

Report 03169-III

ACQUISITION OF INFORMATION ON EXPOSURE AND ON NON-FATAL CRASHES

Volume III - Procedures for an Exposure Survey

Philip S. Carroll
Thomas L. McDole
William L. Carlson
Paul F. Samarco

Highway Safety Research Institute
The University of Michigan
Ann Arbor, Michigan 48104

May 12, 1971

Final Report

Prepared under contract FH-11-7293 for
National Highway Traffic Safety Administration
Department of Transportation
Washington, D.C. 20591

1. Report No. 03169-III		2. Government Accession No.		3. Recipient's Catalog No.	
4. Title and Subtitle Acquisition of Information on Exposure and on Non-Fatal Crashes Volume III - Procedures for an Exposure Survey				5. Report Date May 12, 1971	
				6. Performing Organization Code	
7. Author(s) P.S. Carroll, T.L. McDole, W.L. Carlson, P.S. Samarco				8. Performing Organization Report No.	
9. Performing Organization Name and Address Highway Safety Research Institute The University of Michigan Ann Arbor, Michigan 48104				10. Work Unit No.	
				11. Contract or Grant No. FH-11-7293	
				13. Type of Report and Period Covered	
12. Sponsoring Agency Name and Address National Highway Traffic Safety Admin. Washington, D.C. 20591				14. Sponsoring Agency Code FH-11	
15. Supplementary Notes The opinions, findings, and conclusions expressed in this publication are those of the authors, and not necessarily those of the National Highway Traffic Safety Administration.					
16. Abstract This volume presents detailed procedures for an annual, national survey of driving exposure, based on mailed questionnaires to a random sample of licensed drivers in the 50 states and District of Columbia. The total sample size is 28,000 and subsamples in each state are proportional to the number of licensed drivers. Each questionnaire requires information on driver age and sex, plus a one-day trip record indicating vehicle type, model year, road type, day or night, and trip mileage. Each trip is subsequently classified in one of 26 unique driver-vehicle-road-environment combinations. The sample size assures a minimum significant difference of approximately 30-40 percent between mileage estimates of any two classes. The procedure plans include details of sampling, data collection and analysis, covering a 22 month period: six months preparation, 12 months of data collection during the specified calendar year, and four months of follow-up, analysis and reporting. Required resources in the performing organization include about 14 man years of professional and clerical staff effort, office facilities and materials, and a large computer. The approximate program cost is estimated at \$253,000.					
17. Key Words			18. Distribution Statement NTIS For public distribution		
19. Security Classif. (of this report) none		20. Security Classif. (of this page) none		21. No. of Pages 67	22. Price

Preface

This is Volume III of the final report on Contract FH-11-7293, "Acquisition of Information on Exposure and on Non-Fatal Crashes." It covers Phase III of the contract (Driving Exposure Survey).

Specific details of procedures for a nationwide mail survey of exposure data are presented. The procedures include the sampling plan, the data collection plan, and the data analysis plan. The details presented are an extension of the recommendations given in Volume I for an exposure-survey field test program for the calendar year 1972.

TABLE OF CONTENTS

Section 1 - Introduction	1
Basic Plan	1
Sample Size	2
Section 2 - Summary	4
Detailed Exposure Survey Plan (Task 1)	4
Exposure Survey Instructions (Task 2)	5
Alternative Exposure Survey Methods (Task 3)	5
Exposure Data Analysis Plan (Task 4)	6
Section 3 - Detailed Exposure Survey Plan (Task 1)	7
Basic Plan of Survey Procedures	8
Liaison	8
Sample and Survey Design	9
Scheduling	17
Questionnaire	18
Data Management	26
Data Collection	28
Data Reduction	29
Scheduling, Resources, Costs	30
Section 4 - Exposure Survey Instructions (Task 2)	35
Liaison	35
Bureau of the Budget Approval	35
State Contacts	35
Sample and Survey Design	37
Generate Sample Size	37
Generate Survey Dates	37
Sample Generation Procedure	39
Matching Drivers to Days	40

Assigning Identification Numbers to State Sample	40
Building Master File	41
Prepare Mail Labels	41
Scheduling	41
Specific Survey Dates	41
Master Schedule	41
Questionnaire	41
Prior to Bureau of Budget Approval	41
Reproduction	43
Preparing Questionnaire for Original Mailing	43
Data Management System	43
Log Book	43
Storage	45
Flow Charts	45
Data Collection	45
Data Reduction	47
Section 5 - Alternative Exposure Survey Methods (Task 3)	49
Review of Candidate Methods	49
Selection Criteria	50
Recommended Alternative	52
Section 6 - Exposure Data Analysis Plan (Task 4)	53
Data Input	53
Computer Program	55
Tasks	55
Scheduling, Resources, Costs	57
Section 7 - Conclusions and Recommendations	58

LIST OF ILLUSTRATIONS

Figure 1 - Driver Identification Information to be Acquired in Each State	10
Figure 2 - Typical Questionnaire Cover Letter	20
Figure 3 - Typical Questionnaire	21
Figure 4 - Subject Code Scheme	27
Figure 5 - Suggested Outline for Telephone Contact and Follow-up Letter	36
Figure 6 - Example of Data Collection Schedule	38
Figure 7 - Survey Log Book Sheet	44
Figure 8 - Flow Chart: Questionnaire Preparation	46
Figure 9 - Flow Chart: Data Collection	46
Figure 10 - Flow Chart: Data Reduction	46

LIST OF TABLES

Table 1 - Sample Distribution of Drivers by State and Required Sample Sizes for Each "Day," "Week," and for the Year	12
Table 2 - Example of Random Survey Days and Dates	13
Table 3 - Example of Generation of Random Sample Matching Drivers to Days	16
Table 4 - Variables to Be Used in the Questionnaire	19
Table 5 - Basic Time Line of Survey Events	31
Table 6 - Staff Resources	32
Table 7 - Cost Estimate	34
Table 8 - Example: Assigning Identification Numbers	42
Table 9 - Recommended Exposure Classes	54
Table 10- Required Operations of Computer Program	56

SECTION 1 INTRODUCTION

This volume presents the results of Phase III - Driving Exposure Survey. Its purposes in accordance with the Statement of Work in Appendix B, are to prepare:

1. A detailed exposure survey plan of sampling and data collection procedures, including schedule, resources and costs.
2. Documentation of instructions needed by an organization to conduct the survey.
3. An alternate data-collection plan as a means of checking accuracy.
4. A plan for data analysis.

BASIC PLAN

The plans derived in this volume are based on the recommendation in Volume I for a national field-test exposure survey in 1972. It will use mailed questionnaires with one-day trip records to be completed by subsamples of licensed drivers in each state and the District of Columbia.

The survey plan requires random sampling of licensed drivers in order to produce accurate estimates of exposure at reasonable cost. Although the basic sampling unit is the driver, the basic measurement unit is the trip. Each trip is classified according to six variables: driver age and sex, vehicle type and model year, road type, and day/night. Sampled drivers may produce as many as six classes of trips on a one-day record, among the 26 classes defined in Volume I.

The conclusions of Volume I did not define a sampling plan in terms of the specific days of the year for which exposure should be measured. This decision depends largely on the maximum sample size allowed by cost constraints. Options range from 365 days a year (maximum accuracy) to a single random day (minimum accuracy).

The plan recommended for the first year field test program specifies 28 days, as explained below. The plan provides for representation of the year by season and day of week, and preserves randomness of day selection. It will yield moderate accuracy at reasonable cost. Also, it is a flexible model for extension to more sampling days, larger samples and greater accuracy if additional funds are available in later years.

SAMPLE SIZE

In Volume I, a sample size estimate of 25,000 was derived, based on a rough analysis of its ability to provide significant differences among the 26 recommended classes of exposure, i.e., driver-vehicle-road-environment combinations (see Figure 8 of Volume I). Further analysis indicated a selected sample of 28,000 where an anticipated cumulative response rate of 80% would produce 22,400 returned questionnaires.

In each of the four quarters of the year, seven days will be selected randomly, one each from the seven types of day of the week. Thus, there will be 28 specific sampling days in the year, constituting four representative "weeks"--one in each quarter. The mean mileage estimates for each sample day will be extrapolated and added to produce a total mean annual mileage. The mileage estimates within each day and the total will be broken down into 26 classes.

Because the mileage estimates are independent for each of the 28 days, the minimum sample size must be determined within the context of a single day's subsample. Because of the minimal exposure in a few of the 26 classes (see Table 7 of Volume I) it is estimated that only about 20 classes will be needed (very small classes will be combined with similar classes). It is assumed that the number of drivers (and trips) will be fairly uniformly distributed among the 20 classes (though this does not mean that the mileage will be uniformly distributed). Because of the asymptotic decrease in standard deviation in a class as its subsample

size increases, a subsample size of 25 is often chosen as a minimum with normal distributions. To account for non-normal distributions in some classes, and to ensure the equivalent of at least 25 in each, the required average class size selected is 40.

The 20 classes with average size of 40 require 800 returned questionnaires each day. On the basis of an 80% response, the subsample size each day is 1000. Thus, the total sample size, i.e. the number of drivers to receive questionnaires, is 28,000.

From the pilot survey of Volume I and other exposure studies, it has been determined that the standard deviations of mileage estimates in almost all classes tend to have values within about 10% of the mean mileage. Using this fact, and class sizes of 40, the minimum significant difference between mean values for typical classes is found to be from 30-40% of the means. This is satisfactory in view of the typical differences of 50-100% between adjacent classes, as determined from the AID charts of Volume I.

SECTION 2

SUMMARY

This volume presents the requirements, approaches and findings pertaining to each of the four tasks of Phase III - Driving Exposure Survey (see contract work statement, Appendix A). Summaries of each of the tasks are given under the headings below.

DETAILED EXPOSURE SURVEY PLAN (Task 1)

The requirement of this task is to "prepare a detailed sampling plan and procedures for a nationwide driving exposure survey," including schedule, resources, and costs required. The approach was to begin with recommendations in Volume I for a field-test exposure survey using mailed questionnaires with trip-record forms as a base design. Then, the design details were expanded in a straightforward manner to an extent that would be sufficient for a competent survey organization that does not necessarily have expertise in the technical aspects of exposure.

Seven categories of procedures were identified: liaison, sample and survey design, scheduling, questionnaire, data management, data collection, and data reduction. These categories cover all of the work that must be done from the beginning of the survey program to the point that data is ready for analysis.

Liaison includes work with the Bureau of the Budget to secure their approval of the questionnaire and liaison with officials in each state to obtain access to driver lists for subsequent sampling. Sample construction includes distribution of the 28,000 sample among states according to driver population, randomly selecting the proper number of names in each state, randomly selecting the 28 sampling dates (one of each type of day of the week in each quarter of the year), matching name to dates (1000 each date), assigning identification numbers, and obtaining addresses. Scheduling considerations include lead time for planning and question-

naire approval and materials preparation, 13 months for data collection (final month for followups on December mailings), and a period for data analysis and reporting. Questionnaire development includes design of format and wording to elicit maximum response, writing of cover letters (initial and each of two reminders to be spaced two weeks apart), printing and storage. Data management involves log books, filing system and progress charts to keep track of all questionnaires. Data collection includes assembly and mailing of questionnaires, logging of receipts, identification of non-respondents, and reminder mailings. Data reduction includes checking and coding of individual questionnaires, keypunching, and building of a magnetic tape file.

The program is 22 months in duration. It requires about 14 man months of effort, office facilities and materials, and a large computer. The approximate cost is estimated at \$253,000.

EXPOSURE SURVEY INSTRUCTIONS (Task 2)

The requirement of this task is to "provide documentation so that the organization which actually conducts the survey will have all the instructions needed to perform the job." The results are simply a detailed outline of direct instructions, within the seven categories above. About 100 specific instructions are derived. Many of the instructions are in sequences that must be repeated for each of the 51 sampling areas. Examples of questionnaires, letters and charts are included.

ALTERNATIVE EXPOSURE SURVEY METHODS (Task 3)

The requirement of this task is to "determine means to check the resultant exposure estimates by alternative collection methods." Several alternative methods were reviewed, based on tradeoff data in Volume I. A national home interview method is recommended as an alternative plan, where exposure data is obtained as a gross estimate, i.e., a single value of mileage for a past time period, e.g., three months. This method is selected primarily because of

the potential low costs if it is attached to an existing national survey. The other alternatives considered (independent home interviews, license office interviews, independent mail surveys) would have to be limited in scope because of sampling costs, and they could be compared only with certain state subsamples from the primary survey.

EXPOSURE DATA ANALYSIS PLAN (Task 4)

The requirement of this task is to "prepare a detailed plan for analysis of the data to be collected," including resources and costs. The analysis plan would be an integral part of the total survey program, and would follow directly from the end of the plans prepared in tasks 1 and 2, i.e., completion of the magnetic tape data file.

From the 28,000 drivers in the sample, an 80% response would yield about 22,400 data cases, sorted in the data file according to the 28 survey dates. Each data case includes a number of trip-mileage estimates, with six independent variable levels as classifiers. The purpose of the data analysis is to compute, from the data file, estimates of mean and total national mileages, in aggregate and in 26 driver-vehicle-road-environment classes.

The initial part of the analysis plan is to prepare a computer program to produce the required exposure statistics. The minimum required computer operations are derived. Additional computer programs may be derived for further classifications of the data (e.g., by quarter of year or by state), for re-evaluation of the 26 classes, or for accuracy checking by an alternative data source. In addition to writing, checking and testing the computer programs, tasks of the data analysis plan include performance of computer runs, comparisons and interpretations of results, and final tabulations.

The basic data analysis will be performed during the three months following the end of data collection, and will require approximately 10 man months of effort.

SECTION 3
DETAILED EXPOSURE SURVEY PLAN (Task 1)

This section will discuss the sampling plans and survey procedures for a nationwide driving exposure survey to be conducted by a qualified survey organization. The goal of the survey described here is to provide a reliable measure of the driving exposure for all drivers in the United States. As established in Volume I, it is to be a mail survey of a random sample of licensed drivers in each of the 50 states and the District of Columbia, with basic vehicle-mileage estimation to be done by means of one-day trip records.

The subsample sizes of drivers to be surveyed will be proportional to the number of licensed drivers in each state and D.C. (51 basic sample areas), and thus the total sample will constitute a random selection which is representative of the registered driving population in the whole country. The target population nationwide consists of approximately 105,400,000 licensed drivers¹ of which approximately 28,000 are to be included in the survey (approximately .025%).

The sample drivers will be asked to respond to a mailed questionnaire, including information on vehicle type and model year, driver sex and age, and a trip record including information on road types, light conditions, numbers of trips and miles driven. Non-respondents to the initial mailing will be re-contacted and asked to respond in an effort to maximize final response rate.

The survey will be conducted over the period of one calendar year with the actual survey dates selected so as to provide a representative "week" for each of the four quarters of the year. Thus seven days will be selected at random to correspond to and be representative of each of the seven days of the week in a quarter

¹ 1970 Automobile Facts and Figures, Automobile Manufacturers Association.

(i.e., one of the 13 Sundays in a quarter would be randomly chosen, one of the 13 Mondays, etc.) providing a total of 7 days per quarter (one "week") or 28 representative days per year. Correspondingly, one quarter of the driver sample would be contacted in each quarter.

BASIC PLAN OF SURVEY PROCEDURES

The following paragraphs describe generally the major sets of procedures necessary to perform the sample construction and the data collection. Detailed instructions are covered later in section 4.

Liaison

Depending upon the wording of the contract from the National Highway Traffic Safety Administration, it may be necessary to seek Bureau of the Budget approval for the questionnaire before it can be used to collect data. Should this be the case, a complete copy of the final form of the questionnaire with all the enclosures and reminder letters should be prepared for transmittal via the proper channels to the Bureau of the Budget for approval. This approval can be expected to take 3-4 months. When approval is received, the approval notice must be added to the top of the questionnaire in the space provided on the prototype.

It will also be necessary to seek and receive permission from officials in each of the 51 sampling areas to gain access to their driver license records for the purpose of constructing the sample. As this process is lengthy and may be subject to much negotiation, it is recommended that the process of state level contacts begin early in the project.

A useful procedure to follow in initiating the contacts is to telephone the director (or administrator) of the Bureau of Motor Vehicles or similar office under whose jurisdiction the driver licensing function lies. The names of the appropriate persons can be found by consulting the AAMVA Directory (American Association of Motor Vehicle Administrators).

Telephone contact should then be followed by a letter stating the nature of the request, i.e., access to the driver license records for the purpose of selecting a random sample of drivers in the state. The information desired from the files should include the driver license number, name, address, sex, and date of birth for each selected driver. Figure 1 illustrates the desired information and format. An active campaign of correspondence and telephone contact should be undertaken to assure access to the files, with such activity culminating in a visit to the state at the time of the actual collection of the records. At or near the time of the initial contacts with the states, answers to the following questions must be received: (1) What type of record system does your state have? (manual, computer, etc.) (2) What type of driver license numbering system does your state employ? (alphabetical, soundex, etc.) (3) How many licensed drivers of all types are there in your state? (as accurately as possible) (4) Does one common numbering scheme serve all license types?

When access has been granted, it will be necessary to either (1) have the state generate the sample using guidelines furnished by the contractor or (2) send a representative to the state to perform the task.

Sample and Survey Design

The target population is all the licensed drivers in the United States (105,400,000 in 1968.) Since it is both impractical and unnecessary to ask every driver to participate, a sample of all the licensed drivers will be chosen on a random basis to represent all the drivers in each state. For the purposes of statistical analysis of the data it is estimated that a sample of about 28,000 drivers should be sufficient. The number of drivers to be selected for surveying in each state will be in proportion to the numbers of licensed drivers in that state such that the overall sample totals 28,000 drivers nationwide. For example, the 1967 distribution of drivers by state and the corresponding sample sizes

Driver License Number _____

Name _____

Address: Street _____ P.O. Box _____ Apt. _____ R.R. _____

City _____ State _____ Zip _____

Sex M F

Birthdate / /
 mo da yr

Our Case Number / / /

Figure 1

Driver Identification Information
to be Acquired in Each State

are shown in Table 1. The actual distribution of drivers to be used must reflect the driving population at the time of sampling as determined by questions asked at the time of the state level contacts. An up-to-date revision of Table 1 will yield the number of licensed drivers needed in each state subsample. The total sample will be representative of the nation as a whole with sufficient data to describe the 26 classes of exposure, but it will not be sufficient to yield state averages.

The survey is designed to yield two major sets of information: (1) Exposure information aggregated in 26 classes for an entire year (12 month period) for the United States and (2) Exposure information aggregated for each of the four quarters (3 month periods).

The representative week in each quarter will be generated by randomly selecting a day corresponding to each of the seven types of days of the week within each quarter. To construct a representative week, one Sunday will be selected at random from the 13 Sundays in the quarter and will be designated as the survey date to represent all Sundays in that quarter. Likewise a similar selection will be done for Monday, Tuesday, Wednesday, Thursday, Friday, and Saturday with the result being the representative week of the quarter. The procedure is then repeated for the remaining three quarters generating four weeks or 28 survey days. Typical results of the process are shown in Table 2.

One quarter of the number of drivers as shown in Table 1 for each state will be surveyed in each quarter. On each of the 28 survey days one twenty-eighth of the total national sample will be surveyed.

Since some non-response to the questionnaire is anticipated, additional follow-up dates are identified. These dates are the same day of the week but spaced at two week intervals. Thus, it is possible to have follow-up dates for a given quarter that actually occur in the subsequent quarter. Two such follow-up dates

TABLE 1 SAMPLE DISTRIBUTION OF DRIVERS BY STATE AND REQUIRED
SAMPLE SIZES FOR EACH "DAY", "WEEK", AND FOR THE YEAR

State	Lic. Drivers 1967 (000)	% Distrib.	Daily Sample Base 1000	Week of Quarter x7	Year x4
Alabama	1598	1.53	15	105	420
Alaska	115	.12	1	7	28
Arizona	964	.92	9	63	252
Arkansas	1013	.97	10	70	280
California	10688	10.26	104	728	2952
Colorado	1261	1.21	12	84	336
Connecticut	1898	1.82	18	126	504
Delaware	295	.28	3	21	84
Florida	3336	3.20	32	224	896
Georgia	2153	2.06	22	154	616
Hawaii	393	.38	4	28	112
Idaho	429	.41	4	28	112
Illinois	5801	5.56	57	399	1596
Indiana	2661	2.55	26	182	728
Iowa	1584	1.52	15	105	420
Kansas	1410	1.35	14	98	392
Kentucky	1442	1.38	14	98	392
Louisiana	1622	1.55	16	112	448
Maine	485	.46	4	28	112
Maryland	1867	1.79	18	126	504
Mass.	2791	2.68	27	189	756
Mich.	4514	4.33	44	308	1232
Minnesota	2074	1.99	20	140	560
Mississippi	972	.93	9	63	252
Missouri	2500	2.40	24	168	672
Montana	391	.38	4	28	112
Nebraska	918	.88	9	63	252
Nevada	338	.32	3	21	84
New Hamp.	371	.36	4	28	112
N.J.	3597	3.45	35	245	980
New Mexico	549	.53	5	35	140
New York	7903	7.58	76	532	2128
N. Carolina	2512	2.41	24	168	672
N. Dakota	333	.32	3	21	84
Ohio	5726	5.50	56	392	1568
Oklahoma	1465	1.41	14	98	392
Oregon	1100	1.07	11	77	358
Penn.	5913	5.68	57	399	1596
Rhode Is.	469	.46	5	35	140
S. Carolina	1244	1.20	12	84	336
S. Dakota	404	.40	4	28	112
Tenn.	2060	1.99	20	140	560
Texas	5601	5.39	54	378	1512
Utah	558	.55	6	42	168
Vermont	219	.22	2	14	56
Virginia	2230	2.15	22	154	616
Wash.	1705	1.64	17	119	476
W. Virginia	855	.83	8	56	224
Wisconsin	2280	2.19	22	154	616
Wyoming	221	.22	2	14	56
D.C.	344	.34	3	21	84
total	103172	99.12	1000	7000	28000

TABLE 2 EXAMPLE OF RANDOM SURVEY DAYS AND DATES

Quarter	Day	Sample Size Nat'l	Week of Quarter	Date 1971	1st Follow- up Date	2nd Follow- up Date	Cut Off Date
					2 weeks		
Winter	S	1000	5	Jan 31	Feb 14	Feb 28	Mar 14
Jan 1	M	1000	10	Mar 8	Mar 22	Apr 5	Apr 19
to	T	1000	5	Feb 2	Feb 16	Mar 2	Mar 16
March 31	W	1000	6	Feb 10	Feb 24	Mar 10	Mar 24
	T	1000	8	Feb 25	Mar 11	Mar 25	Apr 8
	F	1000	11	Mar 12	Mar 26	Apr 9	Apr 23
	S	1000	13	Mar 27	Apr 10	Apr 24	May 8
Spring	S	1000	8	May 23	June 6	June 20	July 4
Apr 1	M	1000	11	Jun 14	June 28	July 12	July 26
to	T	1000	1	Apr 6	Apr 20	May 4	May 18
June 30	W	1000	2	Apr 14	Apr 28	May 12	May 26
	T	1000	9	May 27	June 10	June 24	July 8
	F	1000	7	May 14	May 28	June 11	June 25
	S	1000	9	May 29	June 12	June 26	July 10
Summer	S	1000	2	July 11	July 25	Aug 8	Aug 22
July 1	M	1000	1	July 5	July 19	Aug 2	Aug 16
to	T	1000	5	Aug 3	Aug 17	Aug 31	Sept 14
Sept. 30	W	1000	13	Sept 27	Oct 11	Oct 25	Nov 8
	T	1000	2	July 8	July 22	Aug 5	Aug 19
	F	1000	6	Aug 6	Aug 20	Sept 3	Sept 17
	S	1000	1	July 3	July 17	July 31	Aug 14
Fall	S	1000	6	Nov 7	Nov 21	Dec 5	Dec 19
Oct 1	M	1000	6	Nov 8	Nov 22	Dec 6	Dec 20
to	T	1000	3	Oct 19	Nov 2	Nov 16	Nov 30
Dec 31	W	1000	2	Oct 13	Oct 27	Nov 10	Nov 24
	T	1000	11	Dec 16	Dec 30	Jan 8	Jan 22
	F	1000	11	Dec 10	Dec 24	Jan 14	Jan 28
	S	1000	6	Nov 6	Nov 20	Dec 4	Dec 18
totals		28	28000				

are included in Table 2 along with a cut off date. Data from follow-ups will be added to responses corresponding to the representative day of the week. Approximately a 35% response is anticipated to the first mailing, a 40% response to the first follow-up and a 40% response to the second follow-up. Thus, data will be collected over a six week period for each representative day in the survey.

Three basic methods may be used to select a random sample of driving records from state files:

- 1) file counting method
- 2) self-weighting method
- 3) non-self-weighting method.

Every state except Maine maintains at least two files of driver records. Many states have one or more computerized filing systems. The sampling method to be used depends on the characteristics of the various combinations of the filing system.

File Counting Method

Where a large number of records are stored in some smaller number of storage units with a variable number of records in each unit, the file counting method should be used. First, the sampler should obtain a reliable estimate of the average number of records per storage unit. This can be done by computing the average (\bar{X}) over all units or over a random sample of 20 to 30 units. (A unit might be a file drawer or a shelf). Assume that each unit holds \bar{X} records. Then $\lceil m/\bar{X} \rceil$ will be the number of the storage unit (when all units are placed in sequence) that holds the record corresponding to the random number m . (Symbol $\lceil a \rceil$ means "round a up to the next largest integer.") Counting from the first file in the selected storage unit, the number of the record corresponding to the random number will be $m - \bar{X} \lceil m/\bar{X} \rceil + \bar{X}$.

Methods 2 and 3 are appropriate for filing systems in which each record is assigned a numeric or alpha-numeric code. The code for each record need not be unique.

Self-Weighting Method

Select a sample of drivers by obtaining a random sample of all possible alpha-numeric codes. Let each code have an equal probability of being selected. Then include in the sample every record with that code number. Continue to apply numbers to obtain records until the number of records obtained exceeds the number needed for the sample.

Non-Self-Weighting Method

For some states, method 2 will result in the need for records being filled with the selection of all records having the first alpha-numeric code that is randomly selected. An example would be a state that used day of the year born as a code. In any case, when the number of records obtained exceeds the number of records needed, a random sample of all these records could be selected. This would eliminate the excess. This procedure has a serious sampling defect as well as the possibility of the introduction of bias.

After the actual drivers have been selected for a given state, it will be necessary to match them with survey days. A random matching process will be employed to assign drivers to days so that the randomness of the sample is maintained. Outlined below is such a process and Table 3 shows the result of the exercise.

Generation of the Random Sample - Matching Drivers to Days

1. Obtain the list of drivers chosen for a state.
2. Assign an index number to each driver, from 1 to N.
3. Generate N random numbers, enough for the number of drivers selected in the state.
4. Assign an index to each random number.
5. Rank order the random numbers, list corresponding indices.
6. Divide list into 28 equal groups.
7. Match driver indices (step 2) to random number indices.
8. 28 groups give drivers to be surveyed on each of 28 survey dates.

TABLE 3 EXAMPLE OF GENERATION OF RANDOM SAMPLE MATCHING DRIVERS TO DAYS

1 Master list of random numbers		2 Ordered list of random numbers			3 Master list of drivers from state		4 Assignment of drivers to groups		
i	RN	RN	i	$\frac{1}{X}$ divisions	Driver	i	RN (col 2)	i	Driver (Col 3)
				x=28					
1	007	003	23		A	1	23		W
2	415	007	1		B	2	1		A
3	657	113	12	Group 1	C	3	12		L Group 1
4	697	124	8		D	4	8		H
5	985	128	16		E	5	16		P
6	312	137	15		F	6	15		O
7	332	259	14		G	7	14		N
8	124	264	28		H	8	28		BB
9	779	312	6		I	9	6		F
10	892	332	7	Group 2	J	10	7		G Group 2
11	710	360	20		K	11	20		T
12	113	384	27		L	12	27		AA
13	906	390	22		M	13	22		V
14	259	415	2		N	14	2		B
15	137	580	24		O	15	24		X
16	128	657	3		P	16	3		C
17	903	697	4	Group 3	Q	17	4		D Group 3
18	929	710	11		R	18	11		K
19	989	725	26		S	19	26		Z
20	360	779	9		T	20	9		I
21	885	838	25		U	21	25		Y
22	390	885	21		V	22	21		U
23	003	892	10		W	23	10		J
24	580	903	17	Group 4	X	24	17		Q Group 4
25	838	906	13		Y	25	13		M
26	725	929	18		Z	26	18		R
27	384	985	5		AA	27	5		E
28	264	989	19		BB	28	19		S

↓
to
28th group

↓
to
28th group

Once the matching process has been accomplished, identification numbers are to be assigned to each driver (see later) and the entire sample automated for the preparation of mailing labels. At this time the mailing label list should be built from the master driver record of Figure 1.

Mailing Label Format

Name _____ / ____ / ____ / ____ / ____
Address _____
City _____ State _____ Zip _____

Scheduling

The data collection phase of the survey is scheduled to occupy a 12 month period. It is possible that a survey day could be chosen to occur near the end of the 12th month such that time must be allowed into the 13th month to permit the follow-up procedures to be exercised. In general, a six week period must be allowed following the last survey date to permit the follow-up procedure to be completed. Final data analysis also goes well beyond the last data collection.

In addition to the lag time required, lead time is also necessary. Among the considerations of lead time are the length of time required for Bureau of the Budget approval of the questionnaire (currently estimated at 3-4 months), time to contact the states and arrange for access to the driver record files (estimated four months for first time, less in subsequent years), acquisition of driver lists (one month), and preparation of the actual mailing list (one month). Other considerations are time for printing the questionnaire, assigning and applying identification numbers, etc. The times given are not necessarily sequential.

Integrating these factors would require a lead time of approximately six months prior to the beginning of data collection, and four months following.

Detailed scheduling of survey dates, follow-up dates, and lead times for mailing is required using Table 2 as a basis.

To establish the mailing dates (see example in Section 4) a small survey of the postal travel time was undertaken. It was found that an average of 3-4 days was required for a piece of mail to reach its destination and a like amount of time for it to be returned. Considering the time that the survey card was in the hands of the respondent, approximately 8-14 days was required for a piece of mail to be sent out and returned.

Questionnaire

The questionnaire must be developed using the variables previously identified (best predictors of exposure) as the basis for the questions. Table 4 lists the variables and their levels that might be used in a typical questionnaire. A typical cover letter and questionnaire are shown in Figures 2 and 3, as examples of wording and format. Both were pre-tested and found satisfactory. The questionnaire is intended to present the respondent with all instructions necessary to make his task as clear and simple as possible. The trip record form provides an orderly procedure for recording the required mileage information. Thus, the respondent is guided through the process of reconstructing his trips for the given survey day.

Drawing from the experience of prior studies, several response encouraging features are employed. In the cover letter, concepts are presented in brief paragraphs and in clear, concise language. Emphasis is placed upon the importance of the individual's response. The "please-help-us-out" appeal is used rather than one which stresses the rewards of response. Further, the appeal is firm, as opposed to being permissive. The cover letter statement is also

TABLE 4 VARIABLES TO BE USED IN THE QUESTIONNAIRE

<u>Independent Variables</u>	<u>Levels on Questionnaire</u>
Vehicle Type	7 (Car, Small Truck, Large Truck, Tractor Trailer, Bus, Taxi, Other)
Sex	2 (Male, Female)
Road Type	4 (Local Streets, Freeway, State Highway, Rural Road)
Light Condition	3 (Day, Night, Dawn or Dusk)
Age	Open
Model Year	Open
<u>Dependent Variable</u>	
Vehicle Miles Travelled per trip	Open

-Letterhead-

Dear Driver,

The (name of organization) is conducting a nation-wide survey for the National Highway Safety Bureau of the U.S. Department of Transportation.

In order for the Bureau to carry out its highway safety programs effectively and efficiently, it needs to gather information about our nation's drivers, their vehicles, and the amount and type of driving they do.

To save expense, we are not asking every driver in the country to help us, but only a scientifically selected few, chosen randomly from state driver licensing records. Thus, in order to guarantee that our results truly represent a national cross section, your reply is very important.

The answers you supply will, of course, be held in the strictest confidence. They will be used only for counting the number of responses to each question. Only total group averages, using no names, will be utilized in our analysis.

Would you please answer the questions on the questionnaire which follows and then place it in the enclosed stamped, pre-addressed reply envelope and mail it back to us.

Again, it is your reply we need, regardless of whether you are currently licensed, or whether you have done any driving on the day appointed at the beginning of the questionnaire.

We shall be most grateful for your cooperation.

Sincerely,

(Official signature)

(Title)

Figure 2 Typical Questionnaire Cover Letter

Figure 3 - Typical Questionnaire

NATIONAL DRIVER SURVEY

PURPOSE:

The purpose of our survey is to learn more about the driving patterns of drivers, including the numbers of miles and types of roads driven on, and the numbers of trips taken by all drivers in the United States. Such information, when collected and analyzed, will yield valuable data useful in the planning and implementation of future transportation networks and in the vital field of highway safety.

You and a small number of your fellow drivers have been scientifically selected to represent all the registered drivers in this state, therefore your response is extremely valuable to us. A computer analysis of your response will help us to develop a clearer picture of the driving patterns of all drivers. It is important, therefore, that you complete this survey in accordance with the following instructions. Again, it cannot be emphasized too strongly that all information you supply to us will remain confidential.

GENERAL DIRECTIONS:

You will note that the top of the form bears a date. All information you supply should be for that date only. The day listed begins at midnight and continues for the next 24 hours.

Any information supplied should apply only to you, the person to whom this form was addressed. Information can be recorded on the form by someone else, but should be described by and apply only to the addressee.

Pencil or pen may be used to complete the form. Most responses require only a check mark in the appropriate box or the recording of some numbers. Examples are provided. The entire form should not require more than 15 minutes of your time to complete.

Definitions: In responding to the questions, please keep in mind these definitions.

DRIVER: The person who actually drove and controlled the operation of the vehicle and to whom this form is addressed. Do not report times when you were only a passenger.

VEHICLE: Any common vehicle operated on the road or highway including, but not limited to, passenger cars, trucks of all types, busses, ambulances, campers and motor scooters. Do not include off-road vehicles such as farm tractors and other farm equipment, bulldozers, road construction equipment, or bicycles.

TRIP: A journey or excursion made with substantially the same purpose in mind where a considerable amount of time lapses between stops. Intermediate stops to the ultimate destination are not counted as separate trips.

EXAMPLES: A trip is -

- from home to school to pick up or drop off children and home again with a brief stop at a drug store.
- from home to office or place of employment (return journey from work to home is a separate trip).
- from office to several customers' place of business and return to office (a salesman's calls).
- from home to relatives or friends for dinner (return journey counts as a separate trip).

BUSINESS: A trip made during the course of your employment. Driving to and from work is not classified as a business trip.

NATIONAL DRIVER SURVEY

PART I

*** SUPPLY INFORMATION FOR DRIVING DONE ON _____

SPECIAL NOTE *** Be sure that the information you give in response to each of the items pertains to driving done only on the day specified above.

Are you currently a licensed driver (license not currently suspended or revoked)? [] YES
[] NO
[] Don't know

Did you drive on the day given above? [] YES continue with PART II, below
[] NO turn to page 4 and fill out PART IV

PART II

For each vehicle that you drove on the day indicated above, answer the following questions in the appropriate column as in the example below. If you drove more than 3 vehicles, describe the 3 that you operated most.

	EXAMPLE	VEHICLE #1	VEHICLE #2	VEHICLE #3
What is the MAKE of the vehicle? (For example Ford, Chev. Impala, Dart, etc.)	(Ford)	()	()	()
What YEAR is it?.....	19 67	19 _ _	19 _ _	19 _ _
What TYPE OF vehicle is it?				
a. Passenger Car (sedan, stat. wagon, micro-bus, sports car, etc.).....	(X)	()	()	()
b. Small Truck (pick up, panel flat bed, step van, etc.).....	()	()	()	()
c. Large Straight Truck (generally 18,000 lbs. or over).....	()	()	()	()
d. Truck-trailer or Combination Vehicle	()	()	()	()
e. Taxi or Limosine.....	()	()	()	()
f. Bus (school or commercial passenger)	()	()	()	()
g. Other (please specify).....	()	()	()	()

PART III

For each trip that you took on the day indicated, record the information requested. See the example below. Include only times when you were actually driving the car.

Record the miles you list as whole numbers. EXAMPLE: 3 1/2 miles, record as 4 miles.

These definitions may help you describe the road types-

Local Streets and Roads - normal streets and roads through business and residential areas generally inside city and village limits or in built up areas outside cities and villages.

Freeways and Toll Roads - limited access divided highways through both cities and rural areas. EXAMPLE - Interstate or similar highways.

State Highways ----- state numbered highways (2 or more lanes) which are not built like freeways.

Rural Roads ----- numbered county roads, rural roads, local county and township roads, both paved and unpaved.

Continue PART III, next page...

TRIP RECORD SHEET for trips taken on _____
 (from 12:01 AM to 11:59 PM)

TRIP NUMBER	VEHICLE USED	TRIP TYPE	TRIP STARTED IN	TIME OF DAY TRIP BEGAN	TOTAL NUMBER OF PASSENGERS NOT including driver	NUMBER OF MILES driven on different TYPES OF ROADS	TOTAL MILES DRIVEN THIS TRIP
EXAMPLE	<input checked="" type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3	<input checked="" type="checkbox"/> Personal <input type="checkbox"/> Business <input type="checkbox"/> Both	<input checked="" type="checkbox"/> Daylight <input type="checkbox"/> Dawn or Dusk <input type="checkbox"/> Night	<input type="checkbox"/> AM <input checked="" type="checkbox"/> PM	<input type="checkbox"/> 0 <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4+	3 Local Streets 2 Freeway State Highway Rural Road	5 miles
1	<input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3	<input type="checkbox"/> Personal <input type="checkbox"/> Business <input type="checkbox"/> Both	<input type="checkbox"/> Daylight <input type="checkbox"/> Dawn or Dusk <input type="checkbox"/> Night	<input type="checkbox"/> AM <input type="checkbox"/> PM	<input type="checkbox"/> 0 <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4+	Local Streets Freeway State Highway Rural Road	miles
2	<input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3	<input type="checkbox"/> Personal <input type="checkbox"/> Business <input type="checkbox"/> Both	<input type="checkbox"/> Daylight <input type="checkbox"/> Dawn or Dusk <input type="checkbox"/> Night	<input type="checkbox"/> AM <input type="checkbox"/> PM	<input type="checkbox"/> 0 <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4+	Local Streets Freeway State Highway Rural Road	miles
3	<input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3	<input type="checkbox"/> Personal <input type="checkbox"/> Business <input type="checkbox"/> Both	<input type="checkbox"/> Daylight <input type="checkbox"/> Dawn or Dusk <input type="checkbox"/> Night	<input type="checkbox"/> AM <input type="checkbox"/> PM	<input type="checkbox"/> 0 <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4+	Local Streets Freeway State Highway Rural Road	miles
4	<input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3	<input type="checkbox"/> Personal <input type="checkbox"/> Business <input type="checkbox"/> Both	<input type="checkbox"/> Daylight <input type="checkbox"/> Dawn or Dusk <input type="checkbox"/> Night	<input type="checkbox"/> AM <input type="checkbox"/> PM	<input type="checkbox"/> 0 <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4+	Local Streets Freeway State Highway Rural Road	miles
5	<input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3	<input type="checkbox"/> Personal <input type="checkbox"/> Business <input type="checkbox"/> Both	<input type="checkbox"/> Daylight <input type="checkbox"/> Dawn or Dusk <input type="checkbox"/> Night	<input type="checkbox"/> AM <input type="checkbox"/> PM	<input type="checkbox"/> 0 <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4+	Local Streets Freeway State Highway Rural Road	miles
6	<input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3	<input type="checkbox"/> Personal <input type="checkbox"/> Business <input type="checkbox"/> Both	<input type="checkbox"/> Daylight <input type="checkbox"/> Dawn or Dusk <input type="checkbox"/> Night	<input type="checkbox"/> AM <input type="checkbox"/> PM	<input type="checkbox"/> 0 <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4+	Local Streets Freeway State Highway Rural Road	miles
7	<input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3	<input type="checkbox"/> Personal <input type="checkbox"/> Business <input type="checkbox"/> Both	<input type="checkbox"/> Daylight <input type="checkbox"/> Dawn or Dusk <input type="checkbox"/> Night	<input type="checkbox"/> AM <input type="checkbox"/> PM	<input type="checkbox"/> 0 <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4+	Local Streets Freeway State Highway Rural Road	miles
8	<input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3	<input type="checkbox"/> Personal <input type="checkbox"/> Business <input type="checkbox"/> Both	<input type="checkbox"/> Daylight <input type="checkbox"/> Dawn or Dusk <input type="checkbox"/> Night	<input type="checkbox"/> AM <input type="checkbox"/> PM	<input type="checkbox"/> 0 <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4+	Local Streets Freeway State Highway Rural Road	miles
9	If you have taken more than eight trips during the day, give the total number of trips taken here _____ and estimate your total miles for the excess trips						_____ miles

Continue to PART IV, next page . . .

PART IV

What would you estimate as the total number of miles you have operated a motor vehicle during the past 7 days? miles
 () did not drive

What would you estimate as the total number of miles you have operated a motor vehicle during the past MONTH (30 days)? miles
 () did not drive

For the following items, please indicate the information that pertains to the person to whom the questionnaire is addressed.

BIRTHDATE / / SEX [] MALE
 mo. day year [] FEMALE

DATE for which this information applies / /197
 mo. day

Thank you for your cooperation. Please place this form in the envelope for return to us. The envelope is pre-addressed and needs no additional postage.

 / / /

Your comments are appreciated:

written in general terms, rather than specific, on the the theory that a recipient would be less likely to find fault and refuse cooperation with a general highway safety program. For those who oppose unnecessary spending, the economical nature of the survey design is mentioned. The signature on the cover letter is printed in a different color of ink so as to appear hand signed.

For those who fear that by revealing driving information they will also be placing their driver's license in jeopardy, reassurance about the confidential nature of the survey is provided. The questionnaire itself is laid out in such a manner as to encourage response by the use of check marks and numbers to answer the questions. Directive questions are used to pass over unapplicable sections. Return self-addressed, stamped reply envelopes are also provided. The postage stamp is of the single adhesive type as this, too, tends to encourage response. Thus all that is required of the respondent is a writing instrument and a few moments of his time.

Additional questions may be included, e.g., a gross estimate of mileage driven over a past 30 day period, as a check on the accuracy of the trip record. Other interest-arousing questions may be added for the purpose of increasing the response rate.

The questionnaire is designed to be completed on the day following the survey date as a reconstruction of the previous day's driving. Therefore consideration should be given in scheduling the mailing of the questionnaire for arrival on the day following the survey date. Since this factor cannot be rigidly controlled, many questionnaires may be received within two or three days on either side of the desired date.

A complete questionnaire package for mailing must be assembled, including not only the questionnaire and cover letter, but also a return self-addressed envelope and the original mailing envelope with an address label.

Since it is anticipated that approximately 60-70% of the drivers will fail to respond to the first survey package, two follow-up

letters should be prepared (Appendix I). They will serve as substitute cover letters for remailings of the basic survey package. As in the original cover letter, the signature should be printed in a different color ink, and likewise for the postscript on the second reminder.

The questionnaire as shown in Figure 3 contains no provision for coding. A decision must be made whether or not to include coding instructions and keypunching guides as a part of the form. If so, they must be added to the form before it is submitted for approval or printed.

The questionnaire should be reproduced in two colors on a medium weight, white opaque stock. In selecting the paper stock, consideration must be given to the overall weight of the package as postage costs are an important consideration. It is further suggested that both sides of the page be used as this reduces the total number of sheets required. Quantities must be estimated.

For the initial mailing, the numbering of the questionnaires can be done by the printer or by hand. For the follow-up mailing, it will be necessary to hand number the forms using a numbering machine. Also, each form will have to be dated in two places using a hand date stamp.

A subject code scheme has been developed such that each individual is identified by a unique number. This subject code also has provisions for identification of the respondents by state, day, and quarter of the year. The subject code scheme is shown in Figure 4, and state codes are in Appendix J.

Data Management

It will be necessary to develop a record keeping system so that the where-abouts of each interview form is known. Components of the system would include: a log book of the master mail list for checking the forms in and out; checklists for recording the steps in handling of questionnaires; a filing system for the storage of the forms; a progress flow chart to determine where each form

Subject code format:

* * / 0 0 0 0 0 / \checkmark / \wedge

* * state code - see Appendix J

00000 subject no. - see Appendix J

\checkmark quarter code - 1 = Jan. - Mar.
2 = Apr. - June
3 = July - Sept.
4 = Oct. - Dec.

\wedge day code - 1 = Sunday
2 = Monday
3 = Tuesday
4 = Wednesday
5 = Thursday
6 = Friday
7 = Saturday

Figure 4 Subject Code

is in the data reduction process. Further details and examples are given in section 4.

Data Collection

Once the questionnaires have been received and prepared for mailing labels, the data collection phase is ready to begin.

The mailing labels should be prepared and affixed to the envelopes and the proper questionnaires placed inside. Postage should be applied and the envelopes sorted according to zip codes and survey dates. Due to the bulk of the mailing, machine postage is recommended on the outgoing envelopes. It would be wise to check with the post office as to the procedures they recommend.

First class postage is recommended as the questionnaires are dated and should receive the top priority treatment accorded to first class mail.

When the returns begin to come in, they should be checked off and entered into the coding process. As the time approaches for the first reminder, those identified as non-respondents should have mail labels prepared again and the mailing process repeated as above, this time using the first reminder letter in place of the cover letter. In the first reminder letter, it will be necessary to date the questionnaire and affix the proper identification number. The process repeats itself in terms of mailing out and awaiting returns. As the returns come in, they should be catalogued and checked for duplications. Those who still are classified as non-respondents should receive the third mailing. This time the above process is repeated using the second follow-up package. After the cut-off date, those returns which arrive should be so indicated and filed separately.

Any duplicate questionnaires received should be destroyed, retaining only the one completed nearest to the original survey date.

Those questionnaires which have been returned by the post office should be examined for the reason for return. If a faulty address

can be corrected, it should be done and the questionnaire re-mailed on the first follow-up date. If this is not possible, the reason should be noted in the log book and the questionnaire filed in a separate place.

It would be helpful to have built-in inspection procedures at various phases of the assembly and mailing process to insure against errors.

Data Reduction

When a sufficient number of questionnaires has been received the coding process can begin. There will be some overlap between the checking procedures described above and the coding process described here.

As the questionnaires are returned and opened prior to being checked in, any extraneous material should be removed and examined, retained if relevant, and otherwise destroyed. Envelopes which contain return addresses should be retained until the master address list can be checked for accuracy. When this is completed, destroy these envelopes as well. Once the questionnaires have been opened, catalog the returns.

In cataloging or checking in the questionnaires, check the identification numbers as well as the sex and birthdate of the respondent against the master record. Also check the date listed by them against the survey date in the log book. If everything checks, forward the questionnaire to the coding procedure. If not, and the problem cannot be rectified, place the questionnaire aside in a separate file.

As a first step in coding, the questionnaire should be edited for legibility of the responses and correct placement of responses. Any written comments in the comments section should be transcribed onto a record sheet along with the case number.

The actual coding involves the examination of each form to make sure that: (1) the referral questions were correctly followed; (2) that the miles on the trip log add up correctly; (3) that leading

zeros are added where necessary; and (4) the form is in a suitable condition for the data transfer process which follows. If an ex post facto coding scheme is used rather than a self coding form, then the data must be transferred to the coding form.

Once the form has been coded, it is then put through the normal key punching process and the data transferred to punched cards. The cards are verified and the magnetic tape file is built.

At this point, the tape file should be error checked for wild codes, improper responses, etc. When the necessary corrections have been made, the file is ready for analysis. It would be helpful to build in quality control checks at various points by use of spot re-coding.

SCHEDULING, RESOURCES, COSTS

A basic time line of events in a national, mail exposure survey is shown in Table 5. A six-month period (July-December) is prescribed for planning and preparation for the data collection phase. The data collection begins in January and ends in the following January (the 13th month is for follow-up on the original mailings of the 12th month, December). Data analysis would be performed throughout the 12-month period of data collection, especially at the end of each quarter. However, the final data analysis would be done in a three-month period at the end of the 12 primary months of data collection. One final month is provided for reporting.

The required resources include the survey staff, office facilities, computer, mailing materials (including printing subcontract), and travel.

The estimated requirements for survey staff are presented in Table 6. Only the director and typist would be full time for the entire 22 month period. A total of 19 staff positions would be required, full- or part-time. The four liaison staff members would be active for a three month period during the sample construction phase, prior to data collection. Five clerical people would be

TABLE 5

BASIC TIME LINE OF SURVEY EVENTS

<u>Event</u>	<u>Month</u>
Begin program	1
Begin state liaison	1
Questionnaire approval	2-4
Questionnaire printing	4
Construct sample	5
Assemble mailing packages	6
Begin data collection	7
First quarter	7-9
Second quarter	10-12
Third quarter	13-15
Fourth quarter	16-18
End data collection	19
Data analysis	19-21
Reporting	22

TABLE 6
STAFF RESOURCES

	<u>man months</u>
1 Director	22
1 Administrator	6
4 Liaison staff	12
1 Analyst	5
1 Programmer	5
1 Typist	22
5 Clerical	45
2 Coders	15
3 Key punch	<u>30</u>
	162

involved during the two months preceding the beginning of data collection; three clerical people would be involved continually during the data collection, both for preparing mailings and sorting returned questionnaires. This will be sufficient to handle an average rate of about 200 mailings per day and 100 returns per day. One coder, with occasional part-time assistance will be sufficient for the return rate expected. Likewise, two keypunchers with occasional part-time help will be sufficient during a 13-month period.

Office space of about 2000 square feet will be required, with working space for about 10 at a maximum. About 26 drawers of file cabinet space will be required.

It is assumed that the survey project will have access to a large computer within the organization performing the survey. The programmer and keypunchers should be part of the computer services group, but assigned to the survey project.

A cost estimate for the 22-month project is presented in Table 7. The \$253,000 total is sensitive to variations in overhead rate, which depend on the type of organization performing the survey (academic, non-profit research, commercial). An overhead rate of 80% is assumed, being an approximate average of possible rates. The overhead charges include office space, telephone, and other central services of the organization.

TABLE 7
COST ESTIMATE

Salaries and Wages (162 man months)	\$115,000
Overhead (assumed 80%)	92,000
Computer Services	10,000
Materials	8,000
Printing	6,000
Postage	14,000
Travel	<u>8,000</u>
	\$253,000

SECTION 4
EXPOSURE SURVEY INSTRUCTIONS (Task 2)

This section presents in considerable detail the instructions necessary to implement and conduct the nationwide exposure survey described in section 3. It is intended that this section be as representative of the actual tasks as possible. However it is expected that the organization which actually performs the survey will use the procedures as a guide, and that some of the procedures will be modified to suit the circumstances.

LIAISON

Bureau of the Budget Approval

1. Prepare the final form of the questionnaire and cover letters, complete with sample address labels and envelopes. Include coding information if actually printed on form. (See Figures 2 and 3 of preceding section.)
2. Submit questionnaire via contract manager for approval by Bureau of the Budget.
3. Once approval has been received, add approval number to questionnaire in space provided on top right hand corner.
4. Prepare questionnaire for reproduction (see succeeding section on questionnaire).

State Contacts

The procedure outlined below is for one state. Repeat it for all 50 states and D.C. Slight modifications may be necessary, according to circumstances in the various states.

1. Determine the name, address, and telephone number of the administrator under whose jurisdiction the driver licensing function exists. Consult the AAMVA Directory, e.g. Indiana: Commissioner, Bureau of Motor Vehicles; Michigan: Administrator, Office of Driver and Vehicle Information.
2. Make an initial contact by telephone. Explain the nature of the project and the substance of your request. See Figure 5 for suggested format.
3. Write a follow-up letter reiterating the content of the telephone conversation. See Figure 5 for suggested content, modified according to telephone conversation.
4. Continue negotiations until access to driver license records has been granted. A personal visit to the state

- A. Name of organization
- B. Name of contract, contract number, sponsorship
- C. Goal of project: to measure driving exposure of a sample of drivers in the state, by mail survey.
- D. Purpose of contact: to obtain cooperation in generating a random sample of registered drivers in the state.
- E. Nature of request: to gain access to driver license records in order to select a number (N) of names of drivers.
- F. Information sought: name, address, driver license number, date of birth, sex of individual drivers.
- G. All information confidential.
- H. Answers to specific questions
 - 1. Type of record system in state (manual, computer, etc.)
 - 2. Type of driver license numbering system (alphabetical, soundex, etc.)
 - 3. Total number of licensed drivers.
 - 4. Common numbering system for all types of licenses.
- I. Project vital to traffic safety
- J. With letter, include copy of questionnaire

FIGURE 5
SUGGESTED OUTLINE FOR TELEPHONE CONTACT AND FOLLOW-UP LETTER

- may be necessary. Also the NHTSA contract manager can be of assistance.
5. Once access has been achieved, prepare to gather the necessary information for the state sample by either:
 - a) sending guidelines to the state and having them obtain the sample of drivers needed (see following subsection), or
 - b) sending a representative to the state to obtain the sample.
 6. Generate sample. See Figure 1 for format of information.
 7. Prepare sample data for entry into computer file by either editing tape of driver information or keypunching information cards for each driver as received from the state.

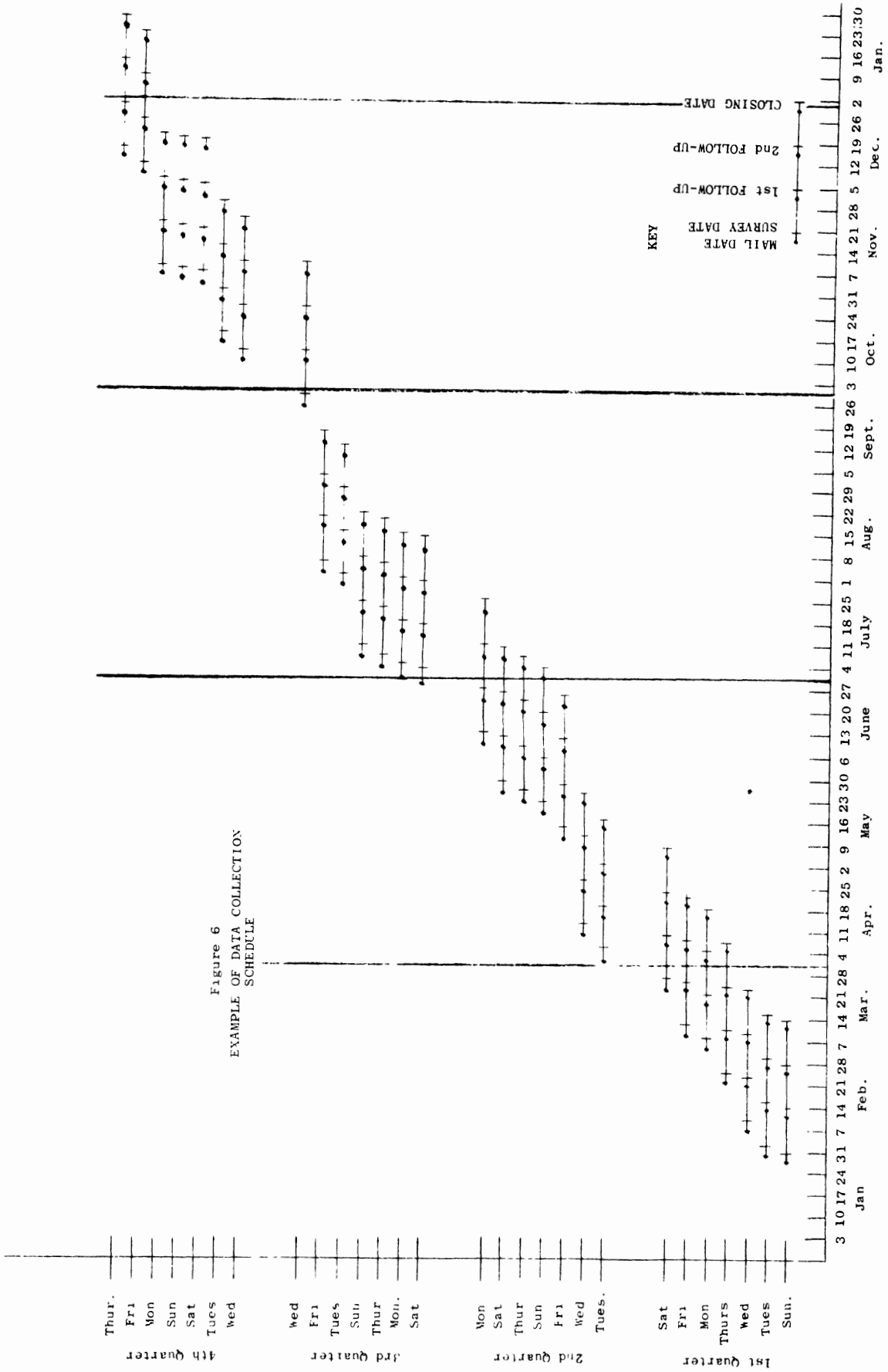
SAMPLE AND SURVEY DESIGN

Generate Sample Size

1. Refer to Table 1 as a guideline.
2. Tabulate answers to question concerning number of licensed drivers in each state.
3. Arrange tabulation in alphabetical order.
4. Calculate a sum total of all licensed drivers in U.S.
5. Calculate a percent distribution of licensed drivers.
6. Round off distribution to nearest tenth.
7. Sum the distributions to 100%, making adjustments if necessary.
8. Move decimal place right one place and round off to nearest whole number. Sum should equal 1000.
9. The result of item 8 is the daily sample required for each state.
10. Multiply the daily sample size by 28 for yearly sample. Sum should equal 28,000.
11. The results of item 10 (yearly state samples) give the total sample size required for each state.

Generate Survey Dates

1. Refer to Table 2 and Figure 6.
2. Set up form as in Table 2.
3. Select each of the 28 days needed randomly.
 - a. For 1st quarter
 1. number all Sundays in a quarter from 1 - 13.
 2. randomly select a number from 1 - 13. Result is Sunday selected to represent all 13 Sundays.
 3. repeat 1 and 2 for six remaining days of the week.
 - b. Repeat a. for remaining three quarters.
 - c. List results as in column "Week of Quarter" in Table 2.



- d. Consult a calendar and write the date that corresponds to the selected day of the quarter as chosen in a and b above for all 28 days.
- e. Also fill in dates corresponding to the days chosen which are 2, 4, and 6 weeks later to correspond to the 1st, 2nd, and cut-off dates respectively as shown in Table 2.
- f. Plot these dates as shown in Figure 6.

Sample Generation Procedure

This procedure is to be done for each state. Two methods are given below. Choose the one that best fits the record keeping system of the state. Let N equal sample size required for a state, and let n equal the number of licensed drivers.

Method A -- Computer selection of licensed

Condition: computer file accessed by driver license number with each license number fairly unique.

- a. use a random number generator to generate 3N random driver license numbers (to allow for vacant numbers).
- b. search the license file for licenses corresponding to these numbers.
- c. accept the first N valid license numbers.
- d. print out name, address, driver license number, date of birth and sex for sample.

Method B -- Manual selection of drivers from file system or computer printout.

Manual file (file drawers, file trays, etc.)

- a. count number of file trays (storage units)
- b. count the number of records in 10% of the file trays; average results to find average number of records per tray (\bar{X} =number of records per tray).
- c. multiply \bar{X} times number of record trays = n.
- d. n equals total number of licensed drivers in state (estimated).
- e. select N random numbers within the range of random numbers from 1 to n.
- f. divide \bar{X} into each random number. For each random number, the quotient equals the drawer number and the remainder is the number of the record in the tray counting from the back.
- g. pull the record and copy off the required information (see Method A, section d).

Computer listing (or other listing) of either all driver license numbers in state or name and address (and other information) of all drivers in state.

Condition: each record entry must be of the same length. i.e.: one license number per line or equal number of lines for name and address.

- a. count the number of records on a page.
- b. count the number of pages.
- c. $a \times b =$ total number of drivers in state (N)
- d. generate N random numbers in the range from 1 to n.
- e. divide each random number by the number of entries per page
- f. quotient gives page number, remainder gives the item number on the page counting from the top to the bottom
- g. either (1) copy the name and address and other information, or
(2) submit the driver license number for each one selected to obtain the names and addresses and other information.

Matching Drivers to Days

Perform for each state. See Table 3.

1. Obtain the list of drivers chosen for a state.
2. Assign an index number to each driver, number from 1 to N.
3. Generate N random numbers, enough for the number of drivers selected in the state.
4. Assign an index to each number.
5. Rank order the random numbers, list corresponding indices.
6. Divide list into 28 equal groups.
7. Match driver indices (step 2) to random number indices.
8. 28 groups give drivers to be surveyed on each of 28 survey dates.

Assigning Identification Numbers to State Sample

Perform for each state.

1. Generate complete identification numbers for sample in state. See Figure 4 and Appendix J.
2. Assign identification numbers with lowest subject number to first driver in first day of first quarter.
3. Continue assigning numbers, changing the subject number consecutively, and the day and quarter codes according to the days and quarters as shown in Table 2. The

revised distribution of Table 1 will give the number of drivers per day and per quarter.

4. Example: Vermont -- 56 drivers per year, 14 per quarter, 2 per day. See Table 8.

Building Master File

1. Arrange each state file in state code (numeric) order.
2. Arrange subjects within state in numeric (subject number) order.
3. Enter each record according to format necessary or as specified in Figure 1, to include subject code number, name, complete address, sex, birthdate, driver license number.

Prepare Mail Labels

1. Print mail labels for entire quarter of entire sample. See format in section 3. Include complete name, address, and subject code number (complete).

SCHEDULING

Specific Survey Dates

1. Refer to Table 2 and Figure 6 for survey-date examples.
2. Establish and mark mailing dates on Figure 6. Mailing dates are set as 4 working days prior to the survey dates.

Master Schedule

1. In addition to Figure 6, establish a master calendar summarizing deadlines, and clarifying overlapping mailing dates.

QUESTIONNAIRE

Prior to Bureau of Budget Approval

1. Prepare questionnaire design in accordance with Figure 3.
2. Prepare letters on official letterhead. Add signature in a different colored ink.
3. Add coding information if necessary.
4. Submit for approval.

TABLE 8

EXAMPLE: ASSIGNING IDENTIFICATION NUMBERS

(Example: Vermont - 56 drivers)

Quarter	Day	Drivers as matched to days	<u>I.D. Number</u>
1	S	1	45/56001/1/1
		2	45/56002/1/1
1	M	3	45/56003/2/1
		4	45/56004/2/1
1	T	5	45/56004/3/1
		6	45/56005/3/1
1	W	7	45/56007/4/1
		8	45/56008/4/1
1	T	9	45/56009/5/1
		10	45/56010/5/1
1	F	11	45/56011/6/1
		12	45/56012/6/1
1	S	13	45/56013/7/1
		14	45/56014/7/1
2	S	15	45/56015/1/2
		16	45/56016/1/2
2	M	17	45/56017/2/2
		18	45/56018/2/2
.	.	.	.
.	.	.	.
.	.	.	.
4	F	53	45/56053/6/4
		54	45/56054/6/4
4	S	55	45/56055/7/4
		56	45/56056/7/4

Reproduction

1. Add Bureau of the Budget approval number if necessary.
2. Submit order to printer for the following quantities:
 - a) questionnaire - 60,000 (have printer number 28,000 copies in accordance with subject identification numbers. Page size $8\frac{1}{2}$ x 11).
 - b) original cover letter - 28,000
 - c) first reminder letter - 18,000
 - d) second reminder letter - 11,000
 - e) return envelope - 60,000 (size to fit $8\frac{1}{2}$ x 11)
 - f) original envelope - 60,000 (size to fit over return envelope)
3. When returned from printers, label cartons and store for later use.

Preparing Questionnaire for Original Mailing

1. Affix postage stamps to return envelope.
2. Affix address labels to original mail envelopes.
3. Date questionnaires with correct dates in two places.
4. Assemble
 1. return envelope
 - 1 questionnaire -- dated and numbered
 - 1 cover letter
5. Stuff into correct envelope. Check questionnaire and mailing label to see if they match.
6. Seal envelope and affix postage.
7. Sort by (1) survey date
(2) zip code
8. Package correctly sorted materials for transportation to post office on assigned day.

DATA MANAGEMENT SYSTEM

Log Book

1. Reproduce sufficient copies of the log book sheet, Figure 7.
2. Enter survey names, arrange according to survey date, state, and numerically within state. Include in each entry the information illustrated in Figure 1.
3. Record flow and status of each questionnaire at appropriate times in log book.

IDENTIFI- CATION NO.	DAY- QUARTER	ORIGINAL DATE		1st REMINDER DATE		2nd REMINDER DATE		CHECKED AND FILED	NON-RECIPIENT			STATUS			DUPLICATE
		OUT	IN	OUT	IN	OUT	IN		POSTAL RETURN	RETURN BY FAMILY, etc.	ADDRESSSEE RESPONSE	NON- ADDRESSEE RESPONSE	REFUSAL		

FIGURE 7--Survey Log Book Sheet

Storage

1. Bulk storage of forms received from printer.
 - a. label all boxes received from printer as to contents
 - b. label all boxes of prepared questionnaires awaiting mailing on specific days
 - c. store in an appropriate secure place
 - d. prepare master list of materials stored
2. Intermediate storage
 - a. all forms which have been received and checked in, and are awaiting coding and other steps, should be held in temporary transfer boxes
 - b. record location and contents of each transfer box
3. File storage
 - a. set up a permanent file storage drawer
 - b. arrange file space by state, with sufficient space for expected return from each state
 - c. place questionnaire within state space numerically by subject number
 - d. prepare a master list of the contents of each drawer
4. Dead storage
 - a. set aside one file space for storage of questionnaires which are unusable (e.g. returned by the Post Office, not completed by correct person, etc.)
 - b. file by subject number
 - c. prepare a master list of the contents of dead storage

Flow Charts

1. Prepare flow charts of activities in questionnaire preparation, data collection, and data reduction (Figures 8, 9, 10).

DATA COLLECTION

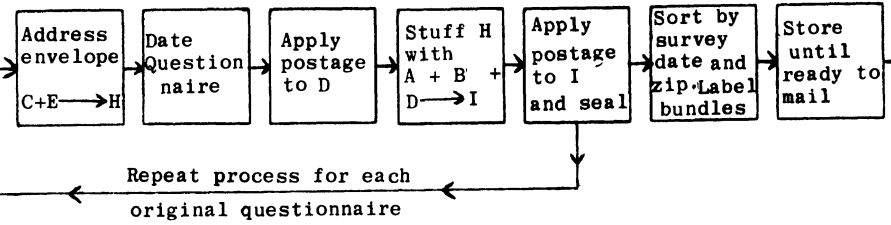
See Figure 9.

1. Mail the questionnaire package as assembled on the correct mail date.
2. Record this mail date in log book under "out" - first wave.
3. After receiving and sorting replies, record the date of reception in the log book under "in" beneath the 1st wave designation.
4. Check off the appropriate "status" of each questionnaire in the log book. (Recipient refusals will not be known until after the cut-off date has passed.)

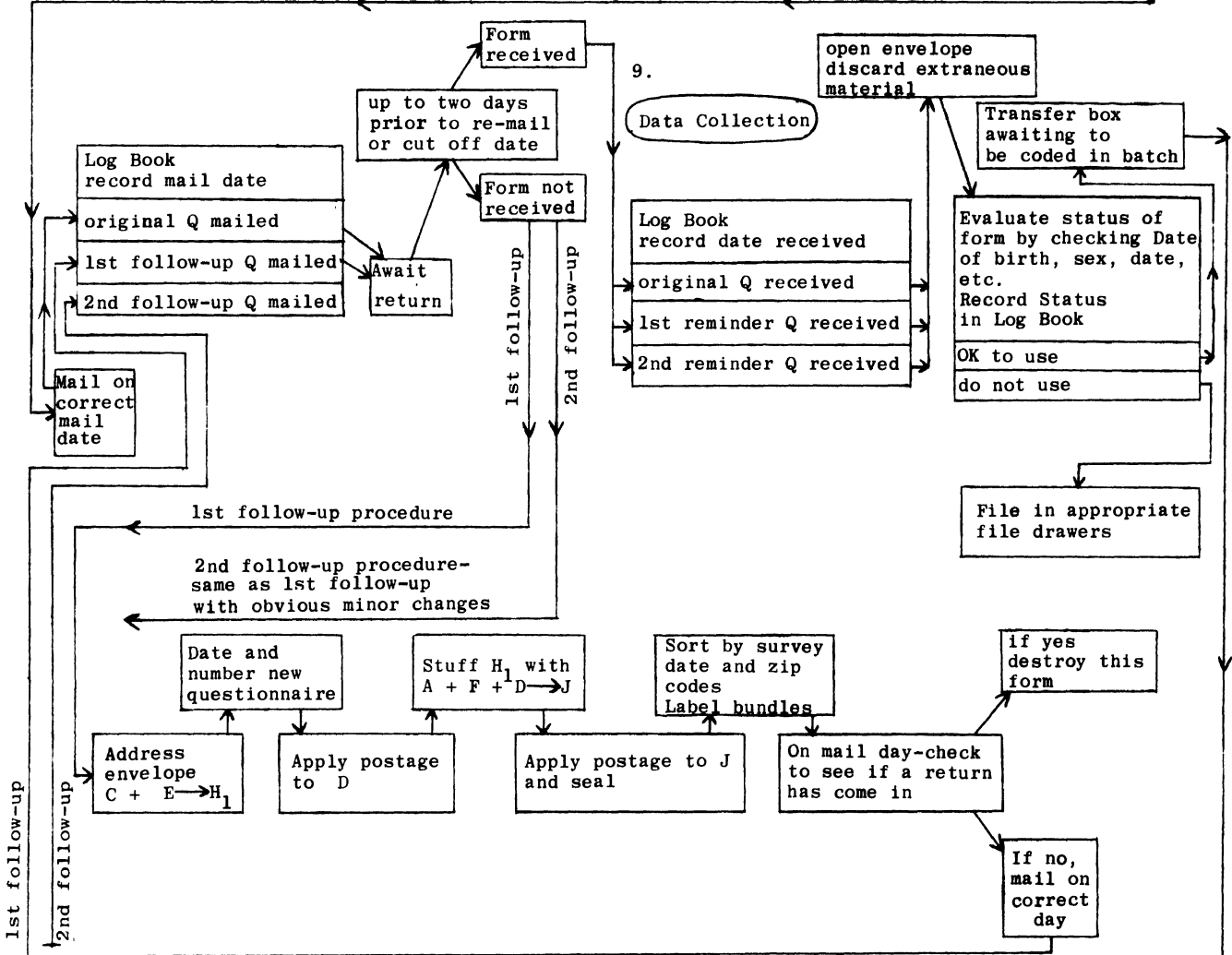
Materials as received from printer-storage until needed

- Item
- A Questionnaire
- B Cover letter
- C Original mail envelope
- D Return mail envelope
- E Mail label
- F 1st follow-up letter
- G 2nd follow-up letter

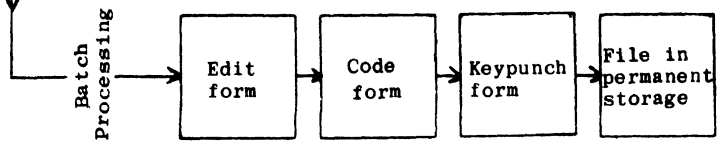
8. Questionnaire Preparation



9. Data Collection



10. Data Reduction



Figures 8, 9, 10: Flow Charts

5. Eliminate all non-recipients and respondents from the original mailing list.
6. Enter valid questionnaires into the coding process.
7. Mail the first reminder questionnaires as before using the first reminder letter in place of the original cover letter. This time the questionnaires must be dated and numbered by hand.
8. Make appropriate log book entries as in 2, 3, and 4 above.
9. Check questionnaires forms as in 5 above.
10. Eliminate all non-recipients and respondents from the first reminder mailing list.
11. Enter valid questionnaires into the coding process.
12. Repeat steps 8 - 11 for the second follow-up, noting the following changes and additions:
 - a) when mailing out the second follow-up use the second reminder letter.
 - b) eliminate all non-recipient and respondents from the second follow-up mailing list. The names that remain are the recipient refusals and should be so designated in the log book.
13. Place all apparent non-response, return by the post office, and duplicates in a special file.

DATA REDUCTION

See Figures 9 and 10.

1. Either as return questionnaires arrive, or at the end of each wave, separate the returned questionnaires from any extraneous material that is with them e.g., returned accompanying letters, or instruction sheets.
2. Retain all envelopes bearing return addresses and compare these with the master address list for accuracy. Then destroy the return envelopes.
3. Compare all filled out questionnaire returns with the master address list for agreement of identification numbers, sex, and date of birth, in order to determine whether the status of the return is to be logged as an "Apparent Addressee Response" or "Apparent Non-Addressee Response." If the latter, log in the book and file separately.
4. Check the date which the respondent listed against the designated day-quarter log book entry. If the dates match, then the questionnaire is ready for coding. If not, then place the questionnaire in a special problems file for special consideration.
5. Examine each questionnaire for legibility and correct placement of responses. Print clarification and corrections where necessary.

6. Examine each questionnaire to see if referral questions were correctly followed.
7. Examine each questionnaire to see if miles entered on trip log add up properly. Make common sense corrections.
8. Examine each questionnaire to see if leading zeroes were added where necessary. If not, add the zeroes in the appropriate places.
9. Fill in the proper codes on each questionnaire after referring to guide sheets for appropriate code entries if form is not self coding.
10. Note the number of people making comments at the end of the questionnaire. Transfer these to a separate sheet along with the identification number.
11. Of those who did make comments, analyze the content of the comments and decide whether the comment was: Very Negative (1), Negative (2), Neutral (3), Positive (4), or Very Positive (5). Use this general feedback, as well as specific suggestions, to plan following year's revisions.
12. Send questionnaires to the keypunchers for transfer of data to punch cards.
13. After keypunching and verification, file the questionnaire in the manual file. File by state and case number within state.
14. Enter the punched card data into a computer file, sorted into 28 groups corresponding to the 28 basic survey dates.
15. Error check the computer file for wild codes, or improper responses and make necessary corrections.

SECTION 5

ALTERNATIVE EXPOSURE SURVEY METHODS (Task 3)

The purpose of this section is to "determine means to check the resultant exposure estimates" of the primary, recommended survey plan (mail questionnaires). Thus, "alternative collection methods" are presented below "so that a check can be made of the accuracy of the methods being used" in the primary survey.

REVIEW OF CANDIDATE METHODS

In section 9 of Volume I, three auxiliary survey methods are suggested for comparison with the primary survey method:

1. Home Interviews (national) -- A random sampling of households, producing quarterly gross estimates of vehicle miles travelled. This method provides a means of comparing national aggregate estimates of exposure (mail survey vs. home interview) but it does not provide a state-by-state comparison.
2. Home Interviews (certain states) -- A random sampling of drivers (within a random sample of counties in a few large states), producing one-day trip records in a selected quarter. This method provides a means of comparing aggregate exposure estimates (mail survey vs. home interview) in the states which have fairly large subsample sizes in the primary survey.
3. Office Interviews (selected areas) -- A random sample of driver license renewal applicants (in a random sample of licensing offices corresponding to the sampling areas of method 1 above), producing one-day trip records in a selected quarter. This method provides a means of comparing aggregate exposure estimates (office interview vs. home interview) in a few sampling areas. However, it does not provide a direct comparison with the primary mail survey method, and hence it cannot be used as an accuracy check.

It is clear that the two home interview methods could be employed for quarterly accuracy checks on the primary mail survey method, but the third auxiliary method (office interviews) could

not be used. The home interview (national) method could check national aggregates of the primary method, and the home interview (certain states) method could check certain state aggregates. The second method is limited in geographic scope, but it provides greater statistical significance in resulting comparisons for a given sample size.

In addition, certain variations of the mail survey method should be considered as alternative checking methods:

- a. Mail survey, at the end of the year, of part of the same sample used in the primary survey, producing one-year gross estimates of mileage.
- b. Mail survey, at the end of the year, of independent samples of drivers in a few large states, producing one-year gross estimates of mileage.

The first version provides a check on national aggregate estimates, and the second provides a check on certain state aggregates. Again, the second method is smaller in scope but better with respect to significance of comparison.

Finally, the annual Federal Highway Administration estimates of exposure, based on state-by-state gasoline sales, will continue to be available, and they should be considered as a means of checking the accuracy of the primary survey method. However, they are not classified by driver-vehicle-road-environment combinations, and hence would be useful only in checking the unclassified aggregate results.

SELECTION CRITERIA

The implication of an "alternative collection method" to check the accuracy of the primary survey method is that the alternative method would have a higher inherent accuracy. But the primary method recommended in Volume I (mail survey with one-day trip records of mileage) is purported to have the highest inherent accuracy within reasonable constraints of sample size (cost). It is

not realistic to propose an alternative plan with a sample size as large as the primary method. Thus, an alternative plan must be based on either of two premises:

1. The alternative plan is intended only to verify the approximate order of magnitude of exposure estimates in the primary method, and the approximate relationship among selected classes of driver-vehicle-road-environment combinations, or
2. The alternative method is intended to "check the accuracy" of the primary method within a very small classification of the total population, e.g., one state, or a certain age group.

The second premise must be rejected because we still do not know the real systematic difference among exposure estimates produced by the various alternative methods (gross estimates, odometer readings, etc.). Even if an adequate sample size were obtained in one state to provide a higher statistical significance for an alternative method, unknown systematic differences such as over-estimation would prevent a valid verification of accuracy in the primary method. Therefore, the first premise is adopted, e.g., that the alternative checking method is intended only to provide approximate verification of aggregate mileage estimates in the primary survey.

The following criteria are proposed in the selection of an alternative exposure survey plan:

1. It must have a substantially smaller sample size than the primary plan.
2. It must represent aggregate exposure in some area represented by the primary method, but not necessarily all states.
3. It must represent aggregate exposure during some time period represented by the primary method, but not necessarily the entire year.
4. It must represent aggregate exposure in some of the 26 classes of the primary method, but not necessarily all of them.
5. Within the areas, time periods and classes represented, it must provide results which have statistical significance of at least the same level as in the primary survey.
6. It must have a reasonably comparable cost with respect to the other alternatives.

RECOMMENDED ALTERNATIVE

Based on the criteria above, the national home interview method is recommended as the alternative data collection plan for checking accuracy of the primary method.

Its characteristics with respect to the criteria above are as follows:

1. Sample Size: About 2000, selected by an existing sampling plan of households, nationwide.
2. Area Represented: The entire country.
3. Time Period Represented: One quarter of the year, preferably the second or third, for convenience and timely completion of data analysis.
4. Classes Represented: Data can be produced for all 26 classes, but perhaps only 10 - 15 would be used for comparison purpose.
5. Statistical Significance: The sample size is adequate, within the limited number of classes, to provide a greater statistical significance of results than in the primary method. (It should be noted that the systematic error in the gross estimates of the sampled drivers prevents the alternative results from being superior to the primary method, in spite of higher statistical significance.)
6. Cost: The sampling costs of this method may be kept very low with respect to the other alternatives if the survey is attached to one of the quarterly national surveys being conducted continually by various survey research centers. The added cost is estimated as only \$10,000.

If it is not possible to attach the alternative exposure survey to one of the continuing research surveys, the national home interview alternative would have a much higher cost, assuming it would be performed independently. In this case, the first alternative mail survey method would be recommended (lower sampling costs and higher response rate than the second mail alternative because sample list would be available.) The home interview method in selected states would be much more expensive than the other three alternatives.

SECTION 6
EXPOSURE DATA ANALYSIS PLAN (Task 4)

The purpose of this section is to present a data-analysis plan which will supplement the exposure survey plans of sections 3 and 4. The analysis plan represents all of the technical work to be done on the survey project from the time that a complete error-checked computer file of survey data is produced on magnetic tape, till documentation of the final report, a period of about two months. Some brief analyses will also be performed at the end of each data-collection quarter, using partial data files. The basic objective of the data analysis work will be to produce statistics representing the driving exposure of the survey sample -- both in national aggregate and each of the 26 driver-vehicle-road-environment classes, as shown in Table 9.

DATA INPUT

From the 28,000 drivers in the sample, an 80% response would yield 22,400 data cases, or about 800 for each of the 28 survey dates. The data file will be sorted into 28 groups according to the 28 survey dates, and hence each group or subfile will have about 800 data cases (1000 maximum). Each data case will include data on the six independent variables and the dependent variable (vehicle miles total and by trip, for one day). Two of the independent variables (age and sex) will be stated just once in each data case. The other four independent variables (vehicle type, model year, road type, day/night) will be stated as many times as there are trips reported. The exposure in each trip will fit into one and only one of the 26 classes, but the total exposure for a given data case (sum of all trip mileages) may be spread over as many as six of the 26 classes.

Each of the 28 survey dates represents either 13 or 14 dates of a given type of day-of-the-week in each quarter of the year. The proper value (13 or 14) must be assigned as data input.

The total number of licensed drivers in the country is also required as data input.

TABLE 9
RECOMMENDED EXPOSURE CLASSES

Classes 1-12: Cars or Small Trucks Driven by Males

<u>Class</u>	<u>Road Type</u>	<u>Light Condition</u>	<u>Age</u>
1	Street	Day	Under 26
2	Street	Day	26-60
3	Street	Day	over 60
4	Street	Night	Under 26
5	Street	Night	26-60
6	Street	Night	over 60
7	Other Road	Day	Under 26
8	Other Road	Day	26-60
9	Other Road	Day	over 60
10	Other Road	Night	Under 26
11	Other Road	Night	26-60
12	Other Road	Night	over 60

Classes 13-24: Cars or Small Trucks Driven by Females

<u>Class</u>	<u>Road Type</u>	<u>Light Condition</u>	<u>Age</u>
13	Street	Day	Under 26
14	Street	Day	26-60
15	Street	Day	over 60
16	Street	Night	Under 26
17	Street	Night	26-60
18	Street	Night	over 60
19	Other Road	Day	Under 26
20	Other Road	Day	26-60
21	Other Road	Day	over 60
22	Other Road	Night	Under 26
23	Other Road	Night	26-60
24	Other Road	Night	over 60

Classes 25-26: Vehicles Other Than Cars or Small Trucks

<u>Class</u>	<u>Model Year</u>
25	5 most recent model years
26	older than 5 most recent model years

COMPUTER PROGRAM

The initial part of the analysis plan is to prepare a computer program to produce the required exposure statistics from the survey data file. It should be possible to create the program from a few existing standard programs of statistical data analysis, with a few modifications. The minimum required operations of the program are given in Table 10. Standard analysis of statistical significance among levels of the predictor variables may also be performed. An alternative model for data analysis, including significance testing, is presented in Appendix K.

If an auxiliary evaluation program is added to the basic survey and data analysis program, then other computer programs may be prepared in conjunction with the basic one described in Table 10. They may include an AID program (see Volume I, Section 4) to re-evaluate the 26 classes (i.e. complete redefinition or simply changing variable-level groupings such as driver age groups, to obtain more uniform distribution of exposure across the 26 classes). Another program may deal with exposure distribution by day of week or quarter of year. Also, some of the state subsamples will be large enough to justify separate analysis within them in order to compare exposure among a few of the large states.

If an alternative source of exposure data is obtained for checking of the basic survey results, a separate but similar computer program will be prepared to obtain the same type of results as indicated in Table 10. An alternative source of exposure data will probably not include one-day trip records, but rather, gross estimates of mileage for one quarter or one year. In this case, it will not be possible to obtain results in all of the 26 classes, but comparisons can be made between total mileages and between mileages of single-variable classes.

TASKS

The data analysis tasks are a straightforward sequence of programming, computer runs, and interpretation. The tasks begin

TABLE 10

REQUIRED OPERATIONS OF COMPUTER PROGRAM

- I. For each of the 28 data groups (survey dates):
 1. Compute sum of total one-day mileages of all data cases (approximately 800) in the data group.
 2. Compute sum of trip mileages in each of the 26 classes, of all data cases in the group.
 3. Compute mean one-day mileage per driver.
 4. Compute mean mileage per driver in each of the 26 classes.
 5. Multiply each of the mean mileages (in items 3 and 4 above) by either 13 or 14, depending on the number of corresponding dates of each type of day-of-the-week in the quarter. These results are:
 - a. mean mileage per driver over the 13 or 14 days of a specific type of day-of-the-week in a given quarter of the year.
 - b. mean mileage per driver (as in a. above) in each of the 26 classes.
 6. Repeat 2, 4 and 5b. for each level of each of the six independent variables separately (age, sex, vehicle type, model year, road type, day/night).
- II. For the entire data sample:
 1. Add the 28 mean mileages of 5a. above to obtain: Mean mileage per driver for the whole year.
 2. Add the 28 mean mileages in each class of 5b. above to obtain: Mean mileage per driver for the whole year in each of the 26 classes.
 3. Multiply the result of 1 above by the number of licensed drivers in the whole country to obtain total annual mileage.
 4. Multiply the 26 results of 2 above by the number of licensed drivers in the whole country to obtain total annual mileage in each of the 26 classes.
 5. Repeat 2 and 4 for each level of each of the six independent variables.
 6. Determine percentages of total mileage in each of the 26 classes, and in each of the levels of each of the six independent variables.

with the availability of an error-checked data file on magnetic tape.

1. Prepare computer program in accordance with the requirements of Table 10.
2. Prepare computer program for analysis of alternative data, as described in the subsection above, if required.
3. Prepare computer program for re-evaluation of data classes, if required.
4. Prepare computer program for special analysis of basic survey data, i.e. by states, by day-of-week, and/or by quarter if required.
5. Test and check programs.
6. Perform computer runs for basic survey data to obtain mean and total mileages required, as in Table 10.
7. Perform computer runs, using programs of 2, 3, and 4 above, if required.
8. Compare results of basic and alternative data sources, if required.
9. Interpret results of computer runs and identify high-exposure and low-exposure classes.
10. Tabulate results for use in the computation of accident rates.

SCHEDULING, RESOURCES, COSTS

The schedule for the data analysis plan is included in the schedule presented in Table 5 of section 3. The basic data analysis will be performed during the three month period following the end of the 12-month data collection period (January-March, months 19, 20 and 21). Preliminary programming and check-out runs may be done at the end of each quarter, using partial data obtained during the preceding data-collection quarter.

The basic resources for data analysis will be a large computer available within the organization performing the survey, and approximately 10 man months of programming and analysis effort. The data-analysis costs are included in the total program costs, as estimated in Table 7 of Section 3.

SECTION 7
CONCLUSIONS AND RECOMMENDATIONS

The basic conclusions of this study are as follows:

1. A national exposure survey concept based on random sampling of licensed drivers in the states--and using mailed questionnaires with one-day trip records--is a practicable basis for the design of a detailed, calendar-year survey plan.
2. The first year of implementation of an official, national exposure survey may be considered a field test, but procedures developed for it should have flexibility to serve in succeeding years when sample sizes may be increased.
3. The level of detail of procedures in the survey plan should be sufficient for proper implementation by a competent survey organization which does not necessarily have expertise in the technical aspects of exposure.
4. A sampling plan in which 28 dates are selected randomly for estimation of trip mileages will provide adequate representation of exposure distribution throughout the year. The 28 dates include one on each of the seven types of day of the week in each of the four quarters of the year.
5. Each trip estimate in the survey should include mileage travelled and data on six independent variables: driver age and sex, vehicle type and model year, road type and day/night. This provides for classification on each trip in one of 26 classes defined by driver-vehicle-road-environment combinations.
6. A sample size goal of 28,000 drivers is proposed for the survey, distributed among the states in proportion to their populations of licensed drivers.
7. A cumulative response rate of 80% is anticipated, based on a 35% response from initial mailings, and 40% response from non-respondents in each of two reminder mailings.
8. The 80% response should produce about 22,400 returned questionnaires, sufficient to achieve minimum significant differences of 30-40% between mean exposure values in classes.

9. Procedures for the survey plan may be organized in seven categories: liaison, sample and survey design, scheduling, questionnaire development, data management, data collection, data reduction.

10. Documentation of procedure instructions for the performing survey organization results in about 100 specific instructions within the seven categories.

11. A home interview survey of exposure, added to an existing national survey, is the preferred alternative method for the collection of data to check the accuracy of the primary survey.

12. A data analysis plan for the computation of required exposure statistics is a straightforward process of programming, computer runs and interpretation of results.

13. The total survey program would be 22 months in duration: six months of preparation, 13 months of data collection and three final months of analysis and reporting.

14. Required resources for the survey program include approximately 14 man months of professional and clerical effort, office facilities and materials, and a large computer.

15. The cost of the survey program is estimated as \$253,000.

16. It will be feasible to conduct a field-test exposure survey, using the derived plans and procedures in the calendar year 1972.

17. Revisions in the survey plan (due to increased sample sizes in the future) may be accomplished with negligible delay for operational use in annual surveys, starting in 1973.

18. Revisions in survey variables and driver-vehicle-road-environment classes will require new questionnaires and computer programs, but they should not delay continuing annual implementation.

19. Results of implementing the survey plan will provide needed exposure data for computation of accident rates, and hence for evaluation of highway safety countermeasures.

Two basic recommendations are:

1. The exposure survey plan presented in this volume should be implemented for data collection in 1972, beginning with a contract to a survey organization in July 1971.

2. Continuing, annual exposure surveys, based on the plan of this volume, should be considered for implementation beginning in 1973.

