NONRESPONSE IN HOUSEHOLD TRAVEL SURVEYS

Michele Zimowski,
Roger Tourangeau,
Rashna Ghadialy, and
Steven Pedlow

Prepared for:
Federal Highway Administration
October, 1997

NORC
1155 60th Street
Chicago, IL 60615

(773) 753-7500

TABLE OF CONTENTS

PREFACE i

1. MEASURING AND REPORTING NONRESPONSE 1-1

1.1 CHAPTER SUMMARY 1-1

Main Recommendations 1-1

1.2 INTRODUCTION 1-2

Selecting a Survey Sample 1-2

Recording the Outcomes of the Field Effort 1-4

1.3 MEASURING THE LEVEL OF NONRESPONSE IN A SURVEY 1-8

1.4 SELECTING THE UNIT OF MEASUREMENT 1-9

1.5 CLASSIFYING UNITS INTO CATEGORIES 1-9

Classification Levels and Categories 1-9

Classification Rules 1-15

An Application of the Classification System to Travel Data 1-19

1.6 CALCULATING RESPONSE RATES 1-24

Calculating Component Response Rates 1-24

Calculating the Overall Response Rate 1-25

Weighted vs. Unweighted Rates 1-28

2. REDUCING NONRESPONSE 2-1

2.1 CHAPTER SUMMARY 2-1

Major Recommendations 2-1

Organization of the Chapter 2-3

2.2 RESPONDENT CHARACTERISTICS 2-4

The Mentally and Physically Handicapped 2-5

Individuals with Limited Literacy Skills 2-5

Individuals with Language Barriers 2-6

The Less Well Educated 2-6

The Elderly 2-7

Urban Dwellers 2-7

Other Groups of Nonrepondents in Travel Surveys 2-8

2.3 INTERVIEWER CHARACTERISTICS 2-9

Interviewer Experience 2-9

Interviewer Gender 2-10

TABLE OF CONTENTS (Continued)

Interviewer Expectations 2-10
Interviewer Style of Interaction 2-10
Interviewer Vocal Characteristics 2-11
Other Interviewer Characteristics 2-11
2.4 METHODS FOR REDUCING UNIT NONRESPONSE 2-12
Interviewer Selection and Training 2-12
Design of Survey Instruments 2-14
Number and Type of Contacts 2-16
Timing of the Contact Calls 2-19
Pretesting Field Procedures and Instruments 2-25
Proxy Respondents 2-30
Special Methods for Reducing Nonresponse 2-32
Subsampling Nonrespondents 2-33
2.5 ITEM NONRESPONSE 2-34
Questionnaire Design 2-34
Retrieval of Missing Items 2-36
2.6 CURRENT PRACTICE FOR REDUCING NONRESPONSE IN TRAVEL SURVEYS 2-37
Pretests 2-37
Incentives 2-38
Follow-up Reminders 2-38
REFERENCES 2-40
3. STATISTICAL METHODS FOR REDUCING THE IMPACT OF NONRESPONSE 3-1
3.1 CHAPTER SUMMARY 3-1
Statistical Methods 3-1
3.2 TYPES OF NONRESPONSE 3-2
Unit Nonresponse 3-2
Item Nonresponse 3-3
3.3 IMPACT OF NONRESPONSE 3-3
Reduced Sample Size 3-4
Introduction of Bias 3-4
3.4 SURVEY METHODS 3-7
Averting Nonresponse 3-7

TABLE OF CONTENTS (Continued)

Following up with a Subsample of Nonrespondents 3-8
3.5 STATISTICAL METHODS 3-10
Weighting 3-10
Calculating Weights 3-11
Alternative Methods for Weighting 3-18
Editing and Imputation 3-22
APPENDICES A-1
Appendix 1: Questions and Answers to Avert Nonresponse A-2
Appendix 2: Examples of Good and Bad Questionnaire Design A-5
GLOSSARY OF TERMS G-1
References G-6

PREFACE

Response rates for household travel surveys conducted within the United States have declined substantially over the past few decades. In recent years, household travel surveys conducted by a combination of telephone and mail have typically obtained rates in the range of 25 to 40 percent. However, some travel surveys have reported response rates as low as 5 percent. In many other parts of the world even the "typical" response rates for U.S. travel surveys would be considered low if not unacceptable.

Nonresponse is of major concern to transportation planners for three reasons. First, often there is a perception that data collected in a survey with low response rates are of a poor quality regardless of the sample's representativeness of the population. Second, for any given method of data collection, the costs of obtaining quality data increase as it becomes more difficult to secure the cooperation of sample members. Finally, if nonrespondents have different travel characteristics than the population as a whole, then data from the survey will be biased.

In response to those concerns, nonresponse was one of five major topics addressed at the 1995 Transportation Research Board (TRB) conference on household travel surveys. In workshops held at that conference, participants developed a research agenda and a set of research problem statements for each topic area. Several of the recommended research projects have since been funded under the Department of Transportation and Environmental Protection Agency's Travel Model Improvement Program (TMIP). This report is one example and is the first in a series of initiatives focusing on nonresponse in household travel surveys. The purpose of this report and the other projects is to improve the quality of survey data gathered by MPOs and state DOTs, and to promote efficient utilization of data collection resources. To accomplish these goals, this report used a three-pronged approach as described in the chapter summaries below.

Chapter 1. Measuring and reporting nonresponse: A standard approach to reporting response rates is recommended. A standard approach, used consistently, can help assess the quality of survey data. Standard reporting also allows users to evaluate different techniques for implementing surveys, thus building a coherent body of knowledge on methods for household travel surveys.

Chapter 2. Reducing nonresponse: To reduce nonresponse, characteristics of respondents and interviewers must be understood. Characteristics of typical nonrespondents to travel surveys are discussed. Procedures to improve response rates are recommended.

Chapter 3. Statistical methods for reducing the impact of nonresponse: Despite our best efforts, all travel surveys are likely to have nonrespondents. Specific methods to adjust survey results to better represent the population are recommended.

This project benefited from the advice and experience of many people outside of NORC and FHWA. We would especially like to thank the members of our advisory panel:

Ed Christopher, Chicago Area Transportation Survey
Bob Griffiths, Metropolitan Washington COG
Ryuichi Kitamura, Kyoto University
Arnim Meyburg, Cornell University
Paul Moore, Research Triangle Institute
Eric Pas, Duke University
Cheryl Stecher, Applied Management and Planning Group
Peter Stopher, Louisiana State University
Johanna Zmud, NuStats

1.

MEASURING AND REPORTING NONRESPONSE

1.1 CHAPTER SUMMARY

This report is written for designers, analysts, and sponsors of household travel surveys, and all other persons who find themselves involved, in one way or another, with the collection, reporting, or interpretation of travel survey data. Its objective is to provide a set of guidelines for measuring and reporting nonresponse in household travel surveys and for reducing the level and impact of nonresponse.

This chapter focuses on the measurement and reporting of nonresponse. It discusses the background for a number of recommendations on how response rates should be calculated and reported in household travel surveys. The recommendations are based on standards established in the survey research literature.

MAIN RECOMMENDATIONS

The main conclusions of this chapter are as follows:

1) Response rates should be reported for each phase of data collection in a multi-phase effort and an overall response rate should be reported that reflects the cumulative effects of nonresponse at each phase.

2) Overall response rates should be calculated using the Council of American Survey Research Organizations (CASRO) definition, which takes into account cases whose eligibility was never determined.

3) Response rates should be reported for both households and persons.

4) A household should be classified as complete when data are obtained for the majority of the eligible persons; a person should be classified as complete if data are obtained for a designated set of critical items.

Two-phase surveys. The first recommendation applies to surveys that include a preliminary phase of data collection during which sample households are contacted to determine their eligibility and to enlist their cooperation in the main data collection effort. In many travel surveys, the preliminary phase is carried out by telephone, and the main phase is carried out through some combination of mail and telephone. It is important in such two-phase studies to report separate response rates for each phase of data collection. In addition, however, a cumulative response rate should be reported that reflects the combined impact of nonresponse at both phases of data collection.

The formula. The second recommendation concerns the formula by which overall response rates should be calculated and the treatment of cases whose eligibility is unknown. Below we describe the recommended procedure for calculating response rates in detail.

The unit of measurement. The third recommendation concerns the unit for measuring response rates. In many travel surveys, data are to be collected for all eligible persons within a sample household. In such cases, response rates should be reported for both households and persons. We describe the procedures for doing so in detail below.

Standards for completed cases. The final recommendation concerns standards for treating households or persons as completed cases. When data are to be collected for several persons within a sample household, the household should be treated as complete if data are obtained for the majority of eligible persons in the household. Similarly, a person should be counted as complete only if data are obtained for all critical items sought.

1.2 INTRODUCTION

SELECTING A SURVEY SAMPLE

The objective of a sample survey is to obtain information on the behavior or attitudes of an entire population of units by gathering information on a representative sample of units drawn from that population. The process of drawing a representative sample consists of three main steps:

- identifying the *target population (*the collection of units of interest to the survey),
- selecting or developing a *sampling frame* (a listing of units that includes the target population), and
- drawing a representative sample of units from that frame.

Target population. In surveys of travel behavior, there are often two related target populations. The first usually consists of all *persons* residing within a specified set of geographic boundaries who meet the age requirements of the survey. The second typically includes all *households* located within those boundaries with one or more age-eligible persons. Households are generally defined as groups of people who live together in a housing unit. In most definitions, including that of the U.S. Bureau of the Census, group quarters, such as dormitories and nursing homes, do not qualify as housing units.

Some travel surveys adopt the Census Bureau definition of a household. Others rely on slightly different criteria to determine the eligibility of a housing unit. Travel surveys should, however, adopt the same definition to the extent possible to ensure comparability of results across surveys. The Census Bureau definition is the recommended convention for these purposes since it is the mostly widely used definition and its use will allow for accurate comparisons with Census data.

According to that definition, a housing unit is a house, apartment, mobile home, group of rooms, or single room that is occupied (or if vacant, is intended for occupancy) as a separate living quarter. To qualify as a housing unit, the occupants must live and eat separately from other persons in the building and have direct access to their unit from the outside of the building or through a common hallway. A family, one person living alone, two or more families living together, and any other group of related or unrelated persons sharing such living arrangements qualify as households. Travel surveys should rely on this definition unless other criteria are necessary to meet the objectives of the study.

Sampling frame. Once the target population is identified, the next step in the process is to select or develop a sampling frame. The sampling frame for a survey will depend on such factors as the target population, the mode of data collection, and the sampling unit. When data are collected by telephone and the sampling unit is the household or persons living within households, the sampling frame often consists of all listed residential numbers within the survey area or all telephone numbers within exchanges in that area. When data are collected by personal interview, the frame often consists of a listing of residential units within the survey area.

Sample selection. The final step of the process is the selection of a representative sample from the sampling frame. The procedures used in this step will depend on the goals of the survey. Smaller subpopulations may be oversampled, for example, if a goal of the survey is to carry out analyses within each subpopulation.

Documentation. Since the procedures adopted in the various steps of the sample selection process will affect interpretation of the survey data and the response rates, the documentation for a survey should include information on the target population, the sampling frame, and the sampling procedures, including any information relevant to the interpretation of the data or the response rates. In a telephone survey, for example, the documentation should indicate whether the sampling frame included business or unassigned numbers and whether such numbers were eliminated through some method of prescreening before the sample was fielded.

RECORDING THE OUTCOMES OF THE FIELD EFFORT

Response Rates

Response rates measure a survey's level of success in obtaining information from all eligible units in the sample. Invariably, some members of the sample do not provide the desired information. There are many reasons why the relevant information may not be obtained.

Typically, some members of the sample are ineligible for the study. In fact, some of the sample selections may turn out not to exist! For example, in a telephone sample, some of the randomly generated numbers that make up the sample may not be actual telephone numbers at all; these are commonly referred to as unassigned or nonworking numbers. Other numbers may be the numbers of fax machines. Still others may be linked to households that fall outside the survey area.

Other reasons for failing to obtain complete data include the refusal by some members of the sample to take part in the study, failure to locate or contact sample members, and language barriers or physical limitations.

Given the wide range of potential outcomes of a data collection effort, it is important to establish rules for tracking and summarizing the success of a survey in collecting data from the members of the sample.

Recommendation

TRACKING THE OUTCOMES OF A FIELD EFFORT— *Establish a system of result codes for tracking the outcomes of the field effort.*

Result Codes/Screening Phase. To illustrate the range of possible outcomes for an individual member, we display in **Table 1.1** typical "result" or "disposition" codes from surveys relying on randomly generated telephone numbers for their samples. These codes are for the initial, or "screening," phase of a survey conducted to identify households eligible to receive a more detailed questionnaire or travel diary in the second phase of data collection.

- **Codes 01 to 08** cover situations in which a screening interview is never attempted. The first five codes are assigned to different types of nonworking or nonresidential numbers; the remainder cover other situations in which the screening interview was never started.

- **Codes 09 to 11** cover what are commonly called "screener refusals." Screener refusals are telephone numbers that may be attached to a household with one or more eligible members, but no one is willing to provide the information to complete the screener.

- **Codes 12 to 15** describe four scenarios in which the screener is completed. Code 12 is for cases in which all members of the household are ineligible for the main questionnaire. The final three codes (13 to 15) cover situations in which the screener is completed and an eligible member is identified.

Table 1.1
Result Codes for Households in a Telephone Survey Sample

DISPOSITION	CODE
NO SCREENER ATTEMPTED	
Business/government number	01
Institutional residence	02
Other non-residence	03
Verified fax/modem line	04
Nonworking telephone number	05
Number changed to new number	06
Closed out after multiple attempts to contact	07
Other	08
SCREENER REFUSALS	
Refused screening	09
Hang up prior to screening	10
Screening terminated before completion	11
SCREENER COMPLETED	
No eligible household members	12
Eligible members consent	13
Eligible members refuse	14
Eligible members with language barriers, or physical or mental limitations	15

Table 1.2 (below) displays a similar set of outcome codes for households screened by personal interview. As in the previous example, the first set of codes, codes 01 to 06, cover situations in which a screening interview is never attempted. Codes 01 to 02 are for sample addresses that turn out to be nonresidential, while codes 03 to 05 cover vacant and inaccessible dwellings, and households where no one is home at the time of the interviewer's visit. As in the earlier example, the second set of outcome codes are for screener refusals, and the last set covers households completing the interview.

Table 1.2
Result Codes for Households Screened by Personal Interview

DISPOSITION	CODE
NO SCREENER ATTEMPTED	
Business address	01
Other non-residence	02
Household not accessible	03
Vacant dwelling	04
Closed out after multiple attempts to contact (not at home)	05
Other	06
SCREENER REFUSALS	
Refused screening	07
Screening terminated before completion	08
SCREENER COMPLETED	
No eligible household members	09
Eligible members consent	10
Eligible members refuse	11
Eligible members with language barriers, or physical or mental limitations	12

Table 1.3 summarizes the procedures used to report outcomes and nonresponse rates in a number of recent personal travel surveys. Each of the organizations carrying out the surveys uses its own system of outcome codes and its own method for computing response rates. Since the systems and methods vary from organization to organization, the response rates cannot be compared in a meaningful way. As a result, it is impossible to evaluate the relative effectiveness of the data collection procedures used by the different organizations, and the relative quality of the data collected. Such evaluations require use of standard reporting procedures.

Table 1.3
Methods for Reporting Nonresponse in Some Recent Travel Surveys

SURVEY FIRM	STUDY	TYPE OF SAMPLE	RESPONSE RATE REPORTED	COMMENTS
E.H. White & Co.	1990 Bay Area Travel Surveys	RDD	Final resp rate = C/Rec Total resp rate = C/Elig	Rec = Recruited = Completes + + Inadequate Completes = Final Refusals Elig = Eligible Recruited + Initial Refusals C = Completes Not clear about UK (unknowns)
NuStats	Eugene/Springfield Area Study Medford ODOT Salem/Portland Salem/Keizer Portland Vancouver Triangle Travel Behavior Survey	RDD	Recruitment rate = Rec/E Completion rate = C/Rec Resp rate = C/E	Rec = Recruited C = Completes Elig = Known Eligible = Rec + C + Breakoffs + Final Refusals E = Estimated Eligible = Elig + e * UK where UK = Unknown eligibility e = Eligibility rate = Elig/Inelig Inelig = Ineligibles = Disconnects + Deaf/Language + Business/gov't + out of area
AMPG	1991 Southern California Origin-Destination Survey	RDD	Participation rate = Rec/HH Completion rate = C/Rec	HH = number of households Rec = Recruited C = Completes
Barton-Ashmon	MAG Regional Planning Survey	RDD	Completion Rate = C/SC	C = Useable Completes SC = Initial Screening Contacts

The wide range of possible outcomes and methods for describing them raises a number of questions about which conventions should be adopted to standardize reporting procedures across travel surveys:

- What categories should be used to track the disposition of each sample member in a survey?
- What is the appropriate unit for monitoring the progress and success of the field effort—the telephone number, the street address, the household, the person, or the trip?
- Which types of cases should be counted as respondents, nonrespondents, eligibles, and ineligibles?
- What criteria should be used to determine whether a sample member has provided sufficient information to be counted as a respondent?
- How should response rates be calculated and reported?

This chapter provides answers to these questions in the form of a set of guidelines for measuring and reporting nonresponse in travel surveys. The guidelines are based on current best practices in the survey literature.

Response rate an indicator of survey quality. Survey nonresponse is important because of its potential for introducing errors—sometimes serious errors—into survey estimates. No matter how representative the sample may be when it is originally selected, it may become unrepresentative because data are collected only from a portion of the sample members. As a result, the response rate is a widely used indicator of the overall quality of the survey. Analysts rely on these rates to assess the level of bias in the survey, to judge the level of confidence to place in the data, and to evaluate the effectiveness of the data collection procedure.

1.3 MEASURING THE LEVEL OF NONRESPONSE IN A SURVEY

The process of measuring the level of nonresponse in a survey sample consists of three steps:

- selecting the unit of measurement for reporting response rates,
- classifying sample units into categories based on their disposition codes, and
- calculating response rates.

The sections below describe the various steps in the process and the recommended set of procedures for standardizing the measurement of nonresponse across travel surveys. If these guidelines are routinely followed, they will allow for meaningful comparisons of response rates from survey to survey and from organization to organization.

1.4 SELECTING THE UNIT OF MEASUREMENT

Most personal transportation surveys attempt to collect travel data for all eligible persons within sample households. The travel data sought from each person may encompass multiple trips. As a result, response rates can be reported for any of several units of measurement--the household, the person, the trip or activity. At each level, the response rates can reflect any of several standards for defining a unit as a responding unit. At the household level, for example, one could count the household as complete if data are obtained for at least one person in the household; at the other extreme, one could count the household as complete if data are obtained for every eligible person. (The National Personal Transportation Survey adopts a middle position, counting a household as complete if data are obtained for the majority of eligible persons.)

When data are to be collected from all eligible household members, response rates should be reported for two units of measurement—the person and the household. The sections below describe how this should be done.

Recommendation

UNITS OF MEASUREMENT FOR RESPONSE RATES—
When data are collected from multiple household members, track and report response rates for both households and persons.

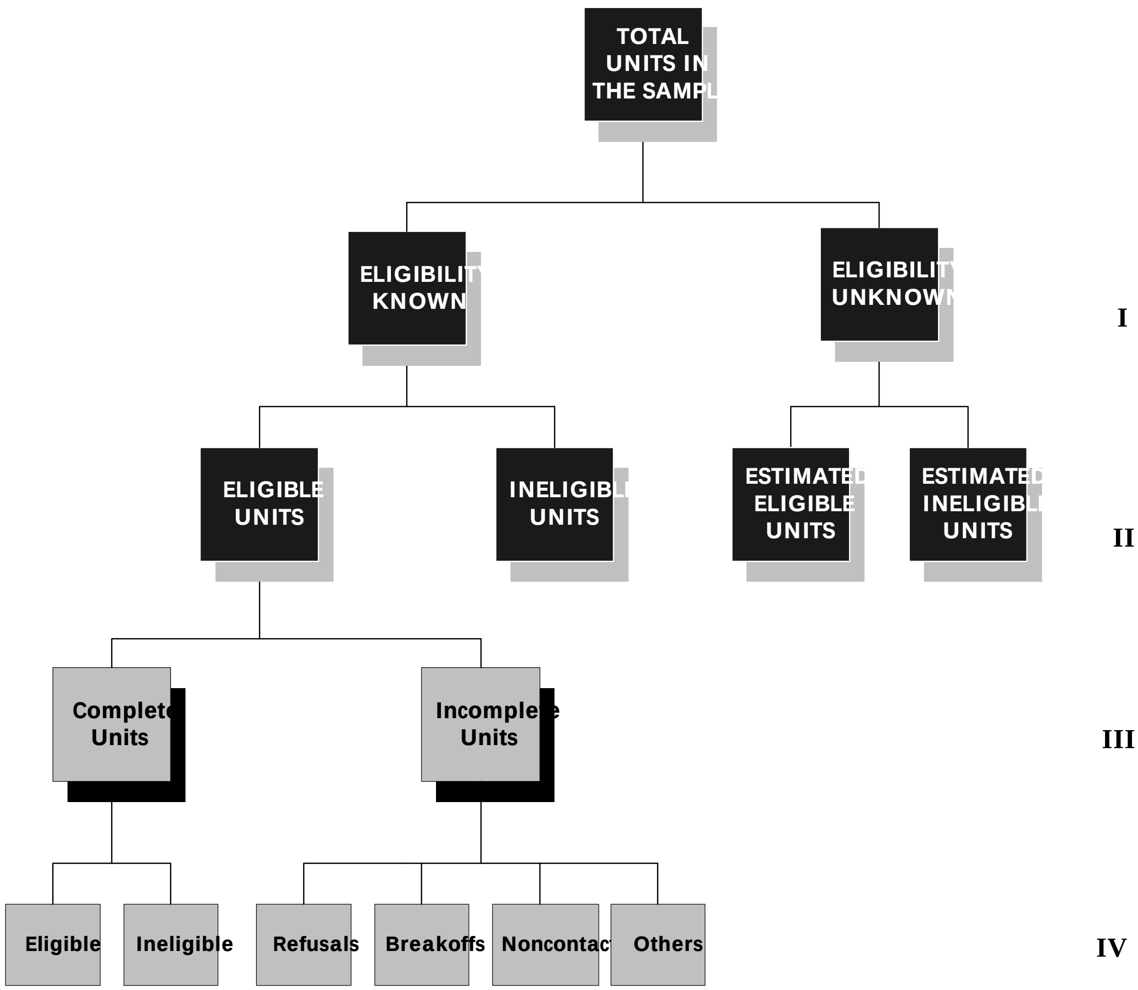
1.5 CLASSIFYING UNITS INTO CATEGORIES

Once the unit of measurement has been selected, the next step in the process is to classify all units in the sample into a set of categories based on their dispositions at the close of the field period. The counts of units in the various categories serve as the data for computing response rates in the final step of the process.

CLASSIFICATION LEVELS AND CATEGORIES

Figure 1.1 shows the various categories required for response rate computations. The categories correspond to four levels of classification, identified by Roman numeral in **Figure 1.1**. The first two levels apply to all units in a sample. The third and fourth level are reserved for units who were eligible to participate in the survey. Information from the first three levels is used to compute response rates. Information from the fourth level is often used to evaluate how various aspects of the survey's design contribute to the overall level of nonresponse in the sample. The categories in all four levels are described below.

Figure 1.1
Required Categories for Response Rate Computations



The classification process begins with all units in the sample. The *total units in the sample* is defined as the number of units fielded, or in other words, the number of units where attempts to collect data are made. In some cases, this number will be less than the number of units drawn from the sampling frame. In telephone surveys, for example, the numbers are often prescreened for business numbers before they are released to the field. In this case, the total units in the sample refers to all numbers remaining in the sample after prescreening. In many surveys, extra units are drawn from the sampling frame and held in reserve to ensure the availability of a sufficient number of units to satisfy the sample size requirements of the survey. The extra units are released to the field only if the yield from the base sample is less than expected. Again, the total units in the sample refers to the number of units released to the field.

In transportation surveys with two phases of data collection—an initial screening and recruitment phase, and a subsequent phase for collecting detailed travel data—the total units in the sample will differ across the two phases. In the initial phase, the total units will include all cases fielded during screening; in the second phase, it will include all cases eligible for follow-up. In such designs, the classification of units detailed below is carried out separately for the two phases of data collection. Thus, a household that completes the initial screener (and is classified as a respondent in the initial phase) may decline to complete the detailed travel diary (and be classified as a nonrespondent in the second phase).

LEVEL I: Eligibility Known vs. Eligibility Unknown

Some travel surveys, such as intercept studies of transit use, select their samples directly from the target population. Since all units in these samples belong to the population under study, the eligibility of the individual units is known before data collection begins.

But more often than not, travel surveys select their samples from populations and sampling frames that include units that do not belong to the target population. The eligibility of selected units must then be determined as part of the data collection process. Most surveys rely on a “screening” interview to identify the subset of units who qualify, but, as a rule, not all units in the sample are successfully screened in the process. Some units are never reached, despite repeated attempts. Others refuse to complete the screening interview. As a result, the eligibility of some units remains unknown at the close of the field period. The first level of the classification system distinguishes between those units and all other units whose eligibility is known by the end of the survey.

The relative number of units in the eligibility unknown category will depend on such factors as the length of the field period, the callback limit (maximum number of contact attempts per sample unit), the efficiency of the contacting procedures, and the properties of the sampling frame. Longer field periods, higher callback limits, and efficient calling protocols tend to reduce the relative size of this category.

LEVEL II: Eligible vs. Ineligible Units

The second level of classification draws the distinction between units that are eligible to participate in the survey and units that are not. A unit's eligibility for a particular survey depends on the survey's definition of the target population. Units satisfying that definition are eligible, while all others are not. Travel surveys, like most other surveys, tend to include units of both types.

The relative number of eligible and ineligible units in a travel survey will depend on the properties of the sampling frame and on the relative size of the target population. Samples of randomly generated telephone numbers, for example, tend to yield higher proportions of ineligible units than other types of samples. Ineligible units in telephone samples typically include nonworking numbers, business and government numbers, numbers outside the survey area, and numbers for households with no eligible members. Ineligible units in samples based on household listings, on the other hand, tend to include demolished or unoccupied dwellings and households whose members fail to meet the eligibility requirements.

When a survey includes both a screening step and a more detailed data collection step, eligibility is usually defined differently for the two steps. If the screening is done by telephone, an unit eligible for the screening interview is any working residential number in the exchanges that define the sampling frame. If the screening is done in person, an eligible unit is any occupied dwelling unit in the designated areas. More stringent criteria may apply to the detailed data collection. Units may have to meet geographical or other requirements to be retained for the collection of detailed travel data.

Some travel surveys report the number of units in each ineligibility category for purposes of evaluating the quality and properties of the sampling frame. Because response rate calculations treat all ineligibles the same, that level of detail is unnecessary for purposes of measuring nonresponse.

Estimated Eligibles vs. Estimated Ineligibles

The second level of classification also breaks down the eligibility unknown category into estimated eligibles and estimated ineligibles. The estimates are typically based on the relative proportion of eligible and ineligible units in the rest of the sample.

LEVEL III: Complete vs. Incomplete Units

The third level divides known eligible units into responding (complete) and nonresponding (incomplete) units. Respondents include all eligible units who have provided "useable" survey information. Nonrespondents include all other known eligibles in the sample.

Whether a unit's response to a travel survey is useable depends on that survey's definition of "completeness." Eligible units supplying partial information qualify as respondents when their level of response satisfies that definition.

At the screening phase of a survey, completeness is usually defined only at the household level; the household is classified as a respondent (as a complete) if enough information is obtained to classify the household as eligible for more detailed data collection.

When the measurement of nonresponse is at the person level, completeness is defined in terms of the items in the questionnaire. Data may be missing for individual items or for entire trips or activities. If enough data are missing, the analyst may decide that the diary or questionnaire is no longer useable. Some studies define a set of "critical items," items that must be answered for the questionnaire to be counted as complete.

When the measurement of nonresponse is at the household level and all eligible members within a household are asked to participate in the survey, completeness is defined in terms of the number or proportion of eligible members who provide useable data.

Some surveys further divide respondents into "refusal conversions" and "others" for purposes of evaluating the success of conversion efforts. Response rate calculations do not require that information.

LEVEL IV: Eligible for Interview, Ineligible for Interview, Refusals, Breakoffs, Noncontacts, and Others

The final level of classification includes all respondent and nonrespondent units in the sample, with the exception of units in the unknown eligibility category. Although the categories in this level are not required for purposes of computing screening response rates, they enable analysts to assess the relative contribution of various sources of nonresponse to the overall level of nonresponse in the sample. This information is often used to identify strengths and weaknesses in the data collection procedures.

The number of categories in this level depends on the design and reporting requirements of the particular survey. In travel surveys consisting of two phases of data collection, for example, respondents who have completed the screening process are divided into two groups, those eligible for the main interview and those not eligible for the main interview. It is often desirable to divide nonrespondents into “refusals to participate,” “breakoffs,” and “failures to contact” for purposes of evaluating the effectiveness of recruitment and contacting procedures.

Although the classification in **Figure 1.1** includes four categories at Level IV, the number of categories may be expanded or collapsed to meet the evaluative and reporting needs of the survey. The categories in **Figure 1.1** are defined as follows:

- **refusals** are eligible units who declined to participate;
- **breakoffs** are eligible units who began to provide data but failed to complete the survey;
- **noncontacts** are eligible units who could not be reached;
- **others** includes all other nonrespondents not included in the categories above.

These same categories are suitable for survey designs that include a screening and a main interview conducted at different points in time. Units included in the sample for both phases of data collection would be classified in both data collection efforts. For example, a case could complete the screener but refuse to take part in the main interview. Generally, of course, cases that were nonrespondents or ineligibles at the screening phase are not included in the sample for the main data collection.

The distinctions among categories in this level also depend on the conventions adopted by the particular survey. In the next section, we provide guidelines for assigning units to categories in travel surveys.

CLASSIFICATION RULES

Recommendation

RULES FOR CLASSIFYING UNITS—
The rules for assigning units to categories should be consistent across travels surveys to allow for meaningful comparisons of response rates.

Unless the rules for assigning units to categories are consistent across surveys, the measures of nonresponse obtained in the studies will have different meanings. Below we discuss the set of recommended conventions for standardizing the assignment of units to categories in travel surveys. Although the examples are largely drawn from surveys in which the data are collected by telephone, the rules are suitable for all types of travel surveys and for all modes of data collection, including mail, telephone, personal interview, and mixed-mode designs.

In a survey with an initial screening phase followed by the collection of detailed travel data, the assignment of cases to categories must be carried out twice, once during each phase. Since the main data collection is generally restricted to cases whose eligibility has been established, the steps dealing with the treatment of cases whose eligibility is unknown may be irrelevant (or greatly simplified) during the main data collection phase.

Recommendation

SURVEYS WITH A SCREENING PHASE—
Units should be assigned two disposition codes when there is both a screening and main data collection phase.

1. During Screening, Assign Noncontacts to the Eligibility Unknown Category (Level I)

Cases never contacted during screening should be assigned to the eligibility unknown category, unless the eligibility of the unit can be reliably determined in some other way. Numbers answered by a fax or modem, for example, belong in this category, unless their status as dedicated fax/modem or business lines is established with some degree of certainty, in which case they belong in the ineligible category.[1]

[1]Many surveys subject such numbers to a verification process in which each number is called at various times during the day and evening on weekdays and weekends. If the number is consistently answered by fax or modem, it is assigned to the ineligible category. If the number turns out to be a non-dedicated residential line, the household is screened in the normal way. If the verification process is still in progress at the end of the field period, the number is assigned to the eligibility unknown category.

Examples of noncontact dispositions typically reported in telephone travel surveys are shown in **Table 1.4**.

Table 1.4
Dispositions in the Eligibility Unknown Category in Surveys Conducted by Telephone

DISPOSITIONS
NONCONTACTS DURING SCREENING
Rings, no answer
Busy signal
Closed out after repeated attempts to contact
Non-verified fax/modem line

Table 1.5 displays similar types of dispositions for surveys conducted by personal interview.[2] Because cases whose eligibility is undetermined are not usually fielded for the main data collection, the eligibility unknown category is likely to be nonexistent or small during the main data collection.

Table 1.5
Dispositions in the Eligibility Unknown Category in Surveys Conducted by Personal Interview

DISPOSITIONS
NONCONTACTS DURING SCREENING
No one at home
Household inaccessible

2. Assign All Units Belonging to the Target Population to the Eligible Category (Level II)

The assignment of units to the eligibility category should be based on the definition of the target population, rather than on the ability of the survey to accommodate the special needs of persons within that population. For example, if the target population for a survey consists of all households located within a one mile radius of a railway station, households that satisfy that definition but are unable to participate in the survey because of language barriers or other limitations should be assigned to the eligible category. The ineligible category should be reserved for units who do not belong to the target population. Examples of dispositions from the screening phase for personal

[2]The "household inaccessible" disposition covers situations where the interviewer is unable to gain access to the housing unit most commonly because the unit is part of a secure building. Other cases that fall in this category include units that are inaccessible due to natural disasters, such as floods and earthquakes.

travel surveys that belong in the ineligible category are shown in **Table 1.6**. Examples of units ineligible for the main data collection are also shown there.

Table 1.6
Ineligible Units in Telephone Surveys

DISPOSITIONS
SCREENING PHASE
Nonworking/Unassigned telephone number
Number changed to new number
Institutional residence/Group quarters
Business/government number
Other non-residence
Disconnected number
Verified fax/modem line
MAIN DATA COLLECTION PHASE
Screened residence, out of study area
Screened residence, unit does not qualify

3. Define Respondents in Terms of a Set of "Critical Items" (Level III)

During the screening phase, households are defined as respondents (completes) according to whether they provide enough information to classify the household as eligible for the main data collection.

During the main data collection, both persons and households may be classified as respondents according to whether they provide useable information. At the person level completeness should be defined in terms of a set of critical items that must be answered for a case to be classified as a respondent. The set of critical items should include all questions that are essential to accomplishing the major goals of the survey. In travel surveys, critical items typically include questions about travel behavior and questions about the determinants of that behavior.

Although the items designated as critical will depend in part on the goals of the particular survey, all travel surveys should include the following set as core items for defining completeness at the person level, unless they play only a minor role in the survey:

- Employment status (employed full-time, part-time, or not employed)
- A geocodable home address
- Total number of trips during the survey period
- Information on each trip:

a) purpose of the trip, b) addresses or identifiable place names for the origin and destination of the trip, c) travel time, d) mode of transportation, e) names of all household members on the trip, and f) number of non-household members on the trip.

In recent years, there has been a growing trend towards conducting activity-based surveys. In such surveys, the critical items would include a set of essential items for each activity rather than for each trip.

Items that typically play a less central role in analysis of travel data, such as the make and model of the automobile, and fees paid for parking, should not be used to establish completeness at the person level unless they are an integral part of the particular survey. In that case, they should be added to the list above as should any other items that are critical to accomplishing the goals of the survey.

The purpose of these rules for defining completeness at the person level is to ensure comparability of response rates across surveys. They are not intended to set a standard for how the data should be analyzed. In some situations, the criteria for determining which cases to include in an analysis may differ from those discussed above.

Recommendation

SURVEYS WITH A SCREENING PHASE—
Units should be assigned two disposition codes when there is both a screening and main data collection phase.

4. During the Main Data Collection, Classify Households as Respondents If Useable Data Are Obtained from the Majority of Eligible Members and for All Critical Items Sought at the Household Level (Level III)

When multiple persons within a household are asked to provide detailed data, completeness at the household level can be defined in at least three distinct ways: 1) households could be counted as respondents if useable data are obtained from all eligible members, 2) households could be counted as respondents if useable data are obtained from the majority of eligible members, and 3) households could be counted as respondents if useable data are obtained from at least one eligible member. The definition used can make a substantial difference in the reported response rate. For example, the Nationwide Personal Transportation Survey (NPTS) uses the second criterion treating households as complete when data are obtained for most of the eligible household members. Within those households, however, data are not obtained for some 7 to 10 percent of the eligible persons. If the NPTS adopted the first criterion instead of the second, the reported household response rate would be lower by 5 percent or more. The second definition—that is, counting households in which useable data are obtained for the majority of eligible members—is the recommended procedure. However, data obtained from households in which the majority of members did not respond should not be discarded. Those data may be useful for some analytical purposes.

For a household to be classified as a respondent, the following critical items must also be obtained:

- Total number of persons in the household
- Number of vehicles
- Number of workers in household
- Household composition (age, sex, and relation to head of each person, plus whether each person has a driver's license)
- A mailing address if diaries will be sent to the household.

This requirement is in addition to the requirement that a majority of household members provide useable information at the person level.

5. Assign Nonrespondents to Categories According to the Reason for Their Nonparticipation (Level IV)

During each phase of data collection, all eligible units who declined to participate should be assigned to the refusal category. Units who provided partial information but did not complete the interview should be classified as breakoffs for that phase of data collection. Eligible cases that could not be reached (including answering machines that were determined to be of households) should be classified as noncontacts.

6. Assign Nonrespondents with Language Barriers or Other Limitations to the "Others" Category (Level IV)

All nonrespondents who were unable to participate because of language barriers or other limitations should be assigned to the "others" category.

AN APPLICATION OF THE CLASSIFICATION SYSTEM TO TRAVEL DATA

In this section, we apply the classification system to data from a fictitious study—the Anytown, USA Travel Survey—to illustrate how units are assigned to categories based on their disposition codes. In the next section, we use the counts in the various categories to illustrate the calculation of response rates. We begin by describing the design and purpose of the Anytown, USA Travel Survey.

The Anytown, USA Travel Survey. The Anytown Travel Survey (ATS) was designed to provide up-to-date information on travel activities and patterns in the Anytown area. The target population for the survey consisted of households in three counties—K, L, and M A household was defined according to the U.S. Bureau of the Census definition.

The sampling frame for the study consisted of listed and unlisted telephone numbers in the survey area. The numbers were generated using a random-digit dialing (RDD) procedure and drawn in proportion to their concentration in the various parts of the survey area. The survey also included an oversample of transit users. The oversample was drawn from telephone numbers collected during an intercept survey of that subpopulation.

Data collection consisted of two phases, a preliminary screening/recruitment interview and a data retrieval interview. Both interviews were conducted by telephone. The purpose of the screening interview was to determine the eligibility of the household, to secure the cooperation of eligible members, and to collect background information on the household. The purpose of the retrieval or main interview was to collect travel and activity information from the eligible units who agreed to participate during the screening interview. In between the first and second interviews, all cooperating households were assigned activity and travel days and provided with diaries for recording their behavior.

Classification of Units in the Anytown, USA Travel Survey. The final dispositions of all 12,568 units in the screening sample are shown in **Table 1.7**. The unit of measurement is the household. The units are sorted into three broad categories—eligible (***A***), ineligible (***B***), and eligibility unknown (***C***)—based on their final dispositions. The eligible category for the screening phase includes six different dispositions:

- **eligible completes** (***a***) are households who provided useable information for the screening phase and met the eligibility requirements for the main data collection phase;
- **ineligible completes** (***b***) are households who completed the screening interview, but failed to qualify for the main data collection phase because they lived outside of the study area, etc.;
- **refusals** (***c***) are households who refused to complete the screening interview;
- **terminations/breakoffs** (***d***) are households who terminated the screening interview sometime after their eligibility was determined, but before the interview was completed;

- **respondents not reached, including residential answering machines** (*e*) are households who could not be reached during the screening phase either because no one was ever at home or because the phones were consistently answered by machine or voice mail;
- **deaf or language barriers** (*f*) includes households who were eligible but unable to participate in the survey because of language barriers or other limitations.

Table 1.7
Sample Dispositions in the Screening Phase for Households in the Anytown, USA Travel Survey

DISPOSITION	FREQUENCY
ELIGIBLE UNITS	
Complete Units	2,777 (a)
Eligible completes	329 (b)
Ineligible completes (outside of study area, other)	
Incomplete Units	
Refusals	4,078 (c)
Terminations / breakoffs	60 (d)
Respondent not reached (including household answering machines)	300 (e)
Deaf or language barriers	213 (f)
TOTAL	7,757 (A)
INELIGIBLE UNITS	
Unassigned / Nonworking numbers	2,000
Disconnected numbers / Number changed	1,902
Business / Government numbers	455
Institutional / Group quarter / Other non-household residence	100
TOTAL	4,457 (B)
ELIGIBILITY UNKNOWN UNITS	
No answer after repeated callbacks	256
Busy signal on every call	98
TOTAL	354 (C)
TOTAL NUMBER OF UNITS IN THE SAMPLE	12,568 (D)

The ineligible category, on the other hand, includes: unassigned or nonworking numbers; disconnected, business, and government numbers; numbers that changed or were out of the survey area; numbers of institutional, group quarters, or other non-household residences; and non-household numbers where answering machines were reached. The unknown eligible category includes units that were never contacted.

Figure 1.2 shows how the screening dispositions in **Table 1.7** translate into the categories of the classification system. The number of units in the eligible and ineligible categories in **Table 1.7** have been summed to yield the count of units in the eligibility known category ($A + B$) in **Figure 1.2**. The eligibility unknown category (C) has been divided into estimated eligible (CE) and estimated ineligible units (CI) by multiplying the count in the unknown category by the proportion of eligible units in the eligibility known part of the sample.

Units in the eligible category (A) have been divided into completes ($a + b$) and incompletes ($c + d + e + f$) according to whether they completed the screening phase of the survey. Incompletes have been further divided into refusals (c), breakoffs (d), respondents not reached (e), and others (f). Units who terminated the interview before completion were placed in the "breakoffs" category. Units who were unable to participate in the first phase of the survey because of language barriers or deafness were assigned to the "others" category.

(**Figure 1-2** on following page.)

Figure 1-2
Classification of Units in the Screening Phase of the Anytown USA Travel Survey

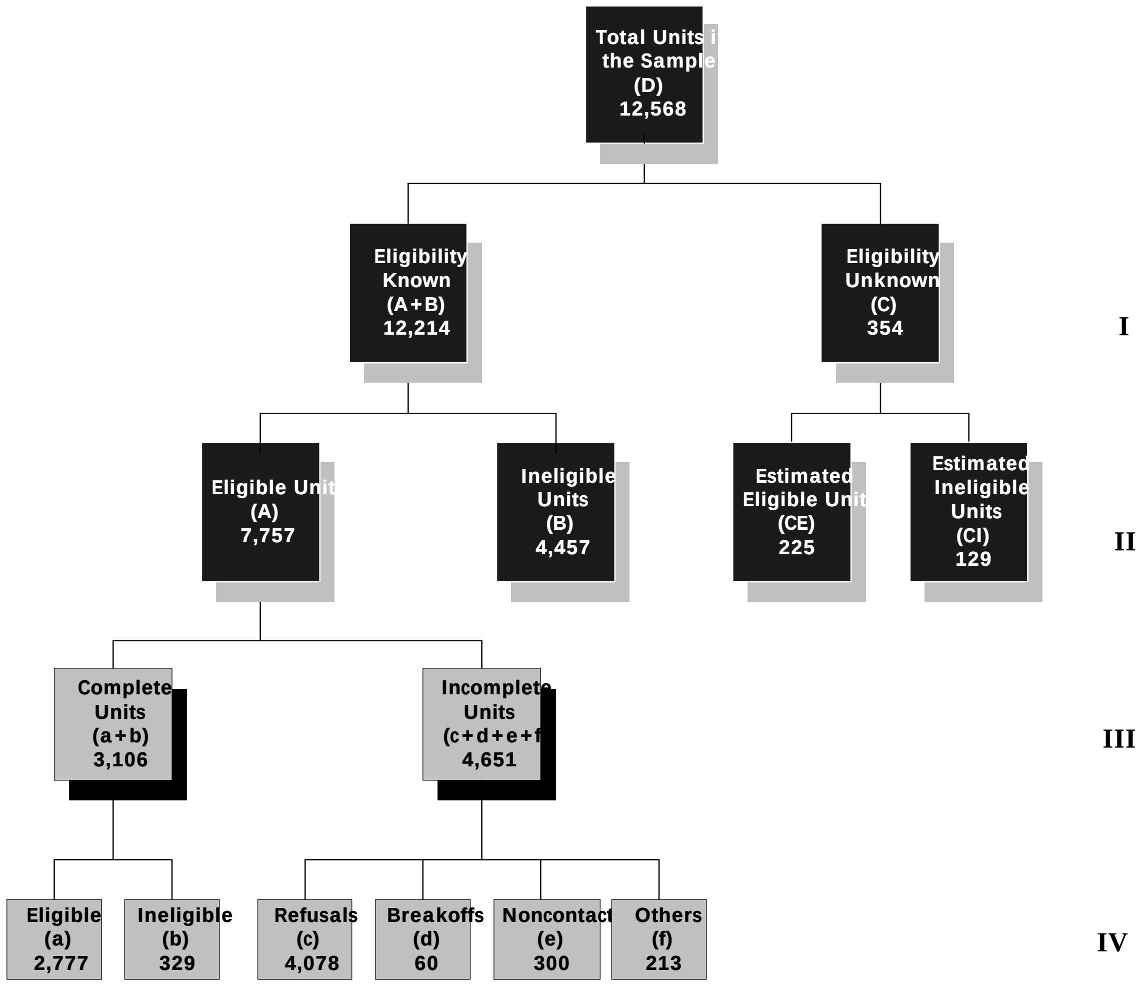


Table 1.8 shows the outcomes of the main data collection phase of the study. The 2,777 eligible households who completed the screening interview are retained for the main data collection effort. Of these, 1,716 completed the main data collection. Fewer categories are required to summarize the outcome of the field effort for the main data collection phase because the eligibility of the households in the main sample was established in the first phase of data collection. Of course, it is possible that some of the households retained for the second phase of data collection could have been ineligible (because they were misclassified during screening or because their eligibility changed between the completion of the screening interview and main data collection).

Table 1.8
Sample Dispositions in the Main Data Collection Phase for Households in the Anytown, USA Travel Survey

DISPOSITION	FREQUENCY	
ELIGIBLE UNITS		
Respondents	1,716	(a)
Refusals	776	(c)
Breakoffs/Partial data	115	(d)
Household not reached	75	(e)
Other	95	(f)
TOTAL	2,777	(A)
INELIGIBLE UNITS		
Screened out	329	
Other ineligibles	0	
TOTAL	329	(B)

1.6 CALCULATING RESPONSE RATES

Recommendation

FORMULA FOR COMPUTING RESPONSE RATES—
Travel surveys should follow the CASRO standards for computing response rates.

Once all units have been classified, it is relatively easy to calculate component response rates for each phase of data collection and an overall response rate for the survey as a whole. The procedures recommended by the Council of American Survey Research Organizations (CASRO) are described below [1]. The formulas are suitable for all modes of data collection, including telephone, mail, and personal interview.

CALCULATING COMPONENT RESPONSE RATES

The recommended procedures for reporting response rates in two-phase surveys consisting of a screening and a main interview follow the CASRO standard for breaking down the overall response rate into two component parts, one for each phase of data collection. To avoid confusion between the two phases of data collection, we will use the subscript 1 to indicate counts from the screening phase of the study (e.g,a_1, b_1, A_1) and the subscript 2 to indicate those from the main interview phase. In terms of our classification scheme of **Table 1.7**, the response rate for the preliminary or screening phase

(RR_1) is defined as the number of respondent units ($a_1 + b_1$) over the total number of eligible units in the sample ($A_1 + CE_1$):

$$RR_1 = \frac{a_1 + b_1}{A_1 + CE_1} .$$

The response rate for the main interview or second phase of the survey is similarly defined as the number of units completing the interview (a_2) over the number of known eligibles units (A_2) in the sample:

$$RR_2 = \frac{a_2}{A_2} .$$

The first response rate measures the survey's success in screening the entire sample. The second reflects the survey's success in interviewing all known eligible units in the sample. When the two rates are multiplied together, they yield the overall response rate, an indicator of the survey's success in measuring the entire sample:

$$RR = RR_1 \text{ x } RR_2 .$$

CALCULATING THE OVERALL RESPONSE RATE

The overall response rate is generally defined as the proportion of eligible members in the sample for whom complete data are obtained. This is the widely cited definition recommended by CASRO and the standard that should be followed in travel surveys.

One difficulty arises in applying this definition—how to deal with units whose ineligibility was never determined. When there are cases of unknown eligibility, CASRO recommends estimating the total number of eligible units based on the proportion found to be eligible among those units whose eligibility was determined.[3] First, the eligibility rates (ER) must be found. For the first stage of data collection, the eligibility rate is:

$$ER_1 = \frac{A_1}{A_1 + B_1} ,$$

[3]In telephone surveys relying on randomly generated telephone numbers this method will tend to overestimate the number of eligibles in the eligibility unknown category. Nonetheless, CASRO recommends this method in the absence of any better method for providing comparable estimates across surveys.

in which A_1 refers to the number of units found to be eligible and B_1 to the number found to be ineligible (see **Figures 1.1** and **1.2**, and **Table 1.7**). Similarly, the eligibility rate for the second phase of data collection is just the proportion of screened units that turns out to be eligible for the main data collection (**Table 1.8**):

$$ER_2 = \frac{A_2}{A_2 + B_2} \quad .$$

Then, the overall response rate (*RR*) is calculated as follows:

$$\begin{aligned} RR &= \frac{a_2}{ER_2 \text{ x } ER_1 \text{ x } D_1} \\ &= \frac{a_2}{ER_2 \text{ x } [A_1 + (ER_1 \text{ x } C_1)]} \quad , \\ &= \frac{a_2}{ER_2 \ (A_1 + CE_1)} \quad , \end{aligned}$$

in which a_2 refers to the respondent units (from **Table 1.8**), D_1 to the total number of units in the sample (from **Table 1.7**), C_1 to the number of units whose eligibility was never established (from **Table 1.7**), and CE_1 to the estimated number of eligibles in the eligibility unknown category (from **Table 1.7**). Of course, if the eligibility of all units in the sample is known, the calculation of an overall response rate is quite simple—it is the number of respondent units over the number eligible—a_2 / A_1 .

Recommendation

COMPONENT AND OVERALL RESPONSE RATES—
In surveys with both screening and main data collection phases, an overall response rate and response rates for each phase should be calculated and reported.

Illustrating the Formulas

To illustrate the application of these formulas, this section applies the equations to data from the Anytown, USA Travel Survey displayed in **Tables 1.7** and **1.8**, and in **Figure 1.2**. The counts of units in **Figure 1.2** serve as the data for the computations of the screening rate. Substituting the counts into the ER formula yields an eligibility rate of 63.51 percent for the first phase of the survey:

$$ER_1 = \frac{\text{known eligibles}}{\text{known eligibles} + \text{known ineligibles}} = \frac{A_1}{A_1 + B_1} = \frac{7{,}757}{12{,}214} = 63.51\% \ .$$

That is, an estimated 63.51 percent of the numbers in the sample were working residential numbers. Similarly, the eligibility rate for units that were screened is:

$$ER_2 = \frac{A_2}{A_2 + B_2} = \frac{2{,}777}{3{,}106} = 89.41\% \ .$$

Multiplying the screening eligibility rate (ER_1) by the number of units in the eligibility unknown category (354) provides the estimate of the total number of eligible units in the unknown category. Note the 225 shown in **Figure 1.2** refers to the total number of units eligible to be *screened*.

The overall response rate for the survey (*RR*) can then be computed in the following way:

$$\begin{aligned} RR &= \frac{\text{final completes}}{\text{estimated total number of eligibles}} \\ &= \frac{a_2}{ER_2 \text{ x } (A_1 + CE_1)} = \frac{1{,}716}{89.41\% \text{ x } (7{,}757 + 225)} = 24.04\% \ . \end{aligned}$$

Since the overall rate can also be derived from the component rates, the calculations above are unnecessary when separate response rates are computed for the screening and retrieval phases of the survey.

From the counts in **Figure 1.2** and **Table 1.7**, the response rate for the screening phase is 38.91 percent,

$$\begin{aligned} RR_1 &= \frac{\text{total number of respondent units}}{\text{estimated total number of eligible sample units}} \\ &= \frac{a_1 + b_1}{A_1 + (C_1 * ER_1)} = \frac{3{,}106}{7{,}982} = 38.91\% \ , \end{aligned}$$

while the response rate for the main phase of data collection is 61.79 percent (see **Table 1.8**):

$$RR_2 = \frac{\text{respondents}}{\text{known eligibles}} = \frac{a_2}{A_2} = \frac{1{,}716}{2{,}777} = 61.79\%.$$

Multiplying these component rates together (38.91 x 61.79) yields 24.04 percent, the overall response rate for the survey as a whole.

WEIGHTED VS. UNWEIGHTED RATES

Recommendation

WEIGHTING—
When the units in a sample have been drawn with unequal probabilities, it is necessary to weight the data in computing response rates.

One complication arises when the units in the sample are not selected with the same probability of selection. Some personal transportation surveys use samples in which all of the units are selected with the same probability. It is not generally necessary to weight the data from such samples. However, most travel surveys use designs in which members of different subgroups are selected at different sampling rates [2]. For example, in a multi-county survey, households from one county may be selected at a higher rate than households in the other counties. Similarly, the NPTS oversamples certain types of telephone numbers to increase the proportion that are working residential numbers. When the sample is not selected with equal probabilities, the data must be weighted to produce accurate estimates for the whole population of interest. In such designs, the cases may be weighted by the reciprocals of their selection probabilities. For example, a case with a selection probability of 1/1000 would receive a weight of 1000. (Chapter 3 discusses the weighting process in greater detail.)

Tables 1.9 and **1.10** below illustrate how weights can be calculated and why they should be used. In this fictitious example, the survey area encompasses three counties, one of them urban (County Y) and two of them rural (Counties X and Z). The sample includes a higher proportion of numbers from the rural counties than from the urban county.

Table 1.9
Hypothetical Sample Selected with Unequal Probabilities

COUNTY	SAMPLE SIZE	POPULATION SIZE	SELECTION PROBABILITY	AVERAGE DAILY TRIPS	RESPONDENTS	RESPONSE RATES
X	220	55,000	1/250	3.25	132	60.0%
Y	700	350,000	1/500	2.80	350	50.0%
Z	300	75,000	1/250	3.20	195	65.0%

In this simplified example, we assume that residents of Counties X and Z were sampled at twice the rate as the residents of County Y . (It is common to "oversample" smaller population subgroups in this way, so that there are enough cases to carry out separate analyses for each subgroup.) The response rates differed in the three counties, with a lower response rate in the more urban County Y area. In addition, residents of the two rural counties appear to make more trips per day on average than their counterparts in County Y. Unless weights are applied to the data, the estimates will be biased in the direction of the rural areas.

The data in **Table 1.10** illustrate how weights can be calculated to compensate for the oversampling of the rural counties relative to County Y. In the first step, we calculate an initial weight for each case that is the reciprocal of the selection probability. For example, cases in County Y were selected with a probability of 1/500; as a result, these cases receive an initial weight of 500. Then, in the next step, the weights are adjusted to compensate for differences across counties in the response rates. The initial weight is divided by the response rate to produce this adjusted weight. In County X, for instance, cases receive an adjusted weight of 416.67 (=250/.60). Notice that the sums of the weights are close to the population sizes. (In the example, the only discrepancy arises from rounding error.) The 350 cases in County Y each get a weight of 1000, for a total of 350,000. Collectively, the cases in each county now receive a weight proportional to—in fact, equal to—the county population.

Table 1.10
Weights for Hypothetical Sample

COUNTY	COMPLETED CASES	POPULATION SIZE	INITIAL WEIGHT	ADJUSTED WEIGHT	SUM OF WEIGHTS
X	132	55,000	250	416.67	55,000
Y	350	350,000	500	1000.0	350,000
Z	195	75,000	250	384.62	75,001

The weighted and unweighted statistics generally differ. In this example, the respondents from County X reported a total of 429 trips (132 respondents x an average of 3.25 trips). Similarly, those from County Y reported 980 trips (350 x 2.80) and those from County Z reported 624 trips (195 x 3.20). In total, then, there were 2,033 trips reported by the 677 respondents, for an unweighted average of about 3.00. By contrast, the weighted average is given by

$$\overline{V} = \frac{\sum W_i V_i}{\sum W_i} \quad ,$$

where V_i is the value of the variable of interest—in this case the number of daily trips—for respondent i -- and W_i is that respondent's adjusted weight. The weighted average comes to 2.91, a number that is closer to the average for County Y than the unweighted average. This reflects the fact that the County Y cases had the largest weights.

When the data are weighted, it raises the issue of whether to use weighted or unweighted data in calculating the response rates. Most statisticians agree that weighted rates give a better indication of the likely impact of nonresponse on the final survey estimates.

To calculate weighted response rates, the sum of the weights for the cases in a given category is used in place of the raw count. For example,*a* in the response rate formula would be the sum of the weights for the respondent units, rather than the raw number. The initial weight (prior to the adjustment for nonresponse) must be used to calculate the weighted response rate.

$$53.5\% = \frac{(132 x 250) + (350 x 500) + (195 x 250)}{(220 x 250) + (700 x 500) + (300 x 250)} \quad .$$

We can illustrate the computation of a weighted response rate using the data from **Table 1.10.** In this example, the overall unweighted response rate is 55.5 percent (677 respondents over 1220 selections). The weighted rate is somewhat lower at 53.5 percent: The numerator represents the sum of the weights for the respondents in each county (there were, for example, 132 respondents from County X, each with an initial weight of 250); the denominator represents the sum of the weights for the eligible selections.

2.

REDUCING NONRESPONSE

2.1 CHAPTER SUMMARY

This chapter describes the potential consequences of nonresponse in household travel surveys and recommends methods for reducing the level of nonresponse. It covers three main topics:

- Respondent and interviewer characteristics associated with nonresponse;
- Procedures that can reduce the level of nonresponse; and
- Procedures that can reduce the level of missing data in otherwise complete questionnaires.

As is customary, we distinguish between two forms of nonresponse—*unit nonresponse* and *item nonresponse*. Unit nonresponse refers to the failure to obtain questionnaires or data collection forms (such as the travel diaries used in many household travel surveys) for a member of the sample. Item nonresponse refers to the failure to obtain a specific piece of information from a responding member of the sample. Item nonresponse is often used interchangeably with the term *missing data*. This chapter contains recommendations for reducing the levels of both forms of nonresponse. Because most household travel surveys use some combination of telephone and mail data collection, it focuses on nonresponse issues for those modes of data collection.

MAJOR RECOMMENDATIONS

The chapter includes a large number of major recommendations based on research findings in the survey and transportation literatures, and on expert opinion in those fields. We summarize them here, grouping them by topic.

Interviewer Recruitment and Training

- Identify and utilize those interviewers and field supervisors who are especially good at converting reluctant participants.

- Use bilingual interviewers when the sample includes a substantial portion of non-English speakers.

Questionnaire Design

- Make the flow of questions in a self-administered questionnaire or diary clear by following the natural inclination of the respondents (e.g., to read from left to right) and by using graphical devices to emphasize the intended path.

Field Work

- Whenever possible, send out an advance letter to members of the sample, explaining the purpose of the study.

- Carry out multiple follow-ups—at least eight calls per number in a telephone survey and at least three follow-ups in a mail survey.

- Select a schedule that optimizes the timing of the contact calls given the length of the field period and the characteristics of the sample. Whenever the length of the field period is relatively short, select a schedule that attempts to produce an early contact through an intensive calling effort.

- Vary the appeals in a mail survey, using a special appeal on the third follow-up attempt.

- Use small prepaid monetary incentives, unless participation is especially burdensome, in which case use larger monetary incentives conditional on participation and commensurate with the level of burden.

Sampling

- Allow proxy respondents if they are likely to have the information sought.

- Subsample the remaining cases if the response rate is unsatisfactory and the budget is running out.

In addition, the report mentions several valuable procedures that are often considered standard survey practice. These include:

- Training interviewers about the purpose and sponsorship of the study, and providing prescripted answers to common questions about the survey;

- Including a brief description of the study in introduction to the questionnaire;

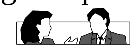
- Translating the questionnaire and all other survey materials when a substantial portion of the sample speaks a language besides English;

- Pretesting the field procedures;
- Pretesting the questionnaire to improve its clarity and estimate its length; and
- Placing sensitive questions at the end of the questionnaire.

Although all these procedures are *effective* for improving response rates, the question arises as to whether they are *cost-effective*. In some cases, studies have examined the impact of a particular procedure—such as the use of incentives—on the total costs of the study. The results suggest that incentives reduce costs by increasing the proportion of the sample responding to early contacts. In other cases, cost data are not available. Still, many of the procedures we recommend are quite inexpensive. An advance letter, for instance, is unlikely to add much to the total cost of data collection, but can have a substantial effect on response rates. Moreover, if the goal is to gather data for a fixed number of persons or households, it may be cheaper to follow up more extensively with fewer sample members than to start with a larger sample but accept a lower response rate. The recruitment of bilingual interviewers, the translation of questionnaires when a large proportion of the target population does not speak English, the design of user-friendly data collection forms, multiple follow-ups by mail or telephone, and the use of proxies (when they are likely to have the information being sought) are all relatively inexpensive steps that are likely to produce a reasonable payoff in terms of reduced nonresponse rates. Adopting special methods to convert hard-core nonrespondents, on the other hand, will also improve response rates, but it is likely to be costly if the number of nonrespondents in the sample is large. In such situations, a more cost-effective approach is to subsample nonrespondents and use special methods to obtain a high response rate in the subsample.

ORGANIZATION OF THE CHAPTER

The chapter begins by describing the characteristics of sample members who tend to be nonrespondents in travel and other surveys. As part of that discussion, it offers specific recommendations for improving their participation rates. The next section discusses characteristics of effective interviewers and strategies for selecting interviewers with desirable characteristics. The remainder of the chapter is devoted to a more general discussion of the various methods for reducing unit and item nonresponse.

Throughout the chapter recommended procedures for reducing nonresponse are enclosed within boxes following the text that discusses them. Whether a procedure applies to face-to-face, mail, or telephone surveys is indicated by the symbols, , , or , respectively.

2.2 RESPONDENT CHARACTERISTICS

Certain sample members are less likely than others to take part in a survey. In travel and other surveys, these individuals tend to include:

- the mentally and physically handicapped,
- individuals with language barriers,
- individuals with limited literacy skills,
- the less well educated,
- the elderly, and
- urban dwellers.

The reasons for the lower participation rates of some of these groups are clear. Individuals with visual difficulties, for example, are unlikely to complete a self-administered questionnaire if the print is too small for them to see. Individuals with language barriers, on the other hand, are unlikely to participate in a survey unless it is conducted in their native language. But apart from such obvious relationships, very little is known about the ways in which, and the extent to which, membership in these groups affects participation largely because direct information on nonrespondents is difficult to obtain.

Information on nonrespondents is important for three reasons. First, it can be used to assess the level of bias in the data due to the exclusion or underrepresentation of certain segments of the population. Second, it can be used to reduce nonresponse to the extent that the survey can be tailored to address the needs and concerns of the nonrespondent groups. Enlarging the print size of a self-administered questionnaire, for example, is likely to increase the participation rate of elderly individuals who often fail to complete questionnaires because they have difficulty reading small print. Third, it can be used to assess the potential payoff of tailoring the survey to specific groups. If the group represents a small portion of the population, the cost of the tailoring effort may outweigh the benefits. However, if the group includes a significant portion of the population, the benefits are likely to be substantial.

The sections below discuss what is currently known about nonrespondents and, on the basis of that information, offer specific recommendations for improving their response rates.

THE MENTALLY AND PHYSICALLY HANDICAPPED

A significant portion of the U.S. population have mental or physical handicaps that may prevent them from participating in a survey depending on what it requires of them [1]. About 7 percent of the adult population report trouble seeing standard newspaper print even with corrective lenses, a factor that is likely to affect their participation in surveys involving self-administered questionnaires. Roughly the same percent have trouble hearing what another person says in a normal conversation, and, as a result, may have difficulty participating in surveys conducted by telephone. Another 3 percent indicate they have a learning disability. As discussed below, these individuals are likely to have difficulty completing a self-administered questionnaire.

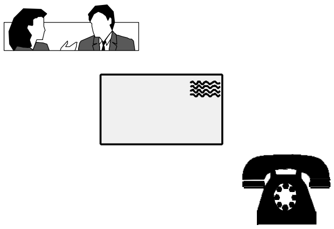
Recommendation

REDUCING NONRESPONSE DUE TO VISUAL DIFFICULTIES —
Use of a font size of 12 points or more for all written respondent materials if the sample is likely to include elderly individuals or persons with visual difficulties.

INDIVIDUALS WITH LIMITED LITERACY SKILLS

According to the National Adult Literacy Survey, the English literacy skills of nearly one-fourth of the U.S. adult population are extremely limited [1]. This portion of the population is able, at best, to read and understand written information when it is brief and uncomplicated. About 19 percent of these individuals who are at the lowest level of literacy report visual difficulties that interfere with their ability to read printed material. Nearly 25 percent of individuals with poor literacy skills are immigrants who may have limited skills in English. Almost two-thirds have less than a high school education. Another 9 percent report having a learning disability. As a whole, individuals at this level of literacy are likely to have difficulty completing all but the simplest questionnaires and, in some cases, even that will be beyond their ability to read and process material written in English.

Recommendation

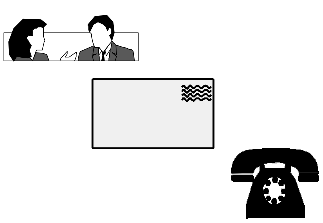


REDUCING NO NRESPONSE DUE TO LIMITED LITERACY SKILLS IN ENGLISH—
Prepare all written respondent materials at the lowest reading grade level possible. Ask all questions as simply as possible, make instructions easy to understand, and follow recommended procedures for questionnaire design (page 2-14). Have a group of experts evaluate the readability of all documents, including self-administered questionnaires, travel diaries, advance letters, correspondence, and all other materials read by respondents.

INDIVIDUALS WITH LANGUAGE BARRIERS

The ethnic and racial composition of the U.S. population has changed rather markedly in recent years [1]. The portion of the population who are of Hispanic origin, for example, grew from about 6 to 9 percent over the past decade, while the percent of individuals of Asian or Pacific Islander descent rose from about 1 to 2 percent in the same period. Recent data also indicate that nearly 10 percent of adults in the U.S. were born in other countries or in U.S. territories, and that roughly 17 percent speak a language other than English. With these changes in the composition of the population, surveys are likely to encounter even larger numbers of individuals with limited proficiencies in English who are unable to participate unless the survey is conducted in their native language.

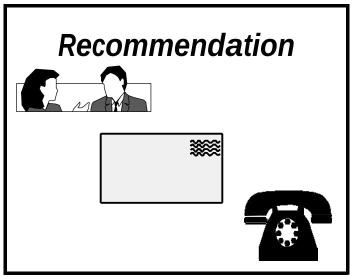
Recommendation



REDUCING NONRESPONSE DUE TO LANGUAGE BARRIERS—
Whenever the size of a non-English-speaking subgroup is expected to exceed 5 percent of the population under study and the sample is expected to include 150 or more members of the subgroup, translate the survey instruments and materials into the appropriate language. Whether the interview is conducted in person or by telephone, use bilingual interviewers who are fluent in English and in the other language.

THE LESS WELL EDUCATED

Individuals with fewer years of formal schooling tend to participate in surveys at lower rates than the rest of the population [2]. That the less well educated tend to be less proficient in English is likely to contribute to their lower participation rates.



IMPROVING THE RESPONSE RATE OF THE LESS WELL EDUCATED—
Follow the recommended guidelines for persons with limited literacy skills in English.

THE ELDERLY

There is considerable evidence in the literature that elderly individuals participate in surveys at lower rates than individuals in other age groups [2,3]. There is some support in the literature that visual and hearing problems, which tend to be more prevalent in the elderly, account for the lower response rates of this group. The lower participation rates may also be related to years of formal education and literacy level, both of which tend to be lowest in the elderly population.

In transportation studies, the lower completion rates of the elderly are probably due, at least in part, to their lower rates of travel. Anecdotal information suggests that the elderly frequently fail to fill out or return travel diaries because they either have no trips to report or feel that the survey is not relevant to them.

IMPROVING THE RESPONSE RATE OF THE ELDERLY —
Follow the guidelines for persons with limited literacy skills and persons with visual difficulties. In contacts with respondents, stress the importance of collecting information from everyone in the sample, including persons with no or few trips to report.

URBAN DWELLERS

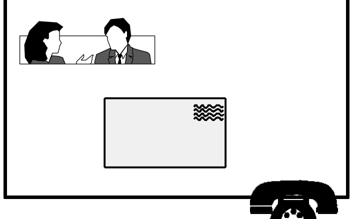
Individuals living in densely populated areas are less likely to participate in surveys than individuals living in rural and other less populated areas regardless of whether the survey is conducted by mail or telephone [2]. This pattern is clearly reflected in data from the 1995 Nationwide Personal Transportation Survey shown in **Table 2.1**. In that survey, which was conducted by telephone, response rates in smaller, less densely populated areas—such as Utica and Elmira, New York—were about three times as high as rates in densely populated areas—such as the New York City boroughs of Queens and Manhattan. As a whole, response rates exhibited a linear relationship with population density. As density decreased, the relative proportion of sample members taking part in the survey increased.

Table 2.1
Response Rate by Population Density for Selected Areas in the States of New York, Massachusetts, and Oklahoma

DENSITY *	RESPONSE RATE
20239	14.5
12594	23.3
11058	25.0
6887	21.9
2496	24.7
1733	27.5
780	30.5
629	30.8
560	33.6
332	32.6
106	44.4
103	35.1
90	50.9
87	42.1
83	49.0
57	38.8
54	43.0
46	51.1
38	44.4

**Number of persons per square kilometer*

Recommendation



IMPROVING THE RESPONSE RATE OF URBAN DWELLERS — *In addition to following the general guidelines for interviewer training and selection, instrument design, and field procedures: (1) delay rostering household members until the end of the screening interview (page 2-14), (2) accept proxy reports of basic background information (page 2-31), (3) use interviewers especially good at converting the reluctant, (4) optimize timing of contact calls, and (5) allow for multiple follow-up attempts (page 2-17).*

OTHER GROUPS OF NONRESPONDENTS IN TRAVEL SURVEYS

In addition to the groups above, nonrespondents in travel surveys tend to include [4]:

- couples with young children, single parents, individuals who hold multiple jobs, and other busy or hard-to-reach individuals, such as 18 to 24 year-olds, and

- individuals who lack the community or civic ties (e.g., renters), which often motivate participation in travel surveys.

The participation of these and the other nonrespondent groups can often be improved through use of special methods, such as increased incentive amounts and special appeals. Such methods are discussed in “Special Methods for Reducing Nonresponse,” pages 32-33 of this chapter.

2.3 INTERVIEWER CHARACTERISTICS

In most surveys, some interviewers are more successful than others in securing the cooperation of sample members. The extent of these differences varies from survey to survey depending on the group of interviewers included in the effort. In one telephone survey, for example, the most successful interviewer achieved a response rate of 94 percent while the least successful interviewer achieved a rate of 58 percent. In two other telephone surveys, response rates for individual interviewers ranged from about 72 to 90 percent and from about 63 to 95 percent. While response rates vary considerably among interviewers, rates achieved by individual interviewers tend to be remarkably consistent across surveys; interviewers with certain characteristics tend to be more effective than others at recruiting sample members.

This section summarizes what is known about the relationship between interviewer characteristics and response rates. This information is important for purposes of reducing nonresponse to the extent that interviewers can be selected or trained to have characteristics associated with high response rates.

INTERVIEWER EXPERIENCE

One of the most consistent findings in the literature is that more experienced interviewers achieve higher response rates than less experienced interviewers. Since better interviewers typically stay at their jobs longer than poorer interviewers, part of this effect is due to self-selection. However, novice interviewers tend to improve over time as they gain more experience in recruiting sample members, suggesting that effective strategies for recruiting sample members can be learned [2,5].

Recommendation

INTERVIEWER EXPERIENCE —
To the extent possible, use experienced interviewers with strong track records in recruiting sample members.

INTERVIEWER GENDER

A number of studies in the survey research literature report a gender difference in response rates favoring female interviewers. A variety of explanations have been offered to account for this result. The explanation receiving the most support is that males attain lower response rates because they have less interviewing experience, on average, than their female counterparts; male and female interviewers with the same level of experience tend to be equally effective in securing the cooperation of sample members [6].

INTERVIEWER EXPECTATIONS

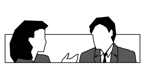
Interviewer expectations appear to contribute to an interviewer's level of success in securing the cooperation of sample members. Interviewers who perceive the task of recruiting sample members as "moderately easy" are more successful at gaining cooperation than interviewers who believe the task is "moderately difficult " [7]. That the task is probably easier for better interviewers could, however, account for this relationship between expectations and response rates.

INTERVIEWER STYLE OF INTERACTION

Other research indicates that successful interviewers tend to be more responsive and adaptive than less successful interviewers. They are likely to address a respondent's reluctance to participate in a survey in a confident and professional manner, and to adapt their approach to the situation as unexpected issues and problems arise [8,9].

Recommendation

STYLE OF INTERACTION —
Train interviewers in effective procedures for persuading reluctant participants and averting nonresponse. Stress the importance of using an adaptive approach that addresses the concerns and objections of individual sample members.




INTERVIEWER VOCAL CHARACTERISTICS

The vocal characteristics of a telephone interviewer can have an impact on response rates. Research examining this relationship reveals the following patterns:

- Interviewers who speak loudly, at a fast pace, and with a standard American accent are likely to achieve higher response rates than interviewers without those vocal characteristics [10].

- Interviewers who use falling intonations (that is, a deeper pitch) on important words in the introductory statement are likely to achieve higher cooperation rates than interviewers who use rising intonations (a higher pitch) [10].

- Interviewers who speak loudly and at a rapid pace with a standard accent and falling intonations are perceived as being more competent and confident than interviewers with less effective vocal attributes [10].

These results are in accord with psycholinguistic research indicating that individuals who talk loudly and rapidly are thought to be more persuasive, confident, and credible than individuals who speak more slowly and softly. A standard American accent may not be preferable for every sample; if the local population has a special accent, interviewers who have the same accent may achieve the highest levels of cooperation.

Recommendation

VOCAL CHARACTERISTICS —
Consider the vocal characteristics of the candidate when recruiting interviewers for a telephone survey.

OTHER INTERVIEWER CHARACTERISTICS

Personal characteristics such as motivation, tenacity, and self-confidence are also likely to contribute to an interviewer's level of success in securing sample members, but little research has been done in these areas.

2.4 METHODS FOR REDUCING UNIT NONRESPONSE

This section of the report examines different steps in the conduct of a survey that will effect the final response rate obtained. It covers:

- interviewer selection and training,
- the design of the survey questionnaire,
- the number of attempts made to contact members of the sample,
- the timing of contact attempts,
- pretesting the questionnaire and field procedures,
- rules for selecting the respondent,
- the use of respondent incentives,
- proxy respondents,
- special methods for converting nonrespondents, and
- subsampling of nonrespondents.

Each section discusses practical measures that can be taken at each step to reduce the level of nonresponse.

INTERVIEWER SELECTION AND TRAINING

Interviewer selection. In both telephone and face-to-face surveys, the interviewer plays a crucial role in contacting and enlisting the cooperation of sample members. As a result, most survey organizations attempt to recruit as interviewers persons who are likely to achieve high response rates. It is often thought to help, for example, if the interviewers are drawn from the same population as persons from whom they will be trying to collect information. In some surveys, the interviewers assigned to a specific member of the sample may match that member on race, sex, or other readily apparent demographic characteristics. The rationale for matching the interviewer and sample persons is not always clear, but at least in one instance matching is likely to yield a high payoff—when the sample includes a large proportion of members who are not native speakers of English. In these cases, it is often essential to recruit bilingual interviewers. Even when a translation of the questionnaire is available, a bilingual interviewer is likely to be more successful at contacting and winning the cooperation of sample members who do not speak English well than an interviewer who speaks only English.

Aside from the demographic characteristics of the interviewers, survey organizations may seek to recruit experienced interviewers or interviewers with a proven track record for achieving high response rates. In fact, some survey organizations maintain rosters of experienced interviewers who are especially successful at persuading reluctant sample members to take part in surveys. Such specialists are sometimes referred to as "converters;" they can play an important part in achieving a high overall response rate.

Recommendation

INTERVIEWER SELECTION —
Identify and utilize those interviewers and field supervisors who are especially good at converting reluctant participants.

Interviewer training. Interviewers must be thoroughly conversant with the study's sponsor, aims, and purposes. They must be trained on these topics in order to respond knowledgeably and confidently to the questions and objections of reluctant respondents. For this reason, most survey organizations train both new and experienced interviewers for each survey. The survey-specific training is likely to cover several topics related to nonresponse such as the following:

- the purpose and sponsorship of the study,
- the provisions for maintaining the confidentiality of the data,
- relevant statutes authorizing the study,
- answers to commonly-raised questions about the study and commonly-cited objections to participation (see Appendix 1 for an example of such questions and answers).

Sample members are more likely to agree to be interviewed when the topic is interesting to them, the study is important, the sponsor is prestigious or is seen as a legitimate authority [10,11]. Aside from covering these basic topics, the training may explicitly describe techniques for averting nonresponse or for persuading reluctant sample members to cooperate.

Recommendation

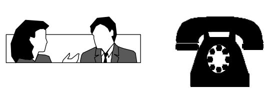
INTERVIEWER TRAINING —
Train interviewers about the purpose and sponsorship of the study. Provide them with prescripted answers to common questions about the survey and to common objections to participating in a survey.

DESIGN OF SURVEY INSTRUMENTS

The design of the questionnaires can have a large impact on the final response rates. In mail surveys, the questionnaire may be crucial in shaping sample members' impressions of the study and thus play a major role in determining whether they take part. Even in a telephone interview, where the questionnaire is not visible to the respondent, the design of the questionnaire may still affect the response rates.

Screening and recruitment scripts. In telephone screening and recruitment interviews, nonresponse often occurs within the first few seconds of the interaction. Thus, the key questionnaire design variables may involve the introduction to the questionnaire, the description of the study (including its purpose and sponsorship), and whether a translation of the questionnaire is available for sample members who have difficulty with English. Most survey organizations use brief introductions that cover such information as the name of the survey interviewer, the company for which he or she works, the name of the agency sponsoring the survey, and the study's purpose. In screening interviews (intended to identify persons eligible for more detailed data collection later on), it may be necessary to obtain a roster listing all household members and to gather information about each person. Respondents may be reluctant to give out detailed information about each family member in the first few minutes of the interview. In fact, some respondents may be reluctant even to list each person by name [13]. It may, therefore, be helpful to defer the rostering of household members to the end of the screening interview, after a modicum of trust has been established.

Recommendation



ROSTER OF HOUSEHOLD MEMBERS —
Postpone the rostering of individual household members until the end of the screening interview.

Self-Administered questionnaires and diaries. Additional design features are likely to become important with self-administered questionnaires and travel diaries. Among these are the apparent length and difficulty of the questionnaires [11,14]. Aside from the actual number and content of the questions, several factors may affect the potential respondent's estimate of the level of effort needed to complete the instrument:

- the sample members' level of interest in the topic;
- the use of photo reduction or other methods for reducing the apparent length of the questionnaire;

- whether or not the questionnaire is attractively designed;
- the flow and ease of following the instructions in the questionnaire;
- the logical grouping and sequencing of questions by topic or chronology.

According to many survey experts, sample members are more likely to complete the questionnaire if the topic is interesting and if the questionnaire is short, attractive, and easy to follow [15]. The findings on the impact of the length of the questionnaire are, however, not completely consistent. According to one study, adding 20 questions to a questionnaire reduces response rates only by an average of 1 percent [11]; according to another study, questionnaires longer than four pages are likely to produce worse response rates—about 8 percent lower—than questions under four pages in length [14]. Another study comparing travel diaries of different lengths found inconclusive results regarding the length of the diary [16]. In addition, efforts to make the questionnaire appear short may create problems. Reducing the type size used in the questionnaire may make it difficult for some members of the sample to read; a substantial portion of the adult population reports difficulty in reading standard news print.

These and other results suggest that respondents are sensitive to a number of features of self-administered questionnaires, all of which contribute to their sense of the overall level of effort needed. If the questions are simple and the instructions are easy to follow, respondents may be more willing to complete a long questionnaire than a short, but difficult one [17]. Unfortunately, what design features make it easy for respondents to complete a questionnaire is not well-established. The design of survey questionnaires is still more of an art than a science. Survey designers have proposed a number of principles for the design of self-administered questionnaires [17]:

- The questions should follow the natural reading order of the respondents (that is, in English, questions should flow from top to bottom and from left to right);
- Questionnaires should use familiar and easily-understood graphical conventions (such as arrows indicating the next question), in which the same design element always cues the same action by the respondent;
- The questionnaire should call attention to key information (via boldfacing and similar methods);
- There should be a clear path for respondents to follow through the questionnaire, and the graphical features should emphasize this path.

Appendix 2 contains contrasting examples of sections from two versions of a questionnaire. One version illustrates the application of the recommended principles. The other version does not follow those principles.

Once again, if language barriers are likely to be an obstacle to participation, questionnaires in the appropriate languages should be used.

Recommendation

DESIGN OF SELF-ADMINISTERED QUESTIONNAIRES —
Ask the questions as simply as possible and make the instructions easy to follow. Use boldfacing and similar methods to highlight key information. Use easily-understood graphic conventions to direct respondents, adopting a reading path that follows established customs for the language in which the questionnaire is written. In English, for example, make questions flow from top to bottom.

NUMBER AND TYPE OF CONTACTS

One of the clearest findings in the literature on nonresponse to surveys is the importance of repeated contacts with sample members. In general, fewer than half of all sample members return a questionnaire after the first mailing in a mail survey. Each additional follow-up contact is likely to yield additional returns [18]. Similarly, in the best telephone surveys, interviewers may attempt 8 to 10 callbacks before giving up on a sample telephone number. One study estimated that more than 10 percent of the telephone sample was reached only *after* the eighth call [19]. We recommend allowing for at least eight calls in planning a telephone survey.

Recommendation

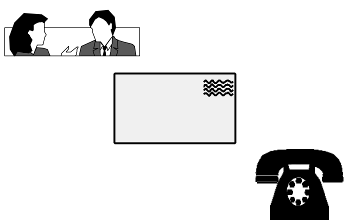
CALLBACK LIMIT—
Allow for at least eight (8) callbacks per sample telephone number when conducting a telephone survey.

Advance letters. Many face-to-face and telephone surveys begin by sending members of the sample prenotification, or advance, letters. Such letters are typically signed by a non-controversial government official or some other prestigious source; they explain the purpose of the study, describe the confidentiality measures to be employed, and alert the sample member that he or she will be contacted shortly by an interviewer. Advance letters accomplish several useful purposes. They help establish the legitimacy of the survey, inform potential respondents about the study's purposes, allay concerns about confidentiality, and serve as an introduction to the interviewer.

Advance letters have also proved valuable in both mail and telephone studies. With mail surveys, studies suggest that an advance letter can be quite useful in obtaining cooperation, increasing response rates by an average of about 17 percent [14]. Even with a telephone sample, it is possible to use advance letters to boost response rates. If the sample is selected from residential lists of numbers, address information usually accompanies the sample telephone numbers. If the telephone numbers are randomly generated, the sample of numbers can be matched to data bases with address information and advance letters can be sent to numbers for which addresses are found.[1]

Because of their proven success, advance letters have become common practice in household travel surveys in recent years.

Recommendation



ADVANCE LETTERS —
Send out advance letters to sample members explaining the sponsorship, purpose, and relevance of the study. Stress the importance of participation and the confidentiality of the results.

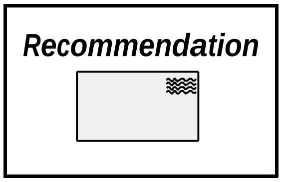
Follow-up contacts. For mail surveys, some experts have recommended specific regimes of follow-up contact efforts. Dillman's *Total Design Method*, for example, calls for an initial mailing and three follow-ups spread over a seven-week period [20]:

- a reminder postcard after one week has elapsed;
- a replacement questionnaire after four weeks; and
- a second replacement questionnaire sent by certified mail.

Although it is not clear that this exact schedule must be followed, the literature does indicate that multiple follow-ups can have a dramatic impact on the final response rates obtained. One study concluded that each contact in a mail survey (including an advance letter) increased response rates by an average of about 11 percent [11]. In line with Dillman's prescriptions, the literature also suggests that a special third follow-up contact—such as a mailing by certified mail or a telephone contact—does boost the final response rates, with an estimated effect

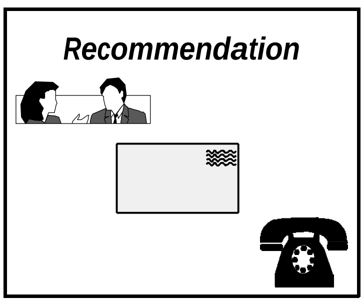
[1] At the national level, about 70 percent of all residential numbers are listed in telephone directories. The percent of listed numbers tends to be lowest in urban areas (about 50 percent) and highest in rural areas.

of about 7 percent [11]. In addition, first class mail and prepaid return postage may increase response rates relative to bulk mail.



FOLLOW-UP CONTACTS IN MAIL SURVEYS —
Allow for at least three follow-up contact attempts in a mail survey. Change the appeal on each contract. Use a special procedure, such as a certified mailing, on the last follow-up attempt.

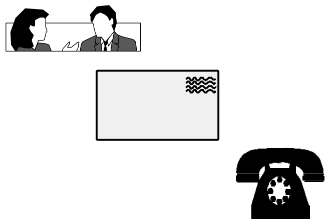
Length of the field period. One implication of the findings on the number of callbacks is that a high response rate will take time to achieve. Multiple callbacks are only possible (and will only be productive) if the data collection period is long enough. For example, the Total Design Method requires a field period of at least eight or nine weeks. Telephone surveys with eight or more callbacks require a field period of similar length. Surveys that must be completed too quickly to allow sufficient follow-up work are likely to have low response rates.



LENGTH OF THE FIELD PERIOD —
Whenever possible, make the field period long enough to permit multiple contact attempts for each member in the sample.

Respondent-initiated contacts. During the course of any survey, some respondents inevitably have questions about the data collection instruments or other aspects of the survey. It is therefore good survey practice to provide a telephone hot line, toll free if possible, that respondents can call at most anytime to obtain timely answers to their questions. The line should be answered by persons who are fully conversant with the purpose and design of the survey, and especially knowledgeable about the content of the questionnaires. A call-in service of this type helps improve response rates and data quality by eliminating misunderstandings, resolving ambiguities, and allaying respondents’ concerns. It also helps establish the legitimacy of the survey. A website or e-mail address can serve a similar function for respondents with on-line capabilities.

Recommendation



RESOURCES FOR RESPONDENTS —

Establish a telephone hot line that respondents can call almost anytime to obtain timely answers to their questions about the survey and its instruments.

TIMING OF THE CONTACT CALLS

Rationale. In telephone screening efforts, attempts to contact and determine the eligibility of sample members usually reach a point of diminishing returns after which additional callbacks no longer produce significant gains in the number of sample members screened. The previous section offers some guidelines for determining the optimal number of callbacks for contacting sample members in travel surveys conducted by telephone. More often than not, other considerations, such as the length of the field period and the availability of resources, determine the callback limit adopted by a survey.

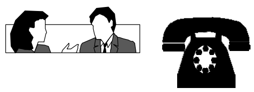
Whenever resources are limited, the survey's level of success in screening the entire sample will largely depend on the timing of the contact calls. The most effective timing for a particular survey will depend on such factors as the sampling frame, the population under study, the length of the field period, and whether the call is the first or a subsequent attempt to make contact with the sample member [21,22]. In most situations, and especially when resources are very limited, the strategy that will lead to the highest response rate for the screening phase is one in which the timing of the contact calls is designed to minimize the overall level of effort required to successfully screen sample members[2] [23,24,25]. In calling protocols of this type, the times are selected to maximize the chances of a successful outcome such that the overall number of calls required to complete the screening effort is kept at a minimum. Examples of such protocols are described below.

In situations where the length of the field period is especially short and there is not enough time to contact sample members at optimal times, the calling protocols must rely on other methods to maximize the response rate for the screening phase. The most effective protocols for situations of this type attempt to produce an early contact through intensive calling efforts [26]. Usually the initial contact call is scheduled for the first available time slot and followed up as quickly as possible with as many callbacks as time and resources allow. The

[2]For a discussion of how failures to screen sample members affect the overall response rate for a survey see the first chapter of this report.

most efficient protocols of this type attempt to optimize the timing of the first and subsequent calls to the extent possible within the time constraints of the survey. At a minimum, they vary the times at which callbacks are attempted for a given case. An example of such a protocol is presented below.

Recommendation



TIMING OF THE CONTACT CALLS —

Optimize the timing of the contact calls to the extent possible, given the length of the field period and the characteristics of the sample.

Optimizing the Timing of Calls. Protocols that optimize the timing of the contact calls vary in their complexity. Some are based on mathematical models or complex rules that require computer algorithms for their use. Others are designed for manual tracking and scheduling of calls. All are based on rules intended to increase the efficiency of the screening phase by minimizing the number of calls required to successfully screen sample members.

Table 2.2 summarizes some of the most effective rules for optimizing the timing of contact calls in household telephone surveys. The extent to which these rules can be applied in practice will depend on the resources and staffing capabilities of the survey organization, and on the amount of time available for the screening phase of the survey.

The first entry in the table refers to situations where the sample is drawn from a frame that consists mostly or entirely of household telephone numbers. Since all or most of the units in the sample will be residential units, the optimal times to initiate contact in samples of this type are days and hours when adult household members are likely to be at home. Research findings confirm the commonly held belief that the likelihood of reaching an adult member at home is highest on weekday evenings and weekends, and lowest on weekdays during daytime hours. Overall, weekday evenings tend to be more productive than weekends. The relative productivity of the individual weekday evenings tends to vary from study to study as does the relative productivity of weekend mornings, afternoons, and evenings.

Table 2.2
General Rules for Optimizing the Timing of Contact Calls

SAMPLE FRAME	MODE OF DATA COLLECTION	TIMING OF INITIAL CONTACT	TIMING OF CALLBACKS
Listed residential telephone numbers	Telephone	First available weekday evening or weekend time slot (defined as a weekend morning, afternoon, or evening)	First callback on a weekday evening or during a weekend time slot, preferably on a different evening or during a different weekend time slot than the initial call. Other callbacks may occur at any time.
Randomly generated telephone numbers	Telephone	First available weekday evening or weekend time slot	First callback on a weekday evening or during a weekend time slot, preferably on a different evening or during a different weekend time slot than the initial call. Other callbacks may occur at any time. At least one callback is scheduled during weekday business hours to screen out business numbers remaining in the sample
Randomly generated telephone numbers	Telephone Two-week field period	First available weekday business hour time slot (defined as a morning or afternoon)	First callback during evening hours on the day following the initial call. Second callback during business hours on the day following the first contact. Subsequent callbacks rotate between evening and daytime shifts. Each unresolved case is called daily until the field period is almost over. Field period ends in an intensified effort to screen all remaining noncontacts.

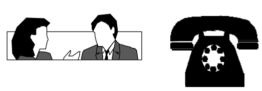
Recommendation

TIMING OF THE CONTACT CALLS FOR RESIDENTIAL TELEPHONE NUMBERS —
When the sample consists largely of residential telephone numbers, schedule the contact calls for evenings and weekends.

The best time to schedule the second contact attempt in samples of this type is also during evening and weekend hours, preferably in a different time slot (a different weekday evening or weekend morning, afternoon, or evening) than the first attempt. Subsequent callbacks can be scheduled for any available time slot since the timing of calls appears to have little effect on the chances of a successful contact at this point in the screening process. When differences are detected they favor evenings and weekends as the most productive calling times, and methods that vary the timing of the calls as the most productive protocols.

The productivity of these protocols, and those discussed below, can be improved by using information collected during previous calls to optimize the timing of subsequent calls. In telephone surveys, for example, the person who answers the phone may be able to provide information about when adult members of the household are likely to be at home. To maximize the chances of a successful outcome, the next call should be scheduled accordingly.

Recommendation



USING INFORMATION COLLECTED DURING PREVIOUS CALLS —
Use information collected during previous calls to optimize the timing of subsequent calls.

Determining the optimal times for initial calls and callbacks for samples drawn from frames that include a relatively large number of nonresidential and other out-of-scope units is more complicated since the optimal times for screening the various types of units in the sample will usually differ, and the effectiveness of any given strategy will depend in part on the relative proportion of residential and out-of-scope units in the sample. A variety of protocols have been developed to handle samples of this type. Some protocols are clearly more effective than others, but research in this area is rather limited and the most effective methods for such samples have yet to be identified.

To illustrate the type of rules associated with the more effective methods in the literature, **Table 2.2** shows two protocols developed for screening samples of randomly generated telephone numbers. Both protocols were developed for the same type of sample, but the rules they apply to control the timing of the calls are quite different. The first protocol gives priority to optimizing the times for household units in the sample. It assumes that the field period is of sufficient length to successfully apply this procedure. The second protocol gives priority to identifying out-of-scope numbers. It attempts to optimize the timing of all calls to the extent possible in an intensive effort to screen all cases within a two-week time frame. The protocol assumes that all telephone numbers identified as nonresidential numbers will be replaced with other randomly selected numbers, which must then be screened before the field period ends.

As mentioned above, the first protocol gives priority to screening the households in the sample. For the most part it follows the rules for screening samples of household units. It schedules the first and second calls for times when household members are likely to be at home. At least one of the remaining callbacks is scheduled during business hours to screen out any business numbers remaining in the sample.

The rationale behind this approach is threefold. First, priority should be given to contacting households in the sample since they are the units of primary interest. Second, most business and nonresidential numbers can be screened out by matching sample numbers against published business directories or through recorded messages or operator intercepts when calls are placed during nonbusiness hours. Third, screening out nonresidential and nonworking numbers during weekend and evening hours reduces the cost of the survey since telephone tolls tend to be lowest during those hours.

The second protocol, on the other hand, schedules initial calls during business hours to expedite the identification of out-of-scope units so that interviewers will have as much time as possible to screen replacement numbers. It schedules the first callback for the evening shift of the day following the initial contact to increase the chances of finding household members at home during the second call. Subsequent callbacks rotate between daytime and evening shifts. Each case is called once daily until the end of the field period approaches. At that time the field effort is intensified. Noncontacts are called more than once daily, sometimes within the same shift.

Recommendation

TIMING OF CONTACT CALLS FOR RANDOMLY GENERATED TELEPHONE NUMBERS —
Select a calling protocol that is likely to lead to the highest screener response rate given the characteristics of the sample and the length of the field period.

If resources are available, response rates obtained with other protocols for randomly generated numbers can be improved by ending the field period with an intensive calling effort of this type. Some methods used for this purpose include: 1) rotating the case through all available shifts (weekday and weekend, daytime and evening shifts) until a successful contact is made or the callback limit is reached, and 2) rotating the case through various time slots (morning, afternoon, and evening hours) within a particular shift, and continuing the procedure in other shifts until the callback limit is reached or the eligibility of the case is resolved [22]. In some protocols, the case may be called as often as every hour.

Recommendation

ENDING THE FIELD PERIOD —
When resources are available, end the field period with an intensive calling effort.

Special rules for telephone samples. To increase the efficiency and productivity of the calling effort, protocols for telephone samples often adopt special rules for certain types of numbers answered by machine or voice mail, and numbers resulting in a busy signal.

Since the majority of numbers answered by machine or voice-mail eventually result in a contact after repeated call attempts[3], most protocols treat such calls as "no answers" and call the numbers again until the callback limit is reached. The exception to this rule is when the recorded message indicates that they are businesses, other non-residences, or non-working telephone numbers, in which case they are classified as such and no further attempts are made [27,28].

The optimal number of callbacks for numbers answered by recorded message tends to be higher than that for other numbers and, as a result, some calling protocols adopt a higher callback limit for these numbers. The optimal limit is about ten calls, after which the chances of reaching a person are very small. Some protocols also adopt a special callback schedule that eliminates Saturday hours from the calling routine since Saturdays are particularly poor days to reach households with answering machines [27,28].

Recommendation

CALLBACK LIMIT FOR NUMBERS ANSWERED BY MACHINE OR VOICE MAIL —
Extend the callback limit to 10 when a number is consistently answered by machine or voice mail. Schedule callbacks to those numbers for days other than Saturdays.

[3] Even though 20 to 30 percent of households within the United States have answering machines or voice mail attached to their telephone numbers, only a small percent of these households consistently use the devices to screen unwanted calls.

Some protocols incorporate special rules for handling calls resulting in busy signals. The rules differ depending on whether a fast or regular busy signal is encountered. Callbacks for regular busy signals—tones that are interrupted at a rate of 60 times per minute and indicate that the line is in use—are usually scheduled for 15 to 60 minutes after the signal is encountered since the chances of reaching someone at home are typically highest during that period [29]. Callbacks for numbers resulting in a fast busy signal—tones that are interrupted at a rate of 120 times per minute and indicate difficulties due to storms or other problems, busy circuits in the calling area, or nonworking telephone numbers—are usually scheduled at varying times during business and evening hours over a period of days or weeks to determine the working or nonworking status of the number [22,29].

Recommendation

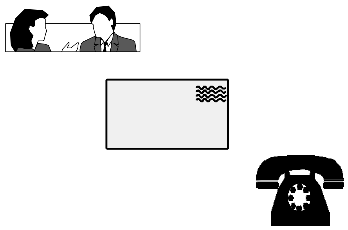
CALLBACK PROCEDURES FOR BUSY NUMBERS —
Whenever possible, call back numbers with regular busy signals within 15 to 60 minutes after the signal is encountered.

PRETESTING FIELD PROCEDURES AND INSTRUMENTS

Whether the survey is to be done over the telephone or by mail, pretesting the questionnaire itself and the field procedures to be used in the main data collection will almost always improve the response rates ultimately obtained. These pretests can take several forms. Many survey firms carry out a small number of interviews (10 to 20) using the proposed questionnaire and attempt to probe for specific problems in the questions, such as comprehension difficulties. In our discussion of techniques for reducing item nonresponse, we describe a variation on this procedure called the cognitive interview. Apart from leading to improvement in the questionnaire, a small-scale pretest can serve a second function, that of providing a realistic estimate of the time needed to complete the questionnaire.

In addition to small-scale pretests of the questionnaire, other types of pretests are useful as well. It is often worthwhile to conduct a dress rehearsal for the main study. Such dress rehearsals are useful for detecting problems in the planned procedures, for estimating the likely response rate in the main data collection, and for experimenting with procedural variations (such as the format of the questionnaire or the amount of incentive to offer). The size of the samples for such field tests often vary widely, depending on such factors as whether multiple procedures are being compared. A review of 55 recent transportation surveys indicated that more than 90 percent used pretest samples of 100 or fewer cases [30].

Recommendation



PRETESTING FIELD PROCEDURES AND INSTRUMENTS— *Conduct a small-scale pretest to test the survey instruments and the field procedures.*

INCENTIVES

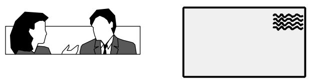
In many surveys, members of a survey sample are offered some form of incentive in an effort to gain their cooperation. In some surveys, the incentive accompanies the request for cooperation and consists of a small nonmonetary gift, such as a pen or a poster, or a small monetary gift of a dollar or two. In others, the incentive is substantially larger, conditional on participation, and offered as compensation for the respondent's time and trouble. The use of incentives is a widespread practice within survey research, especially for mail surveys (where incentives of a dollar or two are commonly employed) and in face-to-face surveys that make particularly onerous demands on the respondents, such as providing a biological specimen (where much larger incentives are the rule). This section provides guidelines for selecting effective incentives for travel surveys based on research in the survey and transportation literatures.

Offering incentives to sample members is one of the most effective procedures for reducing nonresponse in surveys. The effectiveness of a particular offering will depend on such factors as: 1) the monetary value of the incentive, 2) whether the incentive is paid in advance or upon completion of the survey, 3) whether the incentive is monetary or nonmonetary, 4) the mode of data collection—mail, telephone, or personal interview, and 5) the burden the survey places on the respondent [31, 33]. In situations where sample members are recruited by mail or telephone and asked to complete relatively brief self-administered questionnaires—as is the case in many travel surveys, the most effective incentives are usually:

- small in value,
- paid in advance, and
- monetary.

The sections below discuss these research findings in greater detail.

Recommendation



INCENTIVES—
Provide sample members with small prepaid monetary incentives of $2 or less, unless participation is especially burdensome. In that case, offer larger monetary incentives of $10 or more conditional on participation and commensurate with the level of burden placed on the respondent.

Effect of incentives on participation rates. Most research on incentives has focused on evaluating the effectiveness of small monetary offerings, typically $2 or less, or small gifts, such as a pen or a lottery voucher, in surveys where sample members are recruited by mail and asked to complete a self-administered questionnaire, or recruited by telephone and asked to complete a self-administered questionnaire at some later point in time. Surveys of this type generally place a low to moderate level of burden on the respondents; the amount of time respondents are asked to commit to the survey is relatively small, and their participation does not involve threatening or demanding tasks.

The effects of offering small incentives in such surveys have been evaluated under a variety of conditions, including situations where the incentive is paid in advance and sent to the respondent along with the survey materials, or conditional on participation and sent to the respondent after completion of the survey. In cases where respondents are recruited by mail, advance incentives are typically enclosed in the initial packet along with the request for cooperation and the survey questionnaire. In cases where respondents are recruited by telephone and asked to complete a self-administered survey as part of the study, advance incentives are usually sent along with the survey packet after respondents have agreed to participate, typically without knowledge of the incentive. Under these conditions:

- incentives paid in advance are generally more effective in gaining the cooperation of sample members than conditional incentives paid after completion of a survey, even when the promised amount is larger than the prepaid amount.
- incentives paid in advance typically increase response rates about 15 to 20 percentage points, even when the value of the incentive is 50 cents or less.
- monetary incentives are usually more effective than small gifts in gaining the cooperation of sample members.
- incentives conditional on participation often have a negligible or negative effect on response rates.

In situations where the survey places a higher burden on the respondent, larger incentives are often effective, if not required, for gaining the cooperation of certain sample members. Surveys belonging in this category typically place unusual or time-consuming demands on the respondent. They may, for instance, require the respondent to travel to a specific location, to submit to a physical examination, or to complete a battery of tests. In most surveys of this type, the incentives are offered as compensation in return for the respondent's time and effort. In the National Assessment of Educational Progress household survey, for example, youth were asked to complete a battery of four tests and offered $0, $10, $15, or $20 for finishing one, two, three, or four tests. The incentives increased response rates from about 71 percent (no incentive) to 83 percent [33]. Other surveys placing similar demands on respondents show comparable gains in response rates associated with incentives conditional on survey participation. In the 1971 National Health and Nutrition Examination Survey, where respondents were asked to submit to a physical examination, a conditional incentive of $10 increased the response rate from 70 to 82 percent.

Effect of incentives on data quality. Most studies examining the impact of incentives on data quality report a positive effect. Respondents who are offered incentives tend to provide more complete data than other respondents; item nonresponse is usually lower and answers to open-ended questions more complete. Moreover, respondents receiving a prepaid incentive are also more likely to provide complete data than respondents receiving a conditional incentive [32].

Effect of incentives on survey costs. The few studies in the literature evaluating the impact of incentives on survey costs report reductions in the average cost per case when incentives are used [32]. In the field test for the National Adult Literacy Survey, for example, the average cost per case for sample members offered conditional incentives was less than that for members in the no incentive treatment condition (see **Table 2.3**). Participants who were offered incentives also required fewer interviewer contacts on average than participants in the no incentive condition [34]. Other studies report similar reductions in the average number of contacts due to incentives.

Table 2.3
Incentive Amount, Number of Contacts, and Cost Per Case in the Field Test for the National Adult Literacy Survey

INCENTIVE AMOUNT	AVERAGE NUMBER OF CONTACTS PER CASE	AVERAGE COST PER CASE (INCLUDING INCENTIVE)
$ 0	6.7	$130
$20	5.3	$119
$35	5.0	$129

Incentives in transportation studies. **Table 2.3**, adopted in part from a report by Melissa Tooley [35], summarizes the results of some recent studies designed to evaluate the impact of incentives on response rates in household travel surveys. Overall, the results of this work support the conclusions drawn above, which were based on findings in the survey literature as a whole.

In accord with the survey literature, **Table 2.3** shows that promised incentives of larger value tend to yield mixed results in travel surveys. In the San Francisco Study, for example, a conditional incentive of $10 *reduced* the response rate by about 6 percentage points. In the Houston-Galveston Study, on the other hand, a conditional incentive of $10 increased response rates, but only by 6 percentage points.

Travel studies also support the general finding that small prepaid incentives are usually more effective than promised incentives. In the Puget Sound Transportation Panel, for example, a prepaid incentive of $1 yielded a higher response than a promised incentive of greater value ($10). The prepaid incentive increased response rates by about 15 percentage points, while the conditional incentive of $10 increased participation by 11 percentage points.

The results from the Pretest of the Dallas-Fort Worth Household Travel Survey, also shown in **Table 2.4**, support the general finding that monetary incentives are usually more effective than nonmonetary gifts in improving response rates. In that study, a prepaid incentive of $2 led to higher response rates than a gift of a pen in three out of four cases. The differences in participation favoring the $2 incentive ranged from about 8 to 18 percentage points.

Table 2.4
Response Rates by Incentive Condition Comparisons in Five Transportation Studies

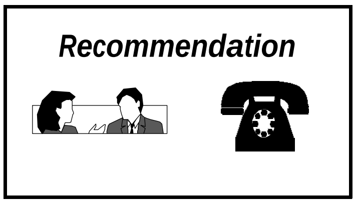
STUDY	YEAR	ADVANCE LETTER	FOLLOW-UP	RETRIEVAL	INCENTIVES	RESPONSE RATES
Puget Sound Transportation Panel	1989	Y	—	Mailback	None $1 (pre) $10 (post)	49% 64% 60%
San Francisco Study	1990	Y	—	Telephone	None $10 (post)	47% 41%
San Antonio Bexar County, Amarillo, and Brownsville	 1990 1990 1991	Y	Y	Mixed	None $5 (post)	28-39% 40%
Houston/Galveston	1984	Y	Y (Multiple)	Mailback	None $10 (post)	32% 38%
Dallas-Fort Worth Pretest	1995	N	Y	Telephone	Pen (pre) $2 (pre)	35%-48% 45%-60%

Note: “Pre” incentives are those included with the initial survey packet; “post” incentives are those offered upon completion of the questionnaires. Response rates may not conform to the recommended procedure for reporting response rates and may vary across the studies. However, comparisons among incentive conditions within a study are valid indicators of the potential effect of incentives on response rates.

PROXY RESPONDENTS

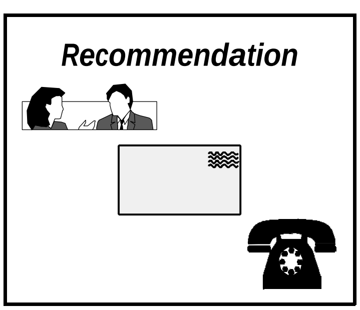
Different surveys impose different rules regarding whom is allowed to supply survey data for a sample member. In some surveys, any adult member of the household is permitted to provide information about other individuals in the household. In other surveys, the acceptability of data from proxy respondents depends on the information sought. In any case, the rules adopted by a particular survey will have an impact on the response rate obtained.

Proxy reporting of basic background information. When the survey requires only basic demographic information about each member of a household, it is common practice to gather this information from any adult member in the household. Proxy reporting of this type is frequently used in travel and other surveys to collect demographic and other background data on household members during the screening interview. This information is usually used to determine the eligibility of household members and to roster individuals within the household. Obtaining screener information in this way usually increases the efficiency and completeness of the screening effort without significantly reducing the accuracy of the information obtained. It enables interviewers to determine the eligibility of household members who are not home at the time of the contact and to collect information on members who are unable to respond for themselves because of mental or physical handicaps, or language barriers. In most cases, the gains in efficiency and response rates outweigh the small losses in accuracy associated with collecting screener information by proxy.



PROXY REPORTING OF BASIC BACKGROUND INFORMATION —
During the screening interview, allow for proxy reporting of basic demographic and background information. Whenever possible, verify the accuracy of the information during the main interview. Resolve any discrepancies that occur.

Proxy reporting of travel behavior and attitudes. The practice of accepting proxy reports of attitudes and behaviors, on the other hand, is less common largely because the gains in response rates are frequently offset by losses in reporting accuracy. In most situations, attitudinal and behavioral information provided by proxy respondents tends to be less accurate than information supplied by the actual respondents. In the 1990 Nationwide Personal Transportation Survey (NPTS), for example, proxy respondents reported, on average, about 25 percent fewer trips per travel day, and about 20 percent fewer miles traveled and driven than persons responding for themselves, suggesting that proxy reports underestimate actual travel behavior [36]. Data from the 1995 NPTS Pretest exhibited a similar pattern; proxy respondents reported fewer trips and fewer miles traveled than self respondents. They were also more likely to respond with “I don’t knows.”



PROXY REPORTING OF TRAVEL BEHAVIOR AND ATTITUDES —
Avoid using proxy reports, unless the accuracy of the information can be assured in one of two ways: 1) the proxy accompanied the actual respondent on the trip or in the activity, or 2) the respondent recorded the information in a travel diary or similar instrument. Always attempt to collect information from the actual respondent. When these attempts fail and it is more than three days since the travel behavior occurred, accept proxy datum when it satisfies at least one of the conditions above.

SPECIAL METHODS FOR REDUCING NONRESPONSE

In any survey, no matter how well it has been conducted, there is usually a group of hard-core nonrespondents who are very difficult to contact, very reluctant to participate, or simply unable to take part in the survey. In travel surveys this group of sample members tends to include [36]:

- the elderly,
- the less well-educated,
- urban dwellers,
- individuals with physical limitations or language barriers,
- couples with young children, single parents, individuals who hold multiple jobs, and other busy or hard-to-reach individuals, such as 18 to 24 year-olds, and
- individuals who lack community or civic ties that often motivate participation in travel surveys.

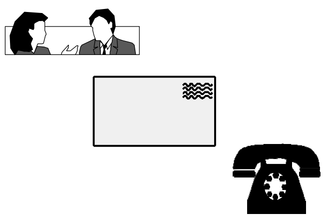
The number of nonrespondents remaining in these groups at the end of the survey can often be reduced by tailoring the survey materials and methods to overcome the obstacles standing in the way of their participation. In Section 2.2, we discussed some general methods for improving the response rates of the elderly, the less-well educated, individuals with poor literacy skills, and individuals with physical limitations or language barriers. The response rates of these and other nonrespondent groups can often be further improved through use of special methods that make participation particularly attractive (or more difficult to avoid), address specific concerns or objections, or call attention to the survey. Some examples of special methods of this type are:

- large incentives conditional on participation,
- persuasive calls from the most successful interviewers or field supervisors,
- special appeals and mailings,
- artfully constructed conversion letters, and
- in-home visits, if the addresses of the sample members are known.

The method selected for a particular group should be tailored to that group. Most 18 to 24 year-olds, for example, are unlikely to be moved by appeals to civic duty, but will often take the time to complete a survey if they are offered a large incentive. Single parent families and families with young children, on the other hand, are often more willing to take part in a study if the sponsor offers to cover babysitter costs while they complete the survey.

While some of these methods for improving response rates can be relatively inexpensive—special mailings and artfully constructed conversion letters, for example, others—such as in-home visits and large incentives, may be too expensive to apply on a frequent basis.

Recommendation



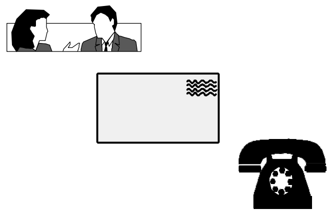
TAILORING THE SURVEY —
To the extent possible, tailor the survey to address the special needs, concerns, and objections of subgroups of individuals unwilling to take part.

SUBSAMPLING NONRESPONDENTS

When it is not possible to adopt special methods for all reluctant sample members, one alternative to simply writing off the final group of nonrespondents is to select a subsample and use special methods to achieve a high response rate among the subsample.

Although formulas have been derived for determining the best point at which to begin subsampling the remaining nonrespondents, subsampling is typically undertaken only when all sample members have been contacted repeatedly, the response rate remains unsatisfactory, and the data collection budget is nearing exhaustion. To illustrate the use of subsampling, consider a hypothetical survey which has achieved only a 50 percent response rate after spending 75 percent of the data collection budget. If a subsample of one in five of the remaining nonrespondents is chosen, the budget level of effort per case can be increased substantially, making it possible to achieve a relatively high response rate among the subsample. For example, if even 40 percent of the subsample provide the required data, the weighted response rate will be 70 percent—that is, within the first half of the sample, 100 percent of the cases are respondents and, within the subsample representing the second half of the cases, 40 percent are respondents. The important point is that the sample respondents include some representatives of the hard-to-complete group.

Recommendation



SUBSAMPLING OF NONRESPONDENTS —

When the data collection budget is almost exhausted, response rates are unacceptably low, and all cases in the sample have been fielded, subsample nonrespondents to boost response rates and to improve the representativeness of the sample.

2.5 ITEM NONRESPONSE

Even if the overall response rate is high, many respondents may be missing data for specific items. The statistical consequences of unit and item nonresponse are identical. Fortunately, the level of item nonresponse tends to much lower than the level of unit nonresponse. In this section, we describe some methods for lowering the rate of item nonresponse.

QUESTIONNAIRE DESIGN

The same questionnaire design features that help reduce unit nonresponse tend to reduce item nonresponse as well. Short questionnaires with clear instructions are easier to complete and more likely to be completed correctly than long questionnaires with difficult to follow instructions. The clarity and appearance of the questionnaire is likely to be particularly important with self-administered questions. Respondents are about twice as likely to incorrectly skip items when they complete the questionnaire themselves than when the questionnaire is administered by an interviewer [37]. This differential can be reduced if the instructions for filling out the questionnaire are clear and simple and if graphics are used to indicate the flow of questions [16].

With a self-administered questionnaire sent to the respondent by mail, the respondent must be able to figure out how to complete the questions without the help of an interviewer. It is, therefore, especially important that the respondents understand the questions and any accompanying instructions. The questions and the instructions must be clearly stated and easily understood.

Techniques for developing questionnaires. There are several commonly-used techniques for developing questionnaires, including focus groups, expert reviews, cognitive interviews, and small pilot studies (typically with fewer than 100 respondents). We especially recommend expert reviews and cognitive interviews as inexpensive but effective means for improving draft questionnaires.

In an expert review, three or four specialists in questionnaire design review the draft, point out problems with it, and suggest improvements. The questionnaire and background information about the study (describing the study's purposes and the analyses in which the data are likely to be used) are typically sent out to the experts prior to a face-to-face or telephone review session. The review session may last 2 to 8 hours depending on the length of the questionnaire.

It is also important to try out the questions with persons drawn from the same population as the sample for the main study. Cognitive interviews may be a particularly valuable tool for detecting respondent problems in understanding the questions or in performing the other cognitive operations (such as retrieval and judgment) that are required to answer accurately. In a cognitive interview, respondents are encouraged to think aloud as they attempt to answer the questions [38]. For example, they may be instructed to:

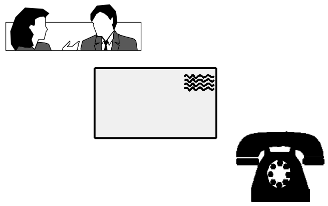
> *Please tell me what you are thinking about as you answer the next few questions. This will help us understand whether the questions we are asking are being understood as intended and whether they have any other problems. Remember to think out loud as you answer. Let's try an example before we turn to the main questions. Tell me what you are thinking as you answer the question How many windows are there in your home?*

In addition, the respondents in a cognitive interview may be asked to:

- point out unfamiliar terms (Are you familiar with [the term]?);
- paraphrase a question or set of instructions (Can you say in your own words what that last question was asking?);
- describe their thought processes after having answered a question (Please tell me how you arrived at that last answer); and
- rate their confidence in their answers (Are you very sure, somewhat sure, or not too sure about that last answer?).

All these techniques are intended to reveal the cognitive processes through which respondents arrive at their answers and to identify any problems with the questions. Typically, interviews are completed with 10 to 15 respondents and a short report is produced that highlights any problems detected with specific items and suggests modifications.

Recommendation

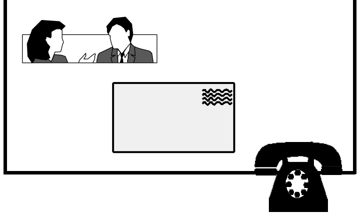


DEVELOPING QUESTIONNAIRES —

Whenever possible, use expert review and cognitive interviews to improve the quality of items in the questionnaire.

Sensitive questions. Aside from comprehension errors or other cognitive problems, other factors may contribute to high levels of missing data. For example, respondents may refuse to answer questions they regard as too personal. Many surveys ask for the annual income of the respondent or the respondent's family, and often this item has higher levels of missing data than any question in the survey. The best that can be done is to ask such sensitive questions toward the end of the questionnaire after the respondent has completed the bulk of the questions. Training interviewers to explain why the question is needed and how the information will be used may also help to reduce nonresponse to sensitive questions.

Recommendation



SENSITIVE QUESTIONS —

Place sensitive questions towards the end of the questionnaire. Train interviewers to explain the importance of the question and how the information will be used.

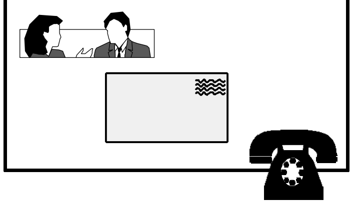
RETRIEVAL OF MISSING ITEMS

The most direct method for reducing the number of items that are missing is to recontact the respondent and obtain any information omitted from the questionnaire. This procedure is called "retrieval" by many survey researchers. Generally, retrieval is reserved for a small number of especially important items, which are predesignated beforehand. Often the "retrievable" items are the same ones used to define whether a questionnaire is regarded as complete. (See the discussion of "critical items" in Chapter 1.)

The recontact is usually made by telephone (if a telephone number is available). The retrieval interviewer explains that an important piece of information has been omitted from the questionnaire and asks the respondent to provide it. If the initial interview was conducted by telephone and the respondent refused to provide the information in the first place, the retrieval interviewer explains the importance of the missing data and points out that the rest of the information that the respondent did provide cannot be used without the missing items.

Retrieval is generally a cost-effective method for salvaging cases that would otherwise be difficult or impossible to include in the analysis.

Recommendation



RETRIEVAL OF MISSING ITEMS —

Recontact respondents who failed to provide a few pieces of key data. This recontact can be made through a less expensive method of contact, such as telephone.

2.6 CURRENT PRACTICE FOR REDUCING NONRESPONSE IN TRAVEL SURVEYS

Current practice for reducing nonresponse in household transportation surveys differs widely across survey firms and particular studies. Some studies incorporate many of the recommended procedures for reducing nonresponse. Others rely on just one or two procedures or employ less than optimal methods.

A recent review of household travel surveys by Peter Stopher and Helen Metcalf [29] reveals the extent to which current practice varies among travel surveys. Their analysis, based on a review of 55 household travel surveys conducted from 1988 to 1994, shows how recent surveys made use of three methods for reducing nonresponse: pretests, incentives, and follow-up reminders. The results of the review are described below.

PRETESTS

About 74 percent of the surveys included in the analysis carried out a pretest prior to the conduct of the main survey. The samples in the studies ranged in size from fewer than 25 to over 500 respondents; the most common size was between 26 and 50 respondents. The objectives of the pretests varied from study to study. All were designed to test the survey instruments. Some were full dress rehearsals for the main survey. Others evaluated particular aspects of the survey's design or its procedures. **Table 2.5** shows the percent of pretests evaluating various aspects of the survey.

Table 2.5

A Summary of Pretest Objectives

ASPECT	PERCENT OF PRETESTS EVALUATING EACH ASPECT
Instrument	100
Management	58
Training of Survey Personnel	48
Sampling Procedures	40
Data Keying Procedures	28
Geocoding Procedures	25
Analysis	23
Effectiveness of Incentives	10

INCENTIVES

Only about twenty percent of the surveys included in the review used some form of incentive to encourage participation. Of the studies employing incentives, nearly half relied on a cash incentive. Of the surveys using cash or lotteries, about half distributed the incentive in advance, while the other half used conditional incentives.

FOLLOW-UP REMINDERS

According to the review, the use of follow-up reminders, that is, contacts to remind sample members to complete their questionnaires after they had received them in the mail, has become common practice in household travel surveys. Nearly 80 percent of the surveys employed this method for reducing nonresponse. Nonetheless, the studies differed in the type, and especially in the number, of contacts used for this purpose (see**Table 2.6**). About 93 percent of the studies contacted the respondents by telephone to remind them to complete the questionnaires or travel diaries. More than half relied on a single reminder, while a good portion, 20 percent, used as many as four or more follow-up reminders.

Table 2.6

Type and Number of Follow-up Reminders

FOLLOW-UP REMINDERS		PERCENT
TYPE	Telephone call	93
	Other	7
NUMBER	One	60
	Two	17
	Three	3
	Four or more	20

References

1. Kirsh, I.S., Jungeblut, A., Jenkins, L., & Kolstad, A. (1993). *Adult Literacy in America: A First Look at the Results of the National Adult Literacy Survey*. Washington, D.C.: National Center for Education Statistics.

2. Groves, R.M. & Lyberg, L.E. (1988). An overview of nonresponse issues in telephone surveys. In R.M. Groves, P.P. Biemer, L.E. Lyberg, J.T. Massey, W.L. Nicholls II, and J. Waksberg (Eds.) , *Telephone Survey Methodology*. New York: John Wiley.

3. Herzog, A.R. & Kulka, R.A. Telephone and mail surveys with older populations: a methodological overview.

4. Couper, M.P. & Groves, R.M. (1992). The role of the interviewer in survey participation. *Survey Methodology*, 18, 263-277.

5. Groves, R.M. & Fultz, N.H. (1985). Gender effects among telephone interviewers in a survey of economic attitudes. *Sociological Methods and Research*, 14, 31-52.

6. Singer, E., Frankel, M.R., & Glassman, M.B. (1983). The effect of interviewer characteristics and expectation on response. *Public Opinion Quarterly*, 47, 68-83.

7. Morton-Williams, J. & Young, P. (1986). Obtaining the survey interview: an analysis of tape-recorded doorstep introductions. *Journal of the Market Research Society*, 27, 35-53.

8. Collins, M., Sykes, W., Wilson, P., & Blackshaw, N. (1988). Nonresponse: The UK Experience. In R.M. Groves, P.P. Biemer, L.E. Lyberg, J.T. Massey, W.L. Nicholls II, and J. Waksberg (Eds.) , *Telephone Survey Methodology*. New York: John Wiley.

9. Oksenberg, L. & Cannell, C. (1988). Effects of interviewer vocal characteristics on nonresponse. In R.M. Groves, P.P. Biemer, L.E. Lyberg, J.T. Massey, W.L. Nicholls II, and J. Waksberg (Eds.) , *Telephone Survey Methodology*. New York: John Wiley.

10. Heberlein, T.A., & Baumgartner, R. (1978). Factors affecting response rates to mailed questionnaires: A quantitative analysis of the published literature. *American Sociological Review*, 43, 447-462.

11. Groves, R., Cialdini, R., & Couper, M. (1992). Understanding the decision to participate in a survey. *Public Opinion Quarterly*, 56, 475-495.

12. Tourangeau, R., Shapiro, G., Kearney, A., & Ernst, L. (In press). "Who lives here? Survey undercoverage and household roster questions." *Journal of Official Statistics*.

13. Yammarino, F.J., Skinner, S.J., & Childers, T.L. (1991). Understanding mail survey response behavior. *Public Opinion Quarterly*, 55, 613-639.

14. Dillman, D.M. (1991). The design and administration of mail surveys. *Annual Review of Sociology*, 17, 225-249.

15. Applied Management & Planning Group. (1995). *Dallas-Ft. Worth household travel survey: Pretest report*. Los Angeles, CA: Applied Management & Planning Group.

16. Jenkins, C., & Dillman, D. (1997). Towards a theory of self-administered questionnaire design. In L. Lyberg, P. Biemer, M. Collins, E. DeLeeuw, C. Dippo, N. Schwarz, & D. Trewin (Eds.), *Survey measurement and process quality*. New York: Wiley.

17. Richardson, A.J., Ampt, E.S., & Meyburg, A.H. (1995). *Survey Methods for Transport Planning.* Parkville, Vic: Eucalyptus Press.

18. Groves, R., & Kahn, R. (1979). *Surveys by telephone: A national comparison with personal interviews*. New York: Academic Press.

19. Dillman, D.M. (1978). *Mail and telephone surveys: The Total Design Method*. New York: JohnWiley.

20. Sebold, J. (1988). Survey period length, unanswered numbers, and nonresponse in telephone surveys. In R.M. Groves, P.P. Biemer, L.E. Lyberg, J.T. Massey, W.L. Nicholls II, and J. Waksberg (Eds.) ,*Telephone Survey Methodology*. New York: John Wiley.

21. Weeks, M.F. (1988). Call scheduling with CATI: Current capabilities and methods. In R.M. Groves, P.P. Biemer, L.E. Lyberg, J.T. Massey, W.L. Nicholls II, and J. Waksberg (Eds.), *Telephone Survey Methodology*. New York: John Wiley.

22. Weeks, M.F., Jones, B.L., Folsom, R.E., & Benrud, C.H. (1980). Optimal times to contact sample households. *Public Opinion Quarterly*, 44, 100-114.

23. Weeks, M.F., Kulka, R.A., & Pierson, S.A. (1987). Optimal call scheduling for a telephone survey. *Public Opinion Quarterly*, 51,540-549.

24. Kulka, R.A. & Weeks, M.F. (1988). Toward the development of optimal calling protocols for telephone surveys: a conditional probabilities approach. *Journal of Official Statistics*, 4, 319-332.

25. Greenberg, B.S. & Stokes, S.L. (1990). Developing an optimal call scheduling strategy for a telephone survey. *Journal of Official Statistics*, 6, 421-435.

26. Tuckel, P.S. & Feinberg, B.M. (1990). The answering machine poses many questions for telephone survey researchers. *Public Opinion Quarterly*, 55, 200-217.

27. Oldendick, R.W. (1993). The effect of answering machines on the representativeness of samples in telephone surveys. *Journal of Official Statistics*, 9, 663-672.

28. Hoagland, R.J., Warde, W.D., & Payton, M.E. (1988). Investigation of the optimum time to conduct telephone surveys. *Proceedings of the Survey Research Methods Section*, American Statistical Association, 755-760.

29. Stopher, P.R. & Metcalf, H (1996). *Methods for Household Travel Surveys: A Synthesis of Highway Practice.* Washington, DC: National Academy Press.

30. Kulka, R.A. (1994). The use of incentives to survey "hard-to-reach" respondents: a brief review of empirical research and current practice. Paper prepared for the Seminar on New Directions in Statistical Methodology, sponsored by the Council of Professional Associations on Federal Statistics (COPAFS), Bethesda, Maryland, May 25-26.

31. Kulka, R.A. (1992). A brief review of the use of monetary incentives in federal statistical surveys. Paper prepared for the Symposium on Providing Incentives to Respondents, convened by the Council of Professional Associations on Federal Statistics (COPAFS) for the Office of Management and Budget, John F. Kennedy School of Government, Harvard University, Cambridge, MA., October 1-3.

32. Chromy, J.R. & Horvitz, D.G. (1978). The use of monetary incentives in national assessment household surveys. *Journal of the American Statistical Association*, 73, 473-478.

33. Berlin, M., Mohadjer, L., Waksberg, J., Kolstad, A., Kirsch, I., Rock, D., & Yamamoto, K. (1992). An experiment in monetary incentives. *Proceedings of the Survey Research Methods Section*, American Statitical Association, 393-398.

34. Tooley, M. (1996). Incentives and rate of return for travel surveys. Paper presented at the Conference of the Transportation Research Board.

35. Moore, P.A. (1995). Using proxy interviews in a nationwide travel survey. Paper presented at the Annual Conference of the American Association of Public Opinion Research, Fort Lauderdale, Florida, May 18-21.

36. Johanna Zmud (1997). Personal communication.

37. Tourangeau, R., Rasinski, K., Jobe, J., Smith, T., & Pratt, W. (In press). Sources of error in a survey of sexual behavior. *Journal of Official Statistics*.

38. Willis, G., Royston, P. & Bercini, D. (1991). The use of verbal report methods in the development and testing of survey questionnaires. *Applied Cognitive Psychology*, 5, 251-267.

3.

Statistical Methods for Reducing the Impact of Nonresponse

3.1 CHAPTER SUMMARY

This chapter describes statistical methods for reducing the impact of nonresponse on survey estimates. Although the best method for reducing the impact of nonresponse is to achieve high response rates in the first place, the methods described here can help reduce the effects of any remaining nonresponse. Unless post-survey adjustments are made, estimates from the survey are likely to be distorted by nonresponse. The magnitude of the bias introduced by nonresponse depends on the proportion of the sample for whom data are not obtained and the amount of difference between the respondents and the nonrespondents. Two forms of nonresponse are usually distinguished-unit nonresponse (in which no data are obtained at all) and item nonresponse (in which one or more items are missing from an otherwise completed interview or questionnaire). The statistical consequences of unit and item nonresponse are the same; both forms of nonresponse can reduce the sample size and introduce bias. However, the most effective methods for compensating for the two forms of nonresponse are different.

STATISTICAL METHODS

Two statistical methods can be used to reduce the effects of unit nonresponse:

- The weights assigned to each case can be adjusted for nonresponse (that is, the weights initially assigned to nonrespondents can be reallocated to respondents with similar characteristics);
- The weights can be post-stratified (i.e., the survey weights can be adjusted to agree with independent population figures).

Nonresponse adjustments should always be made. Post-stratification should be used when accurate population figures are available.

The main tool for reducing the effects of item nonresponse is imputation. It is useful to distinguish “logical” imputation (in which the value of a missing item is inferred from other data) and statistical imputation (in which a statistical procedure is used to predict a value for the missing item). Logical imputation, or editing, should be carried out before statistical imputation is used. “Hot deck” imputation is an especially useful statistical method for imputing missing values in travel surveys. In this form of imputation, similar cases are grouped into cells. Missing values are replaced using the value provided by a “donor” from the same cell; the donor is simply another sample member who provided an answer to the relevant item.

3.2 TYPES OF NONRESPONSE

As we noted in Chapter 1, most discussions of nonresponse distinguish between two forms of nonresponse—unit nonresponse and item nonresponse. [1] Unit nonresponse refers to the failure to obtain questionnaires or data collection forms (such as the travel diaries used in many personal transportation surveys) for a member of the sample. Item nonresponse refers to the failure to obtain a specific piece of information from a responding member of the sample. Item nonresponse is often used interchangeably with the term "missing data."

Although their causes are different, unit and item nonresponse have identical statistical consequences. Unless adjustments are made to the data, the level of nonresponse bias will depend only on two factors—the proportion of the sample for whom data were not obtained and how much the respondents differ from the nonrespondents. It does not matter whether the data are missing because a sample member never responded to the survey at all or because the respondent failed to answer a specific question. However, different methods are typically used to reduce the impact of the two different forms of nonresponse.

UNIT NONRESPONSE

Unit nonresponse refers to complete nonparticipation by a member of the sample. In travel surveys, data are usually collected for both households and persons. There are, therefore, two types of unit nonresponse:

- household-level nonresponse, in which an entire household cannot be contacted or refuses to participate;
- person-level nonresponse, in which one or more persons in the household do not take part but at least one member does.

These two types of unit nonresponse are equivalent for surveys that only select one individual from each sample household. Because most personal transportation surveys collect data on every household member, in the rest of this report we assume that all eligible household members are asked to participate.

For many variables, person-level analyses are appropriate. This results in a straightforward pattern of unit nonresponse, since it is relatively easy to decide whether a person is to be considered a respondent. However, if household-level variables are of interest, it may be unclear how to treat the household if some but not all persons from that household provide data. In Chapter 1, we recommend treating the household as a respondent if the majority of its eligible members are classified as respondents and if a set of critical items are available about the household. Even if the household is considered a respondent, some items could still be missing due to item nonresponse.

ITEM NONRESPONSE

The second type of nonresponse is item nonresponse. This is when there are missing items in an interview or questionnaire that is otherwise completed by a respondent. The data could be missing because the respondent refused to answer (e.g., a question on income), the interviewer failed to ask the question (e.g., he or she may have skipped a question accidentally), or the respondent simply missed the question (e.g., he or she may have forgotten to complete the back of one page of a diary).

Missing trips. When data on one or more trips or activities are excluded from a travel diary, these trips can be treated as missing data. It is often difficult to spot these missing trips or activities unless the previous and next trip or activity form a chain of trips. If this is the case, the ending point for one trip and the starting point for the next trip may be different, suggesting that there is at least one missing trip between them. If undetected, this type of item nonresponse will result in underestimates of the number and distance of trips.

Missing items about a trip or activity. Items missing could include travel times, trip purpose, or the mode of transportation. It may be possible to fill in some or all of the missing information through careful editing of the data. Such editing involves looking at the context in which the trip occurred. Editing will, however, often fail to remove all of the item nonresponse within a reported trip or activity.

3.3 IMPACT OF NONRESPONSE

Nonresponse has two negative consequences for the quality of the estimates derived from the survey. First, it can reduce the number of cases for whom data are available. As a result, the survey statistics will not be as precise. In many surveys, however, the sample size is fixed. Additional cases may be fielded to compensate for those lost through nonresponse. Second, and more important, substantial bias can occur if the nonrespondents are different from the respondents on the characteristics of interest. This bias cannot be corrected with additional cases.

REDUCED SAMPLE SIZE

Reductions in sample size due to nonresponse have a direct effect on the variability in the statistics derived from the survey data. Unit nonresponse rates are as a rule higher than item nonresponse rates. But the effects of the two compound in reducing the sample sizes. This point is illustrated in Table 3.1. For example, if 1000 responses were sought and there was no nonresponse at all, the standard error for a proportion with a mean of 50 percent would be 0.0158. However, if the unit response rate were 70 percent and item response rate were 90 percent, we would only have 630 complete responses for a particular item (variable), giving a standard error of 0.0199, an increase of 26.0 percent. If the item response rate were to further fall to 80 percent, the standard error for the same variable would be higher, 0.0211, which is an increase of 33.6 percent due to nonresponse. Thus both unit nonresponse and item nonresponse contribute to higher standard errors—that is, decreased precision in the estimates from the survey.

Table 3.1
Impact of Nonresponse on Sample Size and Standard Errors

INITIAL n	UNIT RESPONSE RATE	ITEM RESPONSE RATE	FINAL n	S.E. ON 50% ESTIMATE
1000	1.0	1.0	1000	0.0158
1000	0.9	0.9	810	0.0176
1000	0.9	0.8	720	0.0186
1000	0.8	0.9	720	0.0186
1000	0.8	0.8	640	0.0198
1000	0.7	0.9	630	0.0199
1000	0.7	0.8	560	0.0211
1000	0.6	0.9	540	0.0215
1000	0.6	0.8	480	0.0228

Note: S.E. refers to the standard error of an estimate.

INTRODUCTION OF BIAS

The most important effect of nonresponse is the bias it produces. Nonresponse has an impact similar to that of excluding a portion of the population from the sampling frame. In both cases, a possibly nonrandom portion of the population is omitted from the study. This creates the potential for bias, and the effects of this bias can be substantial. Because of the omission of the nonrespondents, the sample no longer represents the entire population of interest. If the analysis ignores the effects of nonresponse, it implicitly assumes that nonrespondents do not differ systematically from the respondents on any characteristic of interest. To the extent that the nonrespondents *do* differ from the respondents, the results will be biased.

Differences between respondents and nonrespondents. Nonrespondents are rarely randomly distributed in the survey population. As we saw in Chapter 2, response rates vary widely across population subgroups and the survey variables are often associated with the characteristics of these subgroups. For example, nonresponse rates are usually much higher in cities than in the suburbs and preferred modes of transportation can differ in the two types of settings. Because they underrepresent city dwellers, travel surveys may underestimate the number of trips made by bus or subway. In one study, nonresponse led to the underestimation of the size of certain segments of the study population by 50 percent[2]

Nonrespondents in transportation surveys. In some respects, nonrespondents in transportation surveys have similar characteristics to nonrespondents in other national surveys. They are more likely than respondents to live in densely populated areas, to have low levels of education, to be elderly, and to have visual or hearing difficulties which prevent them from completing surveys. In addition, language barriers for those whose native language is not English, mental and physical handicaps, and poor literacy skills can contribute to nonresponse. The 1995 NPTS pretest found 4 percent of the age 18 and older respondents used a proxy because of language problems.

In addition, members of other groups may be more prone to become nonrespondents in transportation surveys than in other surveys. For example, the elderly may feel that because they leave the house less frequently than younger persons, their data are not needed for transportation surveys (see Chapter 2).

Level of nonresponse bias. For means (and proportions), the magnitude of the bias resulting from nonresponse depends on the response rate (R_R, as defined in Chapter 1) and the difference in the means (or proportions) for the respondents $\overline{X}_R$) and nonrespondents ($\overline{X}_{NR}$)[1,1]:

$$B = (1 - R_R)(\overline{X}_R - \overline{X}_{NR}).$$

Although there is no hard-and-fast rule about when nonresponse bias represents a serious threat to a survey, many national surveys carried out for the federal government achieve response rates of 80 percent or higher. Thus, when the response rate is low (e.g., less than 60 to 70 percent), the potential for bias is high; the mean for the respondents will be a good estimate for the mean of the whole population only if the respondents and nonrespondents are similar. If this assumption is wrong, then the unadjusted mean will produce biased estimates for the whole population.

It is important to realize that the validity of doing nothing at all about nonresponse rests on the (often implausible) assumption that the respondents and nonrespondents do not differ. Every compensation procedure—including doing nothing—rests on assumptions about the characteristics of the nonrespondents, but some assumptions are more reasonable than others. In estimating means or totals, the simplest assumption is that the mean for the respondents is equal to the mean for the nonrespondents. An alternative assumption is that the missing data are missing at random. Under this assumption, the respondents and nonrespondents may differ in any given sample, but across all possible samples the means for the two groups are the same. Unfortunately, these simple assumptions are rarely tenable in practice. When they are incorrect, even relatively low levels of nonresponse (10 percent to 20 percent) can produce significant bias. **Table 3.2** illustrates how the bias arises. It shows how the proportion estimated from survey data (e.g., the proportion of the study population using public transportation regularly) can be distorted by nonresponse.

When 20 percent of the sample become nonrespondents, and 20 percent of the respondents have some characteristic of interest versus 50 percent of the nonrespondents, the sample estimate will be 20 percent (based on the respondents) but the unbiased estimate (based on both the respondents and nonrespondents) will be 26 percent—a difference of 6 percent. (The unbiased estimate for the entire population is .20 for the 80 percent of the population represented by the respondents and .50 for the 20 percent of the population represented by the nonrepondents. Combining the two groups yields .80 x .20 plus .20 x .50, or 0.26). This margin of error—reflecting the nonresponse bias—is likely to be several times larger than random sampling error. More important, the error produced by nonresponse will not be random.

Table 3.2
Nonresponse Bias, by Level of Nonresponse

NONRESPONSE RATE	PROPORTION WITH CHARACTERISTIC OF INTEREST		BIAS
	Among Respondents	Among Nonrespondents	
.10	0.2	0.3	.01
.10	0.2	0.4	.02
.10	0.2	0.5	.03
.20	0.2	0.3	.02
.20	0.2	0.4	.04
.20	0.2	0.5	.06
.40	0.2	0.3	.04
.40	0.2	0.4	.08
.40	0.2	0.5	.12

Note: Bias is the difference between the expected value of a sample statistic and the population characteristic it is intended to estimate.

3.4 SURVEY METHODS

This section discusses various survey procedures intended to minimize the impact of nonresponse.

AVERTING NONRESPONSE

The best single method for reducing the effects of nonresponse is to have as little nonresponse as possible in the first place. The methods recommended in Chapter 2 all help achieve high response rates, thus minimizing the potential effects of nonresponse bias on survey estimates. Still, every survey incurs *some* nonresponse and the effects of nonresponse at the different stages of data collection cumulate. For example, if 90 percent of the sample households are successfully screened, 90 percent of those return a completed diary, and the diaries include information about 90 percent of the trips on average, then the cumulative response rate across all three stages is only about 73 percent (.9 x .9 x .9 = .729). Despite high response rates at each stage, there is still room for considerable bias. Thus, other procedures are likely to be useful in reducing the impact of nonresponse even when the response rate is quite high.

KEEP NONRESPONSE RATES AS LOW AS POSSIBLE —

Recommendation

Use the methods discussed in Chapter 2 to keep nonresponse to a minimum. The bias from nonresponse is related to the cumulative nonresponse rate, taking into account both unit and item nonresponse.

FOLLOWING UP WITH A SUBSAMPLE OF NONRESPONDENTS

One way to reduce the bias resulting from unit nonresponse is to achieve a high response rate among a subsample of the cases who remain nonrespondents near the end of the regular data collection period. This approach gives some representation in the final sample to the pool of potential nonrespondents.

For example, suppose some combination of callbacks, follow-up letters, and incentives has produced a 60 percent response rate. This means that 40 percent of the cases are, at this point in the data collection effort, still nonrespondents. Within the remainder, it may be better to select a subsample of one case out of four and subject this subsample—representing 10 percent of the original cases—to more intensive follow-up efforts than would be possible if all of the nonrespondents were retained for further follow-up efforts. These "more intensive" follow-up efforts might include:

- Sending specially tailored or personalized follow-up letters designed to persuade reluctant sample members to take part;
- Using more experienced or specially trained data collection staff to "convert" subsample members;
- Moving to more expensive methods of follow-up, such as registered letters or even in-person contacts;
- Offering much larger incentives for participation (e.g., $20 to $50); and
- Greatly increasing the number of callbacks for difficult to reach households or individuals (from, say, 6 or 8 callbacks to 20).

Suppose for the sake of illustration that 60 percent of the subsample selected for additional follow-up ultimately provide data. From a statistical point of view, this is equivalent to achieving a response rate of 84 percent within the entire sample. To understand why this is so, let us consider the initial sample as consisting of two strata, or subgroups. The first stratum, encompassing 60 percent of the cases fielded originally, includes those members of the population who require only standard efforts to be reached and persuaded to take part in the survey. The second stratum, encompassing the remaining 40 percent of the population, includes all of those who require additional efforts.

The response rate within the first stratum was 100 percent and the response rate within the second stratum was 60 percent; thus, the final sample respondents represent 84 percent of the population that the original sample was selected to represent (.60 x 1.00 plus .40 x .60). This calculation assumes that the subsample is a random sample of the initial nonrespondents and that the remainder of this group—that is, the portion not selected for further follow-up—is dropped.

Subsampling of nonrespondents is complicated and it means that the final data set must be weighted. Because of these added costs, the strategy is only useful when two conditions are met.

- First, because subsampling may reduce the total number of cases that are ultimately available for analysis, the projected impact of nonresponse bias must be large enough to make subsampling worthwhile. A large impact reflects either a low response rate (say, 60 percent or less) or large differences between respondents and nonrespondents (i.e., a difference of 10 percent or more in the means of the two groups). Differences between respondents and nonrespondents are sometimes apparent from the screening data. If the standard follow-up procedures are likely to yield either low response rates or large differences between respondents and nonrespondents, subsampling the nonrespondents may yield substantial gains.

- Second, subsampling must be expected to increase the response rate among those who are retained for follow-up. The procedure produces no gain if the response rate for the subsample is 50 percent and 50 percent of all the remaining nonrespondents would have taken part in the survey even without subsampling (had they been pursued using the standard procedures). To be worthwhile, subsampling should produce an increase in the response rate among the outstanding cases of at least 20 percent relative to simply continuing with further follow-up efforts for the entire sample.

It may be difficult to tell whether these conditions are met. If standard follow-up efforts are yielding few additional cases and the projected final response rate is low, subsampling may be worth trying.

CONCENTRATE RESOURCES ON A SUBSAMPLE —

Recommendation

When the potential nonresponse bias is large (because response rates are low or respondents and nonrespondents differ sharply), select a subsample of nonrespondents for intensive follow-up efforts. This method is useful when response rates can be increased by concentrating resources.

3.5 STATISTICAL METHODS

This section describes the three main procedures used to compensate for nonresponse in surveys—nonresponse weighting, post-stratification, and imputation.[1,3]

WEIGHTING

Need for weights. Data from surveys often require the use of weights to produce unbiased population estimates. The weights are typically applied for three main purposes. First, weights are often needed to compensate for differences in the selection probabilities of individual cases. Such differences can arise by design—a specific study may deliberately overrepresent one or more subgroups of a population in order to provide enough cases for separate analyses of the oversampled subgroups. For example, a regional transportation study may oversample a smaller jurisdiction to provide enough cases to allow separate estimates to be made for that jurisdiction. Or differences in selection probabilities may arise as an unintended byproduct of the sampling strategy. For example, in telephone samples, households with multiple telephone lines have more chances of being selected into the sample than households with a single line. Either way, estimates for the entire sample will be biased unless the data are appropriately weighted. Weights are needed to compensate for both deliberate and inadvertent departures from equal probability sampling. In the transportation literature, weighting is sometimes referred to as *factoring*.

Another purpose for weighting is to compensate for subgroup differences in response rates. Even if the sample as selected represents the larger population perfectly, differences in response rates can introduce systematic discrepancies between the population and the sample. For example, in personal transportation surveys, household size may be related to the probability that households will provide the required information (see Section 2.2 in Chapter 2 for a detailed discussion of characteristics related to nonresponse). Differences in response rates across subgroups of the sample can introduce bias into the results. Weighting adjustments can reduce such biases.

A final purpose for weights is to compensate for fluctuations from known population totals. For instance, if one area were overrepresented in a travel survey sample purely by chance, it would be possible to use data from the decennial census or the Current Population Survey (CPS) to adjust for this departure from the population distribution. In addition, adjusting the data to known population totals can help reduce the impact of undercoverage (e.g., the omission of persons in households without telephones) on survey estimates.

CALCULATING WEIGHTS

Weights are often calculated in three steps[4] The first compensates for differences in selection probabilities; the second for differences by subgroup in response rates; the third for differences between the composition of the sample and the composition of the population.

Step 1: the base weight. Typically, the initial, or base, weight (W_{1i}) for a case (e.g., a sample household) is calculated as the inverse of that case's selection probability (Pr_i):

$$W_{1i} = 1/Pr_i \quad .$$

All eligible selections—whether they complete the survey or not—should receive a base weight. The selection probability (or sampling rate) is the proportion of the population selected for the study. In a full random-digit dial (RDD) sample survey, the sampling unit is a telephone number and the selection probability is the percentage of possible numbers within the study area that were actually included in the sample. This total will include both nonworking and nonresidential numbers. Consider, for example, a study of Queens, New York. Suppose Queens encompasses 150 distinct exchanges (e.g., 753-xxxx). Since each exchange includes 10,000 possible numbers (the numbers 0000 through 9999), the total population of possible numbers includes 1,500,000. If 1,500 of these numbers are selected (each with the same probability), then each would have a selection probability of .001 (1,500/1,500,000).

Both the sample and the population from which it was drawn include nonworking and nonresidential numbers, besides the residential numbers that are actually eligible for the study. The sampling rate is the same for the subset of residential numbers as for overall set of possible numbers; it is 1 in 1000, or .001. The presence of ineligible units (such as nonworking numbers) in the sample does not affect the calculation of the weights for the eligible units. In some types of telephone sampling, sampling is restricted to certain subgroups of the possible telephone numbers. When the sample is restricted in this way, the size of the population of possible numbers will be smaller than 10,000 times the number of exchanges in the study area.

In stratified sample designs, the population is first divided into subgroups called “strata” and separate samples are selected within each one. Often different sampling probabilities are used within the different strata. For example, the study area might be divided into areas where different modes of transportation are used more often. High density urban areas might be sampled at a higher rate to increase the number of persons who make trips by walking, riding a bicycle, or taking public transit. Or the sampling rates may be set to ensure that separate estimates can be made for different jurisdictions. If the telephone numbers linked to different areas are subject to different rates of sampling, then the separate selection probabilities will have to be computed for each area. **Table 3.3** shows an example of how this might be done.

Table 3.3
Calculation of Initial Weights in a Stratified RDD Sample

AREA	NUMBER OF SELECTIONS	SIZE OF POPULATION	SELECTION PROBABILITY	W_1
New York County	1,500	1,510,000	.00099	1006.7
Kings County	1,500	1,750,000	.00086	1166.7
Queens County	1,500	1,500,000	.00100	1000.0
Bronx County	1,200	880,000	.00136	733.3
Richmond County	800	280,000	.00286	350.0

Note: Both selections and population represent telephone numbers. W_1 represents the initial weight for sample cases.

Table 3.3 displays a stratified sample of telephone numbers in New York City, with separate sampling rates in each of the five boroughs. Continuing our earlier example, we have assumed that, in Queens County, 1,500 numbers were included out of 1.5 million possible numbers in the exchanges that serve Queens. That gives a selection probability of .001 for each number and a base weight of 1,000.

In an RDD survey, this base weight should then be adjusted to compensate for the fact that people in households with multiple telephones have more than one chance of being selected into the sample. The standard adjustment is quite simple; it is the base weight for household i ($1/Pr_{1i}$) divided by the number of distinct telephone lines (t_i) in the household:

$$W_{1i}^{/} = \frac{1}{Pr_{1i} \; t_i} \quad .$$

To carry out this adjustment, it is necessary to add questions to the interview to determine how many distinct telephone lines the household has (excluding those dedicated to faxes or modems). For example, if a household in Queens County reports 3 telephone lines, then the adjusted weight will be 333.33 (= 1 over 0.001 x 3).

In a survey in which households are first screened and then subsampled for the main data collection, the base weight should reflect the selection probabilities at both phases of selection—selection into the screening sample and retention for the main sample:

$$W_{1i} = \frac{1}{Pr_{1i} \; Pr_{2i}} \quad ,$$

in which Pr_{1i} represents the case's probability of inclusion in the screening sample and Pr_{2i} its probability of retention in the main sample.

If the weights have been properly calculated, their sum represents an estimate of the size of the population from which the sample is selected. For example, in each stratum in **Table 3.3**, the weights sum to the stratum population size.

Recommendation

COMPENSATE FOR DIFFERENCES IN SELECTION PROBABILITIES BY WEIGHTING THE DATA —

Each case should receive a base weight equal to the inverse of its selection probability. In a telephone sample, the weight should also reflect the number of lines to which a household is linked.

Step 2: compensating for nonresponse. The base weight (W_1 or W_1') should then be adjusted to compensate for the effects of nonresponse. Nonresponse adjustments ensure that the sum of the weights is unaffected by nonresponse; they do this by reallocating the weights assigned to the nonrespondents among the respondents. In addition, the nonresponse adjustments can reduce the bias introduced by nonresponse by compensating for differences in nonresponse rates across subgroups of the sample.

Nonresponse adjustments are calculated by grouping cases into nonresponse adjustment cells and finding the (weighted) response rates for cases in that cell. In our hypothetical survey of New York City, the adjustment cells might be the five boroughs. For each cell, the weighted response rate is computed using the procedures described in the first chapter of this report. For cases in cell j (e.g., telephone numbers in exchanges linked to the Bronx), the weighted response rate (R_j) is:

$$R_j = \frac{\sum_{1}^{n_{rj}} W_{1ij}}{\sum_{1}^{n_{ej}} W_{1ij}} \quad ,$$

in which the numerator is the sum of the weights for the respondents in cell j (and n_{rj} is the number of respondents in that cell) and the denominator is the sum of the weights for all eligible cases in that cell (and n_{ej} is the number of eligible cases in the cell). As we noted in Chapter 1, the number of eligibles may have to be estimated if there are cases for whom eligibility could not be ascertained.

The adjusted weight (W_2) is the base weight divided by the weighted response rate (R_j):

$$W_{2ij} = \frac{W_{1ij}}{R_i} \quad .$$

For nonrespondents and ineligible cases, the adjusted weight is set to zero. The sum of the *adjusted* weights for the respondents in cell j should equal the sum of the *base* weights for the eligible cases in that cell.

Table 3.4 illustrates the calculation of adjusted weights for our hypothetical sample in New York City. The table shows the weighted number of eligibles and respondents in each borough, the response rate, and the adjusted weight, which incorporates an adjustment for nonresponse. For example, in Queens County, 700 of the selected telephone numbers turned out to be eligible for the study. The sum of the weights for these cases was 700,000 (700 cases, each with a weight of 1,000). The sum of the weights for the 560 respondents was 560,000. The weighted response rate was, therefore, 0.8 (560,000/700,000). The adjusted weight for each respondent from Queens is 1,250 (= 1,000/0.8); each of the nonrespondents receives a weight of 0. The adjusted weights sum to the same total as the initial weights for the eligible cases (700,000 = 700 x 1,000 = 560 x 1,250). In this example, all the cases in each adjustment cell have the same initial weight. In most surveys, different cases would begin with different initial weights.

Table 3.4
Calculation of Adjusted Weights in a Stratified RDD Sample

AREA	NUMBER OF ELIGIBLE SELECTIONS	NUMBER OF RESPONDENTS	RESPONSE RATE	W_1	W_2
New York County	604,000 (600)	483,200 (480)	.80	1006.7	1258.3
Kings County	787,500 (675)	583,350 (500)	.74	1166.7	1575.0
Queens County	700,000 (700)	560,000 (560)	.80	1000.0	1250.0
Bronx County	352,000 (480)	228,800 (312)	.65	733.3	1128.2
Richmond County	112,000 (320)	67,200 (192)	.60	350.0	583.3

Note: The numbers given in the second and third columns are weighted (using W_1 , the initial weight for sample cases); the parenthetical entries in those columns are raw numbers of eligible and responding cases, respectively. W_2 is the adjusted weight.

Ideally, adjustment cells should be formed using variables that are related both to the likelihood of nonresponse and to the substantive variables of interest in the survey (such as travel behavior). Often, however, the choices are quite limited because so little is known about the nonrespondents and because both respondents and nonrespondents must be classified into adjustment cells. For example, in a telephone survey, the only information available for the nonrespondents may be their area code and exchange (and any geographic information that can be inferred from these). Thus, the nonresponse adjustment cells have to be formed using whatever information happens to be available for the nonrespondents.

When there are two phases of data collection—a screening phase and a main interview phase—separate nonresponse adjustments should be calculated for each phase. The same adjustment cells need not be used in both phases. In fact, the screening data are generally useful for forming adjustment cells to compensate for nonresponse to the main data collection. If R_{1j} denotes the weighted response rate in the first phase of data collection and R_{2k} the response rate in the second phase, then the adjusted weight would be:

$$W_{2ijk} = \frac{W_{1ijk}}{R_{1j}\ R_{2k}}$$

for the respondents and zero for the nonrespondents.

Recommendation

COMPENSATE FOR DIFFERENCES IN NONRESPONSE RATES BY ADJUSTING THE BASE WEIGHT

The base weights should be adjusted for nonresponse. If data are collected in two phases, separately calculate nonresponse adjustments for each phase.

Step 3: post-stratifying to population estimates. As we noted, the sum of the weights represents an estimate of the size of the survey population. Sometimes independent estimates of the size of the population are available (for example, from decennial census data). A technique called post-stratification can be used to bring the survey weights into agreement with the outside population figures. Post-stratification is used to correct for two types of errors in survey estimates—random sampling error and coverage error. Random sampling error refers to chance departures of the sample from the population it is selected to represent. Post-stratification can be expected to reduce random sampling error when a population estimate is derived from the decennial census or from a survey with a much larger sample than one used in the transportation survey being weighted. Coverage errors refer to systematic problems regarding who is included or excluded from the sample. Post-stratification can be expected to reduce the effects of coverage error when the population estimate gives better coverage of the population than the transportation survey sample does. For example, if a telephone survey was used to collect the data, the sample will necessarily exclude households without telephones. The two most frequent sources of figures for post-stratification are the decennial census and the Current Population Survey; these are thought to achieve much higher levels of coverage of the general population than other surveys do.

Post-stratification involves comparing the sum of the weights (i.e., W_2) for a given subgroup with the population estimate for that group. For example, decennial census figures are available for age-race-sex groupings at the level of counties and minor civil divisions. The post-stratification adjustment is calculated by multiplying the adjusted weight of cases in a subgroup, say subgroup j, by the ratio between the population estimate for that subgroup (N_j) and the sum of the weights for sample cases in that subgroup:

$$W_{3ij} = W_{2ij} \frac{N_j}{\Sigma W_{2ij}} \ .$$

The adjustment cells are typically defined in terms of areas (such as townships) and one or more demographic variables (such as household size). For example, in weighting the data from a household travel survey carried out in a large metropolitan area, the analysts took into account household size, the number of vehicles available, and zones (defined by response rates) within townships in the sample area[2].

Table 3.5 illustrates the calculation of post-stratified weights with the data from our hypothetical survey of New Yorkers. The sum of the adjusted weights (W_2) for Queens is 700,000; according to the 1990 census, the total number of households was about 720,149. This produces a post-stratification adjustment factor of approximately 1.029 and a final weight (W_3) of 1285.98 (=1,250 x 1.029).

Table 3.5
Calculation of Post-Stratified Weights

AREA	WEIGHTED NUMBER OF RESPONDENTS	POPULATION ESTIMATE	ADJUSTMENT FACTOR	W_2	W_3
New York County	604,000 (480)	716,422	1.186	1258.3	1492.55
Kings County	787,500 (500)	828,199	1.052	1575.0	1656.40
Queens County	700,000 (560)	720,149	1.029	1250.0	1285.98
Bronx County	352,000 (312)	424,112	1.205	1128.2	1359.33
Richmond County	112,000 (192)	130,519	1.165	583.3	679.77

Note: Population estimates are from the 1990 census and represent the number of households in each county. W_2 is the adjusted weight; and W_3, the post-stratified weight.

Population figures for poststratification adjustments (the values for N_j in the equation) can be obtained from decennial census data, the CPS, or other Census Bureau estimates. Which source to use will depend on how recent the data are, whether they are based on sufficient sample sizes (in the case of the CPS), and whether they provide appropriate grouping variables.

So far, we have emphasized the calculation of household weights. But in many transportation both household-level and person-level weights should be calculated. Typically, the same initial weights would be used (since every household member is selected within sample households). The two sets of weights would, however, incorporate different nonresponse and post-stratification adjustments.

Recommendation

IMPROVE THE ESTIMATES BY ADJUSTING WEIGHTS TO KNOWN POPULATION TOTALS —
Multiply the weights for the cases in a cell by the ratio between the population estimate for the cell to the sum of the weights for that cell.

ALTERNATIVE METHODS FOR WEIGHTING

Factoring to population totals. In some cases, it is possible to skip this three-step process and simply to weight up to population figures instead:

$$W_{ij} = \frac{N_j}{n_j} .$$

In this equation, N_j represents the population total for weighting cell j and n_j the number of completed cases in that cell. The population figures could be based on decennial census data, the CPS, or some other reliable source. To use this method, it is important that the sample be selected with equal probabilities within each group. Suppose, for example, in our hypothetical survey of New York City, we had used this method of weighting. If the cells used for weighting were the different boroughs, then this method would generally yield the same result as the three-step method described earlier. Note, however, that within each borough the sample would overrepresent households with multiple telephone lines. In general, this simple method of weighting ignores any differences in selection probabilities within a weighting cell. Thus, we cannot recommend this approach unless the sample was selected with equal probabilities within each cell.

More complex schemes for weighting. The method of post-stratification described earlier assumes that population estimates are available for each weighting cell. Sometimes data are available for each variable used in defining the cells but not for every combination of these variables. Figures may be available for the total number of households in each township in a county and for each household size by number of vehicle combination but not for the three-way combination of township by household size by number of vehicles. It is still possible for the weights to take all three variables into account using a technique known variously as multidimensional raking, iterative proportional fitting, or the Deming-Stephan procedure. A very similar procedure—the Fratar method—has been used in transportation planning to project the growth in the number of trips over time[5].

The basic principle behind the procedure is simple. The weights are adjusted to bring the survey figures into line with one set of population figures; then they are adjusted to agree with the other set of population figures. They are then readjusted to agree with the first set of figures, and so on, until the survey weights agree with both sets of population estimates.

An example will make the method clearer. Suppose we have population figures for each of four classes of household and for each of two geographic zones, but not for the cells formed by crossing the household classes with the zones. The goal is to bring the sums of the sample weights into agreement with these figures. The census figures indicate an overall total of 200,000 households in the study area.

Table 3.6
Population Data Available for Weighting Adjustment

SUBGROUP	POPULATION FIGURE
HOUSEHOLD CLASSES	
No available vehicles	25,000
One person in hh/one available vehicle	50,000
Two persons in hh/one available vehicle	75,000
Two or more available vehicles	50,000
GEOGRAPHICAL CLASSES	
Low response rate townships	120,000
High response rate townships	80,000

The preliminary weights prior to any post-stratification total 180,000, distributed as shown in **Table 3.7**.

Table 3.7
Sums of Sample Weights by Weighting Class

SUBGROUP	TOWNSHIP GROUP		TOTAL
	Low Response Rate	High Response Rate	
HOUSEHOLD CLASSES			
No available vehicles	12,000	8,000	20,000
One person in hh/one available vehicle	40,000	20,000	60,000
Two persons in hh/one available vehicle	28,000	32,000	60,000
Two or more available vehicles	20,000	20,000	40,000

TOTAL	100,000	80,000	

Not only is the grand total off (180,000 vs. 200,000), but the row and column totals do not match the corresponding figures in**Table 3.6**.

The process of bringing the two sets of numbers into line starts with an adjustment to the row totals. Let T_{jk} designate the sum of the weights for a given cell, T_{j+} the sum across the cells in row j, and T_{+k} the total across the cells in column k. We will use superscripts to denote the different iterations of the process, with 0 representing the initial weights and totals before any adjustment to population figures. The new weight adjusted to the row population figures will simply be the old, unadjusted weight times an adjustment factor:

$$W_{ijk}^{1,r} = W_{ijk}^{0} \frac{N_{j+}}{T_{j+}^{0}} \quad .$$

The adjustment factor (N_{j+} /T^0_{j+}) is the ratio between the population figure for row j and the sum of the current weights in that row. For example, all the weights in the cell for households with no available vehicles are increased by 1.25 (=25,000/20,000). After the application of the adjustment factor, the sum of the weights in each row equals the population figure for the row (N_{j+}). This is shown in **Table 3.8.** Unfortunately, the column totals are still off.

Table 3.8
Sums of Sample Weights after Initial Adjustment to Row Targets

SUBGROUP	TOWNSHIP GROUP		TOTAL
	Low Response Rate	High Response Rate	
HOUSEHOLD CLASSES			
No available vehicles	15,000	10,000	25,000
One person in hh/one available vehicle	33,333	16,667	50,000
Two persons in hh/one available vehicle	35,000	40,000	75,000
Two or more available vehicles	25,000	25,000	50,000
TOTAL	108,333	91,667	

Thus, the next step is to adjust the new weights to the column totals. Once again, this is done by multiplying the current weights by an adjustment factor—the ratio between the population figure for the column and the sum of the current weights for that column:

$$W_{ijk}^{1,c} = W_{ijk}^{1,r} \frac{N_{+k}}{T_{+k}^{1,r}} \quad .$$

That is, the weights of cases in the first column will be adjusted by a factor of about 1.11 (≈ 120,000/108,333) and those of cases in the second column will be adjusted by a factor of about 0.87 (≈ 91,667 / 80,000). **Table 3.9** shows the results. The column totals match the targets (except for rounding error) but now the row totals are off.

Table 3.9
Sums of Sample Weights after Adjustment to Column Targets

SUBGROUP	TOWNSHIP GROUP		TOTAL
	Low Response Rate	High Response Rate	
HOUSEHOLD CLASSES			
No available vehicles	6,615	8,727	25,342
One person in hh/one available vehicle	36,923	14,546	51,469
Two persons in hh/one available vehicle	38,769	34,909	73,678
Two or more available vehicles	27,692	21,818	49,510
TOTAL	119,999	80,000	

The whole process is now repeated, starting with the sums in **Table 3.9** (instead of those in **Table 3.7**). More generally, the weights produced in one iteration (iteration m+1) adjust those produced in the previous iteration (iteration m):

$$W_{ijk}^{m+1,r} = W_{ijk}^{m,c} \frac{N_{j+}}{T_{j+}^{m,c}}$$

$$W_{ijk}^{m+1,c} = W_{ijk}^{m+1,r} \frac{N_{+k}}{T_{+k}^{m+1,r}}$$

For example, the weights in the first row of **Table 3.9** would be adjusted by a factor of 1.01 (≈ 25,342 / 25,000); those in the second row would be adjusted by a factor of 1.03 (≈ 51,469 / 50,000); and so on.

The process generally produces only small changes after three or four iterations. This method can be used with three or more dimensions as well as with two, as in our illustration.

Recommendation

ADJUST WEIGHTS TO POPULATION TOTALS, EVEN WHEN TOTALS ARE AVAILABLE ONLY FOR INDIVIDUAL VARIABLES —
Iterative proportional fitting (also known as frataring or raking) can be used when independent population estimates are not available for every weighting cell, but are available for row and column totals.

EDITING AND IMPUTATION

We have recommended both nonresponse weighting and post-stratification as methods for reducing the impact of unit nonresponse on survey results. When individual data items are missing, we recommend a different approach—the imputation of missing values. With imputation, information that is obtained for a case is used to make a guess about the information that is missing.

Statistical imputation should be distinguished from editing of the data or “logical” imputation. It is sometimes possible to figure out what the missing value should be from an examination of the data that were obtained. For example, the destination for a trip may be missing in the diary for one family member but included in the diary for another. Under these circumstances, it may be reasonable to infer, or “logically impute,” the missing destination. Similarly, when a respondent’s sex is missing but she is listed as the mother of another household member, it is reasonable to fill in her sex as female. Such editing procedures are outside the scope of this report but they can also help reduce the amount and impact of missing data.

Three Statistical Techniques. Three *statistical* techniques are commonly used to impute missing values[6]:

- **Hot deck imputation**, in which the missing value is replaced by the value obtained from another case (the “donor”) with similar characteristics;
- **Regression-based imputation**, in which the missing value is replaced by the value predicted from a regression model;
- **Cell mean imputation**, in which the missing value is replaced by the mean of the values obtained from all other cases that provided data.

With both hot deck and cell mean imputation, cases are first grouped into cells based on variables that are obtained for all (or nearly all) of the cases. The imputation cells serve a function similar to the one played by the adjustment cells in nonresponse weighting. Cases with similar values on the survey variables of interest are placed together in a cell, and the data for the respondents within the cell (or those with complete data) are used to represent the nonrespondents (or those with missing data).

Cell mean imputation is rarely used in surveys, because (except in a few instances) it leads to biased estimates—by replacing each missing value with the cell mean, this method of imputation leads to serious underestimates of the variance of the survey statistics.

Regression-based methods also have problems:

- Unless a "residual" is added to the imputed value, the variance of survey statistics will be underestimated;

- Different models have to be developed for each variable with missing values, and this model-building can be difficult and time-consuming;

- Missing values in the predictors can greatly complicate the imputation process.

Both cell-mean and regression-based imputation have their place. But for most surveys, the hot deck approach will be the most practical method.

It is worth noting that new methods of imputing missing values are being developed all the time. The new methods include multiple imputation (in which each missing value is replaced repeatedly, yielding multiple imputed values) and more sophisticated methods of single imputation. Multiple imputation has the advantage that, when done correctly, it does *not* underestimate the variance introduced by the imputation process[7]. All the other methods tend, to some degree, to lead to underestimation of the variance of the estimates (with cell-mean imputation producing the worst underestimation). However, software is only now being developed for these new methods; as a result, hot deck imputation, which can be carried out with widely available software, remains the most practical method for general use.

Using hot deck imputation. Hot deck imputation is carried out in four steps:

1) **Group cases into imputation cells.** The goal here is to group cases likely to have similar values on the variable in question. Depending on the variable, the imputation cells might take into account household size and composition, the predominant mode of transportation, and the age and sex of the respondent.

2) **Sort the cases within cells.** Often cases are then sorted within the imputation cells. For example, cases may be sorted by the total number of trips they made during the survey period. The sorting allows the imputation to take into account continuous variables as well as the categorical variables used in forming the imputation cells. The final sort may also be done randomly.

3) **Replace the missing values.** The value imputed to a case is just the actual value for the preceding case (the "hot deck" or record) in the cell that has a non-missing value for the variable in question. For example, if the third case in the imputation cell had a valid value for a variable and the fourth case was missing that variable, then the value for the third case (the "donor") would be used as the imputed value for the fourth. Sometimes, a limit is imposed on the number of times a given case can serve as a donor (for example, no more than three times). This limits the impact of any single case on the final survey statistics.

4) **Edit the imputed values.** Imputation can produce values that are inconsistent with other information about the case. As a result, the imputed values should undergo the same editing and consistency checks as other values. Inconsistent or out-of-range values should be reimputed.

As this description makes clear, hot deck imputation is not a simple process. Moreover, it may be useful to form different imputation cells for different variables, making the process even more complex. Several computer programs are available for carrying out imputation and may simplify the work involved.

Still, it is worth bearing in mind the following principles about hot deck imputation:

1) **It is a statistical procedure and it works well on the average.** It should not be used to impute data that are unique to an individual case. For example, it would not usually make sense to use hot deck imputation to impute a purpose for a specific trip.

2) **The more data that are imputed for a given case, the less accurate the imputed values are likely to be.** It makes sense to impute one or two key missing values for a case. It makes far less sense to impute all of the data for a missing person within an otherwise complete household, or to impute all of the data for a missing trip in an otherwise completed diary.

Recommendation

IMPUTE MISSING VALUES FOR KEY SURVEY VARIABLES — *Use hot deck imputation to replace missing values on a few key survey variables. Statistical imputation should not be used to impute all the data for a person or all the data regarding a trip.*

1. For an introduction to the statistical treatment of nonresponse, see Lessler, J.T., & Kalsbeek, W.D. (1992). *Nonsampling Error in Surveys*. New York: John Wiley & Sons.

2. Kim, H., Li, J., Roodman, S., Sen, A.., Sööt, S., & Christopher, E. (1993). Factoring household travel surveys. *Transportation Research Record*, 1412, 17-22.

3. Kalton, G. (1983). *Compensating for Missing Survey Data*. Ann Arbor: Institute for Social Research.

4. Kish, L. (1992). Weighting for unequal P_i. *Journal of Official Statistics*, 8, 183-200.

5. Fratar, T., Brant, A.E., & Buttz, C.W. (1962). Estimating traffic patterns in urban areas. Paper presented at the International Road Federation, Ivth World Meeting, Madrid Spain, October 14-20.

6. Little, R.J., & Rubin, D.B. (1987). *Statistical Analysis with Missing Data*. New York: John Wiley & Sons.

7. Rubin, D. (1986). Basic ideas of multiple imputation for nonresponse. *Survey Methodology*, *12*, 37-47.

APPENDICES

APPENDIX 1:
QUESTIONS AND ANSWERS TO AVERT NONRESPONSE

Q1. I'm too busy.

a. We can reschedule the interview for a time that's more convenient for you.
b. Let's start the interview now and see how far we can get.

Q2. What's the purpose of this call?

a. [If an advance letter has been sent] I'm [NAME] from [FIRM NAME]. This is about the letter you should have received in the past few days.
b. We're conducting research for [AGENCY NAME].
c. This is not market research, we are not selling anything, and I'm not soliciting money.
d. [If necessary, give the respondent the name and number of someone they can call to verify your name and affiliation.]

Q3. How long will this take?

a. On the average, the interview takes about [AVERAGE LENGTH] minutes.

Q4. How did you get my name/number?

a. Your household was selected as part of a scientific sample of [AREA NAME].

Q5. What's the purpose of this study?

a. We are gathering data on transportation. This information is important to area transportation planners.
b. The study is sponsored by [AGENCY NAME].
c. The information from the study will be used by local governments to make plans for new roads, improve commuter services, and for other planning purposes.

Q6. How can I be assured of confidentiality?

a. Before processing the information, we remove any connection between your name or that of your family from the questionnaire.

b. The data are for statistical purposes only. That's why it's not necessary to connect your name with the information you provide.

c. All people working on the study are required to sign forms pledging to keep the data confidential. They will be fired if they violate this pledge.

Q7. I don't do surveys.

a. This is not market research. It is a study sponsored by [AGENCY NAME], which is responsible for transportation planning in this area.

b. Your name will not be sold to any mailing lists.

Q8. Can't someone else do it? There are plenty of other people with more time.

a. You were chosen through a scientific sampling procedure. You cannot be replaced.

b. We can't replace you with anyone else. We need your personal knowledge.

c. You represent many other people. It is important that you be included in the survey.

Q9. I'm not representative of other people.

a. Exactly, and other people cannot speak for you. That's why you are invaluable to our study.

Q10. Why is [AGENCY NAME] interested in this?

a. They are responsible for transportation planning in this region of the country.

Q11. Who's paying for this?

a. This study is being funded by [AGENCY NAME].

Q12. Do I have to do the interview?

a. Your participation is completely voluntary. If you prefer not to answer certain questions, all you have to do is say so. But we appreciate any help you can give us since it is such an important study.

Q13. I don’t want to be part of your data bank.

a. I want to assure you that everything you say is completely confidential. When the data are put into the computer, your name is not included. The information is only for statistical purposes. No individuals or families are identified.

Q14. How can I be sure you won’t give my name to someone else?

a. Our contract to carry out this study expressly forbids us from releasing the names of the people who take part in the study. All employees at [FIRM NAME] sign a pledge to keep the names of respondents confidential.

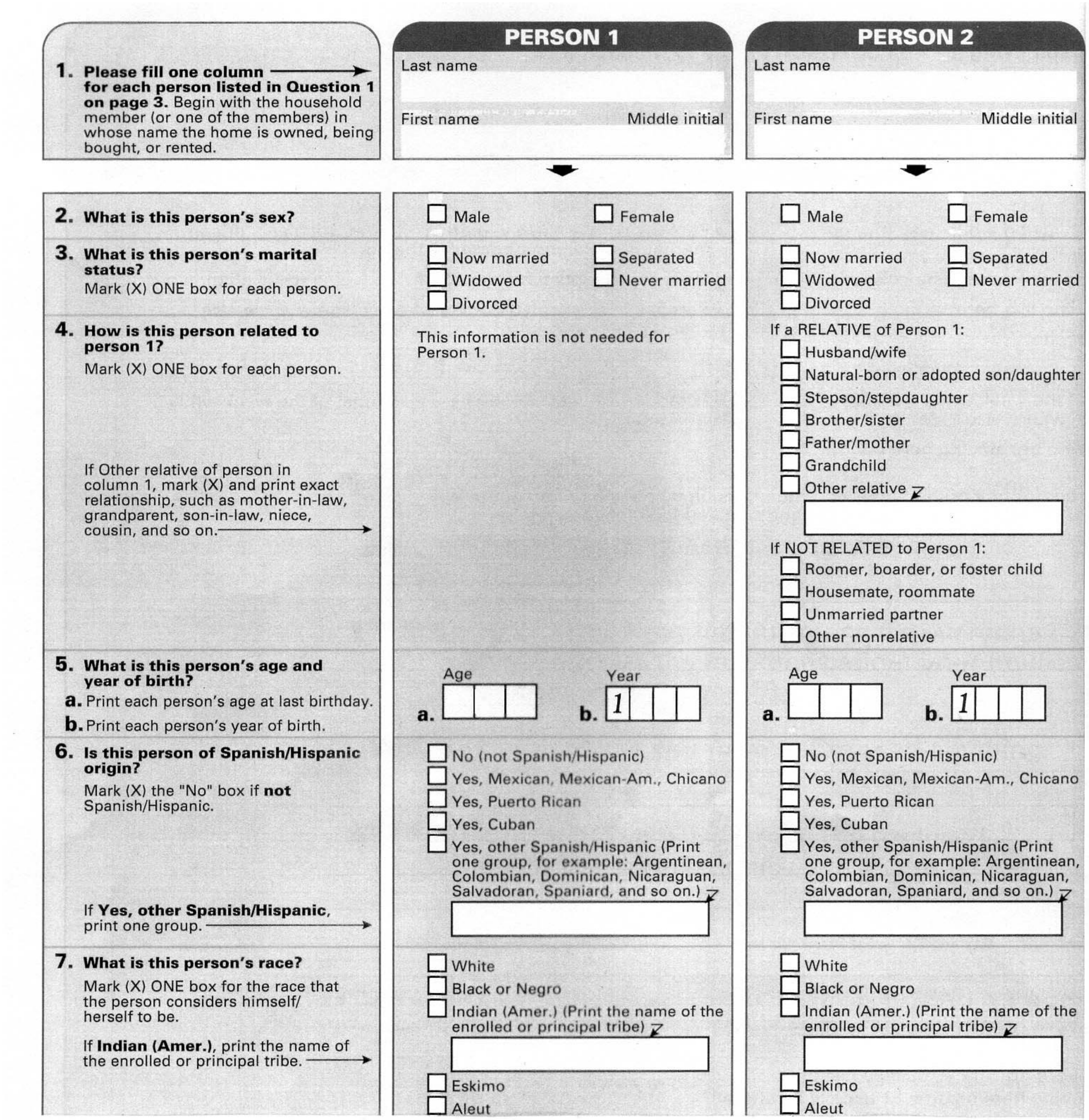
APPENDIX 2:
EXAMPLES OF GOOD AND BAD QUESTIONNAIRE DESIGN

Page 2

PLEASE ALSO ANSWER HOUSING QUESTIONS

	PERSON 1	PERSON 2
Please fill one column for each person listed in Question 1a on page 1.	Last name First name / Middle initial	Last name First name / Middle initial
2. How is this person related to PERSON 1? Fill ONE circle for each person. If **Other relative** of person in column 1, fill circle and print exact relationship, such as mother-in-law, grandparent, son-in-law, niece, cousin, and so on.	START in this column with the household member (or one of the members) in whose name the home is owned, being bought, or rented. If there is no such person, start in this column with any adult household member.	If a RELATIVE of Person 1: ○ Husband/wife ○ Natural-born or adopted son/daughter ○ Stepson/stepdaughter ○ Brother/sister ○ Father/mother ○ Grandchild ○ Other relative If NOT RELATED to Person 1: ○ Roomer, boarder, or foster child ○ Housemate, roommate ○ Unmarried partner ○ Other nonrelative
3. Sex Fill ONE circle for each person.	○ Male ○ Female	○ Male ○ Female
4. Race Fill ONE circle for the race that the person considers himself/herself to be. If **Indian (Amer.)**, print the name of the enrolled or principal tribe. If **Other Asian or Pacific Islander (API)**, print one group, for example: Hmong, Fijian, Laotian, Thai, Tongan, Pakistani, Cambodian, and so on. If **Other race**, print race.	○ White ○ Black or Negro ○ Indian (Amer.) (Print the name of the enrolled or principal tribe.) ○ Eskimo ○ Aleut Asian or Pacific Islander (API) ○ Chinese ○ Japanese ○ Filipino ○ Asian Indian ○ Hawaiian ○ Samoan ○ Korean ○ Guamanian ○ Vietnamese ○ Other API ○ Other race (Print race)	○ White ○ Black or Negro ○ Indian (Amer.) (Print the name of the enrolled or principal tribe.) ○ Eskimo ○ Aleut Asian or Pacific Islander (API) ○ Chinese ○ Japanese ○ Filipino ○ Asian Indian ○ Hawaiian ○ Samoan ○ Korean ○ Guamanian ○ Vietnamese ○ Other API ○ Other race (Print race)
5. Age and year of birth **a.** Print each person's age at last birthday. Fill in the matching circle below each box. **b.** Print each person's year of birth and fill the matching circle below each box.	a. Age 0 ○ 0 ○ 0 ○ 1 ○ 1 ○ 1 ○ 2 ○ 2 ○ … 9 ○ 9 ○ b. Year of birth 1 1 ● 8 ○ 0 ○ 0 ○ 9 ○ 1 ○ 1 ○ 2 ○ 2 ○ … 9 ○ 9 ○	a. Age 0 ○ 0 ○ 0 ○ 1 ○ 1 ○ 1 ○ 2 ○ 2 ○ … 9 ○ 9 ○ b. Year of birth 1 1 ● 8 ○ 0 ○ 0 ○ 9 ○ 1 ○ 1 ○ 2 ○ 2 ○ … 9 ○ 9 ○
6. Marital status Fill ONE circle for each person.	○ Now married ○ Separated ○ Widowed ○ Never married ○ Divorced	○ Now married ○ Separated ○ Widowed ○ Never married ○ Divorced
7. Is this person of Spanish/Hispanic origin? Fill ONE circle for each person.	○ No (not Spanish/Hispanic) ○ Yes, Mexican, Mexican-Am., Chicano ○ Yes, Puerto Rican ○ Yes, Cuban ○ Yes, other Spanish/Hispanic	○ No (not Spanish/Hispanic) ○ Yes, Mexican, Mexican-Am., Chicano ○ Yes, Puerto Rican ○ Yes, Cuban ○ Yes, other Spanish/Hispanic

Example of poor information organization—a trial version of the U.S. Decennial Census questionnaire using a matrix format.



	PERSON 1	PERSON 2
1. Please fill one column → for each person listed in Question 1 on page 3. Begin with the household member (or one of the members) in whose name the home is owned, being bought, or rented.	Last name First name Middle initial	Last name First name Middle initial
2. What is this person's sex?	☐ Male ☐ Female	☐ Male ☐ Female
3. What is this person's marital status? Mark (X) ONE box for each person.	☐ Now married ☐ Separated ☐ Widowed ☐ Never married ☐ Divorced	☐ Now married ☐ Separated ☐ Widowed ☐ Never married ☐ Divorced
4. How is this person related to person 1? Mark (X) ONE box for each person. If Other relative of person in column 1, mark (X) and print exact relationship, such as mother-in-law, grandparent, son-in-law, niece, cousin, and so on. →	This information is not needed for Person 1.	If a RELATIVE of Person 1: ☐ Husband/wife ☐ Natural-born or adopted son/daughter ☐ Stepson/stepdaughter ☐ Brother/sister ☐ Father/mother ☐ Grandchild ☐ Other relative If NOT RELATED to Person 1: ☐ Roomer, boarder, or foster child ☐ Housemate, roommate ☐ Unmarried partner ☐ Other nonrelative
5. What is this person's age and year of birth? **a.** Print each person's age at last birthday. **b.** Print each person's year of birth.	a. Age b. Year 1	a. Age b. Year 1
6. Is this person of Spanish/Hispanic origin? Mark (X) the "No" box if **not** Spanish/Hispanic. If **Yes, other Spanish/Hispanic**, print one group. →	☐ No (not Spanish/Hispanic) ☐ Yes, Mexican, Mexican-Am., Chicano ☐ Yes, Puerto Rican ☐ Yes, Cuban ☐ Yes, other Spanish/Hispanic (Print one group, for example: Argentinean, Colombian, Dominican, Nicaraguan, Salvadoran, Spaniard, and so on.)	☐ No (not Spanish/Hispanic) ☐ Yes, Mexican, Mexican-Am., Chicano ☐ Yes, Puerto Rican ☐ Yes, Cuban ☐ Yes, other Spanish/Hispanic (Print one group, for example: Argentinean, Colombian, Dominican, Nicaraguan, Salvadoran, Spaniard, and so on.)
7. What is this person's race? Mark (X) ONE box for the race that the person considers himself/herself to be. If **Indian (Amer.)**, print the name of the enrolled or principal tribe. →	☐ White ☐ Black or Negro ☐ Indian (Amer.) (Print the name of the enrolled or principal tribe) ☐ Eskimo ☐ Aleut	☐ White ☐ Black or Negro ☐ Indian (Amer.) (Print the name of the enrolled or principal tribe) ☐ Eskimo ☐ Aleut

A revision of the same questionnaire that retains the matrix organization but offers improved navigational aids. The white answer spaces stand out against the shaded background (blue in the original); in addition, the arrows at the top of each column and the white borders between the columns guide the respondent through the questions in the intended order.

GLOSSARY OF TERMS

ACTIVITY: The main business or undertaking engaged in by an individual, alone or with others. Activities include: work and work-related, school, shopping, personal business, preparing a meal, cleaning the home, athletics/exercise, visiting, etc. Activities have the following associated attributes: type, duration, start time, end time, location, and mode of travel to get to the activity location (in the case of an activity that takes place at a location that is different from the previous activity). Participation in activities at different locations is the underlying reason for trip-making. Trip purpose is associated with the main activity type at the destination, for those activities requiring a trip.

ADVANCE LETTER: A letter sent to a sample member (household or person) via mail in advance of the interviewer's attempt to contact the member. An advance letter is meant to "warm-up" the member to the forthcoming call from the interviewer.

CALLBACKS: The re-approach of a sample member after previous contact attempts have been unsuccessful. Callbacks reduce the likelihood of nonresponse error by improving the contact rate.

CALLING PROTOCOL: A calling schedule for contacting sample members. Most calling protocols take into account the outcomes and times of previous calls in an attempt to improve the chances of reaching sample members.

COGNITIVE INTERVIEW: A technique for developing survey questionnaires that focuses on the thought processes respondents go through as they arrive at answers to survey questions. In a cognitive interview, respondents are asked to think out loud as they answer draft survey questions. They may also be asked to respond to a number of follow-up probes to reveal how they arrived at their answers and whether the content or wording of the items should be improved.

COMPUTER ASSISTED PERSONAL INTERVIEWING (CAPI): Face-to-face interviewing performed with the assistance of a computer. In a CAPI interview, the interviewer reads the questionnaire items from a computer screen and records the respondent's answers by entering them into the computer. CAPI systems may also form and manage the sample, display the introductory script, control the wording of the items, check for internal consistency among the respondent's answers, and perform any number of other operations associated with the collection or management of the data.

COMPUTER ASSISTED TELEPHONE INTERVIEWING (CATI): Telephone interviewing performed with the assistance of a computer. CATI systems are similar to CAPI systems in that the questionnaire items are displayed on line and the interviewer enters the respondent's answers with the keyboard or mouse. Most CATI systems perform other operations related to the management or collection of data by telephone. They may, for example, supply the interviewers with instructions, manage the sampling pool, schedule call attempts, record the outcomes of the calling effort, and/or monitor the progress of the interviewers.

CONVERTING: Recontacting initial refusals one or more times in an attempt to persuade them to participate in the survey.

CRITICAL ITEMS: A set of questions that must be completed by the sample member to classify the case as a respondent. The set typically includes all items that are essential to accomplishing the major goals of the survey.

DISPOSITION CODE: A code assigned to each case (member) in the sample that records the most recent or final outcome (e.g., ring-no-answer, nonworking number, respondent refusal, etc.) of the data collection effort. Disposition codes are used to track the status of each sample member, to monitor and manage the field effort, and to compute response rates.

ELIGIBLE UNITS (ELIGIBLES): Sample units who are eligible to participate in the survey because they are part of the target population.

FIELD PERIOD: The time period during which data are collected from sample members.

HOUSEHOLD: The U.S. Bureau of the Census defines a household as all persons who occupy the same housing unit. A household may consist of a family, one person living alone, two or more families living together, or any other group of related or unrelated persons who share living arrangements.

HOUSING UNIT: The U.S. Bureau of the Census defines a housing unit as a house, apartment, mobile home, group of rooms, or single room that is occupied (or if vacant, is intended for occupancy) as separate living quarters. To qualify as a housing unit, the occupants must live and eat separately from other persons in the building and have direct access to their unit from the outside of the building or through a common hall.

IMPUTATION: A statistical or logical technique by which missing data are inferred from other information provided by the respondent and/or other respondents.

INCENTIVE: A monetary or nonmonetary gift or payment offered to sample members in an effort to gain their cooperation.

INELIGIBLE UNITS (INELIGIBLES): Sample units who are not qualified to participate in a survey because they do not belong to the target population.

ITEM NONRESPONSE (MISSING DATA): The failure to obtain a specific piece of data from a responding member of the sample.

LANGUAGE BARRIERS: Language difficulties that prevent or hinder a sample member from participating in a survey. Such difficulties typically arise when the native language of a sample member is something other than the language(s) used in the survey.

LOGICAL IMPUTATION/EDITING: Techniques by which missing data are logically (rather than statistically) deduced from known information about the respondent.

MISSING DATA: See *item nonresponse*.

NONRESPONSE: Failure of the survey to obtain the desired information from eligible sample members.

NONRESPONSE WEIGHTING: Postsampling statistical adjustment (weighting) to partially compensate for possible nonresponse error. Nonresponse weighting is different from the postsampling weighting that is routinely performed to adjust for unequal probabilities of selection.

NONWORKING NUMBER: A telephone number that has not been assigned to a unit by the telephone company.

OPEN-ENDED QUESTION: A survey question that asks the respondent to phrase the answer in his or her own words.

PAPER-AND-PENCIL INTERVIEWING (PAPI): An interview in which the interviewer reads the questions from a printed questionnaire and records the answers directly onto the questionnaire.

POSTSTRATIFICATION: When case weights are adjusted to agree with independent estimates of population totals. Poststratification compensates for differences between the distribution of characteristics in the sample and the distribution of characteristics in the target population.

PRETEST (PILOT TEST): A relatively small number of practice interviews used to test and refine the survey materials and/or the field procedures before the conduct of the main survey.

PROXY: Someone other than the selected sample member who answers survey questions on behalf of that member.

RANDOM DIGIT DIALING (RDD): Techniques that form samples by adding random digits to the telephone prefixes that fall within the sampling area so as to include listed (published) and unlisted numbers in the sample.

REFUSALS: Sample members who refuse to participate in the screening or main data collection phase of a survey.

RESPONDENT RULES (RESPONDENT SELECTION): Rules that are used by the interviewers to choose a respondent from all eligible members within a sampling unit.

RESPONSE RATE: A measure of a survey's level of success in obtaining measurements for all eligible units in the sample; the number of respondents divided by the total number of eligible units in the sample.

RESULT CODE: See *disposition code*.

SAMPLING FRAME: A list of units that includes the target population from which the survey sample is drawn.

SCREENING INTERVIEW: A preliminary interview used to determine the eligibility of sample members. Sample members who meet the eligibility requirements are often recruited for more detailed data collection during this interview.

SELF-ADMINISTERED QUESTIONNAIRE: A questionnaire that is completed by the sample member without the assistance of an interviewer. Respondents to self-administered questionnaires are asked to read the questions and record the answers on their own.

STATISTICAL IMPUTATION: Statistical technique by which missing data for questions are inferred from information provided by the respondent and/or other respondents. In hot deck imputation, data are taken from another case (the donor) similar to the case with the missing data. In regression-based imputation, the imputed value is predicted from other information provided by the same case.

STRATIFICATION: Process in which units with similar characteristics are divided into groups called strata before the sampling process begins. Each unit is assigned to one and only one stratum based on prior knowledge about the unit. Separate samples are then selected within each stratum.

STRATIFIED SAMPLE: The sample that is formed when independent random samples are selected from each stratum (or cell) of a stratified sampling design. (Also see *stratification*.)

TARGET POPULATION: The finite population the survey intends to cover.

TERMINATIONS/BREAKOFFS: Sample members who begin but terminate a interview before they provide enough information to be classified as respondents.

TRIP: One-way travel from one address (place) to another by any means of transportation (e.g., private motor vehicle, public transportation, bicycle, or walking). When the travel includes more than one destination and the travel time between destinations exceeds five minutes, or the purpose for travel to one location differs from that of the other, the travel between locations is counted as a separate trip.

UNIT NONRESPONSE: The failure to obtain questionnaires or data collection forms from an eligible sample member.

WEIGHTED RESPONSE RATE: A response rate that is calculated using the inverse of the selection probabilities as the weight. The weighted response rate is an estimate of the proportion of the target population represented by the respondents to the survey.

WORKING RESIDENTIAL NUMBER (WRN): Telephone numbers that have been assigned to residential housing units by the telephone company.

REFERENCES

Henry, Gary T. (1990). *Practical Sampling*. Newbury Park, CA: Sage.

Lavrakas, Paul J. (1993). *Telephone Survey Methods: Sampling, Selection, and Supervision*. 2nd ed. Newbury Park, CA: Sage.

U.S. Department of Commerce. (1992). Bureau of the Census. 1990 Census of Population and Housing . *Technical Documentation*: *Summary Tape File 3 on CD-ROM*

U.S. Department of Transportation. Federal Highway Administration.*User's Guide for the Public Use Tapes: 1990 National Personal Transportation Survey*. Publication No. FHWA-PL-92-007 Henry, Gary T. (1990). *Practical Sampling*. Newbury Park, CA: Sage.