0)}{.01}$$

$R^2 = 0.38$ S=1.24 N=2705

Note: The value in parentheses ( ) is the partial correlation between AIS and the variable given all other independent variables.

TABLE 57. DISTRIBUTION OF AID FOR CRASH VEHICLE OCCUPANTS

Washtenaw County Occupant Injuries				All Occupant Injuries		
AIS Code		Proportion	Cummulative		Proportion	Cummulative
0	486	.393		727	.270	
1	563	.454	.847	1336	.496	.766
2	71	.060	.907	298	.110	.870
3	54	.044	.951	157	.058	.934
4	10	.008	.959	43	.016	.950
5	15	.012	.971	27	.010	.960
6	17	.014	.985	51	.019	.979
7	3	.002	.987	18	.002	.986
8	13	.011	.998	35	.013	.999
9	3	.002	1.000	3	.001	1.000
	1235			2705		

Details of the regression analysis are presented in Tables 58 and 59.

Y - Accident Injury severity code

X_1 - Vehicle damage index extent code for the case vehicle (e.g., vehicle in which subject was located)

X_2 - Impact vehicle velocity squared for vehicle which is hypothesized to contribute most directly to the injury The value for this variable was determined as follows:

Let

V_1 = case vehicle impact velocity/ $\sqrt{1000}$

V_2 = other vehicle impact velocity/ $\sqrt{1000}$

	COLLISION CONFIGURATION(V59)	AREA OF DAMAGE (V144) CASE VEHICLE
$X_2 = V_1^2$	Single vehicle or unknown (codes 1,2,7)	
$X_2 = V_1^2 + V_2^2$	Head on (3)	--
$X_2 = V_1^2$	Rear Impact (6)	Front
$X_2 = V_2^2$	Rear Impact (6)	Back
$X_2 = V_2^2/4$	Sideswipe (5)	Right or left side
$X_2 = V_1^2$	Sideswipe (5)	Front
$X_2 = V_2^2$	Intersection type (4,8) T or L	Right or left side
$X_2 = V_1^2$	Intersection type (4,8) T or L	Front

TABLE 58: INJURY PREDICTION MODEL FROM WASHTENAW COUNTY CRASHES

```
LEAST SQUARES REGRESSION

ANALYSIS OF VARIANCE OF AIS MDCS            N= 1235

SOURCE              DF     SUM OF SQRS   MEAN SQUARE   F-STATISTIC  SIGNIF

REGRESSION           5      1298.8        259.76         273.75       0.
ERROR             1229      1166.2        .94890
TOTAL             1234      2465.0

MULTIPLE R= .72..8   R-SQR= .52690    SE= .97412

VARIABLE          PARTIAL    COEFFICIENT    STD ERROR    T-STATISTIC  SIGNIF

CONSTANT                     -.20860        .55862 -1    -3.7341      .0002
C-VST(P)          .41621      .34334        .21396 -1     16.047      .0000
SUMVELSQ          .30061      .28548        .25836 -1     11.050      .0000
BOTH EJT          .31067     3.0040         .26217        11.458      .0000
MIND SEP          .18905      .73542        .10696        6.7494      .0000
1 VEHICL         -.00375     -.64352 -2     .64243 -1    -.13130      .8956


< HISTO V=600 INT=10:(0,9) CASES=V950:1-10*V5:1,8,30 >

HISTOGRAM/FREQUENCIES

   AIS MDCS
 0.        +XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX486
 1.0000    +XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX56
 2.0000    +XXXXXXXXXX71
 3.0000    +XXXXXXX54
 4.0000    +10
 5.0000    +X15
 6.0000    +X17
 7.0000    +3
 8.0000    +13
 9.0000    +3
 1235 OBSERVATIONS PLOTTED (EACH X=   7)

< HISTO V=715 INT=20:(-4.0,4.0) CASES=V950:1-10*V5:1,8,30 >

HISTOGRAM/FREQUENCIES

   VAR 715
-4.0000    +
-3.5789    +1
-3.1579    +2
-2.7368    +X7
-2.3158    +X10
-1.8947    +XXX17
-1.4737    +XXXXXXX34
-1.0526    +XXXXXXXXXXXXXXXXXXXXXX92
-.63158    +XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX260
-.21053    +XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX275
 .21053    +XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX173
 .63158    +XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX218
 1.0526    +XXXXXXXXXXXXXXX63
 1.4737    +XXXXX28
 1.8947    +XXX18
 2.3158    +X8
 2.7368    +XX13
 3.1579    +XX9
 3.5789    +1
 4.0000    +2
 1229 OBSERVATIONS PLOTTED (EACH X=   4)
    6 WERE  >   4.0000
```

TABLE 59: INJURY PREDICTION MODEL FROM TOTAI CRASH FILE

```
LEAST SQUARES REGRESSION

ANALYSIS OF VARIANCE OF OVERALL              N= 2705

SOURCE                  DF     SUM OF SQRS   MEAN SQUARE   F-STATISTIC  SIGNIF

REGRESSION               5       2543.3        508.66         331.49      0.
ERROR                 2699       4141.6        1.5345
TOTAL                 2704       6684.9

MULTIPLE R= .61631   R-SQR= .38046    SE= 1.2387

VARIABLE           PARTIAL     COEFFICIENT     STD ERROR     T-STATISTIC  SIGNIF

CONSTANT                       .12291 -2     .50492 -1     .24343 -1     .9806
C-VDI(P)           .33163      .34688        .18995 -1     18.262        .0000
SUMVELSQ           .19801      .22119        .21076 -1     10.495        .0000
BOTH EJT           .20773      1.7064        .15467        11.033        .0000
WBND SEP           .20910      .79247        .71337 -1     11.109        .0000
1 VEHICL          -.00512     -.14717 -1     .55357 -1    -.26586        .7904
```

```
< HISTO V=600 INT=10:(0,9) CASES=V950:1-10 >

HISTOGRAM/FREQUENCIES

  AIS MOCS
0.          +XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX727
1.0000      +XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX1336
2.0000      +XXXXXXXXXXXXXXXX298
3.0000      +XXXXXXX157
4.0000      +XX43
5.0000      +27
6.0000      +XX51
7.0000      +18
8.0000      +X35
9.0000      +3
 2705 OBSERVATIONS PLOTTED (EACH X= 15)
```

```
< HISTO V=614 INT=20:(-4.0,4.0) CASES=V950:1-10 >

HISTOGRAM/FREQUENCIES

   RES CMOD
-4.0000     +1
-3.5789     +4
-3.1579     +3
-2.7368     +X21
-2.3158     +XXXX42
-1.8947     +XXXXXXX57
-1.4737     +XXXXXXXXXXXXX115
-1.0526     +XXXXXXXXXXXXXXXXXXXXXXXXXXXX217
-.63158     +XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX488
-.21053     +XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX556
 .21053     +XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX495
 .63158     +XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX334
 1.0526     +XXXXXXXXXXX101
 1.4737     +XXXXXXXX70
 1.8947     +XXXXXX 3
 2.3158     +XXX29
 2.7368     +XX26
 3.1579     +14
 3.5789     +X10
 4.0000     +11
 2655 OBSERVATIONS PLOTTED (EACH X=  7)
    2 WERE < -4.0000
   48 WERE >  4.0000
```

The basic rationale in this definition was to obtain a velocity vector which is most likely to be associated with occupant injury. For example in rear impact crashes the velocity of the striking vehicle was used. Another likely candidate--relative velocity--was found to result in a worse prediction of injury. The suggestion is thus that the impacting vehicles velocity must be dissipated since the final condition of the vehicles is zero velocity. Side-swipe crashes are belived to result in a glancing blow. If a thirty degree impact angle is assumed, the velocity component perpendicular to the case vehicle is one half of the vehicle veolcity. Hence the use of one-half of the velocity squared.

X_3 - 1 if occupant was ejected either partially or completely
0 if occupant was not ejected

X_4 - 1 if windshield bond separated
0 if it did not

X_5 - 1 for a single vehicle crash
0 otherwise

This model was chosen above a number of others because of its "good" prediction capability-higher percent explained variability (R^2) and symmetric distribution of residuals-and its relatively simple structure. In addition it is consistent with our knowledge of the injury causation process in automobile crashes.

An important problem in multiple regression models is that of multi-colinearity.. Multi-colinearity is the condition in which predictor variables (X's) are not independent. If this occurs both of the correlated predictors may be predicting the same phenomenon. Although this can result in unstable estimates of regression co-efficients, the prediction from the model will not be affected. This results from the fact that the multiple regression model is determined under the restriction of minimizing the squared variability from the predicted value of the dependent variable (Y). Thus, even though two regression models have different coefficients they can still result in similar predicted values when applied to similar data sets. This results from compensating effects of differences in the coefficients. However if multi-colinearity exists the resulting model may be difficult to interpret and it may not predict correctly when applied to other data sets.

Table 60 presents a correlation matrix of the variables used for the Washtenaw County data set. Of some concern are the high correlation between V.D.I. extent code, the impact velocity squared, and the windshield bond separation variables. One method of examining the effect of these predictor variable correlations on the regression model is to examine the partial correlations for each predictor-which are provided in the regression output. These represent the relationship between a predictor and the dependent variable -AIS-after the effects of the other predictor variables have been removed. Thus, they provide an indication of what would happen if this predictor were removed. The partial correlations for the Washtenaw County model indicate substantial individual contributions to the overall predictability for the VDI and velocity variables with a lesser contribution from the windshield bond separation variable. The single vs. multiple vehicle variable could be removed with very little effect on overall explained variability. However, there is a justification for separating single and multiple vehicle crashes based upon crash dynamics. In addition, injuries from single vehicle crashes are predicted better with that variable included.

Several other variables were also used as measures of the vehicle velocity effect. In particular:

1 (Velocity 1) velocity of case vehicle squared
2 (Velocity 2) the sum of both vehicles impact velocity squared
3 (energy) the crash energy rather than impact velocity

$$\text{Energy} = M \quad V^2/2$$

TABLE 60: SIMPLE CORRLEATIONS BETWEEN VARIABLES USED IN REGRESSION MODEL -- WASHTENAW COUNTY OCCUPANTS

CORRELATION COEFFICIENTS

N= 12[illegible] DF= 1[illegible] R0.9500= .0555 R0.9900= .0733

	VARIABLE	Y	X_1	X_2	X_3	X_4	X_5
Y	AIS MDCS	1.0000					
X_1	C-VDI(P)	.0024	1.0000				
X_2	SUBVELSG	.5542	.5115	1.0000			
X_3	BOTH EJT	.0714	.14[illegible]	.17[illegible]	1.0000		
X_4	GRND SEP	[illegible]	[illegible]	[illegible]	.2443	1.0000	
X_5	1 VEHICL	.0010	.1046	.0555	.0364	.1434	1.0000
		AIS MDCS	C-VDI(P)	SUBVELSG	BOTH EJT	GRND SEP	1 VEHICL
		Y	X_1	X_2	X_3	X_4	X_5

The correlations between these variables and AIS are shown in Table 61. In addition each of these variables was used as a substitute for the impact velocity variable in the multiple regression model. This provided for an examination of the partial correlations for these variables. Using both the total correlations - from Table 61 - and the partial correlations it was concluded that injury could be predicted "best" using impact velocity.

Either the velocity squared or the energy variables could be used as substitutes for X_2 in a model that would still provide reasonably good predictability. However, because of the high correlations between these variables they should not be added to the present model.

MODEL APPLICATIONS

The model can be used in two ways. First the variables in the model indicate conditions leading to injury. It should not be surprising that larger impact velocity and vehicle damage index are associated with worse injury. The fact that ejection results in more severe injury has also been shown in numerous studies. However, this model provides a quantification of the magnitude of the effects. As indicated previously regression coefficients are sometimes unstable when the predictors are correlated. However, the coefficients for impact velocity, VDI and windshield bond separation have remained reasonably stable over a large number of alternative regression models and subsets of the data.

The model can also be used as a device for comparing injuries from a wide range of crash types. Specifically, the model computes the expected injury code conditional on crash dynamics and occupant ejection. The combined effect of these variables explains fifty-three percent of the variability as shown in Model I. Differences between the expected and reported AIS codes can be computed yield-

TABLE 61. CORRELATION COEFFICIENTS BETWEEN ALTERNATE MEASURES OF IMPACT VELOCITY, INJURY, AND INJURY PREDICTION VARIABLES--WASHTENAW COUNTY OCCUPANTS

N=540 D.F.=538 Critical Value (α=0.01)=.1108

1. **Overall AIS (Y)**	1.00							
2. Velocity 1	.4408	1.00						
3. Velocity 2	.6376	.5811	1.00					
4. Energy	.6022	.6999	.8056	1.00				
5. SUMVEL SQ(X_2)	.6668	.7236	.9303	.9322	1.00			
6. VDI Extent(X_1)	.6124	.4388	.5284	.6124	.6162	1.00		
7. Ejection(X_3)	.4232	.2222	.2110	.2592	.2382	.1427	1.00	
8. Windshield Bond Separation (X_4)	.4800	.6999	.4170	.3988	.4421	.4452	.2821	1.00
	1	2	3	4	5	6	7	8

ing residual AIS. This new statistic measures injury with the effect of crash dynamics and ejection removed. Thus, it can be used to study the relationships between vehicle, environment or driver components independent of crash dynamics. This approach will be illustrated in the final section of this paper.

SELECTION OF A MODEL

The ultimate goal of this effort is a regression model which predicts injury AIS with minimum variance when applied to any crash or set of crashes. Since the Washtneaw County population contains a representative sample of all injury and non-injury crashes involving late model vehicles in a specific geographic area, it was hypothesized that a better prediction model would result from that population. Thus, many of the non-random selection criteria that are apparent in most in-depth case investigation of crashes are avoided.

Table 62 presents a comparison of Models 1 -- Washtenaw County Occupants -- and 2 -- all occupants -- applied to the entire set of occupant injuries. Since Model 2 is that model having minimum unexplained variance when applied to the entire data set no other model--with the same variables--could have a smaller unexplained variance. The question then is: How close does Model 1 - developed using only Washtenaw County crashes - come to the minimum unexplained variance?

Minimum unexplained variance--from Model 2--is 4035. Model I unexplained variance is 4182 an increase of 147. Percent explained variability was reduced from 38.1% to 35.8%, while mean square error changed from 1.53 to 1.59

The reverse analysis was also performed--Model 2 applied to Washtenaw County data--as shown in Table 63. In this case Model 1 has minimum unexplained error and the question is: How good a substitute is Model 2? Percent explained variability was reduced from 53.3% to 51.8%, while mean square error changed from 0.92 to 0.95.

TABLE 62. COMPARISON OF MODELS DEVELOPED FROM WASHTENAW COUNTY OCCUPANT INJURIES AND FROM ALL OCCUPANT INJURIES APPLIED TO ALL OCCUPANT INJURIES

Model 1 -- Washtenaw County Occupants -- Application

Analysis of Variance

Source	Sum of Squares	Degrees of Freedom	Mean Square	F
Explained by Model 1	2332	5	466	252.7
Unexplained (Error)	4182	2631	1.59	
Total	6514			

Percent Explained = 35.8%

Model 2 --All Occupants -- Application

Analysis of Variance

Source	Sum of Squares	Degrees of Freedom	Mean Square	F
Explained by Model 2	2479	5	496.0	324.2
Unexplained (Error)	4035	2631	1.53	
Total	6514			

Percent Explained = 38.1%

TABLE 63 COMPARISON OF MODELS DEVELOPEL FROM WASHTENAW COUNTY OCCUPANT INJURIES AND FROM ALL OCCUPANT INJURIES APPLIED TO WASHTENAW COUNTY OCCUPANT INJURIES

Model 2 -- All Occupants -- Application

Analysis of Variance

Source	Sum of Squares	Degrees of Feedom	Mean Square	F
Explained by Model 2	1236	5	247.2	260.2
Unexplained (Error)	1149	1213	0.95	
Total	2385	1218		

Percent Explained Variability = 51.8%

Model 1 -- Washtenaw County Occupants -- Application

Analysis of Variance

Source	Sum of Squares	Degrees of Freedom	Mean Square	F
Explained by Model 1	1271	5	254.2	276.3
Unexplained (Error)	1114	1213	0.92	
Total	2385	1218		

Percent Explained Variability = 53.3%

Comparison of Tables 62 and 63 fail to indicate a preference between either model. The conclusion is encouraging since all crashes investigated outside of Washtenaw County were chosen on a case study basis. However, if it is assumed that AIS contains a fixed component, conditional on crash dynamics--as represented by the model--and a uniform random component--unexplained variance--then our procedure is reasonable. The distribution of Model residuals presented in Tables 58 and 59 indicate symmetric distributions about zero. Thus, the models have "explained" certain portions of the differences in A.I.S.--shown by the upper histograms in Tables 58 and 59-and the remaining difference is symmetrically distributed about the predicted value.

Application of Model to Random Subsets of Data

In addition to the above comparison of models it was felt that models developed using randomly selected subsets of the total file should also be compared with regard to their injury prediction capability. To accomplish this the following experiment was conducted. Three random subsets of the occupant injury file were selected. Multiple regression models were fitted in each subset. Slope coefficients, partial coefficients and overall R^2 for these models is presented in Table 64. The percent explained variability for the three models ranges from 0.34 to 0.43, however individual coefficients vary over a much wider range. This is not surprising since the objective of multiple regression models is prediction of injury with minimum squared variability.

A methodology has been presented for analyzing occupant injury data obtained from in-depth crash investigation. There are some potentially serious biases arising from the use of this technique. The most important of these concerns is the crash selection process. The model developed is most applicable to the population of crashes from which it was developed; i.e., will tend to provide lower variance injury predictions when used with data containing the same biases. The ultimate question concerns whether or not this approach provides additional insights into the occupant injury problem.

Specific examples of the use of residual injuries have been presented in Section V.

Conclusion

A crash injury analysis strategy has been developed. Basically it makes use of a mechanism to compute expected injuries which can be compared to actual injuries for various data subsets. The methodology is believed to be in a developmental stage. Clearly a next step consists of applications that test the approach. These should be accomplished by rigorous technical analysis of the results in order to place the approach on a firmer foundation.

Table 64

Comparison of Injury Prediction Model Applied to Random Subsets of the Data

Variables \ Data Subset	All	Independent Subsets: Random Subset 1	Random Subset 2	Random Subset 3
Constant	0.00	.05	.11	-0.14
VDI Extent				
B	.35	.33	.29	.40
partial	.33	.33	.27	.38
Impact Velocity Squared				
B	.22	.18	.27	.22
partial	.20	.18	.25	.18
Ejection				
B	1.71	2.00	.47	2.11
partial	.21	.26	.05	.27
Windshield Bond Separation				
B	.79	.75	.88	.76
partial	.21	.20	.24	.20
Single Vehicle				
B	-0.01	.06	-.15	.03
partial	0.00	.02	-.05	.01
SE	1.24	1.22	1.21	1.27
R^2	.38	0.39	.34	0.43
N	2705	889	895	896

One specific procedure that is likely to provide insight involves extensive model residual analysis. First the model would be applied to specific crashes. Differences between observed and expected injury would then be computed. If the difference exceeded certain values an attempt would be made to "explain" the difference using the investigators detailed knowledge of the crash. This should provide two potential results:

1. A better understanding of injury producing mechanisms.
2. An indication of model shortcomings.

Residual AIS can also be used as a statistic to evaluate the injury reduction effectiveness of vehicle standards. Examples of this approach have been presented in Section V. Similar analyses could be performed for other vehicle, human, or environmental factors.

Appendix F

Choice of Sample Size Controlling both α and β error.

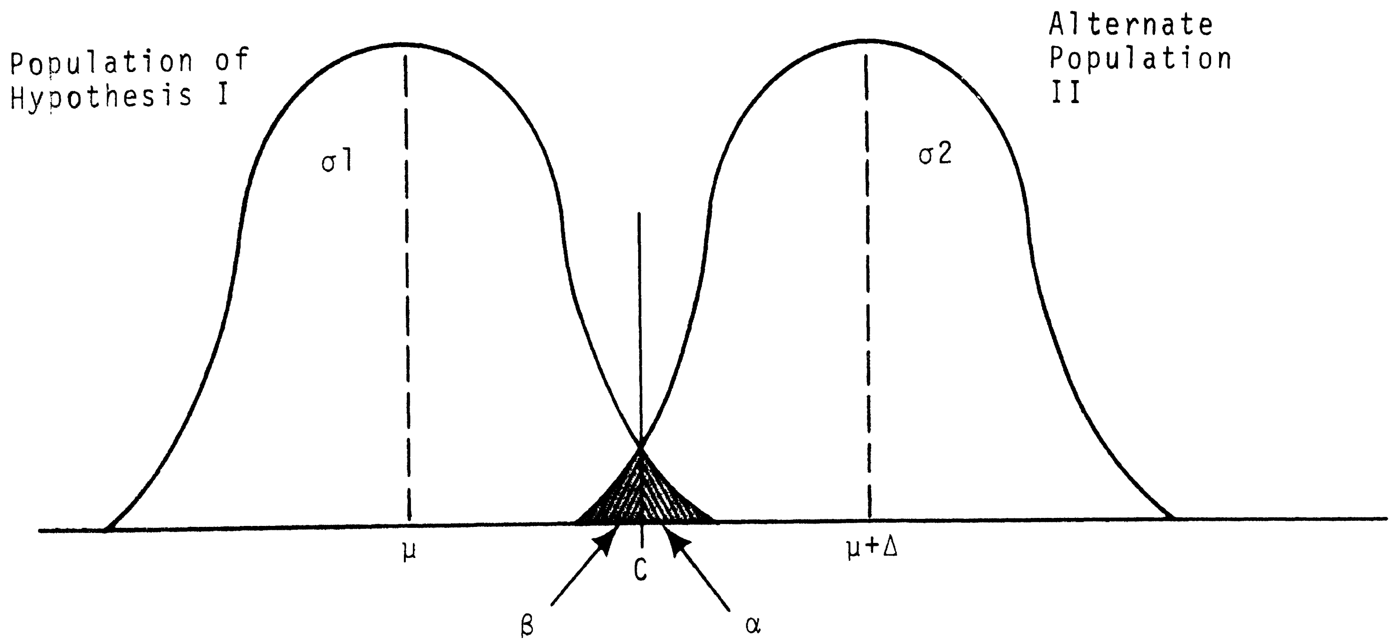


Figure 19. Schematic representation of alpha and beta errors.

A sample of size N is drawn and the objective is to determine which of two populations I or II; is the most likely source of the data. The probability density functions for these populations are shown schematically in Figure 19. To formalize the analysis we state the following:

Null Hypothesis:

Ho: The sample data comes from Population I

The difference between the population means is:

$$\text{Diff} = \mu + \Delta - \mu = \Delta$$

To answer the initial question we will (1) compute the mean of the sample:

$$\bar{Y} = \frac{\sum Y_i}{N}$$

where Y_i is the value of a random observation contained in the sample $\bar{Y}$ is the sample mean.

(2) Select a value C that is between μ and $\mu + \Delta$. If $\bar{Y}$ is less than or equal to C we will accept the hypothesis Ho and conclude that the sample data comes from Population I. If $\bar{Y}$ is greater than C we will reject the null hypothesis and conclude that the data came from Population II.

To aid in the process of choosing C we will define the following <u>rules</u>:

1. If the data really comes from Population I we want to accept Ho with a probability of .95. This implies that 5% of the time we conduct experiments of this type. We will conclude that the data does not come from Population I when it actually does. Thus = 0.05.

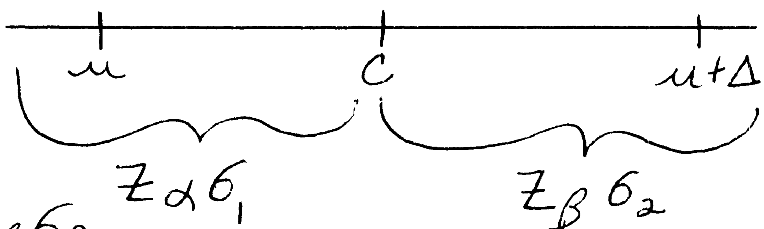
2. If the data really comes from Population II we want to accept Ho with a probability of 0.05, this implies that 5% of the time we conduct experiments of this type. We will conclude that the data does come from Population I when it actually does not. Thus α = 0.05.

We will also make the following assumptions in choosing C.

1. The means of samples drawn from either Population I or II have a normal probability distribution. The central limit theorem insures that this assumption will be acceptable for most real problems.

2. $\sigma_1^2 = \sigma_2^2 = \sigma^2/N$
   Note: the following calculations have the same form if $\sigma_1 = a\sigma_2$. The result would merely contain the factor a.

From these rules and assumptions some additional dimensions can be added to Figure 19.



$\Delta = Z_\alpha \sigma_1 + Z_\beta \sigma_2$

$C - \mu = Z_\alpha \sigma_1$

$\mu + \Delta - C = Z_\beta \sigma_2$

Z_α - Normalized Deviate for a significance level of α

The hypothesis Ho implies that:

$$\mu = E(Y_i) = \bar{Y}$$

In addition assumption 2 implies that:

$$E(\sigma^2) = S^2 = \sum_{i=1}^{N} \frac{(Y_i - \bar{Y})^2}{N-1}$$

$$\therefore E(\sigma_1^2) = E(\sigma_2^2) = S^2_N$$

$$\Delta = Z_\alpha \frac{S}{\sqrt{N}} + Z_\beta \frac{S}{\sqrt{N}}$$

From our rules

e.g., $\alpha = .05$, $\beta = .05$

$Z_\alpha = Z_\beta = 1.645$

$$\Delta = 1.645 \left(\frac{2S}{\sqrt{N}}\right)$$

$$C = \mu + 1.645 \frac{S}{\sqrt{N}}$$

If $\bar{Y} \leq C$

The hypothesis Ho is accepted.

Choice of sample size N

$$\Delta = Z_\alpha \frac{S}{\sqrt{N}} + Z_\beta \frac{S}{\sqrt{N}}$$

$$= \frac{S}{\sqrt{N}} (Z_\alpha + Z_\beta)$$

For a given value of Δ

$$N = \frac{S^2}{\Delta^2} (Z_\alpha + Z_\beta)^2 = S^2 \left(\frac{Z_\alpha + Z_\beta}{\Delta}\right)^2$$

Then for an estimate of S a value of N can be determined. Figure presents the graphical relationship between N and Δ for various values of α and β .

Comparison of proportions

Let:

$$\mu_1 = P_1 \qquad \Delta = \Delta P$$

where P_1 is the proportion of cases in Population I which have a characteristic of interest

$$\sigma_1^2 = P_1(1-P_1) \qquad \sigma_2 = (P_1+\Delta P)\ (1-P_1 - \Delta P)$$

$$\Delta P = Z_\alpha \frac{\sqrt{P_1(1-P_1)}}{\sqrt{N}} + Z_\beta \frac{\sqrt{(P_1+\Delta P)\ (1-P_1-\Delta P)}}{\sqrt{N}}$$

$$\sqrt{N} = Z_\alpha \frac{\sqrt{P_1(1-P_1)}}{\Delta P} + Z_\beta \frac{\sqrt{(P_1+\Delta P)\ (1-P_1-\Delta P)}}{\Delta P}$$

The attached graph - Figure 18 indicates the relationship between sample size N and P for various values of P_1.

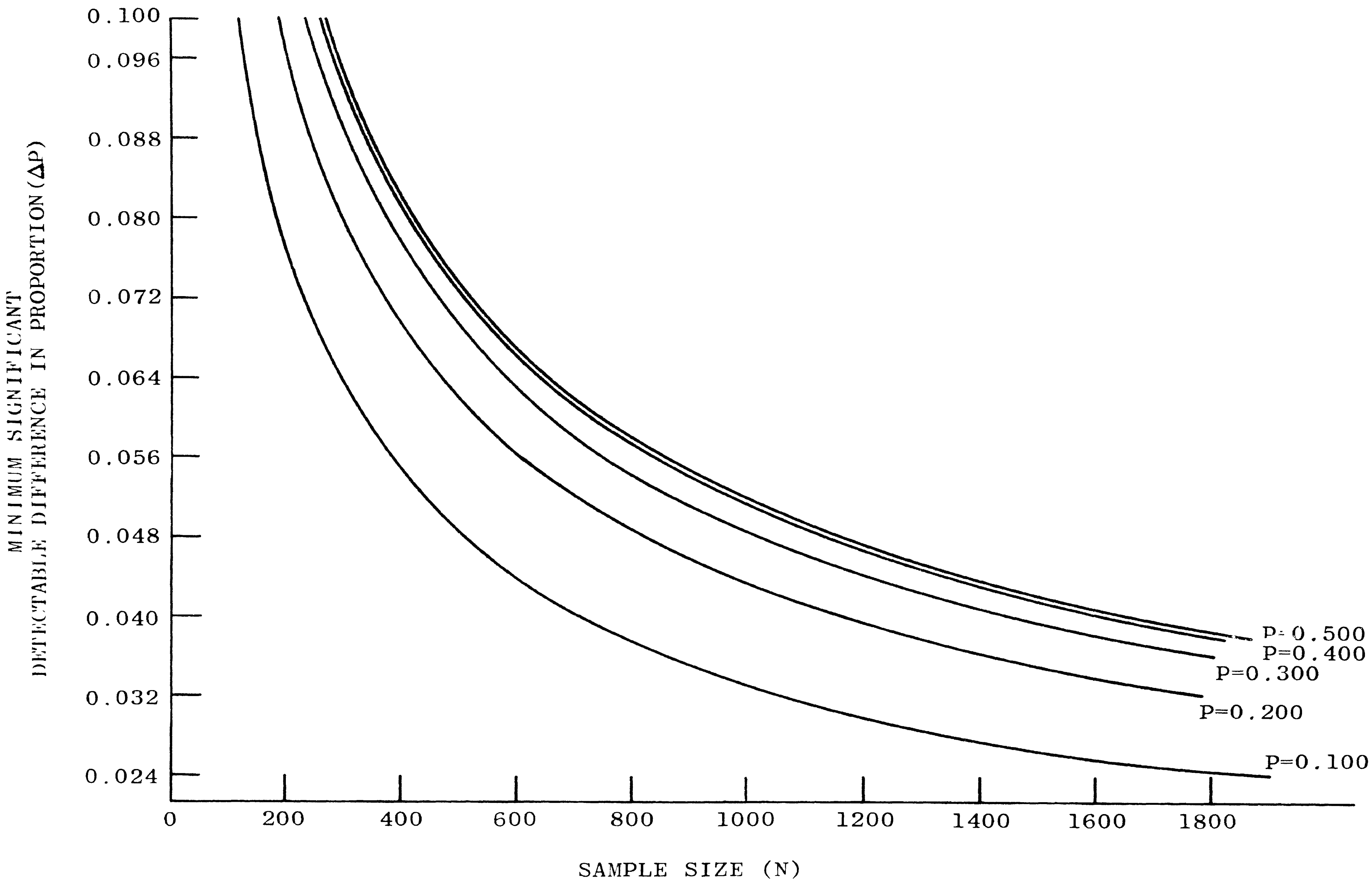


Figure 20. Minimum detectable proportion difference vs same size, when $\alpha = 0.05$, $\beta = 0.05$.

Appendix G

The Effect of Non-Response on Precision of Sample Estimates

Appendix G

The Effect of Non-Response on Precision of Sample Estimates

Consider a sampling survey problem in which only a proportion $\hat{P}_1$ $(0 < \hat{P}_1 \leq 1)$ of the randomly selected subjects agree to participate in the interviewing process. Suppose a characteristic Y is being measured in the interview. Define two subpopulations as follows:

Population	Expected Value of the Characteristic	Proportion of Target Population
1 (participants)	$\bar{Y}_1$	P_1
2 (non-participants)	$\bar{Y}_2$	$1-P_1$

Note that the proportion of the sample who agree to be interviewed $\hat{P}_1$, is taken as an unbaised estimate of P_1. Now the expected value of Y in the population is:

$$E(Y) = P_1\bar{Y}_1 + (1-P_1)\,\bar{Y}_2$$

$$= \bar{Y}_2 + P_1(\bar{Y}_1 - \bar{Y}_2) \qquad (1)$$

However the sample estimate of the expected value of Y is $\bar{Y}_1$. Therefore a bias is introduced as follows:

$$\text{Bias} = \bar{Y}_1 - E(Y) \qquad (2)$$

Substituting equation (1) into equation (2):

$$\text{Bias} = \bar{Y}_1-\bar{Y}_2-P_1(\bar{Y}_1-\bar{Y}_2) \qquad (3)$$

Let $\Delta Y = \bar{Y}_1-\bar{Y}_2$

$$\text{Bias} = \Delta Y-P_1(\Delta Y)$$

$$= (1-P_1)\,\Delta Y \qquad (4)$$

Thus from equation (4) it can be seen that bias is a function of the participation proportion and the difference between participants and non-participants.

In general the difference ΔY between $\bar{Y}_1$ and $\bar{Y}_2$ is not known. Therefore the effect of bias is the introduction of additional uncertainty for the estimate of E(Y). This uncertainty is added to the variance σ_Y^2 that is associated with the estimate of E(Y). Variance is a measure of random variability due to unexplained causes such as measurement error. Combining variance and bias provides a measure of total uncertainty of the estimate of E(Y).

$$MSE = \sigma_Y^2 + (Bias)^2 \qquad (5)$$

MSE - Mean Square Error

Substituting (4) into (5):

$$MSE = \sigma_Y^2 + (1-P_1)^2 (\Delta Y)^2 \qquad (6)$$

If a sample of size N participants is obtained MSE becomes:

$$MSE_{\bar{Y}} = \frac{\sigma_Y^2}{N} + (1-P_1)^2 (\Delta Y)^2 \qquad (7)$$

$MSE_{\bar{Y}}$ - Mean square error for an estimate of E(Y) using a sample of size N.

Notice in equation (7) that only variance and not bias is effected by increases in sample size. Suppose $MES_{\bar{Y}}$ is treated as a measure of unknown variability. Then:

$$S_{\bar{Y}1B} = \sqrt{\frac{\sigma_Y^2}{N} + (1-P_1)^2 (\Delta Y)^2} \qquad (8)$$

Let

$$\sigma_Y^2 = 1$$

$$\Delta Y == 1$$

$$S_{\bar{Y}1B} \sqrt{\frac{1}{N} + (1-P_1)^2}$$

Appendix H

CARPETSWEEP interactive computer conversation

MTS (LA44-0071)
#SIGNON SBZ3
#ENTER USER PASSWORD
?[redacted]

#**LAST SIGNON WAS: 09:39.45 12-13-73
#USER "SBZ3" SIGNED ON AT 08:23.34 on 12-14-73
#source cps8:enter
#EXECUTION BEGINS

CPS8÷ENTRY FORM

WHENEVER YOU NEED HELP, ANSWER WITH A ?

THERE ARE 21 VARIABLES TO BE ENTERED

DO YOU WANT A LIST OF THESE? YES OR NO

yes

LIST OF VARIABLES TO BE ENTERED

VAR NO.	NAME	NO. OF CHARACTERS
1	MAKE AND MODEL OF VEHICLE	72
2	MODEL YEAR	2
3	DRIVER AGE (YRS)	2
4	DRIVER SEX	1
5	ACCIDENT TYPE	1
6	VDI	1
7	IMPACT SPEED (MPH)	2
8	ROAD CONDITIONS	1
9	DATE/TIME OF ACC	13
10	CASE OCC AGE (MO)	2
11	CASE OCC SEX	1
12	CASE OCC WEIGHT (LBS)	3
13	CASE OCC POSITION	1
14	CASE OCC INJ SEVERITY	2
15	ACTIVITY PRIOR TO CRASH	1
16	CHILD RESTRAINT USED	1
IF ANSWER TO VAR #16 IS YES, ENTER THE FOLLOWING		
17	LAP BELT USED	1
18	SHOULDER HARNESS USED	1
19	CHILD SEAT USED	1
20	MFG OF CHILD SEAT	1
21	MODEL NO. OF CHILD SEAT	4

END OF VARIABLE LIST

VAR 1 - MAKE AND MODEL OF VEHICLE

Ford Station Wagon

VAR 2 - MODEL YEAR

72

VAR 3 - DRIVER AGE

29

VAR 4 - DRIVER SEX

F

VAR 5 - ACCIDENT TYPE

?

ACCIDENT TYPE CODES
1 - VEHICLE TO OBJECT
2 - RAN OFF ROAD
3 - HEAD ON TO OTHER VEHICLE
4 - SIDE SWIPE, INTO OTHER VEHICLE
5 - SIDE SWIPE, BY OTHER VEHICLE
6 - REAR END, INTO OTHER VEHICLE
7 - REAR END, BY OTHER VEHICLE
8 - INTO PARKED CAR

VAR 5 - ACCIDENT TYPE

5

VAR 6 - VDI

1

VAR 7 - IMPACT SPEED

15

VAR 8 - ROAD CONDITIONS

Clear, Dry

VAR 9 - DATE/TIME OF ACCIDENT

12-10-73 1030

VAR 10 - CASE OCC AGE

24

VAR 11 - CASE OCC SEX

Male

VAR 12 - CASE OCC WEIGHT

28

VAR 13 - CASE OCC POSITION

?

OCCUPANT POSITION CODES
1 FRONT RIGHT
2 FRONT CENTER
3 BACK LEFT
4 BACK CENTER
5 BACK RIGHT
6 BACK FLOOR
7 REAR (STATION WAGON)
8 UNKNOWN

VAR 13 - CASE OCC POSITION

4

VAR 14 - CASE OCC INJURY SEVERITY

1

CODE NOT ACCEPTABLE

VAR 14 - CASE OCC INJURY SEVERITY

?

INJURY SEVERITY CODES
00 - NONE
01 - MINOR
02 - NON-DANGEROUS, MODERATE
03 - NON-DANGEROUS, SEVERE
04 - DANGEROUS, SERIOUS
05 - DANGEROUS, CRITICAL
06 - FATAL LESIONS IN 1 REGION
07 - FATAL LESIONS IN 1 REGION
08 - FATAL LESIONS IN 2 REGIONS
09 - FATAL LESIONS IN 3 OR MORE REGIONS
98 - UNKNOWN
99 - INJURED, SEVERITY UNKNOWN

```
VAR 14 - CASE OCC INJURY SEVERITY

     01

VAR 15 - ACTIVITY PRIOR TO CRASH

     ?

        ACTIVITY CODES
        1  STANDING
        2  SITTING
        3  LYING DOWN
        9  UNKNOWN

VAR 15 - ACTIVITY PRIOR TO CRASH

     2

VAR 16 - CHILD RESTRAINT USED

     No

ENTER ANY TEXT OR COMMENTS
END WITH "DONE"

done

YOUR CASE WILL BE ENTERED AS FOLLOWS:

             SITE HSRI CASE NO. 33
               12-14-73

        FORD STATION WAGON - 1972
        DRIVER:  29, F
        SIDE SWIPE BY OTHER VEHICLE
        VDI=1  IMPACT=15 MPH   ROAD=CLEAR, DRY
        ACCIDENT OCCURRED  12-10-73 AT 1030
        CASE OCCUPANT:  24 MO, F, 28 LBS
                        SITTING, BACK CENTER
                        MINOR INJURIES
        RESTRAINT USED:  NONE

ENTER "STORE" OR USE "CHANGE" COMMAND

store

OK
```

```
signoff
#OFF AT 08:32.40        12-13-73
#ELAPSED TIME           9.083 MIN.              $.45
#CPU TIME USED          8.277 SEC.              $.65
#CPU STOR VMI           3.583 PAGE-MIN.         $.17
#WAIT STOR VMI          4.599 PAGE-HR.
#DRUM READS            46
#APPROX. COST OF THIS RUN IS     $1.28
#DISK STORAGE      7805     PAGE-HR.          $1.33
#APPROX. REMAINING BALANCE:   $124.47
```

Appendix I
ANALYSIS OF TRANSFORMED INJURY SCALES

Appendix I

ANALYSIS OF TRANSFORMED INJURY SCALES

As indicated previously, there are questions concerning the measurement of occupant injury. Specifically we argued that there is a true injury scale which can be obtained by a monotonic transformation of the present AIS scale. As a test of the effect of this hypothesis 4 different transformations of the present AIS scale were developed and the regression model fitted. These transformed scales are as follows:

Injury I Scale

Value	Range of AIS
0	0
1	1
2	2-3
3	4-5
4	6-9

Injury 2 Scale

Value	Range of AIS
0	0-1
1	2-5
2	6-9

Injury 3 Scale

Value	Range of AIS
0	0
1	1
2	2
4	3
6	4
10	5
12	6-9

Injury 4 Scale

Value	Range of AIS
0	0
1	1
3	2
5	3
10	4
20	5
25	6-9

Injury Scales 1 and 2 are a reduction in the range of the injury scale while 3 and 4 are an extension of the scale. These later scales assign a very high relative value to death and serious injury.

Table 65 presents a comparison of the regression models which use these alternative injury scales. In all cases the multiple correlation coefficient and the partial correlation coefficients are in close agreement over the various models. Since these transformations span the range of potential monotonic transformations of the present AIS there is evidence that a new injury scale would behave in a similar fashion.

Tables 66 and 67 present a comparison of injury prediction model residuals stratified by restraint system usage and by occupant age. In Table 66 , the residuals all indicate less severe injuries with restraints than without. However, only the residuals from Injury 1 Model are significantly different at $\alpha = 0.05$. In particular, the residuals from models 3 and 4 result in very small F statistics.

Severe injuries and fatalities are a small portion of the total reported injuries even when using CPIR data which are biased toward more severe crashes. Therefore, differences in means tend to result more from the large number of moderate or minor injuries that are reduced rather than reductions in severe and fatal injuries. Therefore, placing a high value on severe and fatal injuries adds more to unexplained variability than the differences between restraint system usage strata. In Table 67, the injury differences between age groups all have significant F statistics. In this case, the scaling problem is evidently not as severe.

Table 65

Comparison of Crash Injury Prediction Models

N=2705 Occupant Injuries

Dependent Variable

	V600 AIS	Injury 1	Injury 2	Injury 3	Injury 4
Percent Explained Variability R^2	0.38	0.38	0.38	0.37	0.35
Total S.S.	6685.	2420.	745.	18433.	77680.
Std. Error	1.24	0.74	0.41	2.07	4.33
Model Coefficients					
Constant	.00	0.22	-0.20	-.046	-1.61
VDI-Extent					
Coefficient	0.350	0.24	0.13	0.54	1.02
Partial	0.33	0.37	0.36	0.31	0.28
Impact Velocity Squared					
Coefficient	0.22	0.11	0.06	0.40	0.86
Partial	0.20	0.16	0.17	0.21	0.21
Ejection					
Coefficient	1.71	0.87	0.49	2.86	6.07
Partial	0.21	0.18	0.18	0.21	0.21
Windshield Bond Separation					
Coefficient	0.79	0.49	0.27	1.25	2.44
Partial	0.21	0.21	0.21	0.20	0.19
Single Vehicle					
Coefficient	-0.01	-0.01	0.01	0.01	0.04
Partial	-0.01	0.00	0.01	0.00	0.00

Table 66

Comparison of Restraint System Effectiveness
for Alternate Injury Scale

	N	AIS	Residual AIS Model	Residual Injury 1 Model	Residual Injury 2 Model	Residual Injury 3 Model	Residual Injury 4 Model
Neither	2029	1.39	.043	.032	.010	.034	.027
Seat Belt Only	485	1.06	-.12	-.073	-.026	-.073	-.046
Seat Belt and Upper Torso Belt	57	.89	-.24	-.22	-.057	-.24	-.19
Mean Sq. Within		2.44	1.50	.54	.17	4.22	47082
Mean Sq. Between		26.22	6.19	3.61	.35	4.03	4.30
F-Statistic		10.76	4.14	6.71	2.09	.96	.12
Signif		.000	.016	.001	.12	.38	.89

Table 67

Comparison of Age Group Effects Over
Alternate Injury Scales

Age Group	N	AIS	Residual AIS Model	Residual Injury 1 Model	Residual Injury 2 Model	Residual Injury 3 Model	Residual Injury 4 Model
1-5	61	.72	-.49	-.35	-.12	-.56	-.81
6-10	65	.68	-.44	-.32	-.065	-.53	-.77
11-15	146	.96	-.44	-.28	-.11	-.67	-1.21
16-20	598	1.24	-.16	-.080	-.040	-.25	-.49
21-25	578	1.21	-.057	-.011	-.039	-.12	-.27
26-35	454	1.32	.008	.024	.001	.026	.063
36-45	258	1.45	.21	.11	.046	.29	.51
46-55	254	1.59	.34	.20	.10	.53	1.02
56-65	141	1.65	.43	.28	.12	.64	1.15
66-80	78	2.13	.88	.45	.29	1.24	2.15
Mean Sq. Within		2.33	1.45	7.75	.17	4.18	18.36
Mean Sq. Between		18.61	22.04	.52	1.85	46.19	151.09
F-Statistic		8.00	15.15	14.84	11.15	11.06	8.23
Signif		.000	.000	.000	.000	.000	.000