Hedlund

DOT HS-802 425

A METHODOLOGY FOR ANALYZING GENERAL CATEGORICAL DATA WITH MISCLASSIFICATION ERRORS WITH AN APPLICATION IN STUDYING SEAT BELT EFFECTIVNESS

Contract No. DOT-HS-4-00897

June 1977

Final Report

PREPARED FOR:

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

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
DOT HS-802 425		

4. Title and Subtitle	5. Report Date
A Methodology for Analyzing General Categorical Data with Misclassification Errors with an Application in Studying Seat Belt Effectiveness.	June 1977
	6. Performing Organization Code
7. Author(s) Yosef Hochberg	8. Performing Organization Report No.
9. Performing Organization Name and Address The Highway Safety Research Center University of North Carolina Chapel Hill, N.C. 27514	10. Work Unit No. (TRAIS) 11. Contract or Grant No. DOT-HS-4-00897
12. Sponsoring Agency Name and Address Department of Transportation National Highway Traffic Safety Administration 400 Seventh Street, S.W. Washington, D.C. 20590	13. Type of Report and Period Covered Final Report January 1975 - April 1977 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

In this report, a methodology for analyzing general categorical data with misclassification errors is developed and applied to the study of seat belt effectiveness. The methodology assumes the availability of an original large sample based on a fallible classifier, and requires obtaining a small supplementary sample that is cross-classified by both the fallible and a "true" (usually more expensive) classifying device. For the study of belt effectiveness, the original sample is drawn from North Carolina accident reports for the first eight months of 1975. The "true" classification of a sub-sample of these occupants is assumed to be obtained through hospital reports for the injured occupants and telephone interviews for the non-injured.

Chapter I presents some background material, while Chapter II outlines the mathematical derivation of the proposed methodology. (Details of the statistical methodology are given in Appendix A.) In Chapter III, the data to which the methodology is applied in examining safety belt effectiveness is described. Results of the investigation are given in Chapter IV. An examination of the nature and magnitude of misclassification errors in this police-reported belt and injury data is described in Chapter V. A brief discussion of this work along with suggestions for further research are included in Chapter VI.

17. Key Words	18. Distribution Statement
Misclassification errors Fallible and true classifiers Maximum Likelihood estimators Least Squares estimators Differential risk, Relative risk Seat belt effectiveness	Document is available to the public through the National Technical Information Service, Springfield, Virginia 22161

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	74	

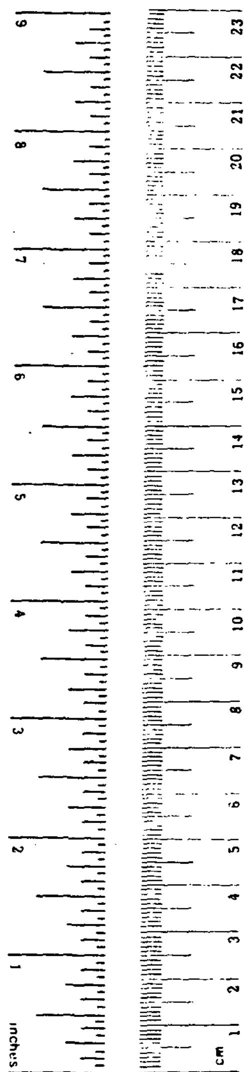
Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

METRIC CONVERSION FACTORS

Approximate Conversions to Metric Measures

Symbol	When You Know	Multiply by	To Find	Symbol
		LENGTH		
in	inches	*2.5	centimeters	cm
ft	feet	30	centimeters	cm
yd	yards	0.9	meters	m
mi	miles	1.6	kilometers	km
		AREA		
in^2	square inches	6.5	square centimeters	cm^2
ft^2	square feet	0.09	square meters	m^2
yd^2	square yards	0.8	square meters	m^2
mi^2	square miles	2.6	square kilometers	km^2
	acres	0.4	hectares	ha
		MASS (weight)		
oz	ounces	28	grams	g
lb	pounds	0.45	kilograms	kg
	short tons (2000 lb)	0.9	tonnes	t
		VOLUME		
tsp	teaspoons	5	milliliters	ml
Tbsp	tablespoons	15	milliliters	ml
fl oz	fluid ounces	30	milliliters	ml
c	cups	0.24	liters	l
pt	pints	0.47	liters	l
qt	quarts	0.95	liters	l
gal	gallons	3.8	liters	l
ft^3	cubic feet	0.03	cubic meters	m^3
yd^3	cubic yards	0.76	cubic meters	m^3
		TEMPERATURE (exact)		
°F	Fahrenheit temperature	5/9 (after subtracting 32)	Celsius temperature	°C

*1 in = 2.54 (exactly). For other exact conversions and more detailed tables, see NBS Misc. Publ. 286, Units of Weights and Measures, Price $2.25, SD Catalog No. C13.10:286.



Approximate Conversions from Metric Measures

Symbol	When You Know	Multiply by	To Find	Symbol
		LENGTH		
mm	millimeters	0.04	inches	in
cm	centimeters	0.4	inches	in
m	meters	3.3	feet	ft
m	meters	1.1	yards	yd
km	kilometers	0.6	miles	mi
		AREA		
cm^2	square centimeters	0.16	square inches	in^2
m^2	square meters	1.2	square yards	yd^2
km^2	square kilometers	0.4	square miles	mi^2
ha	hectares (10,000 m^2)	2.5	acres	
		MASS (weight)		
g	grams	0.035	ounces	oz
kg	kilograms	2.2	pounds	lb
t	tonnes (1000 kg)	1.1	short tons	
		VOLUME		
ml	milliliters	0.03	fluid ounces	fl oz
l	liters	2.1	pints	pt
l	liters	1.06	quarts	qt
l	liters	0.26	gallons	gal
m^3	cubic meters	35	cubic feet	ft^3
m^3	cubic meters	1.3	cubic yards	yd^3
		TEMPERATURE (exact)		
°C	Celsius temperature	9/5 (then add 32)	Fahrenheit temperature	°F

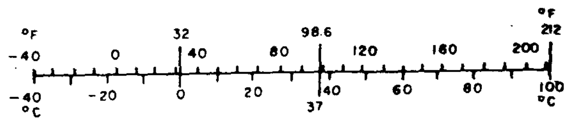


TABLE OF CONTENTS

ACKNOWLEDGMENTS

Sincere appreciation is expressed to those individuals and organizations who supplied the data for this study. In addition to the many accident victims who responded to the survey, the author wishes to thank the North Carolina Division of Motor Vehicles for providing the sample of accident records, and the following 11 hospitals that assisted in the data gathering:

Alamance County Hospital, Burlington, N.C.
Cabarrus Memorial Hospital, Concord, N.C.
Cape Fear Memorial Hospital, Wilmington, N.C.
Forsyth Memorial Hospital, Winston Salem, N.C.
Gaston Memorial Hospital, Gastonia, N.C.
High Point Memorial Hospital, High Point, N.C.
King's Mountain Hospital, Kings Mountain, N.C.
Memorial Hospital of Alamance County, Burlington, N.C.
Onslow Memorial Hospital, Jacksonville, N.C.
Wake County Memorial Hospital, Raleigh, N.C.
Watts Hospital, Durham, N.C.

The author is also very indebted to Ms. Lucy Smith, Research Assistant at HSRC, who was responsible for the coordination of the collection of the supplementary data. As such, she directed the entire effort involved in the hospital and telephone surveys. Special thanks are also due to Ms. Debbie Wood, a programmer at HSRC, whose contributions both in writing original programs and in using a large number of existing ones are evident throughout this work.

A large number of HSRC staff participated in various parts of this project, especially in connection with the supplementary sample. These include Dr. B.J. Campbell, HSRC Director, who contributed throughout the study with advice as to basic issues regarding the hospital and telephone surveys; Ms. Jane Youngblood, who played a major role in conducting the telephone survey; Ms. Anita Leung, who carried out some additional required programming; and Ms. Jane Stutts, who greatly assisted in the preparation of this final report.

A special acknowledgment is due to the participation of Dr. Donald W. Reinfurt, Project Director, in all parts of this work. His advice, guidance, and organizational assistance, as well as his critical reading and suggestions for improving the content and layout of this report, were essential to the final product.

Finally, the author is greatly indebted to Mr. Donald F. Mela, the Contract Technical Manager, for his helpful guidance and suggestions through all phases of the project.

I. INTRODUCTION

There are several major problems which make a precise measure of seat belt effectiveness very difficult. One of these is the presence of misclassification errors in police-reported accident data. To date, most studies on the effectiveness of seat belts in reducing injury have been based on police-level data. Due to the circumstances surrounding the officer's investigation of the crash, however, this data generally contains misclassification errors relating to belt usage and injury sustained. Such errors have the potential of seriously biasing any effectiveness estimates derived from that data.

The problem of misclassification errors in police-reported accident information with reference to studies of seat belt effectiveness was first raised by Mela (1974) and further discussed in Hochberg (1976). The discussion in the latter report supported the need for a methodology for modeling and obtaining unbiased inferences from general categorical data with misclassification errors.

Much has been written on the effects of misclassification errors on studies of association in 2 × 2 contingency tables (see, e.g., Fleiss, 1973, Ch. 3). In Koch (1969), the misclassification errors are assumed to be generated according to a random response error model. As such, the methodology is based on repeated classifications of the experimental elements. Such a methodology, however, can not always be satisfactory because of obvious practical difficulties and since, in many problems, misclassification errors are fixed biases rather than random response errors.

Most studies of the potential effects of fixed bias misclassification errors have severely restricted the number of error parameters examined. In the 2 × 2 table setup, one may theoretically have as many as 12 different parameters for fixed bias misclassification errors. (For example, an element which actually belongs in the first row and the second column may be misclassified into the second row and second column, etc.). In practice, however, many of these parameters are assumed to be zero. Thus, Bross (1954) introduced a model for fixed bias misclassification errors for a 2 × 2 table where only two error parameters are considered; and Hochberg (1976) discussed the effects of six error parameters on three measures of association in 2 × 2 tables of belt usage by level of injury.

While the effects of more general misclassification error structures on inference have been discussed in some recent works (see, e.g., Goldberg, 1975), no methodologies for an improved statistical inference have been presented. The purpose of the present report is to present such a methodology, and to apply it to the study of safety belt effectiveness.

The setup for the methodology is general; i.e., the discussion can be applied to any multidimensional cross-classified data obtained by unrestricted random sampling. The methodology itself is based primarily on the double sampling scheme originally introduced by Tenenbein (1970, 1971, 1972) for estimating the parameters of a multinomial classification when misclassification errors prevail.

The following situation is assumed. There are two classification "devices" available. One is expensive to apply, yet gives "correct" results, while the other is relatively inexpensive but "fallible." As an example, Diamond and Lilienfeld (1962) discuss an experimental situation in public health research where the true classification device is a physician's examination, whereas the fallible classifier is a questionnaire completed by the patient. In other situations, the "true device" and "fallible device" may simply refer to making or not making an extra effort to obtain more reliable data.

The methodology, as developed in this report and applied to the study of belt effectiveness, uses an original large sample based on (fallible) police-reported data, and requires that a small subsample of the data be cross-classified by means of some "true" classifying device. In this case, the true classifier is assumed to be hospital reports on the injured occupants and telephone interviews for the non-injured occupants. The supplementary sample of cross-classified data is then used to adjust the original police-based sample, and inference of seat belt effectiveness is taken from this larger, adjusted sample (the adjusted police data).

In real world problems, it is often the case that the true classification device uses different nominal scales than those used by the fallible device. This is illustrated by the seat belt data presented in this report, where injury is coded by the police using the K,A,B,C,O scale, but reported by the hospitals according to the Abbreviated Injury Scale (AIS). In such instances, use of the two-sample methodology has the additional advantage of enabling one to carry out an efficient study expressing results in terms of the (often) finer scale utilized with the relatively small supplementary sample. Thus, for the present study, final estimates of seat belt effectiveness could be based on the AIS scale, rather than the less precise K,A,B,C,O scale.

In summary, the two-sample methodology proposed in this report represents one approach to resolving the problems of inference arising from classification errors in categorical data. To highway safety researchers concerned with the issue of safety belt effectiveness, it is offered as a viable alternative to drawing inference solely from police report data (which may be biased) or obtaining an independent reliable sample of sufficient size and basing inference entirely on it (a process that is likely to be both costly and time consuming).

The methodology itself is described briefly in Chapter II and in detail in Appendix A. In Chapter III, the original large data source and the supplementary data used to demonstrate the technique are described. Chapter IV presents the results of applying the technique to this data, while Chapter V presents a general discussion of the nature and actual magnitude of misclassification errors in the data. This chapter is the outgrowth of an effort to test certain hypotheses concerning misclassification errors that were made in the Hochberg (1976) report. Finally, Chapter VI provides a discussion of the methodology with suggestions for further research in this area.

II. METHODOLOGY

The statistical methodology developed in this research (see Appendix A for details) pertains to the setup where all variables are subjected to misclassification errors when the fallible device (i.e., police reports) is used. It is assumed herein that the magnitudes of errors within combinations of levels of the correctly reported set of variables are possibly different.

The procedure basically extends Tenenbein's (1970, 1971, 1972) double sampling scheme originally introduced for estimating the parameters of a multinomial distribution when misclassification errors prevail. The procedure simultaneously utilizes the information from a large "fallible" sample (in this case, a large collection of police-reported data) along with a relatively small "non-fallible" supplementary sample (in this case, data from telephone and emergency room respondents) to more efficiently estimate the multinomial parameters ($\pi(\underset{\sim}{i})$) of interest (namely, belt usage by injury category).

The cross-classification of the resulting data by both police reports and non-police reports results in contingency tables with underlying multinomial distributions. The task is to find efficient estimators for the parameters of the resulting distributions along with covariances of these estimators for subsequent hypothesis testing.

The details of the estimation and testing procedures along with the necessary notation are given in Appendix A. In a nutshell, the procedure consists of two stages. In the first stage, Maximum Likelihood techniques are utilized to estimate the overall true distribution of occupants in accidents across the levels of (belt usage) × (injury) × (other variables of interest such as type of car)

This is done by setting up the joint likelihood function (A.1) for the combined sample, differentiating, and setting the partial derivatives equal to zero. This yields the MLE's given in (A.2) which are related to the main parameters of interest ($\pi(\underset{\sim}{i})$) by (A.3).

Asymptotic covariances of the $\hat{\pi}$'s are next derived by Taylor series expansions. For efficiency, these estimates serve as initial input to the asymptotically equivalent Least Squares procedures presented in Grizzle et al. (1969) for additional inferences concerning linear hypotheses involving the $\hat{\pi}$'s.

Thus, the estimates make use of the information in both samples to derive estimates of the π's. If the supplementary (or non-fallible) sample were sufficiently large, it would be optimal to use only this sample. However, the procedure suggested herein allows for a relatively small but expensive "non-fallible" sample supplemented by the large but inexpensive "fallible" sample to carry out improved estimation and testing procedures for such categorical data problems.

III. THE DATA

The data used to demonstrate and evaluate the methodology presented in Chapter II was derived from North Carolina traffic accidents. The original large sample consisted of over 139,000 occupants in accidents involving cars or small trucks for which police report information on belt usage was available. The accidents were those recorded on the HSRC North Carolina accident tapes for the first eight months of 1975. Table 3.1 presents the raw data, broken down by age and sex of driver, car "make" and model year, vehicle damage severity, and accident type. (The "not stated" or "unknown" categories are deleted from the table.)

In order to not only examine the effect of misclassification errors in this police-reported data but also to adjust the belt effectiveness estimates accordingly, supplementary data was obtained for a subsample of over 2,000 North Carolina accidents. For this phase of the study, it was assumed that follow-up telephone interviews would provide "true" information on belt usage and injury level for the non-injured occupants, while special forms completed by participating hospitals would supply the corresponding "correct" information on the injured occupants. During the four month data collection period, over 2100 telephone interviews were successfully completed, and over 900 hospital forms linked with accident reports.

Appendix B presents the cross-classification of the police and non-police data by belt usage and injury level across a number of variables indicated on the accident report form. These include age and sex of driver, car "make" and model year, vehicle damage severity (TAD), and accident type.

Table 3.1 Belt usage and injury level for 1975 North Carolina accident data, broken down by age, sex, car "make", model year, vehicle damage severity (TAD) and accident type

		No Injury			Injury			
		No Belt (U)	Lap Belt (L)	Lap & Sh. Belt (LS)	No Belt (U)	Lap Belt (L)	Lap & Sh. Belt (LS)	Total*
Overall		94834	11287	3006	21127	2010	493	132757
Age	16-55	80777	9579	2703	18316	1712	440	113527
	56+	11460	1404	225	2225	237	40	15591
Sex	Male	63095	7817	2077	12417	1194	286	86886
	Female	29729	3232	868	8253	773	197	43052
Car "Make"	U.S.	52139	6502	1664	11546	1074	241	73166
	Foreign	5369	691	428	1700	189	74	8451
Model Year	1960-1968	26594	1634	68	6296	264	7	34863
	1969-1971	19553	2159	243	4268	384	38	26645
	1972-1973	15950	3052	363	3239	521	69	23194
	1974-1975	8363	1472	1527	1748	276	214	13600
Vehicle Damage Severity	Minor	36931	4355	1140	3876	372	85	46759
	Severe	21958	2530	745	10443	918	224	36818
Accident Type	Non-Collision	11758	1159	388	7100	457	129	20991
	Collision	81150	9897	2560	13593	1510	352	109062

* "Not Stated" cases excluded.

The remaining sections of this chapter describe in greater detail the processes involved in obtaining the supplementary telephone and hospital data, along with the weighting of the supplementary sample required to make it representative of the overall accident sample.

The Telephone Survey

For this phase of the study, the North Carolina Division of Motor Vehicles (DMV) furnished HSRC copies of randomly selected accident report forms recently received from across the state. The initial quota of 250 accident reports per week was gradually increased to 450 per week to build up the sample size for the uninjured occupants. In addition, some 300 supplemental accident reports were obtained from the local police departments in Chapel Hill and Raleigh.

As copies of the accident reports (see Appendix C) were received by HSRC, they were screened to exclude injured occupants as well as accidents involving motorcycles, pedestrians, tractor trailers, etc. Next, as the North Carolina accident report form does not provide the telephone numbers of the drivers involved in the accident, a rather complicated and time-consuming, yet educational process was carried out to locate these individuals. This involved first trying to obtain an appropriate telephone number and then reaching the desired person for the interview. Some of the difficulties encountered included the following:

(a) The names and telephone numbers of females were particularly difficult to locate in the telephone directories since wives are usually not listed separately from their husbands, nor daughters from their fathers. In the case of married women, this problem could have been alleviated somewhat had the police also recorded the husband's name (rather than "same as driver") under the vehicle ownership heading on the accident report. Consequently, considerable effort was devoted to examining the telephone directories to try and match last names with street addresses, or, for the smaller towns, imposing on the telephone operator to perform this task.

(b) In North Carolina, there are a large number of small, private telephone companies operating throughout the state. Not too infrequently, two different telephone companies operated within a given radius of a community, but an individual's telephone number was naturally only listed in the directory of the company that owned his phone. Thus, one could not conclude that a given telephone number was not available after looking in just one directory, unless one was certain that there was not a second telephone company operating in the area. Unfortunately, this information was not always available and thus the "no listing" frequency was inflated.

(c) North Carolina has two different area codes (919 in the East, 704 in the West) with no available listing of which towns are in which area. This only further served to complicate the job of the information operator for addressees in some of the smaller municipalities.

As the telephone numbers became available, the interviewers concentrated their calling during the early evening hours, primarily on weekdays. Initial contact with family members did, however, often require follow-up calls of the accident-involved occupant during office hours, late in the evening, or on weekends. All telephone numbers were attempted for at least three days before being classified as "not reachable".

Appendix D contains a copy of the questionnaire that was utilized, along with the suggested introductory remarks to be used by the interviewer. Over the four-month interviewing period, for nearly half of the accident-involved drivers there was no telephone listing and hence these people were unfortunately not reachable. For an additional few cases, the subject was not available for interview. Of the remaining cases, only 2.9 percent of those contacted flatly refused to cooperate, while an additional 0.5 percent denied being in an accident. The upshot was questionnaire information on 2,132 uninjured occupants, along with the corresponding accident information from the accident report forms.

The Hospital Survey

The second data source used to evaluate the accuracy of police-reported belt usage and injury level required the cooperation of hospitals and, in particular, their emergency room staffs. More specifically, it required that the hospitals submit a completed form (see Appendix E) with information on belt usage and degree of injury for each patient seen in the emergency room as a result of an automobile accident.

Fourteen hospitals across North Carolina were contacted as potential participants in this phase of the study. All but one of the hospitals had assisted during 1972 in a similar type study (see McLean, 1973). The following eleven hospitals participated in this study resulting in statewide information on "correct" belt usage and injury level:

Alamance County Hospital, Burlington, N.C.
Cabarrus Memorial Hospital, Concord, N.C.
Cape Fear Memorial Hospital, Wilmington, N.C.
Forsyth Memorial Hospital, Winston Salem, N.C.
Gaston Memorial Hospital, Gastonia, N.C.
High Point Memorial Hospital, High Point, N.C.
King's Mountain Hospital, Kings Mountain, N.C.
Memorial Hospital of Alamance County, Burlington, N.C.
Onslow Memorial Hospital, Jacksonville, N.C.
Wake County Memorial Hospital, Raleigh, N.C.
Watts Hospital, Durham, N.C.

After each hospital administration agreed to assist in the data collection, a training session with its emergency room staff was held. The keystone of the training session was the stressing of the importance of inquiring about seat belt usage while the patient was being treated. With those unconscious or disoriented cases, the emergency room staffs were encouraged to question witnesses or even the ambulance services personnel regarding belt usage. Because of their expressed interest in the outcome of the survey, all hospitals were promised (and will receive) a copy of the completed report.

Appendix E contains a copy of the form completed by the emergency room staffs on each accident victim. The form (HSR-006) was designed with the goal of being comprehensive and yet easy to complete in the midst of emergency-type pressures and confusion. For ease and accuracy in making the correct injury (AIS) classification, the standard American Medical Association scale definitions on injury categories were incorporated into the form (see page 67). The hospital staff were further instructed to call HSRC (collect) if they had any questions regarding either the forms or some broader aspect of the study. Bi-monthly newsletters were also issued by HSRC to clear up any problem areas as well as to offer encouragement and support to the participating personnel. As a result, there were no major difficulties associated with this phase of the study.

The hospitals collected data on accident victims from March 1, 1975 through June 1, 1975. As forms were completed, they were mailed to HSRC in the pre-addressed business reply envelopes provided. Each week, HSRC compiled a list of the name of each injured occupant reported by the cooperating hospitals along with his birthdate, county of residence, and date(s) of treatment, and then forwarded this information to the Division of Motor Vehicles (DMV). The DMV staff then, to the extent possible, located the accident reports corresponding to the names on the list.

Due to time delays in receiving accident reports from the various police agencies across the state, there were inevitable difficulties in locating the corresponding accident report forms. Two weeks into the study, DMV requested that in addition the driver's name be submitted with each accident victim's name. It was anticipated that this would increase the percentage of linkages with the accident records file. Through the regular newsletter to the hospitals, this additional step was quickly implemented by the hospital personnel.

Even with this additional information with which to link the emergency room data with the police accident data, it was not always possible to locate the corresponding copy of the police report form. This was as anticipated due to occasional lengthy delays in DMV receiving the reports from some of the smaller or more remote police departments. Also, if the accident victim provides false information (names) in the emergency room setting, that case will not likely be able to be used. Nevertheless, the rate of linkage (slightly over 70%) appeared reasonable and no serious biases were evident.

Once a hospital report form was linked with its corresponding accident report from the DMV file, information from both sources was keypunched and placed on file for subsequent analysis. As mentioned previously, a total of 911 emergency room forms were successfully linked and coded during this phase of the project.

Adjustment of the Supplementary Sample

Clearly, the supplementary sample described in the preceding two sections is not a simple random sample from the larger sample of North Carolina accidents for which police-reported information is available. Actually, it is structured as a stratified random sample where its two strata are based generally on injury level. With very few exceptions, those occupants interviewed over the phone did indeed place lower on the injury scale than those for whom information was obtained via the hospital reports.

In order to account for any biases that such a sampling scheme might introduce, the data were adjusted to reflect the overall target population (i.e., all North Carolina accident victims) with respect to certain relevant variables, namely those whose distributions are confounded in the design--age, sex, race, and level of injury.

Table 3.2 compares the police-reported data for the supplementary sample only with all of the police-reported data for the first half of 1975, prior to any adjustment (i.e., weighting).

Table 3.2 Age, sex, race, and injury distributions of supplementary sample and 1975 accident data.

Variable	Level	Supplementary Sample	1975 Accidents
Age	16-55 56+	87.0 13.0	87.8 12.2
Sex	Male Female	60.7 39.3	66.8 33.2
Race	White Non-White	83.9 16.1	77.1 22.9
Injury Level	No Injury C Injury B Injury A Injury Fatality	74.0 10.6 11.9 3.2 0.3	82.7 7.7 7.0 2.3 0.3

As expected, the unadjusted supplementary sample inflates the proportions of injury, except for fatalities. It also oversamples whites and females. Again, this might be anticipated, since these individuals are, for example, generally easier to contact in a telephone survey. Finally, the age deviations between the two samples are small.

As a result of this investigation, the supplementary sample was weighted to match the relevant 1975 accident data with respect to its distribution over the 40 cells in the cross-classification of (age) × (sex) × (race) × (injury level).

IV. RESULTS

In this chapter are presented estimates of injury risk and belt effectiveness based on the North Carolina police-reported data only, the supplementary (non-police) data only, and the combined police and supplementary data (applying the methodology described in detail in Appendix A). This is done for a number of control variables of interest, including age and sex of driver, model year and type of car, vehicle damage severity (TAD), and accident type.

The specific procedure utilized is the modified Maximum Likelihood approach described in Appendix A . This approach is the most convenient to apply, and the results are generally equivalent to those obtained via the complete Least Squares approach.

Due to the relatively small size of the supplementary sample, only two levels of each control variable were considered. These are defined as follows:

Sex: Male
Female

Age: 16-55
56+

Car type: U.S. (e.g., Chevrolet, Plymouth)
Foreign (e.g., VW Beetle, Datsun)

Model year: pre-1972
1972-75

Vehicle damage severity:

Minor (i.e., front center or front left impacts of TAD severity 1; all other impacts with severity $\leq$ 3)

Moderate or severe (i.e., front center or front left impacts of TAD severity > 1; all other impacts with severity > 3)

Accident type:

"Collision" (e.g., collision of motor vehicle in road with another motor vehicle, pedestrian, bicyclist, etc.)

"Non-Collision" (e.g., ran-off-road on the right, overturn)

As previously noted, in the police-reported data, "injured" is defined as having an injury of "C" or worse on the K,A,B,C,O scale, while for the supplementary data, a person recorded as injured has an AIS severity rating of 1 or greater.

Table 4.1 gives the risk of injury and the corresponding belt effectiveness estimates derived from the police-reported data, Table 4.2 for the supplementary (or non-police) data, and Table 4.3 for the combined sample. The measures of belt effectiveness (E) are presented for two cases -- E_{12} for none vs. lap, and E_{23} for lap vs. lap and shoulder -- where

E_{12} = relative decrease in "injury" for lap-belted occupants compared with unrestrained occupants

$$= \frac{[\%\text{ inj. (none)}] - [\%\text{ inj. (lap)}]}{\%\text{ inj. (none)}} \times 100$$

and

E_{23} = relative decrease in "injury" for lap and shoulder-belted occupants compared with lap-belted occupants

$$= \frac{[\%\text{ inj. (lap)}] - [\%\text{ inj. (lap+shoulder)}]}{\%\text{ inj. (lap)}} \times 100$$

Thus, belt effectiveness is viewed as the percentage decrease in injury as one becomes progressively more restrained.

Table 4.1 Estimated injury risks and belt effectiveness, based on police-reported data.

Control Variable		Percent Injured (K,A,B,C) None	Lap	Lap & Sh	Belt Effectiveness None vs Lap	Lap vs Lap & Sh
Sex	Male	16.44 (0.13)[1]	13.25 (0.36)	12.10 (0.67)	19.42 (2.27)	8.66 (5.63)
	Female	21.73 (0.21)	19.30 (0.62)	18.50 (1.19)	11.17 (3.00)	4.16 (6.90)
Age	16-55	18.48 (0.12)	15.16 (0.34)	14.00 (0.62)	17.97 (1.91)	7.67 (4.57)
	56+	16.26 (0.32)	14.44 (0.87)	15.09 (2.20)	11.17 (5.61)	-4.51 (16.47)
Model Year	1960-71	18.63 (0.16)	14.59 (0.53)	12.64 (1.76)	21.67 (2.93)	13.37 (12.47)
	1972-75	17.02 (0.22)	14.98 (0.49)	13.02 (0.72)	12.00 (3.09)	13.05 (5.59)
Accident Type	Collision	14.35 (0.11)	13.24 (0.32)	12.09 (0.60)	7.74 (2.33)	8.68 (5.06)
	Non-Collision	37.65 (0.35)	28.28 (1.12)	24.95 (1.90)	24.89 (3.06)	11.77 (7.58)
Car Type	U.S.	18.13 (0.15)	14.18 (0.40)	12.65 (0.76)	21.81 (2.31)	10.76 (5.94)
	Foreign	24.05 (0.51)	21.48 (1.38)	14.74 (1.58)	10.69 (6.06)	31.36 (8.59)
Damage Severity	Minor	9.50 (0.15)	7.87 (0.39)	6.94 (0.73)	17.15 (4.31)	11.83 (10.22)
	Severe	32.23 (0.26)	26.62 (0.75)	23.12 (1.35)	17.40 (2.43)	13.17 (5.65)
Total Sample[2]		115961	13297	3499	--	--

[1]Standard deviation

[2]"Not Stated" cases excluded

Table 4.2 Estimated injury risks and belt effectiveness, based on <u>supplementary</u> data.

Control Variable		Percent Injured (AIS ≥ 1)			Belt Effectiveness	
		None	Lap	Lap & Sh	None vs Lap	Lap vs Lap & Sh
Sex	Male	26.31	20.42	14.46	22.38	29.21
		(1.30)[1]	(2.39)	(2.73)	(9.87)	(15.73)
	Female	38.09	27.72	23.64	27.21	14.74
		(1.92)	(4.45)	(5.73)	(12.26)	(24.79)
Age	16-55	31.68	22.67	17.28	28.43	23.80
		(1.17)	(2.26)	(2.74)	(7.60)	(14.25)
	56+	21.33	19.51	13.79	8.51	29.31
		(2.82)	(6.19)	(6.40)	(31.44)	(39.75)
Model Year	1960-71	31.13	23.88	11.36	23.29	52.41
		(1.37)	(3.01)	(4.78)	(10.23)	(20.91)
	1972-75	28.53	20.99	18.29	26.40	12.90
		(1.81)	(3.03)	(2.92)	(11.59)	(18.75)
Accident Type	Collision	27.71	21.73	14.12	21.60	35.02
		(1.19)	(2.33)	(2.67)	(9.06)	(14.13)
	Non-Collision	41.71	25.76	22.73	38.25	11.77
		(2.59)	(5.38)	(6.32)	(13.46)	(30.69)
Car Type	U.S.	29.82	21.61	16.88	27.52	21.93
		(1.14)	(2.21)	(2.96)	(7.91)	(15.85)
	Foreign	36.84	27.78	16.95	24.60	38.98
		(3.69)	(7.47)	(4.88)	(21.62)	(24.04)
Damage Severity	Minor	18.86	15.17	12.50	19.59	17.59
		(1.41)	(2.69)	(3.38)	(15.46)	(26.62)
	Severe	42.36	36.36	25.00	14.15	31.25
		(1.82)	(4.37)	(4.51)	(10.96)	(14.92)
Total Sample[2]		1783	384	218	--	--

[1]Standard deviation

[2]"Not Stated" cases excluded

Table 4.3 Estimated injury risks and belt effectiveness based on combined police and supplementary data.

Control Variable		Percent Injured: None	Percent Injured: Lap	Percent Injured: Lap & Sh	Belt Effectiveness: None vs. Lap	Belt Effectiveness: Lap vs Lap & Sh
Sex	Male	26.99 (1.02)[1]	20.17 (2.26)	14.02 (2.71)	25.26 (9.02)	30.48 (15.96)
	Female	38.58 (1.55)	28.88 (4.12)	23.10 (5.64)	25.16 (11.22)	20.01 (22.63)
Age	16-55	32.03 (0.93)	22.81 (2.13)	16.60 (2.70)	28.79 (7.06)	27.23 (14.01)
	56+	23.00 (2.27)	19.97 (5.38)	15.09 (6.73)	13.16 (26.19)	24.46 (40.01)
Model Year	1960-71	31.17 (1.07)	22.77 (2.78)	12.04 (4.88)	26.96 (9.40)	47.11 (22.60)
	1972-75	29.98 (1.47)	21.68 (2.83)	17.91 (2.94)	27.69 (10.30)	17.40 (18.03)
Accident Type	Collision	27.79 (0.97)	22.27 (2.19)	13.11 (2.52)	19.84 (8.53)	41.13 (12.94)
	Non-Collision	46.22 (1.89)	25.97 (4.87)	29.40 (7.88)	43.81 (10.82)	-13.20 (38.35)
Car Type	U.S.	30.50 (0.90)	21.33 (2.05)	16.26 (2.86)	30.09 (7.11)	23.77 (15.66)
	Foreign	38.99 (3.20)	29.77 (7.48)	16.79 (5.05)	23.64 (20.69)	43.62 (22.37)
Damage Severity	Minor	19.65 (1.21)	15.24 (2.67)	11.42 (3.24)	22.41 (14.72)	25.06 (25.47)
	Severe	45.38 (1.41)	38.08 (4.06)	27.50 (4.78)	16.08 (9.45)	27.79 (15.15)
Total Sample[2]		115961	13297	3499	--	--

[1]Standard deviation
[2]"Not Stated" cases excluded

In comparing the results based on the police data (Table 4.1) with those based on the combined data (Table 4.3), one finds that the estimated risks and effectiveness are quite different, with the combined estimates being substantially higher in most cases. Thus, for example, controlling for sex, the estimates of belt effectiveness for males are 25.26% for none vs lap and 30.48% for lap vs lap + shoulder, based on the combined results, compared with only 19.42% and 8.66% for the corresponding estimates based on the police data. A part of this difference can be attributed to the lack of equivalence between the two injury scales employed, with fewer people being classified as injured on the police scale. (Perhaps "injured" on the AIS scale should have been defined as AIS $\geq$ 2!) Most of the difference, however, is indeed probably due to misclassification errors in the police reports.

Since all the data in this study are biased, their quality (i.e., accuracy) cannot be judged solely on the basis of the accompanying standard deviations. A more appropriate measure of the accuracy of the police report estimates is the more general "mean square error" (MSE), where

$$\text{MSE} = \text{Variance} + (\text{Bias})^2.$$

This measure can be applied to both the risk and effectiveness estimates.

In calculating MSE's for the police report data, we assume that the best available estimator for the bias of a given estimate is the difference between that estimate and the corresponding estimate obtained via the "combined" methodology. For example, the police estimate of percentage injury to unbelted males is 16.44 with a variance of 0.017 and the "combined" sample estimate is 26.99 with a variance of 1.04. The bias of the police report estimate is then 10.55 (= 26.99 - 16.44). Thus, the MSE is given by

$$(10.55)^2 + .017 = 111.32$$

This compares with a variance (and approximate MSE) of 1.04 for the "combined" approach.

It should be noted that, in some cases, the MSE's of the police estimates are lower than those of the "combined" approach, even though the estimated biases in the police estimates are quite substantial. This is primarily due to the "relatively" small size of the supplementary sample. If the estimated bias remained the same, but a larger (say, threefold) sample size were available, then all of the estimates based on the combined approach would have much smaller MSE's than those based on only the police-reported data.

While the injury risk and effectiveness estimates based on the police data only differ substantially from the "combined methodology" estimates, a comparison of the supplementary vs. combined data results reveals a generally high level of agreement. This is not unexpected, since both represent consistent estimators, based on the same definition of injury (AIS $\geq$ 1). The positive effect of wearing lap belts and the additional benefit derived from the use of shoulder belts are clearly evidenced in all but a few isolated instances.

In further comparing the supplementary and combined results, one finds that the standard deviations (STD's) of the estimates for the combined approach are usually lower than those resulting from the supplementary data only. However, they are not as much lower as might be anticipated, considering the large increase in sample size. This seems to indicate that the combined methodology may offer only a slight improvement in accuracy over using only the supplementary data, at least when multi-dimensional contingency tables are considered with this considerable amount of data.

By increasing the size of the supplementary sample, one would, of course, decrease the level of error in both the supplementary and combined data results. However, it is suggested that the relative decrease would probably be less for the combined than the supplementary results. That is, as the supplementary sample size is increased, the relative benefit of utilizing the combined samples to estimate the fallible margin in the cross-classified sample will decrease. Such a conclusion should not affect the overall usefulness of the combined methodology, however, since this approach is designed for use in situations where an original large (but fallible) sample is readily available, but where only a relatively small supplementary (non-fallible) sample can reasonably be obtained.

Another important issue is whether the estimates obtained using the combined methodology are the "best" estimates (i.e., those with lowest STD's) obtainable, using only the data available. The answer is that one can probably derive better estimates, even without increasing sample size. If only the supplementary sample is considered, one could use well-known techniques for building models that smooth the original proportions by removing non-significant variations in a multi-factor set-up. The result would be estimates with lower STD's. How to accomplish the same objective with the "combined approach" is a more complicated matter, and will be discussed further in Chapter VI.

V. THE NATURE OF MISCLASSIFICATION ERRORS IN POLICE-REPORTED BELT USAGE AND INJURY LEVEL

The investigation described herein is aimed at exploring the magnitude of the misclassification errors of belt usage and level of injury in actual statewide police data. Table 9 in Hochberg (1976) shows the effect on several measures of belt effectiveness (namely, RIDIT, relative risk, and the odds ratio) for various magnitudes of a combination of misclassification errors. Certain questions remain: How large are these misclassification errors in actual data? Are the simplifying assumptions made in Hochberg (1976) valid assumptions? Do the magnitudes of the errors depend upon other factors such as age and sex of the driver?

This chapter explores questions such as these using the data reported in the text where the police (P) data are "fallible" while the "true" belt/injury status is given by the hospital/telephone interview ($\bar{P}$) data. As before, belt (B) includes the use of any restraint system (lap, lap and shoulder) with $\bar{B}$ indicating no restraint used; injury (I) includes any injury (K,A,B, or C) recorded by the police or an AIS ≥ 1 for the non-police data, while $\bar{I}$ indicates no injury.

Table 5.1 gives the raw frequency data for belt usage and level of injury, cross-classified by the police (P) and non-police ($\bar{P}$) sources.

Table 5.1 Cross-classification of supplementary sample according to belt status and level of injury.

Non-Police \ Police		Injured Belt	Injured No Belt	Not Injured Belt	Not Injured No Belt	Total
Injured	Belt	36	16	33	37	122 (5.1%)
	No Belt	6	305	5	227	543 (22.7%)
Not Injured	Belt	6	6	256	216	484 (20.3%)
	No Belt	2	27	15	1194	1238 (51.9%)
	Total	50 (2.1%)	354 (14.8%)	309 (12.9%)	1674 (70.1%)	2387

It is evident from Table 5.1 that the police are much less likely to report that a seat belt was worn and also much less likely to report that an injury was involved than the non-police source.

More specifically, assuming that the non-police ($\bar{P}$) reports are the "true" classification mechanisms, then, for injured occupants, the police underreported belt use by 43.4% ($= \frac{16+37}{122} \times 100$) and overreported use by only 2.0% ($= \frac{6+5}{543} \times 100$). For uninjured occupants, the respective estimates are 45.9% underreported and 1.4% overreported.

Conversely, for belted occupants, the police underestimated injury by 57.4% ($= \frac{33+37}{122} \times 100$) and overestimated injury by 2.5% ($= \frac{6+6}{484} \times 100$). For unbelted occupants, the corresponding estimates are 42.7% underestimated and 2.3% overestimated.

Clearly this tendency to underreport belt usage and injury level will affect any derived estimates of belt effectiveness. If one defines belt effectiveness as the percentage decrease in risk of injury resulting from wearing a safety belt, i.e.,

$$\text{effectiveness} = \frac{(\%\text{ unbelted injured}) - (\%\text{ belted injured})}{\%\text{ unbelted injured}} \times 100 \qquad (5.1)$$

then, based on the police-reported data, safety belts have a 20.2% effectiveness. This compares with an effectiveness estimate of 34.0% based on the "true" hospital/telephone data. Thus, due to misclassification errors in the data, the police estimates apparently substantially underrate the effectiveness of safety belts in reducing the likelihood of injury. (The extent to which the hospital/telephone data represent the true situation is, of course, unknown, but is believed to be much closer to reality!)

As indicated in Hochberg (1976), there are a total of 12 independent misclassification errors that can arise when classifying individuals into the 2 × 2 table of belt usage $(B,\bar{B})$ by injury level $(I,\bar{I})$. For example, an individual that is actually belted and injured (B,I) could be incorrectly classified as $(\bar{B},I)$, $(B,\bar{I})$ or $(\bar{B},\bar{I})$. From Table 5.1, the "true" number of belted and injured drivers is 122 whereas the reports classified 16 individuals as being unbelted and injured, 33 as being belted and uninjured and 37 as being unbelted and uninjured!

In order to examine further the nature of such misclassification errors in police belt usage and injury data, estimates of the 12 misclassification error probabilities were obtained, along with corresponding estimates of their covariance matrix, following the approach described in Grizzle, Starmer, and Koch (1969). With these estimates, a variety of hypotheses were then tested. Of particular interest were the two "simplifying" assumptions regarding police misclassification errors that were made in Hochberg (1976). These were:

(i) The probability of a double misclassification error (i.e., both on belt use and injury level) is well approximated by the product of the two marginal error probabilities.

(ii) The probability of misclassifying an uninjured occupant (either belted or unbelted) as injured is "unlikely".

The results of the corresponding tests of hypotheses are summarized in Table 5.2 along with estimates $\hat{p}$ of the corresponding misclassification errors based on the data presented in Table 5.1, with the non-police classification representing the "true" condition. It is evident from this data that neither of the two basic assumptions ((i) or (ii)) is tenable.

As a final dimension to this analysis of the magnitude and effect of misclassification errors in this police-reported data, an overall

Table 5.2. Hypothesis tests regarding the various misclassification errors in a 2 × 2 table of belt usage vs injury level.

Ho: Pr {Misclassification error} = 0	$\hat{p}$	d.f.	χ^2
Double misclassification error (i)	.3265	4	52.46
$\bar{I}$ classified as I (ii)	.0238	4	42.00
$\bar{I}$ classified as I given B	.0248	2	12.31
$\bar{I}$ classified as I given $\bar{B}$	.0234	2	29.70
I classified as $\bar{I}$	.4541	4	572.46
I classified as $\bar{I}$ given B	.5738	2	167.94
I classified as $\bar{I}$ given $\bar{B}$	.4273	2	404.51
$\bar{B}$ classified as B	.0157	4	28.46
$\bar{B}$ classified as B given I	.0203	2	11.23
$\bar{B}$ classified as B given $\bar{I}$	.0137	2	17.24
B classified as $\bar{B}$	.4538	4	500.54
B classified as $\bar{B}$ given I	.4344	2	93.13
B classified as $\bar{B}$ given $\bar{I}$	.4587	2	407.42

Note: Corresponding p-values are all $\leq .005$.

Maximum Likelihood Model was fit to the data to examine the dependence of these errors on driver sex (see Table 5.3). The results of the investigation of other factors replacing sex are not detailed herein because it was unfortunately found that the misclassification errors did depend on the levels of these other factors considered (e.g., model year, car size accident type, driver age). Thus, the models did not simplify in a clear-cut manner.

Table 5.3 Analysis of variance table for the Maximum Likelihood Model including sex.

Source of Variation	d.f.	χ^2	p-value
Independence of misclassification errors on sex	12	3.16	> 0.99
Interactions:			
$\bar{P}I \times PB$	1	1.80	> 0.10
$(PB,PI) \times \bar{P}I$	1	1.02	> 0.30
$(PB,PI) \times \bar{P}B$	1	1.05	0.30
$(\bar{P}B,\bar{P}I) \times PI$	1	1.71	> 0.10
$(\bar{P}B,\bar{P}I) \times PB$	1	1.70	> 0.10
$(\bar{P}B,\bar{P}I) \times Sex$	1	1.20	> 0.20
$(\bar{P}B \times \bar{P}I \times PB \times PI)$	1	0.77	> 0.30
Total error	19	12.41	>0.80
Equiprobable model	31	765.84	

VI. DISCUSSION

In Chapter IV, injury risks and safety belt effectiveness estimates were presented for the North Carolina police data only, for the supplementary data only, and for the combined two samples. The variables of interest were age, sex, model year and type of car, damage severity, and accident type. The combined sample results were shown to be the most accurate, although their STD's were not a great deal lower than those associated with the supplementary data only.

The original plan for analyzing the data included considering several levels of some of the key variables (such as TAD and model year), and thus to examine the effects of these variables simultaneously. It was also anticipated that injury level could be defined so as to distinguish between the serious and non-serious and the fatal and non-fatal injuries (in addition to injury-no injury).

However, it soon became obvious that, due to the relatively small size of the supplementary sample and the state of the methodology described in Appendix A, the data could only be meaningfully analyzed using a single variable breakdown and the injury-no injury classification, as presented in Tables 4.1 - 4.3. This is due to the requirement that, in order to use the combined methodology outlined in this report, there must be at least one observation for each level of (injury × police) by (level of belt usage × police) by (level of factor under consideration). This requirement was not able to be satisfied except in the single variable framework for this data set.

The results of this analysis suggest that lap belts alone substantially reduce the likelihood of injury and that lap and shoulder belts together further reduce this likelihood. However, it should be noted that the specific estimates presented are far from satisfactory, due to their large STD's. Also, due primarily to the large STD's, significant differences in belt effectiveness could not be detected between the two levels of any of the factors considered.

The supplementary sample size used in this report to illustrate and test the two sample methodology was somewhat over 2000 cases. As suggested in Chapter III, it took considerable effort and coordination to collect the additional telephone and hospital data for even this "small" a sample. Nevertheless, it now appears that, in order to make statistically significant statements on safety belt effectiveness using this technique, one should probably have had a supplementary sample three or four times as large.

Increasing the sample size to, say, 10,000 would have positive effects beyond decreasing the STD's. For example, it would enable one to simultaneously study the effects of several factors associated with seat belt effectiveness. It would also permit one to examine the effectiveness of safety belts for other occupants besides the driver.

Two additional caveats should be made regarding the supplementary sample used in the present study. First, the sources of supplementary information regarding seat belt usage and/or level of injury were follow-up telephone interviews for the non-injured drivers and hospital reports for the injured. Whether or not these sources did indeed provide "true" information was not examined.

Second, as noted in Chapter III, the combined sample of drivers interviewed over the phone and those reported by participating hospitals was not totally representative of the overall population of North Carolina drivers involved in accidents. While the supplementary sample was adjusted to resemble the overall sample, the effect of these adjustments on the variances and covariances of the estimates was not taken into account in the analysis.

As a result of these limitations and the relatively small size of the supplementary sample referred to earlier, the data in the present study should best be regarded as a mechanism for demonstrating a new technique, rather than as a definitive estimate of safety belt effectiveness. In order to obtain more accurate and reliable results in future applications of this methodology to the study of seat belt effectiveness, one must ascertain that:

1. The supplementary sample is sufficiently large.
2. The quality of the "true" classifier is examined and proven reliable.
3. The supplementary sample is shown to be representative, or if not representative and adjustments are made, these adjustments are accounted for in the statistical analysis.

Finally, it should be noted that, while increasing the size of the supplementary sample will improve the accuracy of the belt effectiveness estimates based on the two-sample methodology, additional research is needed to further improve upon the technique. More specifically, research is needed to incorporate smoothing models for the entries in the supplementary sample, based on relatively few parameters for the misclassification errors. The methodology as it now stands does not allow for using model-predicted estimates of the frequencies in the supplementary cross-classified sample prior to "merging it statistically" with the original sample.

In addition, it is very reasonable to expect that the very large number of misclassification errors (that introduce too many degrees of freedom in the procedures described) could be structured by an appropriate statistical model, resulting in lower STD's for the predicted frequencies. These investigations might well be worth pursuing in the future.

REFERENCES

Bross, I. (1954). Misclassification in 2×2 tables. Biometrics, 10, 478-486.

Diamond, E. and Lilienfeld, A. (1962). Effects of errors in classification and diagnosis in various types of epidemiological studies. American Journal of Public Health, 10, 2106-2110.

Fleiss, J.L. (1973). Statistical Methods for Rates and Proportions. New York: Wiley, Inc.

Forthofer, R.M. and Koch, G.G. (1973). An analysis for compounded functions of categorical data. Biometrics, 29, 143-157.

Goldberg, J.D. (1975). The effects of misclassification on the bias in the difference between two proportions and the relative odds in the fourfold table. J. Amer. Statis. Assoc., 70, 561-567.

Grizzle, J.E., Starmer, C.F., and Koch, G.G. (1969). Analysis of categorical data by linear models. Biometrics, 25, 489-504.

Hartley, H.O. and Sielken, R.L.,Jr. (1975). A "super-population viewpoint" for finite population sampling. Biometrics, 31, 411-422.

Hochberg, Y. (1976). Problems of inference in studies of seat belt effectiveness. Final Report. Chapel Hill: University of North Carolina Highway Safety Research Center. NHTSA Contract No. DOT-HS-4-00897 (PB-249 287).

Koch, G.G. (1969). The effect of nonsampling errors on measures of association in 2×2 contingency tables. Journal of the American Statistical Association, 64, 851-864.

Koch, G.G., Imrey, P.B., and Reinfurt, D.W. (1972). Linear model analysis of categorical data with incomplete response vectors. Biometrics, 28, 663-692.

Landis, R.J. and Stanish, W.M. (1975). GENCAT: A Computer Program for the Generalized Chi-Squares Analysis of Categorical Data Using Weighted Least Squares to Compute Wald Statistics. Chapel Hill: University of North Carolina Department of Biostatistics.

McLean, A.J. (1973). Collection and analysis of collision data for determining the effectiveness of some vehicle systems. Report prepared for Motor Vehicle Manufacturers of the United States, Inc. (MVMA Agreement No. UNC 7301-C19).

Mela, D.F. (1974). Review of safety belt usage and effectiveness in accidents. Report of the National Highway Traffic Safety Administration.

Tenenbein, A. (1970). A double sampling scheme for estimating from binomial data with misclassifications. J. Amer. Statis. Assoc., 65, 1350-1361.

Tenenbein, A. (1971). A double sampling scheme for estimating from binomial data with misclassifications; sample size determination. Biometrics, 27, 935-944.

Tenenbein, A. (1972). A double sampling scheme for estimating from misclassified multinomial data with application to sampling inspection. Technometrics, 14, 187-202.

APPENDIX A

A Methodology for Analyzing General Categorical Data with Misclassification Errors

The methodology outlined in this appendix pertains to the setup where all variables are subjected to misclassification errors when the fallible device is used. In practice one might come across situations in which only a subset of the variables is subjected to errors. Two cases are of interest. One case is when the magnitudes of errors within combinations of levels of the correctly reported set of variables are possibly different, and the other case is when these errors can be assumed to be the same across the corresponding levels. The examples in this report demonstrate the former case, while considerable research is still needed in order to treat the latter case.

The Setup and Notation

Two independent samples are drawn from the target population. Each is an unrestricted simple random sample. If the actual frame for the population is finite, we adhere to the concept of a 'super' population (see, e.g., Hartley and Sielken, 1975). The first sample of n_1 elements is classified only by the fallible device. Let $\underset{\sim}{i}' = (i_1', \ldots, i_d')$ index a specific combination of levels of the d variables under study. The second sample of n_2 elements is simultaneously classified by both the false and the true devices. Here again, we use $\underset{\sim}{i}'$ to index the fallible cell. To index the true classification we use $\underset{\sim}{i} = (i_1, \ldots, i_d)$. Also, let $i_m = 1, \ldots, I_m$ and $i_m' = 1, \ldots, I_m'$, $m = 1, \ldots, d$, with $I_1 \times I_2 \times \ldots \times I_d = k$ and $I_1' \times I_2' \times \ldots \times I_d' = k'$.

Next, we introduce notation for the frequencies and parameters in the two samples. To simplify matters, we use the same letters to indicate similar conceptual quantities in both samples. The distinction, however, is easily made since the second sample will always have two indices corresponding to true and false classifications, respectively. Thus, $n(\underset{\sim}{i}')$ denotes the frequency in the $\underset{\sim}{i}'$-th cell as obtained in the first sample by the false classifier. Similarly, $n(\underset{\sim}{i},\underset{\sim}{i}')$ denotes the frequency in the second sample classified in the $\underset{\sim}{i}$-th cell by the true classifier and in the $\underset{\sim}{i}'$-th cell on the false categorical scale. Likewise, let $\gamma(\underset{\sim}{i}')$ and $\gamma(\underset{\sim}{i},\underset{\sim}{i}')$ denote the corresponding population proportions. We introduce $\gamma(\underset{\sim}{i}|\underset{\sim}{i}') = \gamma(\underset{\sim}{i},\underset{\sim}{i}')/\gamma(\underset{\sim}{i}')$, which is the fraction of times an element actually belongs to cell $\underset{\sim}{i}$ when reported to be in cell $\underset{\sim}{i}'$ by the fallible classifier. In addition, the convention of replacing an index by a period to indicate that summation has been taken over that index will be used throughout, e.g., $n(\underset{\sim}{i},\cdot) = \sum_{\underset{\sim}{i}'} n(\underset{\sim}{i},\underset{\sim}{i}')$.

The intermediate parameters of interest are clearly the $\gamma(\underset{\sim}{i},\cdot)$ for which we use the special notation $\gamma(\underset{\sim}{i},\cdot) \equiv \pi(\underset{\sim}{i})$.

Throughout this work we will use the convention of putting a tilde to indicate a vector. An indexed vector will be used only for the $\gamma(\underset{\sim}{i}|\underset{\sim}{i}')$ where $\underset{\sim}{\gamma}(\underset{\sim}{i}') = \{\gamma(\underset{\sim}{i}|\underset{\sim}{i}'), \text{ for all } \underset{\sim}{i}\}$.

<u>Inference Based on Maximum Likelihood Estimates (MLE) of the $\pi(\underset{\sim}{i})$</u>

Given the data, the likelihood function of the $\gamma(\underset{\sim}{i}')$ and the $\gamma(\underset{\sim}{i}|\underset{\sim}{i}')$ is given by the following:

$$F = g\prod_{\underset{\sim}{i}'} [\gamma(\underset{\sim}{i}')]^{n(\underset{\sim}{i}')+n(\cdot,\underset{\sim}{i}')} \prod_{\underset{\sim}{i}'} b(\underset{\sim}{i}') \prod_{\underset{\sim}{i}} [\gamma(\underset{\sim}{i}|\underset{\sim}{i}')]^{n(\underset{\sim}{i},\underset{\sim}{i}')} , \qquad (A.1)$$

where g is a constant depending on the n_i, i = 1,2, the $n(\underset{\sim}{i}')$, and the $n(\cdot,\underset{\sim}{i}')$; the $b(\underset{\sim}{i}')$ are constants depending on the $n(\cdot,\underset{\sim}{i}')$ and the

$n(\underset{\sim}{i},\underset{\sim}{i}')$. It is now easily verified that the MLE's are given by

$$\hat{\gamma}(\underset{\sim}{i}|\underset{\sim}{i}') = \frac{n(\underset{\sim}{i},\underset{\sim}{i}')}{n(\cdot,\underset{\sim}{i}')}$$

$$\hat{\gamma}(\underset{\sim}{i}') = \frac{n(\underset{\sim}{i}') + n(\cdot,\underset{\sim}{i}')}{n_1 + n_2} \quad . \tag{A.2}$$

Since the $\pi(\underset{\sim}{i})$ and the $\gamma(\underset{\sim}{i}|\underset{\sim}{i}')$ are in 1:1 relation with the set of $\gamma(\underset{\sim}{i}')$ and $\gamma(\underset{\sim}{i}|\underset{\sim}{i}')$, the MLE's of the $\pi(\underset{\sim}{i})$ are given by

$$\hat{\pi}(\underset{\sim}{i}) = \sum_{\underset{\sim}{i}'} \hat{\gamma}(\underset{\sim}{i}')\hat{\gamma}(\underset{\sim}{i}|\underset{\sim}{i}'), \ \forall \underset{\sim}{i} \tag{A.3}$$

Next, we consider the asymptotic variance matrix of $\hat{\underset{\sim}{\pi}}$ which we denote by $V(\hat{\underset{\sim}{\pi}})$. Note that, asymptotically, the set of the $\hat{\gamma}(\underset{\sim}{i}')$ is independent of the set of $\hat{\gamma}(\underset{\sim}{i}|\underset{\sim}{i}')$. A similar statement applies to any distinct vectors $\underset{\sim}{\gamma}(\underset{\sim}{i}')$, $\underset{\sim}{\gamma}(\underset{\sim}{j}')$, $\underset{\sim}{i}' \neq \underset{\sim}{j}'$. This is clear from the block diagonal information matrix which is easily obtained from F. Linearizing the $\hat{\pi}(\underset{\sim}{i})$ by a Taylor approximation around the $\gamma(\underset{\sim}{i}')$ and $\gamma(\underset{\sim}{i}|\underset{\sim}{i}')$, we obtain (for large samples)

$$\hat{\pi}(\underset{\sim}{i}) \approx \sum_{\underset{\sim}{i}'} \gamma(\underset{\sim}{i}|\underset{\sim}{i}')\hat{\gamma}(\underset{\sim}{i}') + \sum_{\underset{\sim}{i}'} \gamma(\underset{\sim}{i}')\hat{\gamma}(\underset{\sim}{i}|\underset{\sim}{i}') - \sum_{\underset{\sim}{i}'} \gamma(\underset{\sim}{i}')\gamma(\underset{\sim}{i}|\underset{\sim}{i}') \tag{A.4}$$

On letting

$$V(\hat{\underset{\sim}{\gamma}}) = \frac{1}{n_1 + n_2}\left[D(\underset{\sim}{\gamma}) - \underset{\sim}{\gamma}\,\underset{\sim}{\gamma}'\right] \equiv ((v_{m,n})) \qquad m,n = 1,\ldots,k'$$

$$V\left[\hat{\underset{\sim}{\gamma}}(\underset{\sim}{i}')\right] = \frac{1}{n_2\gamma(\underset{\sim}{i}')}\left[D(\underset{\sim}{\gamma}(\underset{\sim}{i}')) - \underset{\sim}{\gamma}(\underset{\sim}{i}')\underset{\sim}{\gamma}'(\underset{\sim}{i}')\right] \equiv V_{\underset{\sim}{i}'} \tag{A.5}$$

where $D(\cdot)$ is a diagonal matrix with the vector $(\cdot)$ on the main diagonal, we have asymptotically

$$V(\hat{\underset{\sim}{\pi}}) = \sum_{m=1}^{k'} \sum_{n=1}^{k'} v_{m,n} \gamma(\underset{\sim}{i}'_m)\gamma(\underset{\sim}{i}'_n) + \sum_{\underset{\sim}{i}'} \gamma^2(\underset{\sim}{i}')V_{\underset{\sim}{i}'} \quad . \tag{A.6}$$

When consistent estimators from (A.2) are substituted for the $\gamma(\underset{\sim}{i}')$ and the $\gamma(\underset{\sim}{i}')$ in (A.6), one obtains a consistent estimator $\hat{V}(\hat{\underset{\sim}{\pi}})$ for the dispersion matrix of the vector $\hat{\pi}$.

A Maximum Likelihood test of fit (i.e., $\gamma(\cdot,\underset{\sim}{i}') = \gamma(\underset{\sim}{i}')$ for all $\underset{\sim}{i}'$) is rather straightforward. The unrestricted MLE's are given by

$$\begin{aligned} \hat{\hat{\gamma}}(\underset{\sim}{i}') &= \frac{n(\underset{\sim}{i}')}{n_1} \\ \hat{\hat{\gamma}}(\underset{\sim}{i},\underset{\sim}{i}') &= \frac{n(\underset{\sim}{i},\underset{\sim}{i}')}{n_2} \end{aligned} \tag{A.7}$$

Under the null hypothesis, (i.e., $\gamma(\cdot,\underset{\sim}{i}') = \gamma(\underset{\sim}{i}')$ for all $\underset{\sim}{i}'$) the MLE's of the $\gamma(\underset{\sim}{i},\underset{\sim}{i}')$ are $\hat{\gamma}(\underset{\sim}{i},\underset{\sim}{i}') = \hat{\gamma}(\underset{\sim}{i}|\underset{\sim}{i}')\hat{\gamma}(\underset{\sim}{i}')$. On denoting the Maximum Likelihood Ratio (MLR) statistic by L, we have

$$-2 \log L = -2\left\{\sum_{\underset{\sim}{i}'} n(\underset{\sim}{i}')\log\left[\frac{\hat{\hat{\gamma}}(\underset{\sim}{i}')}{\hat{\gamma}(\underset{\sim}{i}')}\right] + \sum_{\underset{\sim}{i}}\sum_{\underset{\sim}{i}'} n(\underset{\sim}{i},\underset{\sim}{i}')\log\left[\frac{\hat{\hat{\gamma}}(\underset{\sim}{i},\underset{\sim}{i}')}{\hat{\gamma}(\underset{\sim}{i},\underset{\sim}{i}')}\right]\right\}. \tag{A.8}$$

Under the null hypothesis, this is asymptotically distributed as a central Chi-square variate with (k'-1) d.f.

Often, having established the fit, the experimenter will be interested in further inference on $\underset{\sim}{\pi}$ based on the efficient estimator $\hat{\underset{\sim}{\pi}}$. In most practical problems, it is not feasible to obtain simple MLE's of $\underset{\sim}{\pi}$ under further functional restrictions on the $\pi(\underset{\sim}{i})$ (given that the model fits). One can verify this by trying to obtain the MLR test for independence in a 2 × 2 table. Even for this simple problem, the MLE cannot be obtained explicitly and one must call upon numerical techniques. In general, the usual log-linear hypotheses on $\underset{\sim}{\pi}$ (hypotheses such as $C\underset{\sim}{\pi} = \underset{\sim}{0}$ or $C[\log(\underset{\sim}{\pi})] = \underset{\sim}{0}$, where C is a contrast matrix, i.e., $C'\underset{\sim}{1} = \underset{\sim}{0}$) will impose complicated functional relationships among the $\gamma(\underset{\sim}{i}|\underset{\sim}{i}')$ and the $\gamma(\underset{\sim}{i}')$. The MLE's will need to be obtained by some numerical computer subroutines.

The practical approach is to utilize the estimator $\hat{\underset{\sim}{\pi}}$ and the consistent estimator of its variance matrix, $\hat{V}(\hat{\underset{\sim}{\pi}})$, as initial input to the asymptotically equivalent least squares procedures presented in Grizzle et al. (1969) and Forthofer and Koch (1973). This is discussed in greater detail in the final section of this appendix where a convenient technique is given for implementing the Maximum Likelihood approach at the first stage and then proceeding with the Weighted Least Squares approach in the final stage using a single computer program.

Inference Based on Least Squares Estimators (LSE) of the $\pi(\underset{\sim}{i})$

Before discussing the Least Squares approach (which will resemble to some extent that in Koch et al., 1972), additional notation is required. Let $p(\underset{\sim}{i}') = n(\underset{\sim}{i}')/n_1$, $p(\underset{\sim}{i},\underset{\sim}{i}') = n(\underset{\sim}{i},\underset{\sim}{i}')/n_2$, and $\underset{\sim}{p}_1$ be the vector whose elements are all $p(\underset{\sim}{i}')$. Similarly, let $\underset{\sim}{p}_2$ be the vector of length k'k obtained by stretching out all the $p(\underset{\sim}{i},\underset{\sim}{i}')$ in order. Finally, let $\underset{\sim}{\gamma}_i = E(\underset{\sim}{p}_i)$, i = 1,2, and denote $\underset{\sim}{p} = (\underset{\sim}{p}_1',\underset{\sim}{p}_2')'$.

The dispersion matrix of $\underset{\sim}{p}$ is a block diagonal matrix $V(\underset{\sim}{p})$ with $V(\underset{\sim}{p}_i)$ on the diagonal, where

$$V(\underset{\sim}{p}_i) = \frac{1}{n_i}\left[D(\underset{\sim}{\gamma}_i) - \underset{\sim}{\gamma}_i\underset{\sim}{\gamma}_i'\right] \qquad i = 1,2 . \tag{A.9}$$

Next let $\underset{\sim}{F} = A\underset{\sim}{p}$, where

$$A = \left[\begin{array}{c|c} A_1 : (k'-1) \times k' & O \\ \hline O & A_2 : (k'k-1) \times k'k \end{array}\right] \tag{A.10}$$

where A_1 is obtained from an identity matrix of dimension k' by deleting the last row and A_2 is similarly obtained from an identity matrix of dimension k'k. We can now write a model

$$E(\underset{\sim}{F}) = X\underset{\sim}{\theta} \qquad \underset{\sim}{\theta} : (k'k-1) \times 1 . \tag{A.11}$$

where $X = (X_1', X_2')'$ with X_2 being an identity matrix of order k'k-1 and X_1 of dimension $(k'-1) \times (k'k-1)$ has the form $X_1 = [I \otimes \underset{\sim}{1}' | 0]$ where I is the identity matrix of order k'-1, $\underset{\sim}{1}' = (1,\ldots,1)$ and $\otimes$ denotes Kronecker's product.

The variance matrix of $\underset{\sim}{F}$ is consistently estimated by $\hat{V}(\underset{\sim}{F}) = A\hat{V}(\underset{\sim}{p})A'$ where $\hat{V}(\underset{\sim}{p})$ is obtained by substituting the unrestricted MLE's of the $\underset{\sim}{V}_i$ in the expression for $V(\underset{\sim}{p})$. Thus, in large samples, one may use Weighted Least Squares to estimate the vector $\underset{\sim}{\theta}$. The asymptotic LSE of $\underset{\sim}{\theta}$ (which is BAN if (A.11) holds) is given by

$$\hat{\underset{\sim}{\theta}} = (X'\hat{V}^{-1}(\underset{\sim}{F})X)^{-1}X'\hat{V}^{-1}(\underset{\sim}{F})\underset{\sim}{F} \tag{A.12}$$

and the consistent estimator of its dispersion matrix $V(\hat{\underset{\sim}{\theta}})$ is given by

$$\hat{V}(\hat{\underset{\sim}{\theta}}) = (X'\hat{V}^{-1}(\underset{\sim}{F})X)^{-1} . \tag{A.13}$$

A test for goodness of fit is based on

$$X^2 = \underset{\sim}{F}'\hat{V}^{-1}(\underset{\sim}{F})\underset{\sim}{F} - \hat{\underset{\sim}{\theta}}'X'\hat{V}^{-1}(\underset{\sim}{F})X\hat{\underset{\sim}{\theta}} \qquad (A.14)$$

which, under the hypothesis that the model fits, follows an asymptotic Chi-square distribution with (k'-1) d.f.

If the model adequately describes the data, tests of hypotheses with respect to the parameters comprising $\underset{\sim}{\theta}$ can be undertaken. Note that the elements of $\underset{\sim}{\theta}$ are the k'k-1 upper-left elements among the k'k parameters $\gamma(\underset{\sim}{i},\underset{\sim}{i}')$. The last element is obtained from the relation $\Sigma_{\underset{\sim}{i}}\Sigma_{\underset{\sim}{i}'}\gamma(\underset{\sim}{i},\underset{\sim}{i}') = 1$. From $\hat{\underset{\sim}{\theta}}$ and its estimated variance matrix, one can easily obtain the LSE $(\hat{\underset{\sim}{\pi}})$ of $\underset{\sim}{\pi}$ and its estimated variance matrix $\hat{V}(\hat{\underset{\sim}{\pi}})$.

Employing the Maximum Likelihood Approach

Here we use notations from both of the two previous sections. The MLE's of the $\pi(\underset{\sim}{i})$ and their asymptotic variance matrix have already been given. The overall procedure of first obtaining MLE's and then using asymptotic Least Squares theory appears somewhat inconvenient, especially when considering the available computer programs. Here, we discuss a simple technique to implement the MLE methodology, which can be employed using a single computer program. This approach is based on the fact that the MLE's of the $\pi(\underset{\sim}{i})$ can be expressed as compound exponential-logarithmic-linear functions (see Forthofer and Koch, 1973) of the elements of $\underset{\sim}{p}$.

Specifically, we can write $\hat{\underset{\sim}{\pi}}$ (the MLE of $\underset{\sim}{\pi}$) as

$$\hat{\underset{\sim}{\pi}} = Q\left\{\underset{\sim}{\exp}[K\log(A\underset{\sim}{p})]\right\} \qquad (A.15)$$

where

$$\underset{(2+k)k' \times (k+1)k'}{A} = \left[\begin{array}{c:c} a_1 I : k' \times k' & a_2 I \otimes \underset{\sim}{1}' : k' \times k'k \\ \hdashline 0 : k' \times k' & I \otimes \underset{\sim}{1} : k' \times k'k \\ \hdashline 0 : k'k \times k' & I : k'k \times k'k \end{array}\right]$$

$$a_i = n_i/(n_1 + n_2),\ i = 1,2$$

$$\underset{(k'k) \times k'(k+2)}{K} = [I \otimes \underset{\sim}{1},\ -I \otimes \underset{\sim}{1},\ I : k'k \times k'k]$$

where the unspecified identity matrix I has dimension k' and $\underset{\sim}{1}$ is of length k.

$$\underset{k \times k'k}{Q} = I \otimes \underset{\sim}{1}' \ .$$

Thus, on letting $\underset{\sim}{y} = A\underset{\sim}{p}$ and $\underset{\sim}{z} = \underset{\sim}{\exp}[K\log(\underset{\sim}{y})]$, we can conveniently write the asymptotic variance matrix of $\hat{\underset{\sim}{\pi}}$ as

$$V(\hat{\underset{\sim}{\pi}}) = QD(\underset{\sim}{z})KD^{-1}(\underset{\sim}{y})AV(\underset{\sim}{p})A'D^{-1}(\underset{\sim}{y})K'D(\underset{\sim}{z})Q' \ , \qquad (A.16)$$

where $D(\underset{\sim}{y})$ is a diagonal matrix with the vector $\underset{\sim}{y}$ on the main diagonal.

As noted earlier, the vector $\hat{\underset{\sim}{\pi}}$ and the estimated variance matrix $\hat{V}(\hat{\underset{\sim}{\pi}})$ (which is obtained by substituting $\hat{V}(\underset{\sim}{p})$ for $V(\underset{\sim}{p})$) are subsequently used as initial inputs for further modeling based on Weighted (asymptotic) Least Squares procedures as in Grizzle et al. (1969). Thus, one may obtain functions of the $\hat{\underset{\sim}{\pi}}$ which are of interest for further modeling via a repeated chain of linear, log or exponential transformations, and then express a linear model for the resulting functions. The model can be tested for fit; given an adequate fit, linear hypotheses regarding its estimable parameters can be tested. And this entire procedure can be carried out in a single computer run using the new program GENCAT given in Landis and Stanish (1975).

APPENDIX B

The Supplementary (Telephone and Hospital) Data

Table B.1 Belt usage by injury level for supplementary sample, controlling for age.

Non-Police \ Police		16-55 No Injury U	16-55 No Injury L	16-55 No Injury LS	16-55 Injury U	16-55 Injury L	16-55 Injury LS	16-55 Total	56+ No Injury U	56+ No Injury L	56+ No Injury LS	56+ Injury U	56+ Injury L	56+ Injury LS	56+ Total
No Injury (Telephone)	U	1035	12	1	24	1	1	1074	159	2	0	4	1	0	166
	L	122	122	14	5	3	0	266	21	11	1	0	0	0	33
	LS	63	24	69	0	1	1	158	10	6	8	1	0	0	25
Injury (Hospital)	U	210	4	0	279	3	2	498	17	1	0	27	0	0	45
	L	26	19	3	8	18	4	78	2	0	0	3	3	0	8
	LS	6	1	9	5	3	9	33	3	0	1	0	0	0	4
Total		1462	182	96	321	29	17	2107	212	20	10	35	4	0	281

No. missing observations = 13

Table B.2 Belt usage by injury level for supplementary sample, controlling for sex.

Non-Police \ Police		Male							Female						
		No Injury			Injury			Total	No Injury			Injury			Total
		U	L	LS	U	L	LS		U	L	LS	U	L	LS	
No Injury (Telephone)	U	813	10	1	18	1	0	843	381	4	0	9	0	1	395
	L	106	102	12	3	3	0	226	36	32	3	2	0	0	73
	LS	57	24	59	1	1	0	142	16	6	18	0	1	1	42
Injury (Hospital)	U	121	3	0	173	2	2	301	106	2	0	133	2	0	243
	L	19	12	2	8	14	3	58	10	7	1	2	7	1	28
	LS	6	0	6	3	3	6	24	3	1	4	2	0	3	13
Total		1122	151	80	206	24	11	1594	552	52	26	148	10	6	794

No. missing observations = 13

Table B.3 Belt usage by injury level for supplementary sample, controlling for "make" of car.

Police / Non-Police		U.S.							Foreign						
		No Injury			Injury			Total	No Injury			Injury			Total
		U	L	LS	U	L	LS		U	L	LS	U	L	LS	
No Injury (Telephone)	U	1086	13	0	23	2	1	1125	101	1	1	5	0	0	108
	L	130	120	14	5	3	0	272	11	13	1	1	0	0	26
	LS	52	24	55	1	0	1	133	21	6	21	0	1	0	49
Injury (Hospital)	U	197	4	0	273	3	1	478	29	1	0	32	0	1	63
	L	26	18	1	7	20	3	75	2	1	2	3	1	1	10
	LS	7	1	6	3	3	7	27	2	0	4	2	0	2	10
Total		1498	180	76	312	31	13	2110	166	22	29	43	2	4	266

No. missing observations = 26

Table B.4 Belt usage by injury level for supplementary sample, controlling for model year.

Non-Police \ Police		1960-1968							1969-1971						
		No Injury			Injury			Total	No Injury			Injury			Total
		U	L	LS	U	L	LS		U	L	LS	U	L	LS	
No Injury (Telephone)	U	399	1	0	10	0	0	410	368	5	1	6	1	1	382
	L	46	21	0	3	1	0	71	36	40	3	2	1	0	82
	LS	5	1	0	0	0	0	6	14	9	9	1	0	0	33
Injury (Hospital)	U	84	2	0	117	1	0	204	64	2	0	87	1	0	154
	L	8	5	1	4	5	0	23	7	7	1	2	8	0	25
	LS	0	0	0	1	0	0	1	0	0	0	2	1	1	4
Total		542	30	1	135	7	0	715	489	63	14	100	12	2	680

Table B.4 Continued.

Non-Police \ Police		1972-1973 No Injury U	1972-1973 No Injury L	1972-1973 No Injury LS	1972-1973 Injury U	1972-1973 Injury L	1972-1973 Injury LS	1972-1973 Total	1974-1975 No Injury U	1974-1975 No Injury L	1974-1975 No Injury LS	1974-1975 Injury U	1974-1975 Injury L	1974-1975 Injury LS	1974-1975 Total
No Injury (Telephone)	U	296	4	0	8	0	0	308	129	5	0	3	1	0	138
	L	51	58	8	0	1	0	118	9	13	2	1	0	0	25
	LS	10	6	11	0	0	0	27	43	14	57	0	1	1	116
Injury (Hospital)	U	54	0	0	60	1	1	116	25	1	0	35	0	1	62
	L	11	6	1	3	7	2	30	2	1	0	1	2	2	8
	LS	2	1	1	1	1	0	6	7	0	9	1	1	8	26
Total		424	75	21	72	10	3	605	215	34	68	41	5	12	375

No. missing observations = 24

Table B.5 Belt usage by injury level for supplementary sample, controlling for vehicle damage severity (TAD).*

Non-Police \ Police		Minor							Severe						
		No Injury			Injury			Total	No Injury			Injury			Total
		U	L	LS	U	L	LS		U	L	LS	U	L	LS	
No Injury (Telephone)	U	613	6	0	9	0	0	628	401	6	1	17	1	0	426
	L	79	64	6	1	1	0	151	27	40	6	2	2	0	77
	LS	35	14	33	1	0	1	84	24	11	33	0	1	0	69
Injury (Hospital)	U	82	1	0	61	2	0	146	112	2	0	197	1	1	313
	L	12	7	0	4	2	2	27	12	8	2	6	15	1	44
	LS	3	0	4	1	1	3	12	5	1	6	3	2	6	23
Total		824	92	43	77	6	6	1048	581	68	48	225	22	8	952

* Minor = FR-LF,1; OTHER, 1-3
Severe = FR-LF, 2-7; OTHER, 4-7

No. missing observations = 399

Table B.6 Belt usage by injury level for supplementary sample, controlling for accident type.

Non-Police \ Police		Collision							Non-Collision						
		No Injury			Injury			Total	No Injury			Injury			Total
		U	L	LS	U	L	LS		U	L	LS	U	L	LS	
No Injury (Telephone)	U	988	11	1	18	1	1	1020	197	3	0	10	1	0	211
	L	120	113	10	1	1	0	245	20	20	4	4	1	0	49
	LS	65	23	56	1	1	0	146	6	7	20	0	0	1	34
Injury (Hospital)	U	185	4	0	198	3	1	391	41	1	0	108	0	1	151
	L	25	16	1	9	14	3	68	3	3	2	1	7	1	17
	LS	5	1	6	4	1	7	24	3	0	3	1	1	2	10
Total		1388	168	74	231	21	12	1894	270	34	29	124	10	5	472

No. missing observations = 30

APPENDIX C

Accident Report Information
Standard North Carolina Accident Report Form

The following variables from the police accident report form (shown in this Appendix) were utilized in the analysis:

1. Vehicle # (as assigned by police agency)

2. Month of Accident (January - June)

3. Day of Week

4. Hour of Day (e.g., 8:00 - 8:59 a.m.)

5. Accident Type (e.g., collision of motor vehicle in road with pedestrian)

6. Driver's (or Injured Passenger's) Year of Birth (e.g., 1952)

7. Driver's (or Injured Passenger's) Sex

8. Driver's (or Injured Passenger's) Race (white, non-white)

9. Vehicle Year (e.g., 1971)

10. Vehicle Make (e.g., Plymouth)

11. Vehicle Type (e.g., two or four door sedan (passenger vehicle), stations wagon (passenger))

12. V.I.N. (Vehicle Identification Number)

13. First TAD (location and severity; e.g., FD3 = front distributed of relative severity 3)

14. Police Reported Injury (Injuries for driver or injured passenger)

    K - Killed
    A - Serious injury
    B - Moderate injury
    C - Minor injury
    O - No injury

15. Restraint Used (the individual being coded) as Reported by Police

    None
    Lap Belt
    Shoulder Belt Only
    Lap and Shoulder Belt
    Child Restraint System
    Not Recorded

In filling out these items on the back use the following examples:

1. LOCALITY
   1. Business
   2. Residential
   3. School & playground
   4. Open country
2. SPEED LIMIT
3. ROAD FEATURE
   1. Bridge or underpass
   2. Driveway
   3. Alley intersection
   4. Intersection of two roadways
   5. Non-intersection median crossover
   6. End or beginning of divided highway
   7. Other
4. ROAD SURFACE
   1. Concrete
   2. Smooth asphalt
   3. Coarse asphalt
   4. Gravel
   5. Dirt or sand
   6. Other
5. ROAD DEFECTS
   1. Loose material on surface
   2. Holes, deep ruts
   3. Low shoulders
   4. Soft shoulders
   5. Other defects
   6. Road under construction
   7. No defects
6. ROAD CONDITION
   1. Dry
   2. Wet
   3. Oily
   4. Muddy
   5. Snowy
   6. Icy
7. LIGHT CONDITION
   1. Daylight
   2. Dusk
   3. Dawn
   4. Darkness (street lighted)
   5. Darkness (street not lighted)
8. WEATHER
   1. Clear
   2. Cloudy
   3. Raining
   4. Snowing
   5. Fog
   6. Sleet or hail
9. TRAFFIC CONTROL
   1. Stop sign
   2. Yield sign
   3. Stop and go signal
   4. Flashing signal with stop sign
   5. Flashing signal without stop sign
   6. R. R. gate and flasher
   7. R. R. flasher
   8. Officer
   9. Other device
   10. No control present
10. OBJECT STRUCK (first only)
   1. Tree
   2. Utility pole
   3. Fence or fence post
   4. Guardrail or guardpost in median
   5. Guardrail or guardpost on shoulder
   6. Bridge
   7. Underpass
   8. Traffic island, curb, or median
   9. Sign or sign post
   10. Animal
   11. Ditch bank
   12. Parked vehicle
   13. Pedestrian
   14. Other object
   15. None
11. SOBRIETY
   1. Had not been drinking
   2. Drinking-ability impaired
   3. Drinking-unable to determine impairment
   4. Unknown
12. PHYSICAL CONDITION
   1. Ill
   2. Fatigued
   3. Asleep
   4. Other physical impairment
   5. Restriction not complied with
   6. Normal
   7. Condition not known
13. CHEMICAL TEST
14. PEDESTRIAN ACTION
   1. Crossing at intersection
   2. Crossing not at intersection
   3. Coming from behind parked vehicle
   4. Walking with traffic
   5. Walking against traffic
   6. Getting on or off vehicle
   7. Standing in road
   8. Working in road
   9. Playing in road
   10. Lying in road
   11. Other
   12. Not in road
15. VEHICLE MANEUVER
   1. Stopped in travel lane
   2. Parked out of travel lanes
   3. Parked in travel lanes
   4. Going straight ahead
   5. Changing lanes or merging
   6. Passing
   7. Making right turn
   8. Making left turn
   9. Making U turn
   10. Backing
   11. Slowing or stopping
   12. Starting in roadway
   13. Parking
   14. Leaving parked position
   15. Other
16. VEHICLE DEFECTS (List one or more)
   1. Defective brakes
   2. Defective headlights
   3. Defective rear lights
   4. Defective steering
   5. Defective tires
   6. Other defects
   7. Not known if defective
   8. No defects detected

TRAFFIC ACCIDENT REPORT

N. C. Department of Motor Vehicles DMV-349 (Rev. 1-1-73)

ADDED BY ______ (Initial)

MARKS

LOCATION

Date of Accident ______ 19 ___ Day of Week ______ Hour ______ A.M. ☐ P.M. ☐

Do not write in this space

Accident Occurred In ______ County ☐ In ☐ Near City or Town of ______

Outside City or Town ______ Miles ☐ N ☐ E ☐ S ☐ W of ☐ Limits ☐ Center

Patrol Area

On ______ Hwy. No. (I., U.S., N.C., R.P., R.U.) If No., or within corporate limits, identify by name

______ ☐ Miles ☐ Feet (0 Ft. if Intersec.) ☐ N ☐ E ☐ S ☐ W At or From ______ Hwy. No., or Adjacent County Line Toward ______ Hwy. No., City, or Adjacent County Line

ACCIDENT TYPE

Ran off Road			Non-Collision in Road		Collision of Motor Vehicle in Road With:						
1. Right	2. Left	3. Straight Ahead	4. Overturn	5. Other in Road	6. Pedestrian	7. Parked Vehicle	8. Train	9. Bicycle	10. Animal	11. Fixed Obj.	12. Other Obj.

Collision of M.V. in Road With Another M. V.

13. Rear End Slow or Stop	14. Rear End Turn	15. Left Turn Same Roadway	16. Left Turn Cross Traffic	17. Right Turn Same Roadway	18. Right Turn Cross Traffic	19. Head On	20. Sideswipe	21. Angle	22. Backing

No. of Vehicles Involved ☐

VEHICLE NO. 1

Driver: ______ First ______ Middle ______ Last Name

Address: ______

City: ______ State: ______

Is above address same as on Driver's License? Yes ☐ No ☐

Race/Sex: ______ Driver's Lic: ______ State: ______

Date of Birth: ______ Month Day Year Specify Restriction: ______

Member of Armed Forces Yes ☐ No. ☐ Veh. Year: ______ Veh. Make: ______ Veh. Type: ______

Lic. Plate No. ______ State: ______ Year: ______

VIN ______ ODOM. __ __ __ __ __ . __

Owner: ______

Address: ______

City: ______ State: ______

Parts Damaged (TAD) ______ Amount of Damage $ ______

Drivable: Yes ☐ No ☐ Vehicle Removed to: ______

By: ______ Authority: ______

VEHICLE NO. 2 or PEDESTRIAN

Driver: ______ First ______ Middle ______ Last Name

Address: ______

City: ______ State: ______

Is above address same as on Driver's License? Yes ☐ No ☐

Race/Sex: ______ Driver's Lic. ______ State: ______

Date of Birth: ______ Month Day Year Specify Restriction: ______

Member of Armed Forces Yes ☐ No. ☐ Veh. Year: ______ Veh. Make: ______ Veh. Type: ______

Lic. Plate No. ______ State ______ Year: ______

VIN ______ ODOM. __ __ __ __ __ . __

Owner: ______

Address: ______

City: ______ State: ______

Parts Damaged (TAD) ______ Amount of Damage $ ______

Drivable: Yes ☐ No ☐ Vehicle Removed to: ______

By: ______ Authority: ______

Other Property Damaged ______ Amt. of Dam. $ ______ Owner and Address ______

INJURY SECTION INSTRUCTIONS

Give injury class, restraint used, race, sex and age of all occupants in the space corresponding to the seat occupied. Names and addresses are necessary for persons who were injured. For type of Restraint (Res.) used: N – None, L – Lap Belt, LS – Lap and Shoulder, S – Shoulder Belt only, YR – Child Restraint System.

K=Killed	A=Incapacitating	B=Nonincapacitating – Injury other than K or A evident at the scene	C=No visible sign of injury but complaint of pain, momentary unconsciousness	O=No injury

SEAT	Inj cl	Res usd	Race sex	Age	INJURED NAMES AND ADDRESSES First Name Last	SEAT	Inj cl	Res usd	Race sex	Age	INJURED NAMES AND ADDRESSES First Name Last
Left Front					DRIVER 1	Left Front					DRIVER 2 OR PEDESTRIAN
Center Front						Center Front					
Right Front						Right Front					
Left Rear						Left Rear					
Center Rear						Center Rear					
Right Rear						Right Rear					
Total No. Occupants					Total No. Inj.	Total No. Occupants					Total No. Inj.

Injured taken to: ______

WIT- NESSES Name ______ Address ______ Phone No. ______

Name ______ Address ______ Phone No. ______

Arrests: Name ______ Charge(s) ______ (Cit. No.) ______

Name ______ Charge(s) ______ (Cit. No.) ______

Sign Here ______ Officer's Rank and Name ______ Number ______ Department ______ Date of Report

In filling out these items on the back use the following examples:

1. LOCALITY
1. Business
2. Residential
3. School & playground
4. Open country

2. SPEED LIMIT

3. ROAD FEATURE
1. Bridge or underpass
2. Driveway
3. Alley intersection
4. Intersection of two roadways
5. Non-intersection median crossover
6. End or beginning of divided highway
7. Other

4. ROAD SURFACE
1. Concrete
2. Smooth asphalt
3. Coarse asphalt
4. Gravel
5. Dirt or sand
6. Other

5. ROAD DEFECTS
1. Loose material on surface
2. Holes, deep ruts
3. Low shoulders
4. Soft shoulders
5. Other defects
6. Road under construction
7. No defects

6. ROAD CONDITION
1. Dry
2. Wet
3. Oily
4. Muddy
5. Snowy
6. Icy

7. LIGHT CONDITION
1. Daylight
2. Dusk
3. Dawn
4. Darkness (street lighted)
5. Darkness (street not lighted)

8. WEATHER
1. Clear
2. Cloudy
3. Raining
4. Snowing
5. Fog
6. Sleet or hail

9. TRAFFIC CONTROL
1. Stop sign
2. Yield sign
3. Stop and go signal
4. Flashing signal with stop sign
5. Flashing signal without stop sign
6. R. R. gate and flasher
7. R. R. flasher
8. Officer
9. Other device
10. No control present

10. OBJECT STRUCK (first only)
1. Tree
2. Utility pole
3. Fence or fence post
4. Guardrail or guardpost in median
5. Guardrail or guardpost on shoulder
6. Bridge
7. Underpass
8. Traffic island, curb, or median
9. Sign or sign post
10. Animal
11. Ditch bank
12. Parked vehicle
13. Pedestrian
14. Other object
15. None

11. SOBRIETY
1. Had not been drinking
2. Drinking-ability impaired
3. Drinking-unable to determine impairment
4. Unknown

12. PHYSICAL CONDITION
1. Ill
2. Fatigued
3. Asleep
4. Other physical impairment
5. Restriction not complied with
6. Normal
7. Condition not known

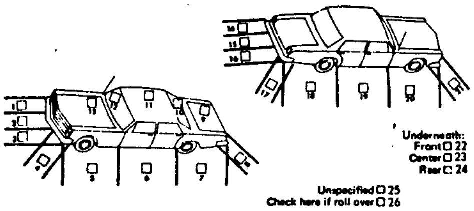
13. CHEMICAL TEST

14. PEDESTRIAN ACTION
1. Crossing at intersection
2. Crossing not at intersection
3. Coming from behind parked vehicle
4. Walking with traffic
5. Walking against traffic
6. Getting on or off vehicle
7. Standing in road
8. Working in road
9. Playing in road
10. Lying in road
11. Other
12. Not in road

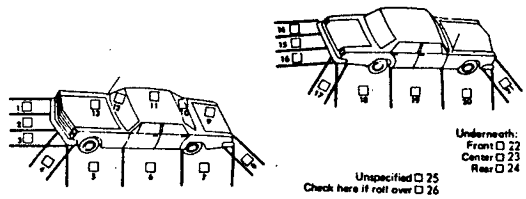
15. VEHICLE MANEUVER
1. Stopped in travel lane
2. Parked out of travel lanes
3. Parked in travel lanes
4. Going straight ahead
5. Changing lanes or merging
6. Passing
7. Making right turn
8. Making left turn
9. Making U turn
10. Backing
11. Slowing or stopping
12. Starting in roadway
13. Parking
14. Leaving parked position
15. Other

16. VEHICLE DEFECTS (List one or more)
1. Defective brakes
2. Defective headlights
3. Defective rear lights
4. Defective steering
5. Defective tires
6. Other defects
7. Not known if defective
8. No defects detected

VEHICLE 1 POINT OF INITIAL CONTACT



VEHICLE 2 POINT OF INITIAL CONTACT



					VEHICLE 1	VEHICLE 2
1. Locality	9. Traffic Control	Not Operating ☐	Not Visible ☐			
2. Speed Limit	10. Object Struck			15. Veh. Maneuver		
3. Road Feature		DRIVER 1	DRIVER 2 or PED.	16. Veh. Defects		
4. Road Surface	11. Sobriety			17. Estimated Speed		
5. Road Defects	12. Physical Cond.			18. Tire Impressions (ft)		
6. Road Condition	13. Chem. Test	YES ☐ NO ☐	YES ☐ NO ☐	19. Distance Traveled After Impact (ft.)		
7. Light Condition						
8. Weather	14. Ped. Action					

INDICATE NORTH

Vehicle 1 was Traveling ☐ N ☐ E ☐ S ☐ W on

Vehicle 2 was Traveling ☐ N ☐ E ☐ S ☐ W on

DESCRIBE WHAT HAPPENED:

Vehicle 1 2 — VIOLATION INDICATED

- ☐ ☐ 1. No. Violation Indicated
- ☐ ☐ 2. Excessive Speed
- ☐ ☐ 3. Yield Violation
- ☐ ☐ 4. Left of Center
- ☐ ☐ 5. Passing Violation
- ☐ ☐ 6. Stop S. or Yield S. Vio.
- ☐ ☐ 7. Traffic Signal Vio.
- ☐ ☐ 8. Safe Movement Vio.
- ☐ ☐ 9. Too Close
- ☐ ☐ 10. Improper Turn
- ☐ ☐ 11. Improper or No Signal
- ☐ ☐ 12. Improper Parking Location
- ☐ ☐ 13. Other Improper Driving

(describe) ____

EMERGENCY ASSISTANCE INFORMATION

INVESTIGATOR NOTIFIED BY ____ ☐ a.m. ☐ p.m.

INVESTIGATOR ARRIVED ____ ☐ a.m. ☐ p.m.

AMBULANCE ARRIVED ____ ☐ a.m. ☐ p.m.

OTHER COMMENTS: ____

RESERVED FOR STATE USE:

20.	21.	22.	23.	24.
25.	26.	27.	28.	29.

RESERVED FOR CITY OR OTHER USE:

APPENDIX D

Telephone Interview: Introduction
Format, Questionnaire

TELEPHONE INTRODUCTION FORMAT

Hello M. ____________________, my name is ____________________ and I am with the University of North Carolina Highway Safety Research Center. The Department of Transportation in Washington, D.C. is continually trying to learn more about seat belt usage and corresponding effectiveness in reducing deaths and injuries in highway crashes. In this connection, we are doing a survey of North Carolina drivers who were recently involved in a traffic accident, primarily to find out how they feel about seat belts in general and whether the seat belt might have helped (or hindered) in the accident in question. Would you mind answering a few brief questions? Thank you.

*　　*　　*　　*　　*

Note:

(1) If the person we need to talk with is <u>not at home</u>, try to find out when a good time to call back and reach him would be. Very generally explain that your name again is ____________________ and that you work for the University of North Carolina. As part of a telephone survey, you are calling people to find out about automobile seat belts and their usage.

(2) If the interviewee wants to know more about HSRC, the explanation can be derived from the following:

> The University of North Carolina Highway Safety Research Center (HSRC) was created by a statute of the 1965 North Carolina General Assembly, and was directed by the Governor to perform three functions:

1) evaluate North Carolina's existing highway safety programs.

2) coordinate and participate in the professional training of persons involved in highway safety.

3) close the gap between knowledge created by highway safety research and its use in saving lives.

(3) If the person needs to know how we know about his accident, explain that all accidents are public record at the Department of Motor Vehicles (DMV) in Raleigh and being a research organization engaged in highway safety research, we often need to have access to these records.

(4) If the individual seems upset, suggest that he feel free to call HSRC collect at (919) 933-2202 and ask for Dr. Campbell or Dr. Reinfurt for further information about this survey.

NAME OF DRIVER ____

1. Was the vehicle you were in during your accident a passenger car ( ); truck ( )? Do you know the make, model and approximate year of the car? ______________________________

   If not, was the car a <u>large</u> car (Olds, Buick), <u>intermediate</u> (Chevelle), or <u>small</u> (Vega).

   (IF TRUCK, YOU ARE FINISHED)

2. Does your car (the one in the accident) have seat belts? Yes ( ) No ( )

   (IF NO, GO TO 5)

3. If so, what kind of belts?
   ( ) Lap
   ( ) Lap and shoulder
   ( ) Don't know
   ( ) Not sure about shoulder part

4. Were you wearing your seat belts?

   Lap only ( )
   Lap and shoulder ( )
   ( ) No belt
   ( ) Shoulder only
   ( ) Unknown or don't remember

   <u>For those who were wearing their shoulder belt</u>:

   Since you were wearing your shoulder belt, can you tell me if yours is the kind that allows you some freedom of movement while you're belted in? (If they need an explanation use turning on the radio or opening the glove compartment as illustrations of freedom of movement).

   Yes ( ) No ( ) Can't say or don't remember ( )

   In your accident did the shoulder belt hold you in place? In other words, did it "lock up" like it was supposed to?

   Yes ( ) No ( ) Can't say or don't remember ( )

   If they don't remember ask, "Did you feel like any part of your chest had been bruised or was sore after the accident from where the shoulder belt went across your chest?" Yes ( ) No ( )
   "Did your waist feel especially sore from the lap belt?" Yes ( ) No ( )

   Did you hit the steering wheel at that time?

   Yes ( ) No ( ) Can't say or don't remember ( )

5. Did the officer ask you if you were wearing a seat belt?

   Yes ( )   No ( )   Can't say or don't remember ( )

6. Were you injured?   Yes ( )   No ( )

   If YES, would you describe your injury as slight ( ), moderate ( ), severe ( )?

   Could you please describe where your injuries were and what types of injuries you had.

   ____

   ____

   ____

   ____

7. If you weren't really injured, can you recall if you had any aches or pains?

   Yes ( )   No ( )

   If yes, can you describe where?

   ____

   ____

   ____

   ____

APPENDIX E

Hospital Survey:
The Hospital Report Form

Highway Safety Research Center
University of North Carolina
Chapel Hill

AUTOMOBILE INJURY AND SEAT BELT DATA
HSR-006

Instructions: Please complete one form for each patient treated for injuries due to an automobile crash. Return the form to HSRC in the attached pre-addressed envelope. No stamp is necessary. If you have any questions please feel free to call collect Ms. Lucy Smith or Ms. Jane Youngblood at (919) 933-2202.

1. Patient's Name ______________________________
   First Middle Last

   Date of Birth ______________ Date Treated ______________

2. Patient's Address ______________________________
   Street or P.O. Box

   ____
   City State Zip Code

3. Safety Restraint Use:
   a. ☐ No Belt
   b. ☐ Lap Belt Only
   c. ☐ Both Lap and Shoulder Belt
   d. ☐ Unknown

Name and Title of
Person Completing Form: ______________________________

ABBREVIATED INJURY SCALE

SEVERITY CODE	SEVERITY CATEGORY/INJURY DESCRIPTION
0 (Zero)	NO INJURY

1	**MINOR**

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**

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)**

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

SEVERITY CODE	SEVERITY CATEGORY/INJURY DESCRIPTION
4	**SERIOUS (Life-Threatening, Survival Probable)**

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)**

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 intracranial pressure (decreasing state of consciousness, bradycardia 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
7	**SEVERITY UNKNOWN**

* 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.

4. Overall Severity of Injuries: (see p.2 for injury scale)

(1) ☐ Minor
(2) ☐ Moderate
(3) ☐ Severe, Not Life Threatening
(4) ☐ Serious, Life Threatening
(5) ☐ Critical
(6) ☐ Fatal
(7) ☐ Unknown

5. For <u>Each</u> Injury:

a. Indicate the <u>location</u> of the injury by marking on the drawing below.

b. Write the <u>degree</u> (e.g. major, slight, compound, 1-in. etc.) and <u>nature</u> (e.g. burise, laceration, abrasion, fracture, burn, internal injury, etc.) of this injury.

EXAMPLE