United States Environmental Protection Agency | Office of Information Analysis and Access Washington, DC 20460 | May 2003 EPA 260-B-03-003 http://www.epa.gov

Survey Management Handbook

Table of Contents

EXECUTIVE SUMMARY

This handbook provides program managers, analysts and planning teams who work with the U.S. Environmental Protection Agency (EPA) with a basic guide to survey management. This clear-cut guide is directed to people who have the responsibility for or an interest in collecting survey data, or who may procure survey research services, rather than whose who already have an advanced knowledge of survey research methodology or statistics. It provides a comprehensive description of all the elements that make up a well designed and well executed survey, building upon firmly established principles of survey research and augmented by a commonsense approach to meet the special needs of EPA. It is to be used as guidance only and does not impose any legally binding requirements upon the Agency.

The first edition of the Survey Management Handbook, published in 1984, received wide use by EPA sponsors of survey research projects. Updated references were added to this volume to provide current authoritative sources for additional clarification. More importantly, this revised volume includes the most up-to-date techniques used to design a survey project and the latest approaches for data collection. Since the last edition, the use of computer technology to assist in conducting interviews is an important development in the field of survey research, specifically computer assisted technologies (CATI) for face-to-face interviews.

Under the auspices of EPA's Office of Environmental Information (OEI), this handbook was prepared under contract with Temple University's Institute for Survey Research (ISR) as part of a series of documents that are being produced to assist agency offices and programs in following Information Quality Guidelines that comply with the Office of Management and Budget (OMB) guideline (RFL-7157-8, March 2002). The content of this document is not intended to cover OMB clearance requirements that may be required for some surveys under the Paper Work Reduction Act of 1980 with amendments. The guidance for obtaining OMB approval can be found in the *ICR Handbook: EPA's Guide to Writing Information Collection Requests Under the Paperwork Reduction Act (PRA*) of 1995 at http://www.epa.gov/icr/

Writers Tom Jabine, Rinaldo Iachahn, Ph.D. and Alan Fox, Ph.D, assisted Mel Kollander, Director of ISR's Washington Office, the lead author.

Questions about this document should be directed to:

U.S. Environmental Protection Agency
EPA West Building
1200 Pennsylvania Ave., N.W.
Washington, DC 20460
Phone: (202) 566-0593
E-mail: Ross.Np@epamail.epa.gov

Phil Ross, Ph.D.
Chief Statistician
Office of Information

Introduction

Statistical surveys play a critical role in Agency decision-making. As policymakers demand more quantitative support for Agency decisions, program managers are giving careful consideration to statistical survey reports and their implications in the framing of regulatory decisions and long-range environmental policies. Reliable survey data on the duration, magnitude, and physical distribution of pollutants in the environment have proven invaluable for determining the precise degree of pollutant control needed to respond to various statutory mandates and the manner in which the Agency should exercise such control.

There have been extraordinary advances in survey methodology in the past few decades, the most striking of them in sampling, computer-assisted interviewing, data processing, imputation, and statistical analysis. This has made large-scale collection of demographic and economic facts easier, faster, cheaper, and more reliable. These advances have motivated survey sponsors to demand increasingly high standards in questionnaire design, data collection methodology, sampling, interviewing, data processing, and analysis.

The growing reliance on high-quality statistical work for Agency planning and policymaking, coupled with the recent advances in survey methodology, in fact, prompted the development a Survey Management Handbook in 1984. This version incorporates advances in survey methodology since then and, unlike the previous version, is contained in 1 volume. In it we examine the methods, procedures, and quality-assurance techniques typically used to collect, process, and analyze survey data, and the actions EPA project officials can take to ensure the technical soundness of all contract work performed during the course of a survey.

It is organized into six chapters, which correspond to the major components of a typical work plan for a statistical survey of human populations. Normally, a large survey research contractor conducts the work plan and the subsequent fieldwork, data processing, and often the analysis, with the EPA sponsoring office playing an oversight role throughout the contract.

The work plan usually consists of:

- An analysis plan
- Specification of the data collection method(s)
- A draft questionnaire and specifications for any pretests
- A sampling plan
- Interviewing procedures
- Data processing procedures

A summary of the topics covered in each of the six chapters is given on the next page.

Chapter Outline—Components of the Work Plan

Chapter 1—Analysis Plan
Examines the steps involved in defining the research objectives of the survey and choosing the analytic approach most appropriate for achieving these objectives.

Chapter 2—Data Collection
Describes the principal methods of collecting survey data and the factors influencing the choice of methods, and suggests ways to evaluate the method proposed for a particular EPA survey.

Chapter 3—Questionnaire Design and Pretesting
Examines the steps involved in developing a sound survey questionnaire, presents criteria for reviewing draft questionnaires, and recommends ways to monitor pretests.

Chapter 4—Sampling Plan
Describes the advantages of sampling, the principal methods of choosing a sample, the components of a sampling plan, and recommends ways to monitor sampling activities.

Chapter 5—Interviewing
Discusses the administrative and quality-assurance procedures typically used to organize, manage, and monitor a survey where interviewing is used to collect the data.

Chapter 6—Data Processing
Examines the steps involved in processing the raw data collected from samples to produce tabulations and analyses to achieve the research objectives. Includes discussion of imputation of missing values.

The survey methods and techniques we discuss are applicable to fairly large-scale surveys. This is because most of EPA's demographic, economic, and social investigations as well as field studies deal with large populations and issues that the Agency necessarily views from a national perspective.

Of course, not every empirical research project EPA undertakes requires the formal apparatus needed for a large-scale survey. Sometimes it is more appropriate to study a handful of cases intensively rather than investigate a representative sample, to interview a few individuals or groups informally rather than use the structured interviews prescribed for major statistical surveys, or to develop in-depth descriptions of a few individuals rather than aim for a set of statistics about a group. In fact, several different approaches may be used to resolve a particular survey research problem. The researcher's challenge is to identify approaches that are most likely to achieve the specific objectives of the project. The purpose of this Handbook is to help you meet this challenge.

Throughout, we discuss theoretical issues in very general terms. No background knowledge of statistics is presumed. In the event you wish to delve further into survey theory, a list of excellent sources is given at the end of each chapter. A complete list of these sources appears at the end of the Handbook, along with a glossary of terms.

Importance of a Consulting Survey Expert. We strongly suggest that you have a survey expert review your survey design and analysis plan early in the planning stage, certainly before you take steps to procure outside technical support. You also may find it necessary to get the advice of experts at various points of the survey in order to effectively apply the methods and techniques we recommend, especially with respect to sampling and data analysis. All too frequently, statisticians are called in after the data are collected, given a stack of completed questionnaires, and asked to make what they can of them. Unfortunately, because of gaps and omissions in the data, flaws in the survey design, mistakes in the questionnaire, and other problems that could easily have been avoided if a survey expert had been called in during the planning stage, there is very little that can be done.

This edition differs from the 1984 version in the following ways:

1. It is self-contained, in one volume; the previous edition came in 2 volumes.
2. It refers to new technology such as:
    - Computer assisted interviewing (CASI)
    - Random digit dialing (RDD) telephone surveys
    - Optical character recognition (OCR)
    - Imputation of missing data
3. It de-emphasizes face-to-face interview surveys in favor of telephone and mail surveys.
4. It contains new sections on the definition of response rates and the payment of incentives.
5. Bibliographic references have been updated to include only works currently available.

This edition is published under contract with the Environmental Protection Agency (Contract EPA-1W-0009-NTEX). Contributors include Thomas B. Jabine and Alan Fox, independent consultants, Ronaldo Iachan (Macro International), and Mel Kollander, with the able assistance of Michael Botts, Joshua A. Chamot, Robert Ricchio and Jonel Haley (Institute for Survey Research, Temple University).

Chapter 1: Analysis Plan

Introduction

In a given research situation, survey designers usually have a choice of research designs, methods of observation, methods of measurement, and types of analysis. All must fit together and be appropriate to the research problem. The choices the researchers make in each case will depend on how much is already known about the problems they are investigating and the specific reasons the information is needed.

Whether, as the survey sponsors, you intend to collect descriptive facts about a population or to delve deeper and attempt to explain certain facts in detail, requires a clear understanding of what you expect the research effort to achieve. Collecting data in the field is no substitute for well thought-out decisions beforehand about what is, and what is not, worth investigating. Without a clear idea of the objectives of your research, the survey is likely to result in much wasted time and money and the accumulation of much unwanted data.

This chapter discusses:

(i) The general approaches survey statisticians use to analyze and interpret survey data;

(ii) How to develop an analysis plan that will clearly define the purpose of your survey, the research objectives, the type of data to be collected, and the most appropriate method of analysis for achieving your research objectives

A. Approaches Used to Analyze Survey Data

In survey research, analysis means categorizing, ordering, manipulating, and summarizing data to obtain answers to research questions. The purpose of analysis is to reduce data to intelligible and interpretable form.

Analyzing data does not provide answers to research questions. Interpretation is necessary. To interpret is to explain. Interpretation takes the results of data analysis, makes inferences relevant to the relationships among the data, and draws conclusions about these relationships. The researcher who makes the interpretation searches the results for their meaning and implications.

A host of analysis techniques are available for studying survey data. However, here the focus is on four main approaches to analysis:

Approaches to Data Analysis

1. Qualitative analysis and evaluation
2. Statistical descriptions
3. Statistical inference
4. Analytic interpretation

Each of these approaches is discussed briefly in the order of their complexity and sophistication. Computer software (SPSS, SAS, and others) can help you do a lot of your own data analyses. Software programs should be consistent with Agency software.

1. Qualitative Analysis and Evaluation

In a qualitative analysis, the researcher's goal is to understand the characteristics of a few individuals, rather than the characteristics of a population or sub-group. A qualitative approach generally is not indicated for sample surveys, which are of major interest in this Handbook, but it may be the most suitable approach in some research situations.

For example, qualitative analysis is often the preferred approach for (a) analyzing the results of case studies or field studies, where a relatively small number of individuals (or specimens) are being investigated; (b) evaluating the results of informal research prior to conducting a full-scale statistical survey; and (c) developing hypotheses to test in a pilot study or a full-scale survey.

2. Statistical Description

Statistical descriptions are by far the most common method of reporting survey data. They often are referred to as "statistical analysis," but this relatively simple approach to the analysis of survey data simply involves working out statistical distributions, constructing tables and graphs, and calculating simple measures such as means, medians, measures of dispersion, percentages, proportions, etc. Statistical description can be used to describe data collected from a probability sample or an entire study population (a “census” survey).

Statistical descriptions are the tabulations researchers prepare, after the data are processed, to aggregate the features of the data file so the analysts can view the database in some intelligible and interpretable form. Statistical descriptions often are done in series, one variable or research question at a time being cross-classified with others, thus producing a descriptive summary of the relationships between the study variables.

3. Statistical Inference

In the broadest sense of the word, inference is the principal approach for analyzing statistical data. Inference is brought into play whenever data are collected from a probability sample rather than an entire population. When a probability sample is used, the researchers estimate the population characteristics from those of the sample as well as estimate sampling errors. Statistical inference is the linking of the results derived from data collected from or about a sample to the population from which the sample was drawn.

4. Analytic Interpretation

This term refers to the statistician’s attempts to explain the relationships between variables using various statistical analysis techniques. For example, researchers may employ multivariate regression to better understand the relationships between exposure to a particular pollutant and the socio-economic characteristics of a study population.

B. Preparing an Analysis Plan

This section will show you how to construct an analysis plan to complement the design specifications you establish for your survey. The basic criteria for the survey design and the analysis plan should be developed simultaneously, early in the planning stage. Constructing a well thought out analysis plan will help you define the design criteria so that you can achieve your research objectives with some desired level of accuracy considering the resources you have available. These design criteria, combined with the analysis plan, provide a sound conceptual framework for whatever work you and the contractor do during the rest of the survey.

The intent of these criteria is to guide the project staff in developing the survey specifications to procure whatever outside technical support may be necessary and to help the contractor prepare a technically and statistically-sound work plan. They may possibly be modified during the contract negotiations before being incorporated into the contract.

Constructing the analysis plan is a five-step process. The project office should develop this with the assistance of Agency statisticians, computer programmers, specialists in the subject area of the research, and systems analysts, as appropriate.

The end-products of the five steps, discussed below, are clear definitions of: (1) the purpose of the survey, (2) the objectives of the research (the main areas of investigation), (3) the data or variables to be investigated, (4) the analytic approaches and methods to be used to achieve the research objectives, and (5) the preliminary tabulations to be prepared from the completed data file after the data are processed.

Steps to Constructing Analysis Plan

1. Purpose of survey
2. Objectives of research
3. Data or variables
4. Analytic methods
5. Preliminary tabulations

Later, after the Agency and the contractor have studied the preliminary tabulations, the analysis plan can be refined to include specifications for additional, perhaps more sophisticated tabulations and the types of statistical analysis techniques that should be applied to fully reveal the informational content of the data base. Usually the contractor does this.

Step 1: Define the Purpose of the Survey

The statement of purpose in your analysis plan should clearly show how the data you plan to collect will result in information that will clarify or resolve some specific environmental problem that some authority has directed EPA to deal with. In other words, you should specify:

Purpose of the Survey

1. How the information is to be used
2. Problems to be addressed
3. Relationship to a specific mandate

Below is a statement of purpose that appeared in a report on an EPA field study of carbon monoxide (CO) using hand-held personal exposure monitors to test levels of CO in a variety of commercial settings. The EPA staff in the Office of Monitoring Systems and Quality Assurance of the Office of Research and Development conducted this survey. The statement clearly shows how the study results would be applied for planning and policymaking purposes, the problems the researchers intended to deal with, and their relationship to a specific EPA mandate.

> *The goal of air pollution control programs in the U.S., as mandated by Federal law and implemented by the States, is to attain National Ambient Air Quality Standards (NAAQS). The NAAQS for carbon monoxide (CO), for example, specify two different concentrations and averaging times, neither of which is to be exceeded more than once per year:*
>
> *35 parts per million (ppm) for 1 hour*
> *9 ppm for 8 hours.*
>
> *Both standards are intended to protect against the accumulation of more than 2% carboxyhemoglobin in the blood. ...*
>
> *Nondispersive infrared (NDIR) monitoring at fixed stations is the usual way for determining a given city's compliance with the NAAQS for CO. During the past decade, a number of studies have revealed that concentrations observed at fixed air monitoring stations have not been representative of concentrations sampled throughout an urban area. Some field studies have shown, for example, that commuters in traffic and pedestrians on downtown streets encountered CO levels above the NAAQS on a given date, while official air monitoring stations reported CO values below the NAAQS at the same time. Furthermore, studies of human activities suggest that most people spend the greatest proportion of any given 24-hour period indoors—in residences, stores, offices, factories, etc. These settings are not necessarily identical to sites selected for fixed air monitoring stations.*
>
> *These studies have raised questions about the usefulness of data generated by today's monitoring stations for protection of public health. An unanswered question is the degree to which conventional fixed stations either underestimate or overestimate the actual exposure of people as they go about their daily activities. The studies have stimulated interest in "exposure monitoring," which treats the person as a receptor and measures the pollutant levels actually contacting the person's body. ...*
>
> *Prior to the late 1970's there was no low cost, accurate means available for measuring CO concentrations to which people ordinarily were exposed in their daily lives. The advent of microelectronics has brought considerable progress in developing reliable, compact air quality monitoring instruments that can operate on batteries. The most dramatic of these are the new miniaturized personal exposure monitors (PEM's).... The present investigation is the first large-scale microenvironmental field study to make use of the new CO PEM instruments....*

Since the kinds of problems EPA has been directed to explore and manage encompass such a wide range of health and environmental issues, you may find it relatively easy to develop an adequate statement of purpose for your survey. What normally is far more difficult is building a set of arguments to justify the expenditure of program funds for your particular project, given the limited resources available to each program to address a mind-boggling number of priority issues. A comprehensive, well-reasoned analysis plan will help you build just such a set of arguments.

Step 2: Define the Research Objectives

Once you have justified the need for the survey from a planning or policymaking standpoint, you can begin to think about how to define its usefulness in "scientific" terms. The desired result should be a clear statement of the research objectives in terms of:

Research Objectives

- Kinds of questions you want answered
- Hypotheses to be tested
- Information to be collected

Questions to be answered

Continuing with the previous example, let's look at how the objectives of the PEM CO study were framed. EPA staff defined several sets of research questions.

The first set of research questions addressed the CO concentrations typically found in commercial settings, for example—

- What levels of CO ordinarily are present in typical commercial settings?
- Are CO levels in typical commercial settings usually zero, negligible, or above the NAAQS?

The second set of questions concerned the variability of CO concentrations and factors that may be associated with that variability. Examples from this set of questions are—

- How do CO concentrations vary over time within and between different cities for a given commercial setting?
- If CO is a street-level pollutant associated with vehicular traffic, do workers have greater protection in offices on the upper floors of a high-rise building?

Another set of research questions addressed the accuracy of the fixed-station monitors operated by air quality management districts to measure the air pollution to which the public is actually exposed, for example—

- Do CO concentrations measured in commercial settings using PEM's correlate with

ambient concentrations measured at fixed stations using NDIR instruments?

There also was a set of questions concerning the research methodology itself, including the following items—

- Is the CO PEM an effective tool for sampling air quality at a variety of urban locations?
- What are the implications of the present study for future research on exposures of the population to CO?

Hypotheses to be tested

Several hypotheses were formed and tested. For example, the researchers tested to see if the indoor concentrations were appreciably less than the outdoor concentrations when the entrance door to each commercial setting was closed.

The information to be collected was identified as—

- 5,000 concentrations of CO at one-minute intervals using PEM's for instantaneous measurement in a variety of commercial settings in several California cities over a nine-month period.

Ultimately five principal objectives were framed:

- To determine the CO concentrations typically found in commercial settings
- To determine the variability of CO concentrations in commercial settings and the time and spatial factors that may be associated with that variability
- To define and classify microenvironments which are applicable to commercial settings
- To determine how accurately fixed station monitors measure the CO settings
- To develop research methodology for measuring CO concentrations in field surveys using PEM's

When you frame your research objectives, be sure they are both specific and answerable. For example, a question like "is water contaminated by aldicarb?" is not answerable, while the following is: *"What proportion of the U.S. population is consuming water that contains more than seven parts per million (ppm) of aldicarb?"* This question, in fact, was an attempt to frame the objectives of an EPA-sponsored field study concerning the pesticide aldicarb, which was believed to be contaminating drinking water in certain communities.

It is impossible to overestimate the importance of framing the research objectives of your survey fully and precisely. No amount of data manipulation later can overcome the problems that may result from poorly defined objectives. Furthermore, once you have defined them, do not attempt to broaden their scope with further research topics or include other types of information unless you are sure of achieving your initial objectives with the resources you have available.

Step 3: Define the Study Variables

Once the objectives are clearly defined, the next step is to define the key variables of the study.

In other words, you will have to identify the specific data items that will be required to meet your stated objectives. A variable is a characteristic of a sample or of a population that varies in magnitude. In surveys of human populations, common variables are age, sex, race, income level, education level, etc.

Returning to our CO PEM example, the basic variable was:

- The average of two simultaneously taken one-minute samples of CO concentrations.

Other variables were developed to test different hypotheses such as those used for comparing indoor and outdoor CO concentrations using different settings of the personal exposure monitor and with the doors of the commercial establishments open and closed, such as—

- Mean CO concentration of indoor PEM setting i with entrance door closed
- Mean CO concentration of outdoor setting i with entrance door closed
- Mean CO concentration of indoor setting j with entrance door open
- Mean CO concentration of outdoor setting j with entrance door open

Step 4: Specify the Analytic Approaches and Methods

Following the guidelines provided in section A of this chapter, the next step in developing the analysis plan is to determine which analytic approach will allow you to achieve your research objectives most efficiently given the time and resources you have available. This means determining which analysis methods are most likely to achieve each of your research objectives. Note that different observation methods, measurement techniques, and analysis methods may be needed to fulfill each of your research objectives.

For most studies of human populations, a questionnaire is the basic information-gathering tool. If you choose this method, you may want to prepare a list of preliminary questions that will measure the study variables you identified in the previous step (see Chapter 3 for details on preparing a questionnaire). You'll also have to decide what level of accuracy (or precision) you will require. The level of accuracy should depend on how you plan to use the results of the survey. And, finally, you'll have to determine what minimally acceptable rate of response (target response rate) is necessary to achieve your research objectives.

You do not have to determine either the measurement techniques or any specific analysis techniques that may be needed to meet your research objectives—that is usually best left to the contractor.

The method of analysis used in the CO PEM study was to use miniaturized personal exposure monitors to measure CO in commercial settings in five California cities and suburbs. Then a number of hypotheses were tested by determining whether there were significant differences between sample results. In all, 588 commercial facilities were visited, including retail stores, office buildings, hotels, restaurants, department stores, and adjacent sidewalk and street intersections. Altogether 5,000 observations were recorded instantaneously at one-minute intervals as the investigators walked along sidewalks and into buildings.

Step 5: Define the Preliminary Tabulations

At a minimum, you should prepare a list of the preliminary tabulations (table shells) describing the form and content of the tables and graphs you want the contractor to generate when the data file is complete. There is nothing statistically sophisticated about tabulations. They are simply counts of the number of responses (or specimens) falling into each of several categories that have previously been defined.

The list of preliminary tabulations should include the title of each table and graph you want the contractor to prepare from the completed data file, and you should define the horizontal and vertical headings of each. Later, the contractor will total all the responses, specimens, or other items falling under each heading. Note that it is rarely possible to draw up a list of the final tabulations during the planning stage, especially if the subject matter is complex. Usually, most of the tabulations and analyses are not decided on until the results of the data file are in some intelligible and interpretable form.

One example of the tabulations created for the CO PEM study was the number of commercial settings by type of setting and geographic location. The following is a slightly abbreviated version of the table shell used for this study—

	Geographic Location			
Commercial Setting	Union Square San Francisco	University Avenue Palo Alto	Castro Street Mountain View	TOTAL
Indoor				
Restaurants				
Hotels				
Theaters				
Indoor subtotal				
Outdoor				
Arcade				
Intersection				
Mid-block				
Outdoor subtotal				
GRAND TOTAL				

For additional information, see “Steps in Processing Survey Data” in Chapter 6.

Bibliography: Chapter 1

U.S. Environmental Protection Agency, EPA Quality Manual for Environmental Programs, Washington, DC, EPA, 2000.

U.S. Environmental Protection Agency, Office of Environmental Information, EPA Quality Manual for Environmental Programs (EPA Manual 5360 A1), Washington, DC, EPA, 2001.

Chapter 2: The Data Collection Method

What data collection method should be used for a particular Agency survey? There is no general answer, and in many cases, any one of the major traditional collection methods—face-to-face interviews, telephone interviews, self-administered mail questionnaires, or some form of computerized data collection—may be equally suitable as the primary method.

Researchers no longer arbitrarily consider face-to-face interviews the most effective way of obtaining reliable survey data. If open-ended questions and extensive probing are used, the presence of a skilled interviewer may motivate the respondents to provide the richest and the most comprehensive data. However, in many other research situations, phone interviews or mail surveys may be just as effective in eliciting the needed data or even more so—and at a lower cost. There even are times when the presence of an interviewer may detract from the quality of the responses.

In some cases, the nature and scope of the problems the survey proposes to address may not be defined well enough to begin designing an effective questionnaire and systematically collect data from the target population. This is especially true when the Agency is dealing with an emerging problem, a new field of science or technology, or a population that has never been studied before. Using exploratory research techniques such as focus groups or in-depth interviews with a few of the potential respondents may identify key topics for subsequent investigation using more traditional statistical techniques.

The remainder of this chapter looks at:

- The methods most often used to collect survey data for EPA;
- The factors used in determining the most appropriate method for a particular Agency-sponsored survey; and
- How to assess the suitability of the proposed collection method(s).

A. Principal Data Collection Methods

This section examines the most frequently used methods of collecting survey data. First it looks at the main traditional methods used in statistical research and then at two exploratory research techniques that are applicable when the study objectives are not defined precisely enough to begin a systematic data gathering effort.

In fact, most surveys use a combination of data collection methods (known as “mixed-mode.”) For example, exploratory techniques may be used early on to clarify key topics. Or, if a mail survey is chosen as the primary collection method, telephone or face-to-face interviews may be used later to contact respondents who do not reply within a certain time limit. A combination of mail and telephone interviewing may be used, whereby respondents are mailed background information and a telephone interview is scheduled later.

Principal Data Collection Methods—

Traditional Methods:

- Face-to-face interviews
- Self-administered mail questionnaires
- Telephone interviews
- Random-digit dialing telephone surveys
- Computer-assisted interviews

Exploratory Research Methods:

- Individual in-depth interviews
- Focus group interviews

1. Traditional Survey Research Methods

The data collection instrument for all traditional collection methods is a "structured" questionnaire, which may be on paper or on a computer. The questions, their sequence, and their wording are fixed in a structured questionnaire. If interviewers are used, they may be allowed some leeway in asking the questions, but generally very little.

Face-to-Face Interviews

Face-to-face interviewing was the mainstay of survey research methodology for more than 50 years, and was used for many EPA surveys. Coupled with a well-designed, well-tested questionnaire, the face-to-face interview is a powerful, indispensable research tool. It is adaptable to a wide variety of research situations and is uniquely suited to in-depth explorations of issues. The problem is that it is very expensive and does not always produce better results when compared to other methods. In a face-to-face interview, the members of the sample are visited in their homes or workplaces by trained interviewers and asked to respond to a fixed set of questions. The interviewers record the respondents' answers on a printed questionnaire.

Self-Administered Surveys

In the most basic form of self-administered survey, researchers mail printed questionnaires to the respondents at their homes or businesses. The respondents complete the forms and return them by mail. Like face-to-face interviews, self-administered mail questionnaires have been used for decades to collect survey data. EPA relies heavily on this traditional survey research method to collect complex technical and scientific information from business and industry. A well-designed mail survey can achieve virtually the same degree of respondent cooperation as a personal interview survey, at a far lower cost. Careful design is especially crucial here—poorly designed mail surveys will likely yield biased answers and low response rates.

Self-administered surveys can also be hand-delivered; they do not necessarily have to be mailed. An in-house employee survey at EPA was one such example.

Internet Surveys

With the introduction of the Internet, a new way of collecting survey data has been created. Internet surveys are self-administered surveys in which researchers send a questionnaire via e-mail to the respondent. Similar to mail surveys, the respondent has the responsibility of completing the questions and returning the data. Internet surveys have been a recent introduction in survey research. The mechanism has been used for feedback after EPA conferences.

Telephone Interviews

Telephone interviewing is rapidly becoming the principal method of collecting survey data in research situations where probing or in-depth exploration of the issues is not required, and where an accurate list of phone numbers is available. This is most commonly true of establishment surveys, where accurate lists of establishments are maintained by an organization such as EPA.

There are two kinds of telephone interviewing techniques: (1) traditional and (2) computer-assisted telephone interviewing (CATI). (A third form of telephone interviewing—random digit dialing—is discussed in a separate section; this section covers situations where the telephone numbers are known to the research organization.)

Traditional telephone interviews are similar to face-to-face interviews. The interviewers pose questions to individual respondents at their homes or workplaces by telephone and record the answers directly onto a printed questionnaire. The interviewers generally work from one central location under the supervision of an experienced researcher. This is not commonly used any more.

However, computer-assisted telephone interviewing (CATI) is more commonly used. A printed questionnaire is not used, except maybe while the survey is being developed. Instead, the questions are programmed into a computer. The interviewer sits in front of a monitor and reads the questions to the respondents over the telephone as they appear on the screen. The interviewer types the respondent's answers and they are automatically entered into the computer. This radically different interview technique not only speeds up the collection and processing of respondent information, but also avoids the human errors normally associated with handling, checking, and transferring data from a printed questionnaire into machine-readable form.

CATI also has other advantages. It permits the use of very complex "skip" patterns. Depending on the response to one question, the computer can be programmed to determine the next question to present on the screen. It also provides the interviewer with instant feedback if an impossible or out-of-range answer is entered.

In addition to the traditional problems caused by the lack of complete and current lists of telephone numbers, the use of CATI has recently been complicated by a proliferation of call-blocking and screening devices, cell phones, phones used for faxes and Internet, and general resistance to telephone solicitations.

Computer-Assisted Surveys (CASI, CAPI)

Apart from "centralized" computer-assisted telephone interviewing (such as CATI, mentioned above), computer assisted survey information collection (CASI) methods are increasingly used in personal surveys, with the computer located in the field rather than centrally.

Computer-assisted personal interviewing (CAPI)—using portable computers to conduct face-to-face (household) interviews—was first tested in the late 1980s, and adopted by major survey organizations in the 1990s as hardware (laptop) limitations were resolved. These methods rely on trained interviewers to administer the questionnaires. By contrast, in computerized self-administered data collection, respondents read the survey questions and record the answers by themselves using electronic questionnaires.

The evolution of CASI methods for personal interviews has seen variations such as Computer Assisted Personal Interviewing (CAPI) and Computer Assisted Self-Interviewing (CASI) used in numerous national surveys. The privacy afforded by the latter methods, especially in its Audio-CASI (ACASI) form, has led to its recommendation in several surveys involving sensitive issues. More recently, interviewer-administered methods (CAPI) have been found to yield significantly better reporting than self-administered methods (CASI, ACASI) for sensitive issues.

Several forms of computerized self-administered questionnaire (CSAQ) technology were introduced initially for industrial and business surveys including disk-by-mail (DBM) surveys and electronic mail surveys (EMS). With these methods, the respondent answers the questions in his or her computer (or terminal), and returns the completed answers either by mailing back the disk (DBM) or by modem (EMS).

The advantages of CASI methods generally include:

- Automated, "optimal" scheduling of calls
- Skip patterns that accommodate complex question structures
- On-line error check and resolution (editing)
- Automatic and prompt data entry

Random-Digit-Dialing Telephone Surveys

Various forms of random digit dialing (RDD) sampling have been designed to overcome the limitations of telephone lists, mentioned in the section above. The following sequence reflects the order in which these methods were introduced for residential telephone samples:

1. Sampling from lists and directories (covered above)
2. Unclustered random digit dialing (RDD) methods

3. Clustered RDD methods, including Mitofsky-Waksberg

4. List-assisted, density-stratified sampling

Each of these methods has continued to be used as its successor(s) have become popular; in addition, some surveys use combinations of these methods.

1. Sampling from lists and directories. As explained above, sampling from published directories has some major weaknesses, including the failure to cover unlisted numbers, and the difficulties in rapidly and continuously updating the sampling frame. These problems have become increasingly serious as unlisted numbers are more and more prevalent, and the population of residential telephones has become more dynamic. In some cities about one-half of all phone numbers are unlisted. Currently, only low-budget or special population surveys use this method. RDD samples for telephone surveys were introduced to combat these weaknesses.

2. Pure RDD samples. In RDD sample designs, telephone numbers are selected totally at random from a specified frame of numbers. The frame is typically restricted to those prefixes known to contain working residential numbers ("WRNs") within the geographic area of interest. These 6-digit prefixes[1] in current use are available from commercial sources such as Bell Communications Research (Bellcore.) However, a key inefficiency of strict RDD methods is that a large proportion of the sample numbers called are ineligible or not in service. Pure RDD samples are almost never used now.

3. Mitofsky-Waksberg clustered RDDs. This form of sampling improves on strict RDDs by capitalizing on the fact that working numbers (WRNs) tend to be clustered within the same "100-blocks" (8-digit blocks of numbers.)[2] If one working number is found, there tend to be others that differ only in the last 2 digits. These 8-digit blocks, each with 100 possible telephone numbers, are the clusters in the Mitofsky-Waksberg-type of design.

In Waksberg-type methods, telephone numbers in different clusters (100-blocks) are selected and called to determine their eligibility. Once a cluster with an eligible number is identified, additional numbers are selected and called within that cluster.

Because numerous calls (follow-ups) may be necessary to determine eligibility, the method presents some problems in scheduling and in achieving the desired cluster sample sizes. An additional weakness of clustered samples is that clustering effects tend to increase the sampling variability of the survey estimates. Several variations of the method have been developed to mitigate these problems.

4. List-Assisted RDDs. Another way to enhance the efficiency of telephone sampling designs is to stratify by the number of working residential numbers in each 100-block. The assumption here is that if there is at least 1 or 2 directory-listed residential numbers in a 100-

[1] The area code and first 3 digits of the phone number itself; for example, (301) 654-xxxx.
[2] For example, (301) 654-88xx.

block, other numbers in the block are likely to be WRNs as well. Blocks in high-density strata (e.g., those with 2 or more listed numbers) are sampled at higher rates than blocks in low-density strata. Lists of 100-blocks by number of directory-listed phone numbers are available from several national suppliers.

2. Exploratory Research Methods

The Agency occasionally explores emerging problems about which little is known. It may be determined that only a statistical survey will allow us to explore the central issues of the emerging problems; however, some aspects of the issues may not be defined well enough for us to begin constructing a structured survey questionnaire. In such cases, "unstructured" survey research methods may prove effective in clarifying key issues.

Focus Group Interviews

Focus group interviews are perhaps the most common "unstructured" research technique. The participants are members of the target population who are called together for informal discussions focused on specific issues or specific parts of the proposed survey questionnaire. Focus groups often will unearth aspects of emerging problems that might not surface in individual, in-depth discussions. Focus groups are especially appropriate for exploring the attitudes, opinions, concerns, and experiences of selected segments of a population of interest; identifying key concepts; helping to phrase questions so they will be clear to all potential respondents; and evaluating drafts of survey questionnaires. Focus groups also may be used early in the development stage of a research project to help the Agency determine whether a quantitative survey is feasible.

Probability-sampling techniques generally are not used to select the study participants. Instead, several relatively homogeneous groups of six to twelve people are selected from various subgroups of the target population. From two to as many as twelve groups may be formed, each led by a skilled moderator knowledgeable about the study objectives. The moderator interacts with the participants and "focuses" the discussion on a few topics of special interest to the researchers.

A topic outline is prepared at the beginning of the study. Usually, fairly general topics are identified for the first group to discuss, with researchers gradually focusing the discussions on more specific subject matters in subsequent group sessions. The groups usually meet for about two hours. Although the topic outline is used as a general discussion guide, the participants are given ample opportunity for spontaneous comment, provided they do not stray too far from the material in the outline.

Individual In-depth Interviews

Another valuable tool is the unstructured survey that involves individual, in-depth discussions with individuals who are knowledgeable about, or involved in, the issues the Agency proposes to study. A topic outline, rather than a fixed set of questions characteristic of a structured questionnaire, guides the interviews.

With in-depth, individual interviewing, probability selection methods generally are not used to choose who will be interviewed. Instead, the Agency selects a "convenience" sample, representative of different segments of the target population. Any number of individuals may be chosen to participate in the study. The interviewers should have experience in conducting in-depth interviews, and most importantly, knowledge of the subject matter.

In-depth individual interviews are particularly valuable when researchers are unsure about:

1. Which topics are most relevant to the research objectives
2. Whether members of the target population are likely to have the kinds of information the Agency needs
3. How to phrase certain items on the survey questionnaire
4. What type of question format is likely to be most effective for obtaining specific information on certain topics (e.g., open or closed questions)
5. Which topics the members of the target population are likely to consider threatening or particularly sensitive
6. Which subgroups in the target population are most likely to be able to supply specific data the Agency needs?

B. Comparing the Collection Methods

This chapter states that no collection method is intrinsically better than any other. However, certain methods are clearly more appropriate in certain research situations and just as clearly contraindicated in others. This section highlights some of the principal distinguishing features of each of the traditional collection methods.

1. Special Characteristics of Face-to-face Surveys

Face-to-face interviewing used to be the most frequent method used at EPA for collecting survey data from the general public. Moreover, it used to be considered as the only viable approach for collecting highly complex, sensitive, technical information from business and industry. Although no longer the predominant data collection method, it is a standard against which other methods are judged.

Face-to-face interviews have many advantages:

- They generally achieve a higher response rate, greater cooperation, and more complete and consistent data, especially when in-depth exploration of the issues is desirable.
- They are uniquely suited to probing—a technique used to study the respondent's knowledge of key issues, attitudes or, more typically, to clarify and learn the reasons for

their answers.

- They are the only viable data collection method when first-hand observations of the respondents or the interview site are necessary. Both telephone interviews and mail surveys are inappropriate when eyewitness reports are desirable.

- They permit the use of visual aids, which may make respondents more cooperative and willing to give less biased replies.

  For example, interviewers can show respondents a calendar to refresh their memories about specific events or time intervals. Or, instead of reading a long list of possible replies, interviewers can hand respondents a checklist (or "prompt card") of suggested answers to elicit an appropriate reply. When an interviewer verbally gives respondents a choice of three or four possible answers, they often have difficulty remembering all of them. The net result is a bias towards the first or last item mentioned. In addition, if interviewers are required to question respondents about their income or other topics that many people consider too sensitive to discuss with a stranger, prompt cards listing the reply categories tend to cut down on inaccuracies and outright refusals to answer the question.

  Similarly, in a survey of the general public where respondents are required to evaluate a product or other object (a new pollution-control device, for example), face-to-face interviews may be the only viable data collection option. However, if interviewers are given products for business or industrial respondents to evaluate, it may be feasible to mail the firms a sample of the item (or different versions of the product) in advance, and schedule a follow-up telephone or mail interview to get their reports or opinions.

Face-to-face interviewing has many disadvantages, however:

- Geographic dispersion - Setting up a complex field operation in a large number of sampling areas to interview only a few respondents in each area obviously is prohibitively expensive. To hold down costs, researchers "cluster" respondents in a few selected geographic areas and set up mobile field units to collect the data. Field supervisors remain at a more central location. Clustering does increase the sampling error of the survey, however. Even with clustering, face-to-face surveys have higher costs and personnel requirements, given the need for extensive training of field staff and close supervision of widely dispersed interviewers throughout the data collection period.[3]

- Cost - Face-to-face surveys cost much more than either telephone or mail surveys of similar complexity. Cost differences alone can tip the balance against this survey.

[3] However, widely dispersed samples have little effect on both telephone and mail surveys because they are generally operated from a centrally located office.

- Paperwork - Moreover, the paperwork is much more involved in this type of survey. In addition to the questionnaire, it may be necessary to use as many as 20 different forms and documents to coordinate and control the fieldwork and processing operations: confidentiality agreements, prompt cards, interviewer calling cards, press releases, interviewer progress reports, interviewer evaluation forms, respondent verification and evaluation forms, and letters giving respondents advance notice of the survey.

2. Special Characteristics of Self-Administered Surveys

Like face-to-face interviews, self-administered mail questionnaires have been used effectively for decades to collect survey data. Mail questionnaires are particularly appropriate for obtaining detailed technical and scientific data, and they are the least costly of the collection methods for medium-to-large amounts of data. Specific advantages include:

- They are indispensable for collecting certain kinds of detailed technical data, especially if respondents need to consult their records or other people for the necessary data. Self-administered questionnaires allow respondents great flexibility in preparing replies. Respondents have time to think about the questions, gather information from their files, and get advice from others at their own convenience. Particularly for household surveys, the ability of the respondent to see all answer choices before checking his or her answer will improve survey results. In personal interview situations, respondents tend to remember only the first or last answer choice read by the interviewer.

- Mail questionnaires are the least costly of the traditional collection methods for gathering medium-to-large amounts of data, largely because there is no cost for interviewers, or expenses are limited to telephone call-backs to assure an acceptable response rate.

- Broad geographic coverage is possible with comparatively little effect on the overall cost of the survey.

- The sampling variability may be low because there is usually no need to cluster the sample in small geographic areas (clustering causes increased sampling variability).

- “Interviewer bias” is minimized in self-administered surveys—respondents generally are most honest in self-administered surveys. In the presence of an interviewer, respondents tend to give more socially-acceptable, less critical replies. For example, if respondents are asked if they like living in their community, they tend to say they do, even though on the whole they may dislike it greatly. The same question on a mail questionnaire will elicit more truthful responses. Likewise, many respondents feel uncomfortable giving responses that the interviewer might find insulting.

Self-administered surveys have some limitations. For example:

- The questionnaires should be very carefully designed to compensate for the lack of

social interaction that other collection methods provide. Researchers must depend entirely on the questions and written instructions to elicit satisfactory responses and motivate the respondents to cooperate.

- Questions that are suitable for self administered questionnaires are relatively limited, especially for household surveys. Open questions should be used sparingly—more than a few requests for lengthy answers may result not only in refusal to answer particular questions but also may cause respondents to abandon the questionnaire altogether. Generally, if respondents are required to read any but the simplest language, or to write out answers in their own words rather than circle or check a printed response, the results tend to be very poor. Of course, these concerns are less likely to be a problem if the respondents are representatives of businesses or industries.

- Language barriers can be a problem; in some cases the questionnaires may need to be available in Spanish or other language common to the area, which increases costs. (Of course, this is a problem with personal or telephone interview situations as well.)

- In addition to language barriers, cultural norms may make certain populations averse to filling out any official-looking paperwork. Face-to-face interviewers may be able to convince sample members that there is no connection between the survey and any official or law-enforcement function.

- Self-administered surveys may be inappropriate if the researchers want respondents to complete the questionnaire with no involvement from others. When questionnaires are self-administered, it is impossible to know the circumstances under which they were completed.

- A substantial follow-up effort is almost always necessary to achieve a reasonable response rate in any voluntary mail survey. To increase the response rate, researchers sometimes give respondents the option of telephoning their replies rather than mailing back the completed questionnaire. Any mail survey must explicitly plan for reminder cards and letters, and telephone or personal interview follow-ups, in especially difficult cases.

3. Special Characteristics of Internet Surveys

Some researchers have lauded the Internet as the primary arena where the most significant social research will be conducted (Bainbridge, 1999; American Association of Public Opinion Research, 1998). There are several advantages to Internet surveys:

- Using the Internet to distribute a questionnaire can be a quick, inexpensive, time-efficient mechanism if the research population can be easily accessible—such as a listing of EPA employees who recently attended a conference.

- Real time results can be major advantage of utilizing such an approach. At any time

during the response period, descriptive statistics can be elicited.

- Other positives of using Internet surveys include "facilitative interaction between survey authors and respondents collapsed geographic boundaries, user-convenience, and arguably, more candid and extensive quality" (Smith, 1997, p. 2).

There are also certain drawbacks in utilizing such an approach:

- One being the sample derived from an e-mail listing cannot be representative of the population being studied. Calibrating the results may be necessary for gender, ethnicity, and class. The Internet tends to under represent women, minorities and the poor. These deficiencies can be compensated. The General Social Survey, for example, does not utilize a random selection of people, but utilizes a quota sample (Bainbridge, 1999).

- Another deficiency that may be encountered in using an Internet survey includes providing monetary incentives (Smith, 1997).

In conducting an Internet survey, one may create one by oneself, but for a more professional appearance and greater acceptance, one may consult an agency such as Zoomerang, which specializes in e-mail surveys. In the age of SPAM and computer hacking, one should also be conscious of the hesitancy of the sample to respond to a hoax. Be sure to state whom the survey is for and why it is being distributed. Legitimacy is a necessity.

One may also want to be cognizant of the importance of completing the survey. Many computer users are bombarded with e-mails. It has become a major daily task to shift through which messages are worthy of one's precious time. The respondent needs to know the importance of the study and the approximate time to complete the questionnaire. The researcher may give the respondent the option of obtaining preliminary results. Such a measure elicits a tone of importance and rewards the respondent with a completed outcome.

4. Special Characteristics of Telephone Surveys

Some of the advantages of telephone interviews are:

- Telephone surveys cost about one-half as much as face-to-face surveys of comparable size (Frey, 1989). They are also easier to manage, produce faster results, and with few modifications, can be used in most research situations where face-to-face interviewing is indicated.

- Cost savings result from the fact that about one-quarter as many interviewers are needed to reach the same size sample, and the cost of training the interviewers is about one-fifth as much. Moreover, travel costs for interviewers and field staff are virtually nonexistent.

- Telephone surveys are easier to administer. Monitoring, administration, and quality control are simpler than in face-to-face surveys because no field operation is necessary. Moreover, it is easier to correct interviewer mistakes quickly. People on the contractor's

staff who review and edit the completed questionnaires are typically close to the interviewer, and can quickly provide feedback to the interviewers about errors and omissions.

- Re-contacts are easier. Respondents can easily be re-contacted after the initial interview to correct inaccuracies, inconsistencies, and omissions.

- Telephone surveys are faster. Results can be obtained more quickly from telephone surveys than from face-to-face or self-administered surveys. Interviewing, monitoring, training, editing, and coding operations are usually centralized in one location. If any changes in the questionnaire or interviewing procedures have to be made because of problems encountered in the pretest, the researchers can quickly incorporate them into the main survey. Even after the interviewing in the main survey is under way, it is easy to notify the interviewers immediately about any needed changes. Follow-up interviews to check the interviewers also are much easier. If computer-assisted telephone interviewing is used, time-consuming manual screening, editing, coding, and data entry operations required for the other data collection methods (including traditional telephone interviewing) are unnecessary.

- Access is easier. Telephone interviews permit access to respondents located in areas where face-to-face interviews are especially difficult, such as locked apartment or office buildings, sub-divisions with security guards preventing access, or dangerous neighborhoods.

- With a few modifications, telephone interviews can be used in almost all research situations where face-to-face surveys are suitable, provided correct addresses and phone numbers are available.

  For example, if pictures or products are shown to the respondents to motivate or enable them to answer certain questions, these can be mailed to the respondents and an interview scheduled at a later date. The combined mail/telephone technique is widely used in marketing surveys.

  The "prompt cards" that face-to-face interviewers use to motivate respondents are not applicable for phone surveys. However, the questionnaire can be modified to obtain the same information. The most common procedure is to break questions with multiple-choice replies into a series of simpler questions and offer the respondents a set of Yes/No alternatives until all possible answers are covered.

Telephone interviewing also has several disadvantages.

- Response rates for telephone surveys are five to fifteen percent lower than comparable face-to-face surveys, despite considerable improvements in interviewer training, feedback procedures, and monitoring techniques during the past few years (de Leeuw and van der Zouwen, 1988). The reason is that respondents generally find telephone interviews more tedious and less rewarding than face-to-face interviews, and hence tend

to be less cooperative over the phone.

The prevalence of commercial telephone solicitations has prompted increased public resistance to legitimate phone surveys, and answering machines and Caller-ID make call screening easy. Compounding these problems is the fact that many phone numbers are unlisted at the request of the customer (not just that they have not yet been listed.) In some areas as many as 50-60% of numbers may be unlisted (combining unlisted and not-yet-listed.) [4] While repeated calls, letters, and publicity can alleviate part of this problem, response rates for phone surveys remain relatively low.

In all, response rates for telephone surveys can be below 70% even in extremely well designed surveys. This problem is especially acute in large urban areas, particularly in the West.

- Telephone interviews are not the best way to collect factual data if respondents have to search their records or consult with others. However, it may be possible to mail respondent's background information in advance and schedule a follow-up phone interview to obtain the needed data.

- Telephone surveys normally should be relatively short to avoid excessively burdening respondents who may have other things to do. While arranging for callbacks at more convenient times may help, less information is generally gathered by phone surveys than by mail surveys or face-to-face interviews.

- Interviewers may have to work at odd hours to obtain interviews from people who work during the day. Because the interviewing "window" is relatively short, phone surveys are most efficient when they cover several time zones, thereby lengthening the peak interviewing times.

- Of course, interviewers cannot reach people who have no phones. This means that important subgroups such as low-income people will be underrepresented in surveys of the general public if telephone interviews are the exclusive collection method.[5]

Random-digit-dialing (RDD) sampling is a way to overcome the problem of unlisted phone numbers. This is described in the next section.

5. Special Characteristics of Random-Digit-Dialing Surveys

The various types of RDD surveys have several advantages—

[4] Many residents may not have their phone numbers listed in a current directory, either because they moved in after the directory was last published, or because they do not want their phone numbers listed. In most areas up to 30 percent of households do not have their phone numbers listed; in large metropolitan areas this proportion approaches 50 percent. [Piekarski, 1989]

[5] The overall telephone coverage rate in the U.S. is around 95%, which appears to be nearly universal. However, among certain subgroups the rate is substantially lower—as low as 70% (among rural households in the South.) Telephone non-coverage should be considered carefully before deciding to conduct a telephone survey.

- They don't rely on lists of specific telephone numbers, and hence can be used to identify eligible households where no list exists. The problem of incomplete or nonexistent lists from which to draw a sample is more common than many people think; RDDs solve this problem by not requiring them.

- Because little or no time is spent on sampling, RDDs often can be conducted quickly.

Disadvantages include—

- As is true for telephone surveys, the number of questions that can realistically be asked is limited.

- Because no address is associated with the phone number, there is no opportunity for advance notification and mailing of background materials, or for mail questionnaire follow-ups. [6]

- Noncontact rates tends to be high in metropolitan areas (Steeh, Kirgis, Cannon, and DeWitt, 2001).

- The increased prevalence of cellular phones renders many geographic concepts moot—a phone number might not belong in its presumed geographic location. A similar problem is caused by area codes that cross geographic boundaries; if only a specific area is to be surveyed, geographic screening questions are necessary.

6. Special Characteristics of CASI Surveys (Computer Assisted Survey Information)

CASI methods have many advantages:

- They can accommodate complex skip patterns reliably

- They allow immediate error checks and resolution (editing)

- Data are entered automatically and promptly

- They can be relatively long.

At the same time there are several disadvantages:

- Setup times are likely to be longer than paper-and-pencil or telephone surveys

- Costs are higher, except for the largest surveys where the setup, programming, and testing costs are spread out over many respondents.

- Not all contractors have the required expertise.

[6] Using "reverse directories" is not normally a good solution, as most of these are incomplete.

- Not all respondents are comfortable using computers, potentially biasing results.
- Likewise, not all interviewers (for CAPI surveys) are comfortable using computers.
- Security can be a problem, because laptop computers are easily stolen or lost.

C. Factors Affecting the Choice of Collection Methods

A host of interrelated design factors, as well as the time and funds available, affect the contractor's choice of the primary data collection method for a particular survey.

The remainder of this section briefly examines the selection factors that normally determine the choice of the primary data collection method for a statistical survey. They are:

Major Selection Factors:

1. Characteristics of target population
2. Data requirement
3. Obligation to reply
4. Definition of response rate
5. Target response rate
6. Improving response rates
7. Available time
8. Available funds

1. Characteristics of the Target Population

The characteristics of the target population often are an important consideration in selecting the primary data collection method. For example, mail surveys of the general public have lower response rates than any of the direct interviewing techniques. However, with careful design and execution mail surveys can approach the response rates of other techniques. Conversely, most surveys of business populations use mail questionnaires as the primary collection method and follow-up incomplete or incorrect responses with telephone interviews.

Face-to-face interviews are generally the preferred approach for elderly respondents and those with limited education. Low-income respondents, and those with limited command of English, also do best in face-to-face interviews.

The location and distribution of the target population are also factors. Face-to-face interviews are more cost-effective when the target population is concentrated in a small geographic area, such as a particular city or county. However, if the target population is widely dispersed travel and administrative costs may make a face-to-face survey prohibitively expensive and time-consuming. Under these conditions self-administered questionnaires or telephone interviews are more realistic options. Mail and telephone surveys are the least affected by a widely dispersed sample.

2. Data Requirements

The general nature, extent, and complexity of the data requirements are important determinants in choosing the primary collection method. It used to be thought that mail questionnaires should be kept very short; research has shown that this is not necessarily true if questionnaires are carefully designed.

The data requirements of many establishment surveys require respondents to consult their records, or other people, in order to prepare adequate replies**.** A self-administered mail questionnaire may be the only feasible way of getting the necessary data in such cases, possibly supplemented by telephone reminders or actual interviews.

It may be preferable in establishment surveys to use face-to-face interviews, if it is necessary to ask many questions that respondents may consider threatening or unusually sensitive. To minimize the impact of what may be perceived as potential threats to their operations, establishment respondents, for example, may furnish inaccurate or incomplete replies. If it is necessary to collect highly sensitive technical data, the contractor may recommend using trained investigators to make first-hand observations of records or physical facilities to ensure that the Agency obtains complete and valid data. In addition, the respondent may not furnish sensitive information unless compelled by law and if given assurances that the provided information will not be shared by others, except in statistical terms.

3. Respondent's Obligation to Reply

The respondent's obligation to provide information to the Agency often has a critical impact on the choice of the primary collection method. In some cases, the Agency can make responses from businesses and other organizations mandatory, where the respondents must provide the required data or face civil or criminal sanctions. Whenever a mandatory response is required, a relatively high response rate is ensured, no matter what collection method is used, and even self-administered mail questionnaires become a viable option. On the other hand, a well-designed, self-administered, mail questionnaire can yield a good response rate even in a voluntary survey, as long as extensive follow-up of nonrespondents is provided.

4. Definition of Response Rate

While seemingly trivial—"the response rate is the proportion of the sample that responds to the survey"—actually measuring it can be difficult and subject to definitions that make a response rate appear to be higher than it actually is. Therefore, this section discusses various definitions of response rate and recommends one that should be used in all EPA solicitations so that contractors all use the same measure in their bids and actual surveys.

There have been several attempts at methodically defining response rates and disposition categories. One of the best of those is the 1982 *Special Report on the Definition of Response Rates*, issued by the Council of American Survey Research Organizations (CASRO). As defined by CASRO, the response rate is the number of complete interviews with reporting units divided

by the number of eligible reporting units in the sample. Several response rates are described below, based on the following components:

RR: Response rate; followed by number suffix (1-6)
I: Complete interview
P: Partial interview
R: Refusal and break-off in mid-interview
NC: Non-contact
O: Other
UH: Unknown if household or occupied housing unit
UO: Unknown, other
e: Estimated proportion of cases of unknown eligibility that are eligible

RR1 = I/[(I + P) + (R + NC + O) + (UH + UO)]

Response Rate 1 (RR1) is the number of complete interviews divided by the number of interviews (complete plus partial) plus the number of non-interviews (refusal and break-offs plus non-contacts plus others) plus all cases of unknown eligibility (unknown whether eligible plus unknown for other reasons.) Of all the choices here, RR1 is the lowest, most conservative one.

RR2 = (I + P) /[(I + P) + (R + NC + O) + (UH + UO)]

Response Rate 2 (RR2) counts partial interviews as respondents. This is higher than RR1. The way "partial interview" is defined is critical here—obviously, answering only 1 or 2 questions out of 50 should not count as a partial interview, but what about answering 40 out of 50?

RR3 = I /[(I + P) + (R + NC + O) + e (UH + UO)]

Response Rate 3 (RR3) estimates what proportion of cases of unknown eligibility are actually eligible. In estimating e, one is guided by the best available scientific information on what share eligible cases make up among the unknown cases and must not select a proportion in order to boost the response rate. The basis for the estimate should be explicitly stated and explained

RR4 = (I + P) / [(I + P) + (R + NC + O) + e (UH + UO)]

Response Rate 4 (RR4) allocates cases of unknown eligibility as in RR3, but also includes partial interviews as respondents, as in RR2. The same cautions about the factor "e" apply.

RR5 = I / [(I + P) + (R + NC + O)]
RR6 = (I + P) / [(I + P) + (R + NC + O)]

Response Rate 5 (RR5) is a special case of RR3 in that it assumes that e=0 (i.e. that there are no eligible cases among the cases of unknown eligibility, or the rare case in which there are no cases of unknown eligibility.) Response Rate 6 (RR6) makes that same assumption and also includes partial interviews as respondents. RR5 and RR6 are only appropriate when it is valid to assume that none of the unknown cases are eligible ones, or when there are no unknown cases. RR6 represents the maximum response rate.

Response rates for random digit dialing telephone surveys are even more complicated because the possible outcomes are even more numerous. For example, how does one handle problems such as answering machines, call-blocking, and cell phones that are not tied to specific geographic areas? These issues are beyond the scope of this guidebook.

5. Target Response Rate

The collection method likely to produce the highest response rate given the available funds is preferable. Face-to-face surveys tend to have the highest response rate, other factors being equal, but they are the most expensive. Telephone surveys can produce response rates nearly as high, if they are skillfully designed and carried out. Recent research has shown that mail surveys, formerly considered to yield poor response rates, can achieve the 75 percent minimum response rate that is recommended for all Agency-sponsored surveys.

For bidding purposes, it is recommended that RR1, the most conservative response rate, be specified, and that it be required to achieve a 75 percent level. Note that this target response rate should be measured after all follow-ups have been completed.

6. Ways to Improve Response Rate—Follow-ups and Incentives

Research has shown that response rates improve substantially for each follow-up, although less for the last follow-up than for the first. However, in many instances the difficult cases differ substantially from those obtained at the first try. So even if a "final" follow-up only gets a few additional cases, it may change the survey results substantially. Additional contacts with the sample population improve results, where "contact" includes an advance letter, a questionnaire mailing, a reminder card or phone call, or another copy of the questionnaire.[7]

Another way to improve response rates is to increase the respondent's interest in the topic or it's perceived importance. A well-crafted advance letter is very important here, as are multiple follow-ups—each helps to establish the importance that the sponsor places in getting responses from the sample member.

The use of cash incentives is controversial, and in fact is explicitly discouraged by OMB for most government-sponsored surveys. Not only can cash incentives increase the total cost of conducting a survey (and not improve response rates appreciably), but also their use can bias the results. In well-designed surveys with multiple follow-ups, cash incentives are rarely needed.

Non-cash incentives might be considered. These might include (a) gifts or gift certificates; (b) a promise to provide the results to respondents; or (c) in some environmental surveys, mitigation measures. EPA-related examples include:

- In radon and other indoor air studies, offers to provide mitigation if elevated levels are found.

[7] OMB 1999, Section FASQ #1.

- An incentive package for soliciting in-use vehicles for laboratory testing. This includes a loaner vehicle, gasoline, a free tune-up, and a cash payment. In this case the survey results are completely determined by the characteristics and condition of the tested vehicle; they cannot be biased by the respondent's attitude toward the incentive.

Examples where cash incentives might be considered include studies involving multiple questionnaires or bio-monitoring (e.g., urine or blood specimen collection,) where relatively high levels of monetary incentives may be necessary to produce acceptable levels of respondent cooperation. For instance, both the Agricultural Health Study (AHS) panels and the Children Total Exposure to Persistent Pollutants (CTEPP) have used incentives in the $100 to $150 range.

7. Available Time

The length of time the Agency can wait to get results also may be a deciding factor in the selection of the data collection method. Computer-assisted telephone interviews and RDDs have by far the fastest turn-around time. Conventional telephone surveys also can be done more quickly than face-to-face surveys. Mail surveys are generally not appropriate if time is critical.

8. Available Funds

The amount of money available for the survey is almost always a critical factor in choosing the primary data collection method. As indicated earlier, individual face-to-face interviews are the most expensive way of collecting survey data, other factors being equal. Personnel costs (for interviewers, supervisors, trainers, and quality control staff at different field locations) are approximately double that of a comparable telephone survey, where the interviews are usually conducted at one central location. Mail surveys usually are the least costly option, largely because the cost of interviewers is limited to some follow-up calls to increase the response rate or to correct inconsistencies and missing or inaccurate replies

Nevertheless, the least expensive option should not be selected unless it will produce results of acceptable quality. Sometimes it is better to use a higher-cost method and reduce the size of the sample. For example, a mail survey using face-to-face or telephone interviewers to follow-up incomplete or unanswered questionnaires usually produces higher quality results than a "pure" mail survey, even if a smaller sample is used to hold down costs.

Summing Up

It is recommended that you leave the selection of the collection method(s) up to the contractor. However, as the representative of the sponsoring office, you will have to approve the contractor's choice. The previous discussion of the special features of the traditional data collection methods and the influence of various survey design factors will help you assess the appropriateness of the proposed method. To further guide your assessment, Exhibit 1 on the next page indicates the methods most likely to produce satisfactory results under a variety of circumstances.

Although one or a combination of the traditional collection methods will ultimately be selected for testing purposes and for the main survey, using one of the exploratory research techniques discussed in section A could considerably improve the survey design. At a relatively low cost, either individual in-depth interviews or focus group interviews can clarify problems that may be difficult and costly to correct once the survey is under way.

Exhibit 1: GUIDE FOR CHOOSING A DATA COLLECTION METHOD

	AGENCY REQUIREMENTS	LIKELY TO BE THE BEST CHOICE
GENERAL	• Fast turn-around • Lowest possible per unit cost • Highest possible response rate • Fewest possible errors and biases	• Telephone* • Mail • Face-to-Face • Face-to-Face or Telephone
SPECIAL DATA	• Complex technical data (in a mandatory survey) • Detailed data (in a voluntary survey) • Respondent's opinion of a product or device • Highly sensitive information	• Face-to-Face or Mail • Face-to-Face • Face-to-Face** • Face-to-Face or Mail
COVERAGE	• Coverage of all sub-groups in population • Coverage of widely dispersed sample • Coverage of high-crime or remote areas	• Face-to-Face or Mail • Mail • Telephone or Mail
SPECIAL AIDS OR TECHNIQUES	• Extensive probing • Third-party observation of records or facilities • Respondent diaries • Respondent consultation with others or record searches • Visual aids (calendars, scales, etc.)	• Face-to-face • Face-to-face • Face-to-face or Mail • Mail • Face-to-Face or Mail**

* CATI is especially effective.

**Telephone may be satisfactory if visual aids are mailed to the respondents in advance.

Bibliography: Chapter 2

Traditional, Structured Techniques:

Beimer, Paul P. et.al. Measurement Errors in Surveys, New York, John Wiley & Sons, 1991.

Cochran, W.G. and G. M. Cox. Experimental Designs, New York, John Wiley & Sons, 1957.

Couper, Mick P., Reginald Baker, Jelke Bethlehem, Cynthia Clark, Jean Martin, William Nicholas, and James O'Reilly (editors). Computer Assisted Survey Information Collection, New York, John Wiley & Sons, 1998.

Dillman, Don A. Mail and Telephone Surveys: The Total Design Method, New York, John Wiley & Sons, New York, 1978. *Especially Chapter 2.*

Dillman, Don A. Mail and Internet Surveys, the Tailored Design Method, New York, John Wiley & Sons, 2000.

Fowler, Floyd J., Jr. Survey Research Methods (3rd ed). Thousand Oaks: Sage Publications, 2002.

Groves, Robert M. and Mick P. Couper. Nonresponse in Household Interview Surveys, New York, John Wiley & Sons, 1998.

Groves, Robert M., Paul Biemer, Lars Lyberg, James Massey, William Nicholls, and Joseph Waksberg (editors). Telephone Survey Methodology, New York, Wiley, 1988.

Kalton, Graham. Introduction to Survey Sampling. Newbury Park: Sage Publications, 1983.

Lyberg, Lars, Paul Biemer, Martin Collins, Edith De Leeuw, Cathryn Dippo, Norbert Schwarz, and Dennis Trewin (editors). Survey Measurement and Process Quality, New York, John Wiley & Sons, 1997.

Office of Management and Budget. Implementing Guidance for OMB Review of Agency Information Collection. Office of Information and Regulatory Affairs.

Salant, Priscilla and Don A. Dillman. How to Conduct Your Own Survey. New York, John Wiley, 1994. *Chapter 4, "Choosing a Survey Method."*

Sudman, Seymour, Norman M. Bradburn and Norbert Schwarz. Thinking about Answers, the Application of Cognitive Processes to Survey Methodology, San Francisco, Jossey-Bass, 1996.

Exploratory Techniques:

Goldman, Alfred E. and Susan Schwartz McDonald, The Group Depth Interview: Principles and Practice. Englewood Cliffs, Prentice-Hall, 1987. *Mainly focused on market research, but contains much useful information about how to run focus groups.*

Hoinville, G. and R. Jowell, and associates, Survey Research Practices, London, Heinmann Educational Books, 1978. *Chapter 2, "Unstructured Design Techniques. "*

Works Cited:

American Association for Public Opinion Research. "Proceedings of the Fifty-third Annual Conference of the American Association for Public Opinion Research," Public Opinion Quarterly, 62, p. 434-440, 1998.

Bainbridge, William S. "Cyberspace: Sociology's Natural Domain," Contemporary Sociology-

A Journal of Reviews, 28 (6), p. 664-667, 1999.

de Leeuw, Edith D. and Johannes van der Zouwen. "Data Quality in Telephone and Face-to-Face Surveys: A Comparative Meta-Analysis." In Telephone Survey Methodology, ed. Robert Groves, Paul Biemer, L. Lyberg, J. Massey, W. Nicholis, and J. Waksberg, pp. 283-99. New York, Wiley, 1988.

Smith, Christine B. "Casting the Net: Surveying an Internet Population," http://www.ascusc.org/jcmc/vol3/issue1/smith.html, 1997.

Steeh, C.; N. Kirgis; B. Cannon and J. DeWitt. "Are They Really as Bad as They Seem? Nonresponse Rates at the End of the Twentieth Century." Journal of Official Statistics, 17 (2), pp. 227-247, 2001.

Chapter 3: Developing the Questionnaire

A well-designed, thoroughly tested questionnaire is the most basic tool in survey research. Developing a valid questionnaire for an Agency-sponsored survey requires close collaboration by the sponsors and the contractor throughout the design and testing process. This is true regardless of how the questions are to be asked—in person with paper and pencil, by phone, or using one of the computer-assisted methods described in Chapter 2.

This chapter discusses:

A. The principal steps in developing a good survey questionnaire, and the roles of the project officer and contractor in designing and testing it

B. Reviewing questionnaire drafts

A. Developing the Questionnaire; Roles of Project Officer and Contractor

This section discusses the steps normally involved in developing a structured questionnaire for a statistical survey. The process involves 16 steps, the majority of which are performed by the contractor. Agency-sponsored surveys that are largely repetitions of earlier studies may shortcut many of the steps, but for surveys that address new environmental concerns, a thorough questionnaire-development effort is strongly recommended.

Preparing a survey questionnaire appears to be an easy task, but in fact it is extremely difficult – even for an experienced questionnaire designer. In no case should you, or the contractor, begin to draft the questionnaire until the Agency's data requirements have been clearly framed. The reason is that each question should have an obvious link with the data requirements. The requirements then are transformed into operational concepts and expressed in a logical series of questions.

Usually several drafts of the questionnaire—one or more pretests drafts and a pilot test replicating the actual conditions of the main survey— must be prepared and reviewed before a final version is ready to be printed for the main survey. If several versions of the questionnaire have to be designed to accommodate the needs of different types of respondents, more drafts may be necessary.

A summary of the questionnaire-development process is given in Exhibit 2 on the next page. The check marks (✓) indicate the six steps in which the EPA sponsor plays the primary role. This role is generally limited to (a) specifying the research topics, (b) reviewing drafts, and (c) monitoring the overall design and testing process.

Let's look now at the individual steps in the development of the questionnaire.

Exhibit 2: Sponsoring Office's Tasks in Questionnaire-Development Process

Agency Responsibility	Activity	Step
✓	Prepare analysis plan	1
✓	Draft list of topics or suggested questions	2
	Conduct exploratory group or individual interviews	3
	Prepare first draft of questionnaire	4
✓	Review and approve draft of questionnaire	5
	Prepare plan for pretest	6
✓	Initiate OMB clearances for pretest and main survey	7
	Conduct and observe pretest	8
	Debrief pretest interviewers and assess findings	9
	Revise questionnaire and prepare plan for pilot test	10
✓	Review revised questionnaire and pilot test plan	11
	Recruit interviewers and prepare training materials	12
	Pilot test final questionnaire	13
	Revise procedures and questionnaire for main survey	14
✓	Review and approve procedures for main survey	15
	Print or program final questionnaire	16

✓ indicates item is primarily an Agency responsibility

1. Prepare Analysis Plan ✓ [Agency responsibility]

The first step in constructing a questionnaire for an EPA-sponsored survey is to determine the analysis requirements. Because you, as sponsors of the project, are likely to have greater expertise in the subject matter of the research, the contractor should not prepare the analysis plan. As discussed in Chapter 1, the analysis plan should define:

a. The purpose of the survey

b. The research objectives

c. The key variables

d. The analytic approaches and methods to be used

e. A list of preliminary tabulations

You should include at least a draft analysis plan that the Agency issues for the survey. Then, if a contract is awarded, the contractor can refine the draft and submit it for approval along with the other components of the work plan.

2. Draft List of Topics or Suggested Questions ✓ [Agency responsibility]

It is suggested that you prepare a comprehensive list of research questions and an informal list of the items you would like to see on the final questionnaire. Keep in mind that all questions should be clearly relevant to the objectives of the research—you should not ask for information that may be "nice to have." If you decide to draft an informal list of questions, therefore, as you write each item, ask yourself, "Why do I want to know this?" "It would be interesting to know" is not an acceptable response. Also, do not try to write the questions verbatim or to format the questionnaire; it's best to leave those tasks to the contractor (see Step 4).

Before preparing your list of research topics and preliminary questions, it is suggested that you look for questions or scales that have been used in earlier Agency surveys to explore various environmental issues**.** In addition, you may find questions or scales used in other (non-EPA) survey reports helpful in framing your research questions. This is especially true if you may want to compare your survey results with other surveys or the decennial Census.

A search of this type may seem time-consuming and tedious, but it is often time well spent. Even if you find only a few good items, this may cut down on the time required to test the questionnaire. Moreover, the search will generally give you a better perspective on your analysis needs.

If you do find usable questions, they are unlikely to cover all aspects of the problems the new survey is intended to address, especially since EPA often deals with evolving issues on which little research previously has been done. No doubt you will have many new questions you expect the contractor to explore.

Keep in mind that any list of topics or questions prepared at this stage should be regarded as preliminary. Only after completion of exploratory studies and one or more advance tests of the data collection instrument can you be confident that the questionnaire will meet your data and analysis objectives. Some compromises in the data requirements may be necessary if respondents in exploratory interviews (Step 3) are unable or unwilling to answer certain kinds of questions.

3. Conduct Exploratory Group or Individual Interviews

Even if you succeed in preparing a reasonably complete list of topics or preliminary questions, you may find that there are still gaps in your understanding of the issues. If so, before the contractor begins the initial draft of the survey questionnaire, explore some of the key issues with a few members of the populations you plan to investigate.

A series of focus group interviews or in-depth interviews may prove fruitful in resolving uncertainties at this early stage of the questionnaire's development. These have been found to be highly effective in resolving a range of conceptual problems that would be prohibitively costly or impossible to resolve later. Individual in-depth interviews or focus groups can be used to explore attitudes, opinions**,** concerns, and experiences of potential respondents; develop data specifications; test the wording of questions; or even to evaluate an entire draft of a

questionnaire.

These techniques are suitable for both household and non-household surveys. For example, sometimes it is essential for the sponsors to know the record-keeping practices of the industries they intend to survey so they can determine what kinds of questions the respondents may reasonably be expected to answer accurately.

Either of these exploratory research techniques is likely to add two to six weeks to the overall development process. If OMB clearance is necessary, it may take somewhat longer.[8] However, because the final questionnaire undoubtedly will require fewer refinements and less testing, you may be able to recover lost time before the main survey begins. These interviews normally are the contractor's responsibility, although it is highly recommended that you attend them.

4. Prepare First Draft of Questionnaire

The contractor can begin to draft the questionnaire using (a) the data and analytic requirements you formulated in Steps 1 and 2; (b) the findings of the exploratory interviews, if any (Step 3); and (c) other specifications in the work plan concerning the data collection, processing, and analysis procedures. A structured questionnaire typically consists of:

- Introductory information explaining the objectives of the survey and the reasons the respondent's cooperation is solicited. (In a self-administered questionnaire, this information is usually stated in a cover letter).

- Identification and control information showing the name of the survey sponsor, the name of the organization collecting the data, the authority for collecting the data (e.g., any applicable statutes), the OMB control number and expiration date of the clearance, code numbers identifying the individual response unit (the household, business, individual, etc., and where the unit is located), and any additional information needed for control purposes.

- A set of standardized questions addressing the research problem.

- Instructions to the person entering the data into a computerized file.

- Definitions of all technical and unusual terms. (An EPA-sponsored survey of businesses or industries frequently will include an entire section of definitions.)

In most cases, once you have formulated the basic content of the questionnaire and approved the work plan, it is best to let the contractor construct the questionnaire. The content and wording of the individual items as well as the overall organization and format of the questionnaire will be major factors in determining whether the survey ultimately produces timely, reliable, and useful information.

The questions should be worded so they can be clearly understood, arranged in the best possible

[8] OMB clearance is needed whenever you plan to interview 10 or more respondents.

order, and capable of eliciting objective, unbiased answers. If the questionnaire is to be self-administered, it has to be designed in a way that will motivate the respondents to make the necessary efforts to retrieve, organize, or report the required information in the specified format. If it is to be administered by a trained interviewer, the design and format should facilitate the work of the interviewers in asking questions and recording responses. The format should also expedite the coding and data entry operations during the processing phase.

5. Review and Approve First Draft of Questionnaire ✓ [Agency responsibility]

Extensive reviews of the first draft of the questionnaire (and all subsequent drafts) are vital to ensure that:

- The content is relevant to and focused on the research objectives;
- The wording is clear and unambiguous; and
- The overall organization and format of the questionnaire will facilitate the data collection, processing, and analysis activities.

As project officer, one of your principal responsibilities during the development process is to ensure that the questionnaire is constructed so that it will achieve the objectives of the study. Criteria for a systematic review of draft questionnaires are given in section B of this chapter; therefore, it will not be further discussed here.

In addition to circulating drafts to key people on the project staff, you should have computer programmers, systems analysts, and statisticians review them, as well as people outside EPA who are knowledgeable about the subject matter or the intended uses of the data. After the contractor incorporates changes in the draft, make sure the comments of all reviewers are accounted for.

6. Prepare Plan for Pretest

While the Agency is reviewing the initial draft of the questionnaire, the contractor should prepare a plan to pretest it informally on one or more subgroups of the target population.

The pretest plan should cover:

- The scope of the test (whether the entire questionnaire or only certain questions will be evaluated).
- The size and composition of the test sample.
- The techniques to be used in administering the test (e.g., face-to-face or telephone interviews).
- Procedures for training the interviewers and observers.

- Procedures for conducting and evaluating the test.
- The kinds of tabulations and analyses that will be done.

Pretesting is essential for all structured questionnaires, regardless of the data collection method proposed for the survey itself. However, the techniques used to pretest a face-to-face or telephone survey (involving an interview) and a mail survey are quite different.

Face-to-face or Telephone Surveys

For a face-to-face or telephone survey, one or more informal pretests are mandatory. However, rigorous analytic techniques are not normally used. Instead, interviewers, observers, and respondents subjectively evaluate various aspects of the questionnaire. At a relatively low cost, pretests can determine whether changes in the wording of the questions, their sequence, or the length of the questionnaire are likely to improve the quality of the survey data. Pretests also may indicate a need for adding or eliminating certain questions.

Usually the contractor will do a few informal tests; then, when the wording and format of the questionnaire have been refined, they will conduct a formal test, called a "pilot test," to evaluate the data collection procedures as well as the questionnaire. For a major interview survey, a full-scale pilot test should be done. (Step 13)

Some of the techniques used to evaluate pretests of an interview survey are:

- Observations by trained supervisory staff
- Discussions with respondents immediately after the questionnaire is administered
- Daily interviewer debriefings
- Interviewer records of call-back rates and the duration of the interviews
- Recordings of a few test interviews[9]
- Written reports by interviewers on the difficulties encountered in collecting the data, and suggestions for improving the questionnaire, control forms, or the interviewing procedures
- Debriefings at the conclusion of the pretest with the interviewers, questionnaire designers, field supervisors, and observers
- Preliminary tabulations of the pretest data

Mail Surveys

Techniques for pre-testing a mail survey generally follow the steps outlined above for face-to-

[9] By law, you must get approval in advance to make an audio or video tape of any interviews.

face surveys. Usually, a draft of the questionnaire is mailed to a small subset of the target population. The results are then tallied and evaluated, possibly with telephone follow-ups.

A less formal method of testing a mail questionnaire is to mass-administer it to a group of respondents "classroom-style," with a moderator and several observers in attendance. Some face-to-face interviews may also be used for testing mail questionnaires at an early stage of their development.

When the contractor submits the pretest plan for Agency review, make sure:

(a) The pretest sample adequately represents all important subgroups of the target population;

(b) The size of the sample is adequate for a valid test;

(c) The test conditions approximate those of the actual survey; and

(d) Enough time has been allowed to analyze the test results and incorporate any necessary revisions in the questionnaire.

Submit the plan along with the draft questionnaire to your office's paperwork clearance officer.

7. Initiate OMB Clearance for Pretest and Main Survey ✓ [Agency responsibility]

If data are to be collected from ten or more members of the public, a major responsibility of the project officer is to obtain OMB clearance(s) for all pretests and the main survey in a timely way. Clearance is mandatory per the Paperwork Reduction Act (PRA) of 1995.

The purpose of the OMB review is to ensure that (a) the information that agencies propose to collect is in the public interest; (b) the reporting "burden" (the length of time it takes a respondent to complete a questionnaire or be interviewed) is reasonable; (c) certain statistical standards are met; and (d) privacy and confidentiality are maintained.

The OMB clearance process is time-consuming and you should allow approximately 4 months—two weeks for each Agency office that must review the clearance package before it goes to OMB, and at least two months at OMB. **OMB clearance must be started early in the design process.** You may submit a combined clearance request for the pretest (or pretest and pilot test—Steps 10-13) and for the main survey.

Because the OMB clearance paperwork may have to be approved by many Agency offices, the project manager should closely follow the materials to ensure that they have not been "lost" on somebody's desk.

8. Conduct and Observe Pretest

While awaiting the OMB clearance, the contractor will sometimes organize and train the interviewers and other staff to be used for the pretest, but usually it is best to wait until the

clearance is granted. Clearance is mandatory if 10 or more identical interviews are to be conducted, according to the Paperwork Reduction Act (PRA) of 1995.

Preparing for Pretest (Contractor)

The contractor's principal responsibilities in preparing for the pretest are:

- Selecting the agreed number of respondents from the target population. For an informal pretest, 20 to 50 respondents usually will suffice. Generally, a "purposive" sample rather than a probability sample is drawn so that all subgroups in the target population or specific subgroups of concern are represented.

- Choosing interviewers for the test. Some survey research firms maintain an experienced team of interviewers solely for pretests. Others use only supervisors so they can gain experience that will be useful in training and overseeing the interviewers picked for the main survey. Still others use interviewers with education and experience similar to that of the interviewers to be used for the main survey. In all cases, it is best to use as many interviewers as possible, provided each of them has a sufficient workload to justify the cost of their training and travel.

- Selecting and training one or more field supervisors to oversee the interviewing.

- Training the interviewers in the general purposes of the survey and the specific objectives of the pretest. This kind of training is vital for all the interviewers who participate in the test—even the most experienced. If the interviewers do not have a thorough understanding of the questions, it will be impossible for the questionnaire designers to determine whether problems with the questionnaire are due to poor interviewing or to the questionnaire itself.

- The interviewers also should be thoroughly trained in the proper way to administer the questionnaire (e.g., not to arbitrarily reword questions and to effectively probe and ask questions when respondents' first answers are inappropriate, inaccurate, or incomplete.)

The pretest itself is frequently conducted under conditions similar to that planned for the main survey.

<u>During Pretest</u> (Agency staff)

Once the pretest is in progress, it is recommended that you or members of your staff:

- Observe several pretest interviews to gain first-hand experience in how the questionnaire works in practice. Discussions with respondents following each pretest interview—a major feature of informal pretests—provide important feedback to questionnaire designers. Discussions reveal how respondents interpreted various questions; difficulties respondents experienced in replying to certain items; how respondents would ask certain questions or their feelings about questions to which they responded "Don't

Know," etc.

- Attend some of the daily debriefings with the interviewers. The purpose of these debriefings is to get immediate feedback from field personnel on problems they have had with the questionnaire so the contractor can make on-the-spot refinements for testing during the next day's interviewing. Things to cover during these debriefings might include:

  - Difficulties interviewers encountered in locating respondents.
  - Questions that made respondents feel embarrassed or uncomfortable.
  - Questions that were awkward to read.
  - Items respondents refused to answer and the reasons given for the refusals.
  - Difficulty interviewers had in maintaining rapport with respondents.
  - Whether the respondents became impatient or bored.
  - Whether respondents seemed to want to rush through any part of the questionnaire, particularly the ending.
  - Whether the format of the questionnaire was particularly hard to follow.
  - Whether any items required further explanation.
  - How long the interviews took.
  - If there was enough space to record answers, especially to open questions.

9. Debrief Interviewers and Assess Pretest Findings

When the pretest is over, the contractor generally will hold one or more debriefing sessions with all the interviewers, supervisors, and observers who have participated in the pretest.

You and members of your staff should attend these sessions so that any necessary changes in the questionnaire or training procedures can be jointly agreed to and quickly implemented. The format of these sessions generally is similar to that of focus group discussions (see Section A of Chapter 2). Based on the outcome of the final debriefings and any preliminary tabulations, you and the contractor should determine if further revisions or tests of the questionnaire are needed.

The contractor should revise the questionnaire after each pretest until all problems are resolved. In a major survey, another pretest should be done after each revision because the revisions may cause new problems.

Note: Steps 10-13 may be omitted if no further tests are planned; these are referred to as "pilot tests" to distinguish them from "pretests" (Steps 6-9).

10. Revise Questionnaire and Prepare Plan for Pilot Test

The last step in the testing process should be a full-scale pilot test—a more formal type of pretest. A pilot test is, in effect, a "dress rehearsal" for the main survey. Normally, it should duplicate the field procedures as closely as possible, and the questionnaire should approximate the one that will be used in the main survey.

The first step in preparing for the pilot test is to develop a planning document clearly delineating the objectives of the test. Pilot tests can be used to:

- Evaluate the wording, content, and format of the questionnaire, and test alternative versions, if necessary.

- Identify and correct weaknesses in the proposed interviewing procedures—the interviewer's instructions and training manuals, the length of the interviews, and the logistics of the field operations.

- Provide a realistic body of data to test the proposed processing procedures—the specifications and instructions for coding, data entry, computer editing, and tabulation operations

If the test is carried through to the analysis phase, the preliminary tabulations can provide a final check on the analysis plan

It takes considerably longer to conduct, process, and evaluate the results of a pilot test than results from an informal pretest. From 5 to 10 months may be required for the pilot test, after the Agency approves the questionnaire. This includes the time required to obtain OMB approval (up to 4 months).

In the pilot test of a face-to-face survey, at least 50 respondents and several interviewers at different skill levels are generally used. It is not unusual to have up to 300 respondents and as many as 20 interviewers. Potentially "difficult" respondents or "hard-to-reach" population groups should be included.

The interviewers should also be selected and trained in the specifics of the test, and one or more field supervisors appointed to keep track of the interviewers' workload and evaluate their performance.

11. Review Revised Questionnaire and Pilot Test Plan ✓ [Agency responsibility]

You and your staff should critically review the revised questionnaire and pilot test plan, giving special attention to the proposed tabulations and analyses. Circulate it to computer programmers and system analysts, if necessary.

The contractor should allow enough time to analyze the pilot test data and apply the findings

before the main survey begins. Important benefits of pilot tests frequently are not realized because the analysis is not planned in enough detail and insufficient time and resources are committed to it.

If you have not yet applied for OMB clearance of the pilot test, you should do so at this time. For assistance, contact the office or refer to the Internet. It is recommended that you combine it with the clearance request for the main survey (Step 7) so the contractor can proceed with the main survey as soon as the pilot test results are analyzed.

12. Recruit Interviewers and Prepare Training Materials

The quality of the interviewing in the pilot test and the actual survey will be greatly influenced by the amount of care taken in selecting and training the interviewers. Typically, a great deal of effort goes into the development of the questionnaire so it will effectively yield valid, unbiased data. To achieve satisfactory results in an interview survey, the data must be collected in a systematic, uniform manner from all respondents.

The interviewers selected for the pilot test usually work in the main survey as well. If the contractor has a permanent field staff in the sampling areas, there probably will be no need to recruit new interviewers. Most large survey research firms maintain a permanent cadre of interviewers located throughout the United States. Having a permanent interviewing staff does not guarantee the quality of the fieldwork, but experienced interviewers are far more likely to collect good data than a group of new interviewers recruited solely for one survey.

In addition to selecting the interviewers, the contractor should: (a) develop procedures and materials for training the interviewers and a field supervisor; (b) determine how many training sessions will be needed; and (c) where the session will be held. This can be done while awaiting the OMB clearance for the pilot.

Interviewer training for the pilot test should cover the objectives of the survey, the content and concepts of the questions, interviewing techniques, the procedures to be used to control the quality of the fieldwork, and practice interviews. Instruction manuals and other training materials also should be prepared so their effectiveness can be assessed before the interviewers for the main survey are trained. (See section B of Chapter 5 for detailed information on training.)

13. Pilot Test Final Questionnaire

Once the interviewers are recruited and trained, the interviewing phase of the pilot test should proceed much like any other data collection operation using a structured questionnaire. The techniques used to observe and evaluate the test are similar to those used in informal pretests (see Steps 8 and 9) with one major difference—a greater focus on statistical evaluation of the data.

For example, debriefing sessions should be held with the interviewers and observers following the test. The debriefings may alert the project management team to problems with specific questions, the order of the questions, or the length of the questionnaire. As a result, it may be

necessary to change or discard certain questions. If the average length of the interviews is too great, some questions may have to be dropped to stay within the established time and budget constraints.

14. Revise Procedures and Questionnaire for Main Survey

When the pilot test is concluded, the questionnaire should require few revisions. By gradually fine-tuning the data collection instrument, the contractor should be able to begin the main survey with clear assurance that the resulting data will meet the Agency's objectives.

In addition to modifying the questionnaire, the contractor should submit a revised data collection plan to the Agency for approval before the actual survey begins. The plan should include: (a) provisions for training and supervising the interviewers, (b) rules for respondent eligibility (respondent rules); (c) rules for following up the initial contacts with respondents; (d) rules for verifying and evaluating the interviews; and (e) the quality-control measures that will be used to ensure that the target response rate for the survey and for individual items are achieved. (See section A of Chapter 5 for detailed information on preparing for the interviews).

15. Review and Approve Procedures for Main Survey ✓ [Agency responsibility]

The project staff, data processing specialists, and systems analysts should critically review the final draft of the questionnaire and the proposed data collection procedures. It is strongly recommended that you have a survey expert review these materials (whatever collection method is planned) before granting approval to proceed with the survey. If you have not submitted the OMB clearance request for the main survey, do so at this time in coordination with EPA's Office of Environmental Information's Information Strategies Branch.

16. Print or Program and Test Final Questionnaire

The questionnaire for the main survey should not be printed until the results of the pilot test indicate there are no more serious problems. The questionnaire should not go to the printer until you have received an OMB control number; both the number and the expiration date of the clearance must appear on the form.

Make sure that the contractor orders enough questionnaires. It is best to get 50-100 percent more than the number of respondents. The extra copies can be used for training purposes and practice interviews; some are lost during the distribution process, others are wasted in the field; and some may be needed for follow-up interviews.

Check proofs of the questionnaire received from the printer for spelling and typographical errors. When the printed version arrives, batches should be spot checked for poor print quality, missing pages, etc. For computer-assisted surveys, this step involves programming and testing the final survey instrument.

B. Reviewing Questionnaire Drafts

This section provides instructions for systematically reviewing a survey questionnaire. The instructions are intended to help you critique drafts submitted by the contractor for Agency approval during the development process, as shown in Exhibit 2, above.

The instructions are presented in three parts, which should be reviewed in order:

(a) The form, content, and wording of each question individually

(b) The content and organization of the questionnaire as a whole

(c) The overall format

A checklist of the suggested criteria for this three-stage review is given in Exhibit 3. Use it, along with a copy of the analysis plan (see Chapter 1), to guide your reviews. Also, be sure to circulate review drafts to others with expertise in questionnaire design, data processing, and statistical analysis, as appropriate.

1. Reviewing Individual Questions

Begin your review of the questionnaire by critically examining the following elements for each question:

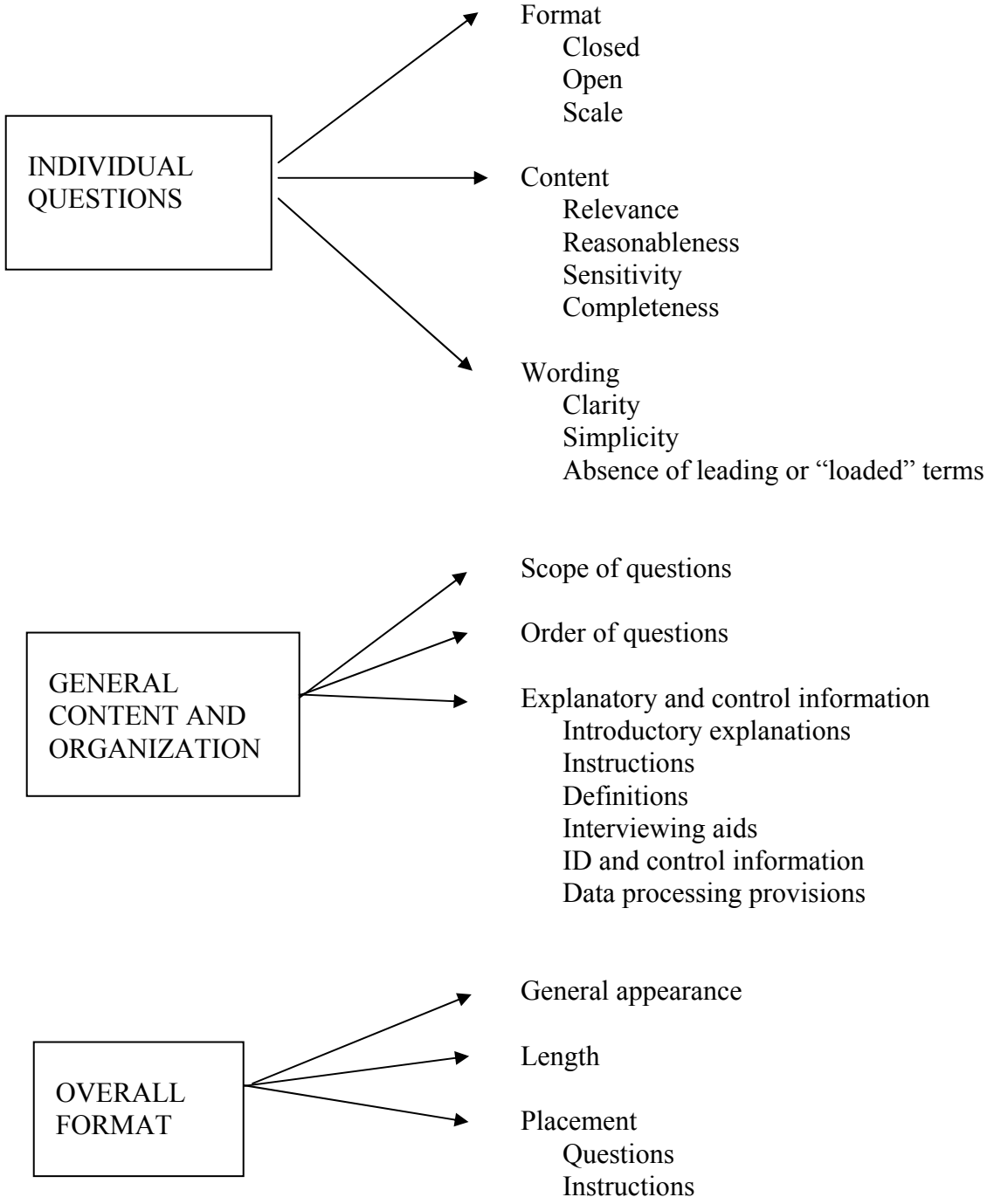
Individual Questions:

(a) Format
(b) Contents
(c) Wording

(a) Format

You will want to look first at the appropriateness of the answer format of each question. There are three reasons**:** (a) Survey questions are classified by their answer format, (b) the form is the most immediately visible aspect of a question, and (c) the proposed form of the question may affect your review of the content and wording. The following information clarifies the basic types of survey questions and the advantages and limitations of each.

Exhibit 3: Criteria for Reviewing Survey Questionnaires



Types of Survey Questions

There are three basic types of survey questions, closed, open, and scale:

(i) Closed Questions

Closed (or closed-ended) questions offer respondents a choice of two or more response options, the most common of which are “Yes/No” and ”Agree/Disagree.” Sometimes a third option, "Don't know" or "Undecided," is used. Multiple-choice questions are also classified as closed; these permit respondents to choose their answer(s) from several response categories.

(ii) Open Questions

Open (or open-ended) questions ask respondents to reply in their own words. Traditional open questions allow respondents to give their opinions fully, in language comfortable to them, without restriction. However, open questions do not necessarily call for a verbal response. They are often used when very short numerical answers are sought—age in years, expenditures in dollars, volume in cubic feet, etc. [10]

Open questions are further classified as fully-open (the traditional open question) or partially-open. When a question is fully open, the interviewer simply records the reply verbatim. The questionnaire will include a blank space for the interviewer to write in the respondent's answer.

Partially-open questions are more similar to closed questions. They appear to be open to the respondent, but they actually provide a fixed set of response options. The interviewer selects the response option(s) closest to the respondent's answers, or sometimes will guide the respondent to an answer within certain limits. Partially open questions on self-administered questionnaires provide several fixed response options as well as an "Other-Specify" category, which prompts for a written answer.

(iii) Scale Questions

Scale (or ranking) questions permit respondents to rank their responses according to (a) preference or interest, (b) degree of agreement or disagreement, or (c) some other scale of measurement. Scale questions are actually a special form of closed questions.

Scale questions are good for measuring attitudes and values because they allow researchers to identify the intensity of respondents’ feelings, beliefs, or preferences. For example, you might devise an intensity scale to measure a community’s preference for air quality strategies.

[10] Often the typography indicates the format of the answer. For example, $ □,□□□,□□□ where an answer in whole dollars is requested.

Closed or Open Questions?

Many survey research firms have a decided preference for closed questions. There are three reasons: (1) closed questions tend to be more reliable; (2) they are easier for interviewers, coders, and analysts to deal with; and (3) unlike open questions, they generate no irrelevant, unintelligible responses to complicate the data processing and analysis phases. Nevertheless, closed questions can have certain disadvantages, most notably their superficiality. A questionnaire containing only closed questions might not get to the heart of complex or new issues.

Closed questions also tend to force replies. Sometimes respondents choose any answer to conceal their ignorance about the topic or they may pick a response that does not reflect their true opinion—only because the respondent feels compelled to check or circle one of the fixed responses.

Open questions have many advantages:

- They put a minimum of restraint on respondents' replies and the manner in which they express them.

- The open format permits interviewers to probe the respondents' knowledge of a subject and their frames of reference, and to clarify or ascertain the reasons for the answers they give.

- Open questions are also appropriate when the potential responses are nominal, e.g., questions asking for a single-word response such as the respondent's age or income.

The richness of open-question data can be a disadvantage, however:

- A major challenge for coders is reducing a large number of varied responses to a few categories that can be treated statistically. Coding a complex set of open responses is not only time-consuming and costly, but also introduces some amount of coding error. If the data categories are extensive, the contractor needs to develop complex coding instructions, train staff in the proper use of the codes, and make periodic reliability checks to estimate the amount of coding error. (See Chapter 6 for more information on coding.)

- Open questions take more time to answer than closed questions. This tends to increase the response burden of the survey and may lead to greater item nonresponse or complete refusal to cooperate.

- Open questions also require greater interviewer skill in recognizing response ambiguities, and in probing or drawing out respondents – particularly respondents who are reticent or not highly verbal – to make sure answers are codable. This aspect of the open format has made some researchers wary about using it except in situations where they are sure of getting well-trained, well-supervised interviewers.

In sum, the open format is an invaluable tool for exploring a topic in depth, and is essential if you are beginning work on a new research topic and need to explore all aspects of the subject. However, because of their complexity, from both the interviewer's and the respondent's viewpoint, open questions are more useful during the development and pretest phases than in a survey's final implementation, by which time the likely answer choices should have been formulated.

When lists are used, complete information can be obtained only if each item is responded to with a "Yes/No," "Applies/Does not apply," "True for me/Not true for me," and the like, rather than with instructions such as "Circle as many as apply."

Rating-scales with more than four or five verbal points should not be used. Numerical scales are preferable if more detailed measurement is desired. Respondents should not be asked to rank their preferences among a number of options unless they can see or remember all the options. In face-to-face interviews where prompt cards are used, respondents can rank no more than four or five options. In a telephone interview, rankings can be obtained by a series of paired comparison questions. However, respondent fatigue limits the total number of alternatives that can be ranked.

(b) Content

Next, you'll want to review the content of the individual items. Each question should be (a) relevant to the Agency's informational or analytical objectives; (b) reasonable, given the respondents' probable knowledge and experience; (c) sensitive to the respondent's self-interest; and (d) complete. More specifically:

(i) Relevance

Each question should be clearly relevant to the informational and analytical objectives of the survey, as defined in the analysis plan. Except for the first one or two questions – which may be designed simply to orient the respondents or put them at ease – each item on the questionnaire should yield a particular piece of data that will contribute to the objectives of the survey. Of course, more than one question may be needed to get a complete perspective on a single research question or variable.

(ii) Reasonableness

The question should ask for information the respondents can reasonably be expected to provide, given their probable knowledge and experience. The extent to which people can respond to the question will affect both the quality and quantity of their responses. Rather than admit their ignorance, respondents may give a false reply or no reply at all. Therefore, in reviewing the question, consider the difficulty of the question from the respondent's perspective.

For example, is the respondent required to recall events or transactions that happened weeks or months ago? Periods of a year (or sometimes longer) are applicable for highly salient topics such as the purchase of a new house, the birth of a child, or a serious auto accident. On the other hand, periods of a month or less would apply for items with low saliency, such as the purchase of

clothing or minor appliances.

If detailed information on frequent behavior of low saliency is required, respondents can be asked to keep diaries. Diaries will provide more accurate results than memory. In a business survey, the use of records (if available) and direct observation by interviewers will improve reporting of the desired information. In addition to diaries, records, and direct observation, other techniques can be used to motivate respondents to supply accurate data. For example, (a) probes or follow-up questions, (b) verbal reinforcement by interviewers, and (c) interviewing aids such as pictures, calendars, checklists or prompt cards.

(iii) Sensitivity

In addition to being unable to answer, the respondents may not want to reply to a particular question because they feel some harm may come to them, or they will be embarrassed, or that the information is too personal to divulge to others. The net result is the same as for unreasonable items—many inaccurate or missing responses, or refusal to cooperate.

Therefore, in reviewing the content of individual questions, it is important to consider the sensitivity of each question. Topics many people regard as sensitive are income, assets, profit, religion, political affiliation, and beliefs. Any question dealing with such topics needs to be well justified. (In fact, OMB requires additional justification for questions that are likely to be considered intrusive or damaging to respondent self-esteem.)

If the question is not essential, it may be best to drop it. If it is essential, there are ways of minimizing the possibility of inaccurate or missing responses:

1. Careful placement helps. Locating a sensitive question towards the end of the questionnaire, or grouping it with related questions of a non-threatening nature, tends to improve the reliability of the response. (See "Placement" at the end of this section.)

2. For obtaining information on the frequency of socially undesirable behavior, open questions are better than closed questions, and long questions are better than short questions (Gilgun, 1995).

3. If respondents are being asked to rank attitudes or behavior, the scale should start with the least socially desirable response options. Otherwise, the respondent may choose a socially desirable answer without hearing or reading the entire set of responses.

4. In asking about socially undesirable behavior, it is better to ask respondents whether they have ever engaged in the behavior before asking them about their current behavior. Also, it is better to ask about "current" rather than "usual" behavior.

(iv) Completeness

Each question should have all the necessary elements for obtaining the desired information. There are several tests you can apply to each question to determine whether it is complete. For

example:

1. If the respondent is to check only one response category out of a fixed set, the categories must be exhaustive, i.e., cover all possible alternatives. If not, then an "Other-specify" category should be added. Response categories also need to be mutually exclusive—overlap might confuse the respondent.

2. If the question contains a time reference, the period or date should be specified.

3. If you want the respondent to reply with a numerical amount, clearly indicate the desired units, such as days, tons, or dollars. [11]

4. If the respondent is asked to give an opinion on a particular issue, a "Don't know" or "No opinion" response category may be needed. Including such a category will often affect the results. Whether or not to include the category is dependent upon the necessity of the respondent's opinion—even though he or she may have little knowledge of the pertinent issues.

5. Questions should be phrased so that the analysts can distinguish between no response and a response of "Zero" or "None." For example,

   If "*Annual volume of chemical waste products _____ (metric tons)*" is left blank, it will not be clear to the analysts whether the firm's waste products were zero tons or whether they simply did not answer the question. This can be remedied by changing the item to— *"Annual volume of chemical waste products: ☐ None or ____ (metric tons)"*

(c) Wording

The last set of review criteria for individual questions concerns wording. Each question should be (a) clear and unambiguous, (b) simple and specific, and (c) totally free of any leading or "loaded" language.

In reviewing the wording, read each question slowly, preferably aloud, and assess the following:

(i) Clarity

To keep response errors and biases to a minimum, each question should be clearly and unambiguously worded so there is no way for anyone in the sample to misinterpret it.

Words that can change the entire meaning of a question if they are not correctly interpreted should be bold-faced, underlined, or italicized. For example, any change in the frame of reference from a previous question should be clearly indicated—a request for "total gross sales last month," rather than a request earlier in the questionnaire for total gross sales last year; or

[11] The layout can indicate the format of the desired response. For example, $ ☐☐☐,000 for hundreds of thousands of dollars, $ ☐☐☐.☐☐ for dollars and cents.

"monthly net income," rather than "monthly gross income." If necessary, the question should be reworded to eliminate any chance of misinterpretation, or a brief introduction should be given as a transition.

Note that **boldface** or underlining "jumps out" more than does *italicizing*.

Words with multiple meanings are especially problematic. For example, in a question like "Do you think EPA has treated the chemical industry fairly?" "fairly" could mean "justly," "equitably," "not too well," "impartially," or "objectively." In cases like these you should describe exactly what you mean rather than rely on a single word to convey what might be a complicated concept.

Any unusual words should be defined. (See Definitions later in this section.) Slang and colloquialisms should be avoided, not because they violate good usage, but because many respondents may not know what they mean. On the other hand, there is no reason not to use contractions; if a sentence "reads" more naturally with a contraction ("it's" rather than "it is"), there is no reason not to use the contraction.

(ii) Simplicity

Simply worded questions also help to reduce the number of inaccurate and missing responses. Compound questions giving two or more frames of reference—so called "double barreled" items—confuse respondents and result in many invalid responses. For example, a question like "Do you feel that air pollution is a serious problem and that dust from construction sites is the major cause?" would confound many respondents, who may agree with only half the question.

Making questions as specific as possible tends to make the respondent's task easier, which, in turn, results in fewer invalid replies. Normally, a question should tap a specific opinion, not a general attitude. Items should be directed toward specific rather than general concerns.

(iii) Absence of leading or "loaded" terms

Respondents generally want to be thought of as good people. Even where they might be expected to strongly oppose something or someone, respondents tend to choose an answer that is most favorable to their self esteem, that they think makes them look intelligent or thoughtful, that they think the interviewer would like them to give, or that is in accord with social norms ("politically correct.") A further factor leading to bias is a desire to be polite to an interviewer, who usually is a stranger. In being polite, respondents will hesitate to say unkind things they believe might offend the interviewer. Therefore, any question asking about socially desirable or undesirable behavior or attitudes tends to produce bias and needs to be worded with care. In fact, one of the most common traps questionnaire designers fall into is to use leading or "loaded" words, particularly words that are loaded with "social desirability." Even without deliberately wording the questions in a leading way, an interviewer's voice inflexions can produce a situation where the respondent is encouraged to answer in a particular way.

However, there are instances where leading questions may be necessary. For example, you

might ask the question, "When was the last time your exhaust filtration equipment failed to function properly?" The equipment actually may have never failed. On the other hand, if the researchers believe the respondents may have a tendency to underreport such failures, asking the question this way may result in more accurate statistics.

2. General Content and Organization

Next, examine the questionnaire as a whole, specifically looking at the:

General Content and Organization:

(a) Scope of questions
(b) Order of questions
(c) Explanatory & control information

(a) Scope of the Questions

Of course, the questionnaire should cover all aspects of the problem. Since you, as the survey sponsor, will have contributed the basic substance of the questionnaire, your review of the overall content should be a simple matter of making sure that the draft includes all of the Agency's data requirements. The analysis plan will be invaluable for guiding this part of your review.

(b) Order of the Questions

Questions should be logically ordered and grouped into coherent categories. The categories do not necessarily have to be labeled, but similar items should be grouped together. A transition statement should mark significant change in topics.

Whether respondents complete the questionnaire on their own or in the presence of an interviewer, they are less likely to become fatigued and will make fewer mistakes if they don't have to constantly shift mental gears. Most respondents are not experts at questionnaire design, but they certainly can distinguish between a questionnaire that is well organized and one that is poorly ordered, duplicative, and repetitive – and they are less likely to be cooperative in responding to a poorly constructed one.

The order of the questions should consider:

- First, the respondent; then
- The interviewer (if any); then
- The processing personnel; and lastly
- The analyst.

Sequencing questions in favor of the respondents tends to improve the quality of their answers. The least sensitive, most general, and simplest questions should be placed first. Beginning the questionnaire with a few non-threatening or easy-to-answer items tends to promote a more

positive attitude on the part of the respondent. Moreover, if at all possible, socioeconomic questions should not be located at the beginning of the questionnaire since some respondents may find them threatening; these include questions about age, race, income, and employment status. Usually it is best to place them close to the end, so that refusals won't affect answers to earlier questions, unless, of course, these questions are critical to the survey's goal, in which case they may be placed earlier in the questionnaire.

Because open-ended questions require more thought than closed questions they are best put at or near the end, unless to do so would seriously break up the subject-matter sequence.

(c) Explanatory and Control Information

In addition to the actual questions, survey questionnaires contain a variety of explanatory and control items to guide people who will be handling the forms—respondents, interviewers, and data processing personnel. Do not neglect these items in your review.

Below are suggestions for critiquing the following "special" questionnaire items:

(i) Introductory explanations to respondents or interviewers

(ii) Instructions to whoever completes the questionnaire

(iii) Definitions

(iv) Interviewing aids

(v) Control numbers to identify the questionnaires and control their flow

(vi) Codes and directives for processing personnel

Virtually all questionnaires contain a few explanatory remarks at the beginning, either for the respondent or to suggest the interviewer's opening remarks. Introductory information should include: (a) what the study is about; (b) its objectives; (c) why respondent cooperation is important; (d) how responses will be used and who will have access to them; and (e) how to get help if respondents have any problems (for a mailed questionnaire.) A good introduction is particularly important in a mail survey where no interviewer will be present.

Respondents also should be told at the outset that accurate and complete answers are desired and that they should think carefully, search their memory, and if appropriate, take time to check their records. If any questions are particularly sensitive or threatening, a few additional comments may be necessary.

For a mailed survey, introductory information should be included in a one-page letter accompanying the questionnaire. The letter should be individually addressed and signed, if possible. (The mail-merge capability of most word processors makes this feasible at little extra cost.)

A mail questionnaire also should advise respondents what to do with the questionnaires when they have completed them. Should they be returned in self-addressed envelopes? What's the deadline for completing them? (Note that deadlines will increase the response rate.) A return address should appear on both the cover letter and the questionnaire.

Suggestions for the interviewer's opening remarks are usually stated at the top of the questionnaire. These should be brief, because long explanations tend to make respondents uncomfortable. The interviewers should simply identify themselves and the organization they represent, and state the purposes of the survey in one or two sentences.

(ii) Instructions

Instructions to respondents or interviewers on how to complete the questionnaire need to be carefully phrased to prevent errors and omissions. Review the instructions as carefully as you do the questions.

All instructions should be uniform in style and clearly distinguishable from other material on the questionnaire, e.g., set off in capital letters. For most surveys, only instructions applicable to all interviewing situations should appear on the questionnaire.

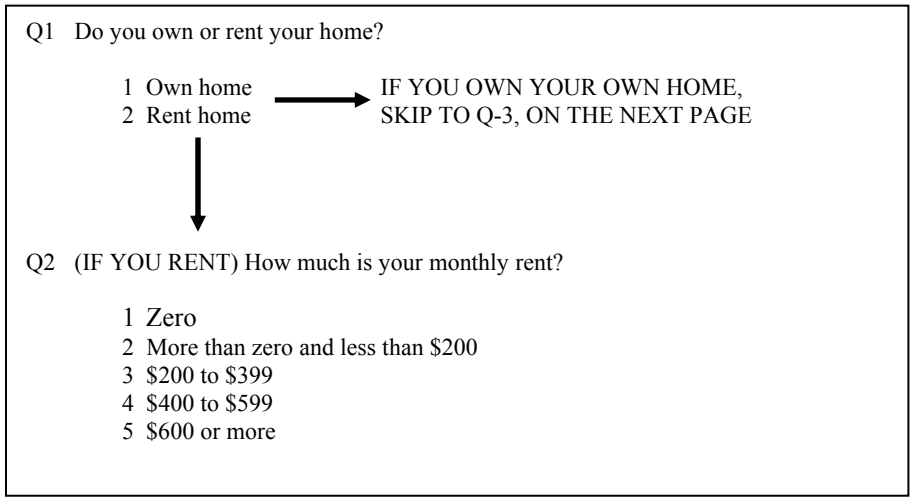
There are two basic kinds of instructions:

1. Directions on how to answer an individual question.

2. Skip instructions, which instruct the person completing the form where to go next, depending on how they answer the current question.

Skip instructions should be (a) worded positively and (b) refer to a later question. They tell the person completing the form where to skip to when a particular reply is given, not where to go when no answer is given. Skip instructions should never ask the respondent to skip backwards to a previous question. They can successfully be combined with arrows, as in the following example: [12]

[12] From Salant & Dillman 1994, page 116. The example also illustrates the recommendation to pre-code the answer choices rather than simply use check boxes.



Complex skip patterns should be avoided, especially on mail questionnaires. However, they are easily managed in a computer-assisted telephone interview because the system can be programmed to present the next question correctly, based on the last answer keyed in by the interviewer.

Note that, in addition to the instructions printed on the questionnaire, interviewers are given separate question-by-question written instructions. These are commonly more detailed and cover unusual interviewing situations. Many surveys incorporate the instructions into a manual and use them both for training and reference purposes. The instructions are not read to the respondent.

(iii) Definitions

In the interest of clarity, any unusual terms on the questionnaire should be defined. For example, if manufacturers are asked to estimate the "value of goods sold" last year, the questionnaire should indicate whether answers should be expressed in current dollars, the depreciated book value, or something else. Definitions should also indicate what units are to be used—dollars, millions of dollars, etc.

Definitions of technical terms are often a major component of questionnaires for Agency-sponsored surveys. It is not unusual for an entire section to be devoted to definitions. Be sure to have the most knowledgeable project personnel review all definitions.

(iv) Interviewing Aids

Although the visual aids that interviewers show respondents to encourage more accurate replies are not strictly a questionnaire component, you should review them along with the questionnaire to make sure they contain an appropriate range of alternative answers.

(v) ID and Control Information

Every questionnaire should contain information to identify it and control its flow through the

collection and processing stages. At a minimum, the first page or cover page should include the following: (a) the title of the study; (b) the name of the organization conducting the study; (c) the OMB control number and expiration date; and (d) a space to insert a code number identifying the response units for follow up, evaluation, cross-referencing purposes, or for determining sample weights (see Chapter 4.) (Since it is possible for the questionnaire to come apart, each page should be numbered and include some information identifying the form.) [13]

In addition, in face-to-face or telephone surveys, there should be a space to record the date and time the interview began and ended. The contractor also may include a place to rate the performance of the interviewer or processors.

Make sure that proper identification and control information is included on the final draft of the questionnaire. Check these items again when you review proofs of the final questionnaire.

(vi) Data Processing Provisions

If at all possible, the format of the questionnaire should be arranged so it is easy for the transcribers or the data entry clerks to proceed from one item to the next. Certain formats and coding schemes can simplify the processing operations and, at the same time, facilitate the tasks of the respondents or the interviewers.

Closed questions can be "pre-coded" to facilitate processing and ensure that the data are in proper form for analysis. Pre-coding involves assigning a code number to every response option. The response options are either explicitly stated in the question or are printed on a card handed to the respondent. When they appear on the questionnaire, the respondents select their replies by checking a box, circling a coded answer, underlining a preprinted response option, or writing in a code or a number. Provisions also may be made for "No answer" or "Don't know" replies.

When the completed questionnaires are processed, the data entry clerks simply key the appropriate numerical codes directly into the computer. This eliminates one processing step because the replies do not have to be coded or transcribed onto a coding or keying sheet before being entered into the computer.

Electronic coding is increasingly being used to process all manner of surveys, including some that are filled out by individual respondents in their own handwriting or by checking boxes. If your survey uses any form of optical character recognition (OCR), you should consult with the contractor about any special layout requirements.

3. Reviewing the Overall Format

The last step in your review should be devoted to the general format of the questionnaire, specifically its general appearance, length, and question placement.

[13] Separation of pages is minimized by printing the questionnaire in booklet form, stapled or bound through the middle.

Overall Format:

(a) General appearance
(b) Length
(c) Placement of questions and instructions

Although the contractor should have designers experienced in the proper formatting of questionnaires, a final review by Agency subject matter and data processing specialists may suggest revisions that will improve the questionnaire's effectiveness.

A well-formatted questionnaire can significantly reduce response errors. If the questionnaire is designed to be self-administered, your review of the format should have high priority. The format should give primary consideration to the respondents, then the interviewers, and lastly the data processors.

(a) General Appearance

The general appearance of the questionnaire, the kind of paper it is printed on, the size and style of the type, and the amount of open space all influence how well the respondents or the interviewers are able to follow instructions and complete the questionnaire. Appearance is very important in a self-administered questionnaire and will influence the response rate and accuracy.

The questionnaire form should look professionally designed and easy to answer. If the form is more than four pages long, a booklet format is desirable. It should be printed on good stock because it will be subjected to a great deal of handling during the course of the collection and processing operations.

Colored paper or color-shaded sections may be helpful in a complex questionnaire. Shading can be used to direct attention to answer spaces, to highlight certain topics, to indicate transitions between sections, and to reserve space for office use. The reduction in respondent and clerical errors is well worth the small additional expense for two-color printing.

Large, clear type should be used throughout. Different type styles should be used for questions, instructions, and data processing notations. Instructions should be in bold type or capitals so they are clearly distinguishable from the questions. Type styles should be consistent throughout the questionnaire.

Above all, the questionnaire should not look crowded. Ample white space should be allowed because it will make the questionnaire look easier to complete, and generally will result in fewer errors by both interviewers and respondents. Response formats should be consistent, and adequate space should be allowed for replies to open questions, arithmetical calculations, and general remarks by respondents or interviewers.

(b) Length

Survey literature abounds with recommendations on questionnaire length. The general consensus

is that setting an arbitrary limit on length is unnecessary and unrealistic. Much depends on the method of administration, the respondent's obligation to reply, the subject matter, and the way the questionnaire is constructed. The ideal length of an interview, regardless of the type of survey, is between 20 and 45 minutes.

Since no social interaction is involved, self-administered mail questionnaires sent out to the general public are directly affected by length. If the subject matter is interesting and relevant, and the respondents are generally well educated, the questionnaire may be 12-16 pages long and there will be no serious loss of cooperation. However, if the topics are likely to be of little interest to the respondents, the questionnaire should not exceed four pages. Anything longer is likely to induce fatigue and result in a considerable number of response errors and a lower completion rate. Poorer response can be expected if efforts to cut length include crowding questions, using oversize paper, or reducing the print size.

The length of a self-administered questionnaire is not as important in a business survey. In fact, EPA relies heavily on long, complex, self-administered questionnaires for obtaining detailed technical information from business and industry. Whether replies are voluntary or mandatory, a long mail questionnaire is often less burdensome than a lengthy face-to-face interview. The questionnaire is less disruptive of office routines and each organization has an opportunity to discuss the questions and search its records, as necessary.

The length of the data collection instrument directly affects the total “response burden” of the survey, which is the time it takes to complete the data collection instrument. The estimated amount of time it takes to complete the proposed questionnaire, multiplied by the number of respondents in the sample, is the total response burden you reported to OMB in your clearance request. The burden should not exceed the allowance provided for the survey in your office's OMB Information Collection Budget, according to the Paperwork Reduction Act (PRA) of 1995.

(c) Item Placement

The placement of the questions, instructions, and other items on the questionnaire can make the task of respondents and interviewers easier and more enjoyable. The placement of response categories also should be consistent. In some cases, good placement helps to minimize response errors, refusals, and non-completions.

Below is a discussion of some general rules for the placement of (a) questions and (b) instructions. Placement “rules” for other items (i.e., introductory material, definitions, and ID and control information) were covered earlier in this section.

(i) Questions

The questionnaire should start with a few short items that are relevant, interesting, non--threatening, and necessary. As previously mentioned, placing questions the respondent may perceive as threatening at the beginning of the questionnaire may result in defensive—and frequently invalid—responses, or total refusal to cooperate. It is best to put potentially threatening queries close to, but not at, the very end of the questionnaire (Dillman, 2000).

Important questions should be placed towards the beginning. The last items in a questionnaire rarely get the same degree of attention as earlier ones, hence the least significant items should be placed last.

It is generally best to start a mail questionnaire with a few short, simple, closed questions. Never begin with an open question requiring a lengthy response. Writing long answers may be difficult and embarrassing for some respondents, who may worry about making spelling and grammatical errors. Finally, include space at the end for general comments.

Questions (and associated answers) should never be split between two pages. The person completing the form may think the question is complete and inadvertently provide a premature, inaccurate response.

(ii) Instructions

Instructions on how to answer a question or a series of questions should be placed before the items they refer to, not grouped at the beginning of the questionnaire.

Instructions for responding to individual items should be placed either immediately before the question or immediately to the right, prior to the space provided for the answer.

Skip instructions should be placed immediately after the answer space allowed for the question. Words or arrows, or both, can be used to advise respondents or interviewers which question they should answer or ask next, depending on how the current question was answered. (This was illustrated in the example above.)

Coding or probing instructions for interviewers should be placed after the question. Notations for coding personnel should be in small type and located so they will be as unobtrusive as possible to respondents or interviewers.

Bibliography: Chapter 3

Dillman, Don A. Mail and Internet Surveys: The Tailored Design Method (2nd ed.). New York, John Wiley, 2000.

Dillman, Don A. Mail and Telephone Surveys: The Total Design Method. New York, John Wiley & Sons, 1978.

Gilgun, Jane F. "We Shared Something Special: The Moral Discourse of Incest Perpetrators." Journal of Marriage and the Family, 57, 1995.

Kish, Leslie. Survey Sampling. New York. John Wiley & Sons, Inc, 1995.

Robinson, John P and Philip R. Shaver. Measures of Social Psychological Attitudes, Revised Edition, Ann Arbor, Institute for Social Research, University of Michigan, 1973.

Salant, Priscilla and Don A. Dillman. How to Conduct Your Own Survey. New York, John Wiley, 1994.

Sudman, Seymour and Norman M. Bradburn. Asking Questions: A Practical Guide to Questionnaire Design, San Francisco, Jossey-Bass, 1990.

Kornhauser, A. & P. Sheatsley, et al. "Questionnaire Construction and Interviewing Procedures," in Research Methodology in Social Relations, 4th Edition, New York, Holt, Rinehart, and Winston, 198.

Chapter 4: Sampling

Sampling is selecting some portion of a target population—sometimes called a study population or universe—and investigating just this portion, or sample.

Until the late 1940s many statisticians felt that collecting information about every member of a population they wanted to investigate was the only acceptable way to conduct a survey. Today, as a result of technical advances in sampling theory and its applications, sample surveys are now widely accepted as an efficient and reliable way of studying individuals, land areas, or even extremely unstable environmental media such as surface water or air.

This chapter provides an overview of the basic concepts of sampling theory and some practical tips on monitoring the sampling activities of a survey contractor. It covers two general types of sampling: probability sampling, which refers to the selection of sample members by chance, and non-probability sampling, where the units selected for study are chosen according to some purposive or convenient scheme, often by expert judgment. Specifically, it examines:

A.	The advantages of using sampling for an Agency-sponsored survey
B.	The relationship between sampling errors and sample size
C.	The methods used to design survey samples
D.	The major components of a sampling plan
E.	Ways the sponsoring office can ensure the quality of the sampling activities

A. Advantages of Using Sampling

Why collect information from only a sample rather than everyone in the population?

In most research situations, investigating the entire study population (taking a "census") is both impractical and inefficient. The most important reason for investigating a sample of the population is that it is cheaper to collect information from a small number of people, land areas, processes, etc., than to collect it from the entire population. In addition, fewer staff are needed to collect the information and process it into a form suitable for analysis. Using sampling for studies of human populations also reduces the burden of survey respondents and provides faster and more accurate results (because a smaller volume of information has to be collected and processed.) Finally, sampling enables one to concentrate limited resources on obtaining answers from everybody in the sample rather than relying on those who choose to reply to a census, often a poor representation of the total population. These concepts can be grouped into four main advantages for sampling.

1. Lower costs

If the population of the proposed study is very large or national in scope—collecting information about the entire population is simply out of the question from a cost standpoint. For example, the cost of taking a census of the U.S. population in 2000 was over $6 billion.[14] A high-quality sample survey of a large human population requires a small fraction of the resources needed to collect data from everyone in the population.

The per-unit cost of a sample is normally higher than complete enumeration of the population because more highly trained staff and more stringent quality control throughout every phase of the survey are required. However, the total cost is far lower, and in many cases counterbalanced by greater accuracy of the data.

Similarly, if you plan to use an expensive measurement procedure to collect certain environmental data, studying a sample of the population often is the only feasible way to keep costs reasonable. For example, the cost of using an expensive monitoring device to measure ambient air quality in more than a small number of communities may be prohibitively expensive—as well as unnecessary—given the advantages of sampling techniques.

2. Reduced Paperwork

The Office of Management and Budget, in accordance with the Paperwork Reduction Act (PRA) of 1995, imposes limits on all Federally sponsored information collections. Using sampling to study a population of interest helps to minimize the paperwork demands that Federal agencies impose on the public, particularly on business and industry.

3. Faster Results

The Agency often needs the results of its survey research projects quickly. Because fewer respondents or specimens have to be investigated in a sample survey, the time required to collect and process the data is generally lower.

4. More Accurate Results

Since survey researchers use carefully controlled procedures to collect and process sample data, it is common for a well-chosen sample to produce more accurate results. Although sampling introduces a source of error in the data—sampling error—that would not occur if all members of the population were studied, that error is identifiable and measurable.

At the same time, because the investigators focus available resources only on a portion of the population, there is less chance for human error and, therefore the data quality tends to be higher. Human errors can creep in at any stage of survey—during the data collection phase, during the editing and coding of the questionnaires, and during the tabulation and analysis operations.

[14] And even here the term “census” is a not entirely accurate, as many questions in the Decennial Census are asked only of a sample of 1 in 6 households.

Because there are fewer data to deal with in a sample survey, greater quality control can be exercised throughout each stage to guard against all manner of errors.

Another reason sample surveys tend to be more accurate is that a representative sample of the study population is more likely to respond when there are sufficient resources to aggressively follow-up non-respondents. This is not always the case with censuses, with the result that only the "easiest," most cooperative respondents are included in the results. Due to budget limitations, when covering a large population with follow-up attempts there is a tendency for completing interviews with only the most cooperative population members. These respondents tend to have different characteristics from those who are more difficult to survey.

Given these advantages, are there any research situations where sampling may not be appropriate for collecting environmental and health data that EPA needs to effectively fulfill its mission? In studies of human populations, if the study population is small, or if separate detailed data for small subsets of the population are desired, collecting data for the entire population may be appropriate for at least some parts of the investigation. For example, if your target population is all U.S. chemical manufacturers, it may only be feasible to study a sample of them to get the information you need. However, if you were interested in a specific chemical that is produced at only ten plants in the United States, it probably would be best to collect data from all of them. Similarly, if you were interested in all the chemical manufacturing plants in a single county, it might be best to survey all plants within the county.

B. Sampling Errors and Sampling Size

In establishing the minimum design criteria for your survey, it is recommended that you include an acceptable level of sampling error for the key statistics. Since this task should be done in the planning stage, before a contractor is hired, sampling errors will be discussed before considering other aspects of sampling. You will also learn how sampling errors are measured and the relationship between sampling errors and sample size.

The purpose of most surveys is to measure certain characteristics of a population. When only a portion of a population is used for study purposes, survey statisticians need a way of estimating the extent to which this portion—the sample—and the entire population differ from each other. Studying a sample rather than every member of a population means abandoning mathematical certainty and entering the realm of inference and probability ("statistics"). Furthermore, the values of the estimates or statistics derived from the data collected from the sample will differ from the actual values that would have resulted had data been collected for the entire population using exactly the same methodology.[15] The difference between these two sets of values for every statistic is called the sampling error, as defined in the following section.

1. Sampling Errors Defined

Sampling errors are measures of the extent to which the values estimated for the sample (means, totals, or proportions) differ from the values that would be obtained if the entire population were

[15] That is, using the same questionnaire, follow-up procedures, data processing procedures, etc.

surveyed. Since there are inherent differences among the members of any population, and since data are not collected for the whole population, the exact values of these differences for a particular sample cannot be known. Moreover, different samples give different results. Therefore, to compute sampling errors statisticians measure the average differences between sample estimates and population values.

When a probability sample is used, sampling errors can be estimated with some precision. A probability sample is one in which each member of the target population has a known, positive probability of being selected. Without probability sampling, there is no way to know how much error there is in the data and, hence, how much confidence one can place in the survey findings.

Non-sampling errors. Unlike sampling errors that statisticians can measure and take into account in reporting the survey findings, other sources of data errors in a survey are (a) estimation biases, (b) systematic errors caused by inaccurate measuring devices, (c) exclusion of part of the population due to a faulty sampling frame or nonresponse, and (d) failure of the interviewers to ask all the questions. All produce errors that are much more difficult to measure than sampling errors, and which can significantly affect the survey results.

2. Measuring and Expressing Sampling Errors

Let's look now at the ways statisticians measure and report sampling errors when probability methods have been used to select the study population.

Suppose you have contracted for a survey to determine how many households in a particular city—for example, City X—are getting their drinking water from contaminated sources. Now, after completing the survey, let's say the contractor estimates that 40 percent of all households in City X are using contaminated sources. The contractor tells you that the standard error, or standard deviation, of this estimate is 2 percentage points; that this estimate is likely to be within 4 percentage points of the true proportion of households in City X using contaminated water. What does this mean?

The standard error is a measure of the probable accuracy or precision of any one estimate derived from sample data. To relate the standard error of this particular statistic—that 40 percent of all households in City X are using contaminated sources—to the true value, the contractor formed a 95 percent confidence interval, which is approximately defined as:

Sample estimate ± **twice** the standard error (S.E.)

The confidence interval in this example is the interval from 36 to 44 percent, i.e. 40 percent ± (2 x 2) percentage points.

Provided the contractor has used a reasonably large sample of households in City X to collect data on the quality of its drinking water, chances are 19 out of 20 that this confidence interval would include the value you would get if you surveyed all the households in City X. [16] If you

[16] The 95% confidence level is the most common criterion used in survey research.

were willing to accept lower odds, or if you wanted higher odds, other multiples of the standard error could be used to attain other confidence levels. For example:

Confidence Interval	Approximate Level Of Confidence
Estimate ± (1.0 x S.E.)	68%
Estimate ± (1.6 x S.E.)	90%
Estimate ± (2.0 x S.E.)	95%
Estimate ± (2.6 x S.E.)	99%

Sampling errors may be expressed either in absolute or relative terms. To illustrate the difference, let's suppose City X has a total of 5,000 households. The 40 percent estimate of households using contaminated drinking water translates to a total of 2,000 households. Stated in absolute terms, the standard error of this estimate is 100 households.

Exhibit 4 shows the absolute and relative sampling error of this estimate expressed in three ways.

EXHIBIT 4: ABSOLUTE AND RELATIVE SAMPLING ERRORS AND CONFIDENCE INTERVALS:
Households Using Contaminated Drinking Water

Survey Results:
Population of city = 5,000 households
Survey estimates that 2,000 households contaminated
Sampling error (SE) = 100 households or 2% of total

Type of Estimate	Calculation	95% Range
Absolute	2000 ± (2*100)	1,800 - 2,200
Relative—		
Proportion	.40 ± (2*.02)	.36 - .44
Percent	40 ± (2*2)	36% - 44%

The absolute sampling error is plus or minus 200; one can say with 95% confidence that the number of households using contaminated sources is between 1,800 and 2,200. Expressing this in relative terms, one can say that between 36 and 44 percent (or 0.36 to 0.44) use contaminated sources.

A common way of expressing sampling errors is the coefficient of variation, which is the sampling error divided by the estimate. In this case, this would be 100 ÷ 2000 or 5%. The coefficient of variation is often abbreviated as "CV." [17]

Therefore, when you establish the Agency's minimum design specifications, be sure to state

[17] Note that the coefficient of variation is always based on one standard error. By contrast, confidence intervals, as shown in Exhibit 4, are based on some multiple of the standard error. In most cases this multiple is 2, for a 95% confidence interval.

whether you are referring to absolute or relative sampling errors or the coefficient of variation (CV). This is especially important for estimates of percents or proportions. In addition, you should be aware of the distinction between the standard error or standard deviation and the confidence interval.

3. Determining Sample Size

How large a sample is needed for a particular survey? Questions about sample size seem to be simple ones, but answering them is not so simple. It is recommended that you specify the desired level of sampling error in the survey specifications, and calculate the required sample size, rather than try to specify the sample size in the survey specifications.

The level of sampling error (or level of precision, as it is sometimes called) is closely related to the number of units in the sample, but only distantly related to the sample size as a proportion of the size of the population. For example, in estimating percents or proportions, the sampling error associated with a sample of 1,000 units taken from a population of 100,000 is almost the same as the error for a sample of the same size from a population of 100,000,000. Seemingly very small samples can get precise results from very large populations.

It is recommended that you specify the level of precision you need for the key estimates (statistics) and leave it to the offeror to propose a sample design that meets this specification at the lowest possible cost. If you specify both precision and sample size, the offerors may find it impossible to meet both your requirements.

To achieve the most efficient sample design, the contractor determines a sample size that:

(1) Will achieve a fixed level of precision for minimum cost; or
(2) For a fixed cost, will achieve the greatest estimation precision.

In virtually all EPA survey contracts (1) will apply. In other words, the contractor starts with a requirement to attain a given level of accuracy (precision) and needs to satisfy this requirement at minimum cost.

How many sample members are taken from where? An example of the difficulty a contractor may encounter in allocating a sample in an environmental study is the following: If a contractor has the capacity to chemically analyze 1,000 specimens of lake water, how many sample lakes and how many specimens per lake are most efficient in answering the questions?

When you draft the specifications for the survey, be sure to consult a sampling specialist to ensure that the precision levels you set are reasonable given the resources you have available.

Survey specifications should include the following items:

- The levels of precision for the key statistics, as discussed above.

- The level of geographic detail for which estimates are needed. If the target population is

the entire U.S. population, getting estimates at a specified level of precision for each State would require a sample roughly 50 times larger than that required to get estimates with the same level of precision for all 50 States combined.

- Variability of the characteristics of the target population, based on prior knowledge. The greater the differences between the units in the target population, the larger the sample needed to achieve a specified level of precision. In fact, the level of precision in sample surveys is based on sample variance, which measures the lack of homogeneity among the data collected from the sample.

- The methods used to design the sample. Survey designers use many sampling methods and combinations of methods to design a survey sample. The levels of precision for a sample of a given size will vary, depending on the sample design.

  Cluster sampling, a method of choosing a survey sample in which all the sampling units are clustered in one or more geographic areas rather than across the entire area in which the population is located, has perhaps the greatest impact on the precision of the statistics. (See section C below for more about cluster sampling.) Estimates derived from a sample of 1,000 households chosen at random from throughout the city would give considerably higher precision than those derived from a sample of only 50 households chosen from each of 20 randomly selected city blocks.

- Expected level of nonresponse. In almost all sample surveys, researchers will not succeed in obtaining responses for every unit in the sample. There are many reasons for this, which will be discussed in Chapter 5. For example, a respondent may refuse to be interviewed, or an interviewer may fail to contact an acceptable respondent, or the person designing the sample may include ineligible units in the sampling frame (such as a business that is no longer active.)

  Often, survey designers increase the sample size to compensate for the anticipated rate of nonresponse. This will reduce sampling errors, but it will not reduce the bias in the estimates that arises because eligible units provide no data or incomplete data (assuming that non-respondents differ from respondents in some way.)

- Cost and time. As indicated above, the resources the Agency has available to do the survey place constraints on the size of the sample—generally, the larger the sample, the more the survey will cost. Moreover, if there is a deadline for obtaining the results, the time it will take to collect and process the sample data also may limit the size of the sample.

C. Sampling Methods

This section briefly describes the methods most commonly used to design survey samples involving face-to-face or self-administered mail surveys of both households and establishments. Knowing something about the different methods used to construct a sample will give you a better understanding of sample designs you may have to review. To illustrate the different methods, the

City X example will continue to be used in section B.

Our focus in this section is on probability sampling methods. Probability sampling, also called random sampling, is an objective process used throughout the world. Also described here are three types of non-probability samples.

1. Probability Sampling Methods

Probability samples are those in which the members of the population (the sampling units) are selected at random—solely by chance. "Random" is not equivalent to "haphazard." A true random selection is independent of human judgment. The two distinctive features of probability sampling are:

- The use of some random device (such as a table of random numbers) to determine which units in the population are included in the sample. This prevents the person designing the sample from biasing the selection (consciously or unconsciously) towards a sample that will produce some desired result.

- The sample can be used to make estimates of the sampling errors associated with the survey findings. Hence, anyone using the survey data can determine how accurate the data are and how much confidence to place in any conclusions based on the sample data.

Let's look at six of the most common methods of probability sampling used today:

(a) Simple Random Sampling

In simple random sampling, each unit in the target population has an equal chance of being selected. Simple random sampling is particularly appropriate for small studies where the sampling units are approximately the same size or importance, or if there is no measure of size available. A study of hospital medical records to review diagnoses of pesticide poisoning is a situation where simple random sampling may be appropriate. Simple random sampling is seldom used by itself in Agency surveys, but it is frequently used in combination with one or more of the other sampling methods described in this section.

Let's see how a simple random sample of 500 from the 5,000 households in City X would be drawn. First, it would need to prepare a list of all 5,000 households. The list might be obtained from property tax records, by canvassing the area, or from some other means. It would then list all the households by address, and number them in sequence from 1 to 5,000.

To begin the selection of the sample, you would pick a random number between 1 and 5,000—254 for example. The household with that number would be the first unit included in the sample because 254 is less than 5000. You would continue to randomly select numbers until the desired number of sample units had been chosen.

What if the same random number comes up more than once? Usually, numbers that have already been picked are set aside so that no number (254 in this example) shows up more than once. This

is known as simple random sampling without replacement—a number, once selected, is not returned to the sampling frame. [18]

(b) Stratified Sampling

It is often useful to divide the population into subgroups for sampling purposes. If you propose to sample from every subgroup, then the subgroups are termed strata. In stratified sampling, the population is divided into two or more strata, and the sample is selected separately from each subgroup or stratum.

Stratification does not imply any departure from probability selection. It only means that before any units are selected, the population is divided into two or more strata. Then a random sample is selected within each stratum.

Continuing with our example, suppose there is reason to suspect that contamination is more likely to occur in some parts of City X than in others. If so, geographic stratification could be used to select the survey sample. Separate samples could be drawn from each of the city's seven wards. This would ensure the selection of some sampling units in each ward, whereas if it were not stratified, the sample could—purely by chance—be heavily concentrated in one or two wards.

How should the overall sample be allocated among the strata, or wards? If there was no clue as to the likelihood of contamination in different strata, some sampling fraction, say 1 in 10, would probably be used in each of the wards. This is called proportional stratified sampling because the distribution of the sample households in each ward would be proportional to the number of households in each ward.

It is not necessary to use the same sampling fraction in each stratum. If information indicated that the drinking water contamination problem was much more serious in three of the seven wards, a higher rate in those three wards could be sampled.

The primary reason for using stratified sampling is to make the sample more efficient – to produce estimates with smaller sampling errors. How well this objective is met depends on the criteria used to define the strata.

(c) Cluster Sampling

In cluster sampling, groups or "clusters", of nearby units in the population are formed and a random sample of the clusters is selected. In other words, within a particular stratum, rather than selecting individual units without regard to where they are located, whole clusters of units are selected.

To illustrate cluster sampling, one way of selecting a probability sample of households in City X

[18] Note that sampling with rep1acement, where the numbers are returned to the frame, is sometimes used for probability samples, including simple random sampling.

would be to first select a sample of city blocks at random and then pick a sample of some or all of the households living in those blocks. If City X has a total of 100 blocks, you might use simple random sampling to choose 10 blocks and then interview some or all the households in only these 10 blocks.

Estimates derived from a cluster sample are likely to have considerably larger sampling errors than estimates from a simple random sample of the same size. The reason is that adjacent sampling units tend to have similar characteristics. This similarity, or correlation, reduces precision by producing a degree of redundancy in the data collected from members of the same cluster.

Why then use cluster sampling? First, there is a considerable savings of time and expense in compiling a frame that lists only the units in the sampled clusters rather than all the units in the population. Second, if face-to-face interviews will be used to collect the data, by concentrating them in a smaller geographic area, the overall cost savings can be enormous—especially in a national sample.[19] Thus, cluster sampling is usually used in the relatively few face-to-face surveys that are still conducted.

(d) Systematic Sampling

In systematic sampling, researchers first list the sampling units (which may or may not be individual members of the population) in some specific order. Then, they select units for the sample by computing an appropriate sampling interval (I) and taking every I^{th} unit in the sampling frame. The starting point is chosen at random from the first I units; this is called a random start.

To select a systematic sample of 500 households in City X from the 5,000 households in the frame, you might use a sampling interval of 10 (5,000 divided by 500) and a random start between 1 and 10 (I). For example, if our random start were 7, the households included in the sample would be those numbered 7, 17, 27, and so on, up to the household with the number 4,997.

Systematic sampling is widely used in survey research, especially in combination with other methods. It has two main advantages—

- Only one random number need be picked during the selection process, rather than one for each unit needed to complete the sample.

- If the sampling units are listed in some meaningful order—for example, by block in City X—the effect of using systematic sampling is essentially the same as using stratified sampling—certain i.e., certain types of units are assured adequate representation in the sample.

Another version of systematic sampling is based on the ending digits of identification numbers.

[19] This is also true for some telephone and mail surveys where non-respondents are to be followed up by personal visits if necessary.

In this method, the last digit of a set of serial numbers that constitute the sampling frame is chosen at random, and all the units in the frame with ID numbers ending in those digits are included in the sample.

For example, suppose the Social Security Number (SSN) of the head of each household in City X was listed. you could select the 1-in-10 sample by including all households with SSNs ending in "4." This method would yield a sample of approximately 500 households, although the exact size would depend on which ending digit was chosen as the random start.

Caution should be used in selecting a series of ID numbers for sampling purposes, because they are not always assigned randomly. Social Security numbers frequently are used for sampling human populations based on ending digits, and these should be suitable because the ending digits are assigned randomly. By contrast, for business surveys, IRS employer identification numbers (EINs) may not be appropriate because EINs were initially issued in a non-random way.

(e) Sampling with Probability Proportional to Size

Up to now, all the methods described have involved sample designs where every member of the population, or at least the stratum, has an equal chance of being chosen as part of the sample. However, in some sample designs, not all the sampling units have the same selection probability. If the population characteristics in which the researchers are interested are related to the size of the sampling unit, and it is possible to obtain some measure of the size of the units, greater precision usually can be achieved by giving larger units a greater probability of selection. This is sampling with probability proportional to size (PPS).

For example, in sampling the U.S. population, researchers typically select Metropolitan Statistical Areas (MSAs), counties, or other sampling units with probability proportional to the number of individuals residing there. In a soil study, counties may be selected with probability proportional to the crop acreage as the size measure. Or, for a study of rivers, hydrologic units may be selected with probability proportional to the miles of river they contain.

To illustrate, suppose a sample of 10 of the 100 blocks in City X was selected. You could simply select 10 blocks with equal probability using either a simple random sample or a systematic sample. However, if a count of the number of households in each city block was done (from a recent census, a local telephone directory, or some other source), and the blocks varied quite a bit in size (number of households), a more efficient sample design might result if the more populous blocks had a greater chance of selection. ("More efficient" sample design means one in which the statistics will have smaller sampling error.

The selection procedure would be as follows:

(1) First, you would list all 100 blocks in some order, and alongside each block, list the count (the number of households residing there) and the cumulative total of these households, as in the table below.

(2) Then, the total number of households in City X (5,000) would be divided by the number

of blocks to be chosen -- 10 in this case. The result, 500, is the sampling interval that would be used for selection purposes.

(3) Next, you would select a random start number between 1 and the sampling interval (500), for example, 213. You would then form a series of sample-selection numbers by beginning with the random start and adding the interval as many times as needed, i.e., 213, 713, 1213, 1713, . . . 4713.

(4) Finally, for each sample selection number (e.g., 213 or 713) You would choose the first block whose cumulative total equals or exceeds that number until 500 units are chosen for the sample. The table below shows how the first 4 blocks were selected, e.g., blocks 2, 6, 9, and 10. If there is a selection of the same block more than once, the block can be divided appropriately.

Block	Households		Sample Selection	
Number	In Block	Cumulative	Number	Selected
1	120	120		
2	220	340	213	✓
3	50	390		
4	170	560		
5	90	650		
6	130	780	713	✓
7	310	1090		
8	40	1130		
9	300	1430	1213	✓
10	600	2030	1713	✓
11	150	2180		

PPS sampling is especially applicable for selecting the first-stage units of a multi-stage design (discussed next.) To use PPS sampling, it is necessary to have measures of size for all the units in the target population or frame, e.g., counts of households by block in City X. The measures of size need not be exact; it is sufficient for them to be reasonably close to, or correlated with, their actual sizes.

(f) Multi-Stage Sampling

Previously discussed was a sampling method called "cluster sampling," where groups of units rather than individual units are used to form the sample. Multi-stage sampling refers to the process of selecting subgroups within the clusters chosen at a previous stage. In fact, all multi-stage designs are cluster samples. For practical purposes, virtually all large Agency-sponsored surveys use some form of multi-stage sample selection. Multi-stage designs are essential for any face-to-face survey of a widely dispersed sample.

Continuing with the City X example, suppose you did not have a current listing of the 5,000 households in the city. You might decide to use a multi-stage design to select the sample. Here is how a two-stage sample design could work. In the first stage, you might select a sample of blocks using probability proportional to size, as discussed above, based on approximate block

counts from the best available source such as the latest Census. Next you would prepare lists of all the households in the sample blocks. Then, by simple random sampling or systematic sampling, you would select a sample of households from the list of households residing in each of the blocks selected in the first stage.
The most important advantages of multi-stage sampling are:

(1) Researchers can concentrate on a smaller number of areas, with a consequent reduction in time, staff, and dollars.

(2) Researchers need only obtain listings of the sampling units chosen at the previous stage, rather than a complete list of the population. In the above example, lists need to be created only for the households in the blocks selected in the first stage, instead of all 5,000 households in City X.

(i) Multi-Stage Sample—Household Survey

Most multi-stage samples involve four or five stages of selection. An example of a household survey of this type is the Panel Study of Income Dynamics, a longitudinal survey conducted by the University of Michigan's Survey Research Center. The stages of selection are:

Stage 1: Selection of "primary areas," usually counties or groups of adjacent counties such as Metropolitan Statistical Areas. In the Survey Research Center's design, 74 primary areas consisting of individual counties were selected; these are also known as "Primary Sampling Units" or PSUs.

Stage 2: Selection of "sample locations" (cities, towns, and rural areas) within primary areas.

Stage 3: Selection of "chunks" (areas such as city blocks or rural townships, each containing from 16 to 40 housing units) from each sample location.

Stage 4: Selection of "segments," of 4 to 16 housing units, in each sample chunk.

Stage 5: Selection of "housing units" from the sample segments.

(ii) Multi-Stage Sample—EPA Establishment Survey

EPA's 1990 National Pesticide Survey is a good example of what was essentially an establishment survey, where the subject being interviewed was a "thing" rather than a "person." This survey used a very complex design, summarized (and considerably simplified) as follows:

Stage 1: Selection of "primary" areas. All U.S. counties considered to be rural, were selected with probability proportional to: (1) the estimated number of domestic wells in each (2) using a 12-stratum classification based on the amount of pesticide use (4 levels) times the vulnerability of each county to groundwater contamination (based on auxiliary information from the U.S. Department of Agriculture and elsewhere). This "PPS" scheme assured that counties most likely to have contaminated wells were the most likely to be in the sample.

Stage 2: The second stage used small Census Bureau geography (now called Block Groups), again classified by the 12-stratum scheme used for Stage 1, and further classified by whether certain crops were grown in the area (based on USDA Extension Agent reports for the 90 counties selected in the first stage.)

For each area chosen at that stage, data from the 1980 Census on the number of housing units using wells for drinking water, and direct observation to update these numbers to current levels, were combined to choose a total of 500 clusters with the greatest number of domestic wells.

Stage 3: Each household selected in Stage 2 was interviewed to determine the number and location of all domestic wells in each property. Finally, a fixed number of domestic wells per second stage unit was sampled with equal probability, and data collected from each one.

The data collected at each sampled site included: (a) interviews of household members; (b) separate interviews of farm operators found in the sample; (c) water sample collection from each well; and (d) local area characteristics. The water sample collection used great care to identify and avoid contaminating each sample, and all sample containers were shipped to EPA for detection and analysis several dozen different contaminants.

Our discussion of probability sampling methods has merely scratched the surface of the techniques survey statisticians use to construct samples, and the ways they apply them to investigate various populations. Frequently, complex combinations of the methods described are used, along with variations such as double or sequential sampling, replicated sampling, and controlled selection.

There are several references at the end of this chapter that will help you expand your knowledge of probability sampling methods.

2. Non-Probability Sampling Methods

Non-probability sampling methods are characterized by a subjective selection procedure. Unlike probability sampling, the choice of the sample members is not random, but, consciously or unconsciously, is influenced by human choice—usually by expert judgment.

Nonprobability Samples:

- Convenience or haphazard samples
- Judgment or purposive samples
- Quota samples

The problem with all non-random selection schemes is that even the most conscientious individuals make unconscious errors of judgment that may be considerable. These errors, which are very difficult to measure, are called “biases.”

Because non-probability samples do have applications in some environmental research situations, several types will be briefly examined. Non-probability samples are also sometimes used in the final stage of selection of some environmental studies where strict probability sampling is not feasible, such as obtaining specimens for chemical analysis (house dust from a sample household, or water specimens from a segment of a sampled stream.) They also are sometimes suitable for small-scale qualitative exploratory studies, and for pretests or pilot tests of EPA-sponsored surveys where the intent is to use probability methods to select the sample for the survey itself.

Note that when non-random methods are used to select pre-test or pilot-test samples, the choice should not be restricted to "easy-to-get" units. If pretest samples include only units for which it is easy to collect information, it will be difficult to anticipate the kinds of problems that may occur in the main survey and how much the survey itself is likely to cost in time and dollars. When using non-random samples to obtain a set of answer choices to convert open-ended pretest questions to closed-ended final questions, "easy-to-get" units may give very different answers than the total sample, thus throwing off your answer choices.

In any research situation where non-probability sampling is used, keep in mind that the results only pertain to the sample itself, and should not be used to make quantitative statements about any population – including the population from which the sample was selected.

Now, let us look at the most common non-probability samples.

(a) Convenience or Haphazard Samples

Convenience or haphazard samples are samples selected from populations for which it is relatively easy to collect information on a particular topic. Another feature of these samples is that the population groups from which they are selected do not reflect, with any measurable degree of error, the characteristics of some larger, well-defined group of which they are a part.

The following are examples of convenience samples of human populations—

- Voters interviewed in a shopping center;
- Volunteer subjects for experiments (e.g., households responding to a radio or newspaper appeal for volunteers to try out a new kind of water purification equipment in their homes);
- People answering a reader opinion questionnaire;
- People writing to their representatives or senators about a particular issue.

No matter how many choose to respond, these "surveys" almost invariably are seriously biased—they represent nobody except those who choose to respond.

(b) Judgment or Purposive Samples

These are samples that an investigator or another subject-matter expert considers to be "representative" of some study population. Like convenience samples, judgment samples are often used by EPA for pre-testing purposes. For example, to pretest a survey of chemical plants that manufacture sulfuric acid, an expert researcher in the field might arbitrarily choose for preliminary investigation a few plants where all the manufacturing processes commonly used in the industry are represented.

Judgment sampling is most usefully applied to early, exploratory phases of research involving extremely small samples. In environmental studies, judgment sampling and probability sampling are sometimes combined in a multi-stage sample, the final stage being a judgment sample.

There is nothing inherently wrong with well-conducted judgment-sample surveys, as long as their limitations are recognized.

(c) Quota Samples

In some national surveys, investigators use probability sampling to choose the first one or two stages of a sample, and use quota sampling for subsequent stages. Therefore, quota sampling is a version of stratified sampling in which the selection within strata is non-random.

Quota samples are frequently used in marketing and opinion research. For example, in an opinion survey, the interviewers will each be given a quota of interviews to conduct with various classes of individuals, households, businesses, etc. An interviewer's quota might consist of a specified number of individuals in each of a set number of age-sex categories. Within these categories, and in the assigned area, the interviewer is free to decide how to locate and interview the specified number of individuals.

However, since the selection process is subject to human judgment, there is no guarantee that biases will not occur. For example, an interviewer may fill his or her quota in the top age group mainly with people 65 or 66, thus under-representing the very old.

Quota sampling has two main advantages:

- It is less costly than random sampling—perhaps one-third as much; and
- There is no need to develop a frame for selecting respondents in the sampled area, which means that callbacks are avoided. If an eligible respondent is not available at a dwelling when the interviewer calls, the interviewer simply proceeds to the next dwelling.

As with all other non-probability samples, the non-randomness in the selection of the sampling units is the main disadvantage of quota sampling. Thus, it is impossible to estimate the sampling variability from the sample and to know the possible biases, which may be sizeable.

Even the best-designed probability sample can degenerate into a seriously-biased convenience

sample if sample members who are "hard to get" are simply ignored. Following up on non-respondents is absolutely essential to the conduct of a successful survey.

D. Major Components of a Sampling Plan

The starting point for developing a sampling plan is the development of the five minimum survey design specifications that are recommended for all Agency surveys. These design specifications, which the sponsoring office should clearly define in the survey specifications, are: (a) the research objectives; (b) the target population and coverage; (c) the required level of precision (sampling error); (d) the target response rate; and finally, (e) the use of probability sampling throughout the selection process, whenever feasible.

Components of a Sampling Plan:

1. Sampling frames
2. Sample selection
3. Estimation procedures and weighting
4. Sample error calculations

If a contractor is conducting the survey, their technical proposal will usually include a preliminary sampling plan.

1. Sampling Frames

A sampling frame is a listing of population elements—geographic areas, manufacturing plants, crop acreage, telephone numbers, city blocks, households, factories, etc.—from which the survey sample is drawn. The frame is the most important component of the overall sample design because it identifies the population elements from which the sample is chosen. The population elements listed on the frame are called the sampling units. Often these are groups or clusters of units rather than individual units.

The choice of sampling frames, and the steps taken to assure their completeness and accuracy, affects every aspect of the sample design. Ideally, a sampling frame should:

- Fully cover the target population;
- Contain no duplication;
- Contain no "foreign" elements (elements that are not members of the population);
- Contain information for identifying and contacting the units selected for the sample; and
- Contain other information that will improve the efficiency of the sample design and the estimation procedures.

If the sample design calls for a multi-stage selection, a separate frame is prepared for each stage (or stratum) of the sample design. For example:

- In the two-stage sample design for City X that was used earlier to illustrate multi-stage sampling, the frame for the first stage would be a listing of the blocks in City X. The frame for the second stage would be listings of all the households living in each block selected for the sample.

- In a survey of plants manufacturing sulfuric acid, the sampling frame of the first stage might consist of a list of all U.S. chemical companies that manufacture sulfuric acid at one or more of their plants. After selecting a sample of these companies, you could make a list of all the sulfuric acid plants belonging to the companies chosen at the first stage. This list would serve as the frame for the second stage of selection.

Developing the frame can be a major undertaking involving substantial effort and expense. Complete, current frames do not always exist. Many frames have missing units and some frames contain duplicate listings. Both of these frame imperfections cause biases if they are not detected before the selection is done.

Illustrating several of these points, a city telephone directory is a poor frame for a telephone survey of all local households. Studies show that as many as 30 percent of U.S. households have unlisted numbers or no telephones.[20] Using the telephone directory, therefore, would result in undercoverage of the population. Moreover, some households would be over-represented because they have more than one listed number. Finally, most directories also include business and other nonresidential numbers, some of which are hard to distinguish from residential numbers.[21]

For surveys of businesses, it is especially difficult to obtain complete and current lists. Probably the best lists are those maintained for Federal programs like Social Security, income taxes, unemployment insurance, and the economic censuses. Unfortunately, these lists generally are not available to EPA and other Federal agencies, so other sources should be utilized—commercial business lists or lists that EPA maintains of organizations that are required to comply with certain Agency regulatory requirements.

In general, perfect or ideal frames are seldom available. The sampling plan should always specify what steps the contractor will take to evaluate the frames and deal with any deficiencies such as missing or inaccurate elements.

2. Sample Selection Procedures

The sampling plan provides complete specifications for procedures to be used for selecting units from the frame at each stage of sampling.

[20] This includes households that have not yet appeared in any directory because they have recently moved in. Unlisted and not-yet-listed numbers can approach 50 percent in some large metropolitan areas. In poor, rural areas up to 30 percent of households may not have phones.

[21] Commercially available lists of business phone numbers can help locate residential numbers, but these lists are not always accurate and ignore the possibility of combined business and residential use of a telephone number.

Most sampling is done at a central location, usually the contractor's main office. However, for some of the later stages of sampling, the selection may be done in the field. For example, in a face-to-face survey, the field supervisors may select sample housing units from block or segment listings prepared by the main office. Similarly, in a mail survey, if the contractor intends to conduct follow-up interviews with some of the people who do not send back questionnaires, procedures for selecting the follow-up sample should be described in the sampling plan.

The selection procedures in the sampling plan should specify—

- Any tasks necessary for reorganizing or otherwise refining the frame prior to selection, such as:
- Screening to eliminate units that clearly are not in the target population; and
- Transforming information about individual units into measures of size (necessary for sampling with probability proportional to size).
- Whether the selection of sampling units (at each stage) will be with equal or variable probability. If variable probability is to be used, the basis for assigning selection probabilities to individual units must be included.
- The sample sizes or intervals. If stratified sampling is used, sizes or intervals may vary by stratum. For some designs it may be necessary to obtain preliminary counts or other tabulations from the sampling frame to determine the most appropriate size or intervals.
- The specific probability mechanism to be used to select the individual sampling units or, for systematic sampling, the random starting point.
- Any steps that will be taken to screen out ineligible sampling units, obtain better addresses, etc., after the initial selection is made.

3. Estimation Procedures and Weighting

Estimation procedures are the methods used to convert sample data into estimates for the population—totals, means, proportions, and other statistics. The actual preparation of the estimates (and the calculation of sampling errors, discussed below) is done towards the end of the data processing phase, but the procedures that will be used to obtain the estimates should be included in the plan. The approach used for the estimations also plays a role in determining the size of the sample—another reason for determining the estimation procedures early in the process. In addition, some estimates require the capture of certain data when the sample is selected, during the data collection phase, or during the processing phase of the survey.

Estimation procedures:

(a) Applying weights
(b) Adjusting for nonresponse
(c) Using auxiliary information—ratio estimation

The estimation procedures should specify how the contractor proposes to derive the most precise estimates possible from the sample data using statistical techniques such as:

- Applying "weights" to give greater relative importance to some sampled elements than to others;
- Making adjustments to reduce the bias caused by eligible sampling units for which no data were collected; and
- Using auxiliary information obtained from the questionnaires, the sampling frames, or other sources such as administrative records, other surveys, etc.

What follows is a brief discussion of the three methods of enhancing data quality.

(a) Applying Weights

When analyzing complex samples, statisticians assign weights (or multipliers) to adjust for: (a) sampled elements for which the probability of selection was in some way unequal; (b) eligible units for which no data were collected (total nonresponse units); and (c) sampling units not included in the sampling frame (non-coverage errors.)

For example, if all the sampled elements had the same probability of selection (sometimes called a "self-weighting sample"), survey analysts can obtain valid estimates of some statistics such as proportions, means, percents, and medians without weighting the data obtained from the sample. However, to estimate totals for the sample, all units are weighted by the reciprocal of the sampling fraction. There are two scenarios:

1. Single Stage probability of selection. If a simple random sample of 1 in 10 housing units has been selected, population totals could be estimated by applying a weight of 10 to the data for each housing unit sampled, or by tabulating the sample data and multiplying the sample counts or aggregate by 10.

2. Multiple Stage probability of selection. If, for example, a multi-stage sample were used and a sample of 10 city blocks were selected from a total of 50 blocks, and then every tenth household in these 10 blocks were selected for interviewing, the overall selection probability for these households would be 1 in 50:

   $10/50 \times 1/10 = 1/50$

   A uniform sampling weight of 50 would then be used to estimate totals from the sample data.

(b) Adjusting for Nonresponse

The techniques used to adjust for total nonresponse (eligible members of the sample that provide no data) are usually incorporated in the estimation procedures. The techniques used to make these kinds of adjustments are:

Reweighting the sampled units by the inverse of the proportion of units that responded. For example, if 80 percent of the sample responded (0.80), a reweighting factor of 1.25 (1.00 ÷ 0.80) would be used to adjust for the nonresponse. Reweighting factors are often computed separately by stratum or for each member chosen at the first stage of selection. This allows for variations in the proportions of different categories or areas of the sample that responded.

Duplicating the values reported by the sampled units to compensate for eligible units that did not respond. Information from all sampled units can be used in selecting the units that are duplicated. For example, the units to be duplicated could be selected from the same size or industry category, or from the same geographic area, as the non-responding units.

$$\text{Ratio estimate of unemployed individuals} = \frac{\text{Unbiased estimate of unemployed individuals}}{\text{Unbiased estimate of total population}} \times \text{Independent estimate of total population}$$

These kinds of nonresponse adjustments will reduce nonresponse biases but will not eliminate them entirely. The use of nonresponse adjustments is not an acceptable substitute for diligent efforts to collect data for all eligible units in the sample.

Note that different techniques are used to adjust for missing data from single questionnaire items (these are called item nonresponses, and are discussed in Step 7 of Chapter 6.)

(c) Using Auxiliary Information—Ratio Estimation

Survey analysts often can improve sample estimates by using auxiliary information about the population, which may be taken directly from the sample (from the questionnaires, for example), from the sampling frames, or from independent sources. Auxiliary information is most often used to construct ratio estimates. Suppose, for example, that you want to estimate the number of unemployed individuals in a national household survey.

Simple unbiased estimate: One way to do this is to tabulate the unemployed people in the sample and assign them appropriate weights based on their selection probabilities, a procedure known as simple unbiased estimating.

Ratio estimate: However, suppose you have an estimate of the total population from an independent source at the time of the survey (the U.S. Census, for example). This independent estimate could be used to construct a ratio estimate of unemployed individuals as follows:

In other words, the sample data would be used to estimate the proportion of unemployed individuals and apply that figure to an independent estimate of the total population to derive a more precise estimate of the number of unemployed individuals in the population. If there were

independent estimates of the population by age and sex, you could make separate ratio estimates of the number of unemployed individuals in each age-sex group and add them up to get an estimate of the total number of unemployed individuals in the population.[22]

Several different kinds of ratio estimation procedures are available, as are other procedures that make use of auxiliary information, such as regression estimation. The choice of procedures will reflect the survey designer's judgment about how all relevant data from the sample, the sampling frames, and other sources can be used to develop the most precise survey estimates, i.e., how to make the best use of all available information.

In practice, weighting can be a complex task because a combination of adjustments is often necessary. Weights may first be assigned to adjust for unequal selection probabilities. Then, these weights may be revised to adjust for varying levels of response within the sample. Still further revisions may have to be made to adjust the sample to known distributions in the population. Therefore, the sampling plan should fully describe the estimation methods, formulas, or procedures the contractor plans to use to produce the survey estimates.

4. Sample Error Calculations

Of all aspects of sampling, calculating sampling errors is the most technically complex. Most surveys collect data on a large set of variables and produce estimates for both individual variables and their relationships to each other. It is impractical and usually impossible to calculate standard errors for all estimates. Therefore, survey analysts normally compute standard errors for only key statistics and a few selected estimates. From these calculations, they develop generalized models from which other standard errors can be inferred.

The sampling plan should specify:

- The estimates for which sampling errors will be calculated. (Standard errors should be computed for all key variables and a selection of other statistics.)

- The approach that will be used to calculate the sampling errors (formulas, methods, or software packages).

- Any assumptions or approximations implicit in the proposed approach.

The extent of sampling error depends on the design of the sample. The formula for calculating standard error found in most over-the-counter software packages is applicable only to simple random sampling with replacement designs. It will produce overestimates, or, more often, underestimates of sampling error if applied indiscriminately to other sample designs.
The sample designs for most of the surveys EPA sponsors are complex, often involving a combination of multi-stage and stratified sampling methods. For these complex designs, survey designers use a variety of approaches for calculating sampling errors such as the "Taylor

[22] This is likely to be more accurate than the simple unbiased estimate because it adds precision derived from Census information, which is more accurate than any figure derived from a relatively small survey.

expansion method," "balanced repeated replications," "jackknife repeated replications," and so forth.

In addition, several software packages have been developed recently for calculating sampling errors of estimates that are based on complex sample designs. The selection of suitable software poses difficulties because most packages treat the sampling units chosen at the first stage as being sampled with replacement when, in fact, this is rarely the case. An exception being SUDAAN which has the WOR (without replacement) function available; however, most packages limited capacity for a finite population correction or can be assumed to be with replacement.[23] In any case, calculation of sampling errors is best left to a contractor.

E. Monitoring Sampling Activities

The sponsoring office's greatest impact on the development and faithful execution of a sound sampling plan occurs in the design stage of the survey. Therefore it is suggested that you, as project officer, do the following **before** the contract is awarded—

- Specify in the survey specifications what should be included in the sampling plan. The main components of a sampling plan—selection and development of the sampling frame, sample selection procedures, estimation procedures, and procedures for calculating sampling errors—are discussed in section D of this chapter.

- Make sure the technical evaluation panel reviewing the responses includes someone qualified to evaluate the sampling plan. Expertise in survey sampling theory is necessary to spot defects such as:

  - Any (unnecessary) departures from probability sampling;

  - Imprecise descriptions of the sample selection procedures;

  - Sample sizes or sampling allocation rates that will not achieve the levels of precision;

  - Incorrect estimation formulas or methods; and

  - Inappropriate formulas or methods for calculating sampling errors.

After contract award, you should monitor the execution of the sampling plan:

- Be sure the contractor tests the validity of the sampling frames before starting to select the sample for the survey itself. Missing and duplicate sampling units can cause difficulty if they are not detected.[24] Frame counts, broken down by geographic area and other characteristics, should be checked against information about the population that may be available from other sources. For example, the accuracy of totals for various

[23] See Step 8 in Chapter 6 for more information on the application of these approaches to the calculation of sampling errors after the data are processed.
[24] Minor misspellings can mask duplications.

kinds of industrial establishments may be crosschecked with the most recent economic census. Sometimes, especially when using commercial business lists, it may be desirable to contact a small sample of the units in the frame to determine what proportion are currently active members of the population, and to check the accuracy of names, addresses, and other identifying information. While the contractor normally will perform the validity tests, the results should be fully documented for Agency review.

- Compare sample selection procedures in the work plan with the results of the sample selection operations actually carried out at each stage of the survey itself.

- If any sampling is to be done in the field, the contractor should pre-test the selection procedures and provide counts of the number of units selected at each stage, broken down by categories for which frame information is available. Agency experts, or the contractor, should check these counts against the anticipated sample sizes. Frame totals can be checked by (a) applying appropriate sampling weights to the sample counts, and then (b) using tolerances based on estimated sampling errors, comparing them with actual frame totals. Make sure these checks are made before giving the contractor permission to start collecting data for the main survey.

- Review the specifications for preparing the sample estimates. Later, when the contractor has completed the preliminary tabulations, check the key statistics against (a) data from prior surveys or other sources and (b) known totals from the sampling frames that were used. (For further details, see ”Preparation of the Outputs” in section A of Chapter 6.)

- Review the specifications for calculating sampling errors. Check the actual estimates of sampling errors for plausibility as soon as they are available. An easy way is to compare them with the sampling errors that would have been obtained if a simple random sample had been used. The ratios of the contractor’s estimates to the corresponding values of the sampling errors for the simple random sample generally range from slightly less than 1 to about 2 or 3, depending on the sample design used. If all the ratios are much larger or smaller, there is likely to be a programming error or an error in the estimation formula (or method).

  Another method is to plot the estimated sampling errors against the corresponding estimates obtained from the sample data (totals, percents, means, etc.). The values usually will follow a fairly regular pattern, with larger sampling errors normally associated with smaller totals or percentages close to 50%. Any extreme values may indicate processing errors for the items in question. If the plotted values for a particular class of estimates do follow a regular pattern, a curve can be fitted to these calculated values. This curve can be used to estimate sampling errors of items for which sampling errors were not actually calculated.

Bibliography: Chapter 4

Cochran, W.G., Sampling Techniques, New York, John Willey & Sons, 1963.

Environmental Protection Agency. National Pesticide Survey: Summary Results of EPA’s National Survey of Pesticides in Drinking Water Wells. Washington, D.C.: United States Environmental Protection Agency, Office of Water, Office of Pesticides and Toxic Substances, 1990.

Kalton, Graham. Introduction to Survey Sampling; Quantitative Applications in the Social Sciences. Beverly Hills, Sage Publications, 1983.

Satin, A. and W. Shastry. Survey Sampling: A Non-Mathematical Guide. 2nd Edition, Ottawa, Statistics Canada, 1993.

U. S. Environmental Protection Agency, Office of Environmental Information. Guidance on Choosing a Sampling Design for Environmental Data Collection, for Use in Developing a Quality Assurance Project Plan, (EPA QA/G-5S) Washington DC, EPA, 2002.

Chapter 5: Interviewing

A survey interview is a conversation between an interviewer and a respondent for the purpose of obtaining certain information from the respondent. Coupled with a well-designed, well-tested questionnaire, personal interviews are a powerful, indispensable survey research tool. Whether conducted at the respondent's home or place of business, or over the telephone in a centralized, supervised environment, interviews have been used effectively to collect survey data for more than 50 years. They are especially appropriate for sounding out people's opinions, future intentions, feelings, attitudes, and reasons for behavior, and are adaptable to a wide variety of research situations.

Most surveys no longer rely on traditional face-to-face interviews using paper questionnaires; data collection is primarily done by mail, telephone, or some form of computer-driven interviewing. Nevertheless, personal interviews are still commonly used for follow-ups of hard-to-find respondents, or for clarifying answers to specific questions, so knowing the principles of good interviewing techniques is important.

This chapter looks at:

A. Quality-assurance procedures

B. Organizing and staffing field operations

C. Conducting the interviews

D. Monitoring the interviewing process—role of sponsoring office

Our emphasis throughout this chapter is on face-to-face surveys. However, much of the text is relevant to telephone interviewing, and to the extent that interviews are used for follow-up or quality control purposes, to mail surveys as well.

A. Quality-Assurance Procedures

It is vital for the contractor to establish a set of procedures to assure the quality of the work done throughout the data collection phase. The quality-assurance procedures should cover:

1. Respondent rules—who is to be interviewed at each sampling unit;

2. Follow-up procedures—how much effort the interviewers should exert to secure an interview.

3. Quality control strategies—The strategies that are to be used to ensure the collection of high-quality data. These are intended to reduce data errors for which interviewers are primarily responsible

The respondent rules, follow-up procedures, and quality control strategies should be incorporated into the work plan and approved by the sponsoring office before any data for the

main survey are collected. They should be revised as necessary following any pretests or pilot tests. The contractor should highlight these procedures and strategies in all training programs and instructional materials prepared for the interviewers, supervisors, and support staff. The three types of quality-assurance procedures are explained in greater detail below.

1. Respondent Rules

Respondent rules specify which individual or individuals are eligible, acceptable, or most desirable as respondents for each unit of observation. These rules also specify whether the respondents are to be interviewed alone, or with other respondents at the same unit, and whether individuals who are not respondents may be present.

How stringent or flexible the respondent rules should be depends on the questions to be asked and the conditions under which the interviews are to be conducted. Obviously, the more inflexible the respondent rules, the more "call-backs" the interviewers will have to make to reach the designated respondents. Conversely, the more flexible the rules, the higher the interviewers' completion rates will be.

Respondent rules usually include eligibility criteria such as age (in household surveys) and title or type of responsibility (in business surveys.) Sometimes the rules designate only one person in the sampling unit as an acceptable respondent; this may be the head of the household, the board chairperson, or the supervisor of public works. In other cases, anyone who meets the eligibility criteria may be designated as the respondent. For some surveys, the interviewers may be required to talk with several individuals at each unit (all responsible adults, for example), with each respondent supplying answers to different parts of the questionnaire. In other surveys, a particular type of respondent may be identified as the "most desirable" respondent, but the interviewer may be allowed to interview any other responsible adult if this person is not available.

Respondent rules also specify whether interviewers may talk with an alternate respondent—a "proxy"—after they have made a certain number of unsuccessful attempts to interview the designated respondent. However, using proxies may produce a marked deterioration in data quality. Usually, some information about the units of observation is best supplied by one particular person (the head-of-household or the plant manager, for example). If data are obtained from someone other than the designated respondents, there are likely to be serious gaps, inaccuracies, and biases in the information the interviewer gets. Nevertheless, if it is imperative to obtain some information about the unit of observation, the rules may allow the interviewer to collect data from neighbors, co-workers, or others if the designated respondents cannot be reached.

2. Follow-up Rules

Follow-up rules prescribe the amount of effort to complete an interview with the designated respondent(s) for each sampling unit. Follow-up rules should specify:

- The number of attempts to secure an interview from a single unit or a cluster of units;

- The time of day the interviewers are to make the initial and subsequent visits (or attempts, in the case of a phone survey); and

- Any allowable deviations from these rules (for example, to hold down costs, the interviewer may make fewer personal visits to units in sparsely populated areas.)

For a particular survey, the stringency of the follow up rules will depend on (a) how vital the researchers believe it is to obtain information directly from the designated respondents rather than from proxies; (b) the survey budget (call-backs are costly); (c) how soon the data are needed (inflexible follow-up rules may unnecessarily delay the project); (d) the characteristics of the target population (some types of respondents are difficult to reach during the day); and (e) the characteristics of the areas to be surveyed (for example, widely dispersed units, inner-city neighborhoods).

3. Quality Control

Guarding against missing and inaccurate data is a major objective in any survey. Strategies should be developed to control three principal types of non-sampling errors that occur during the data collection phase, all of which can seriously compromise the results:

(a) Coverage errors, which result from interviewing ineligible units or failing to interview eligible units;

(b) Nonresponse errors, which result when no data or incomplete data are obtained from eligible units; and

(c) Response errors, which are incorrect reports by the interviewer or the respondent, whether inadvertent or deliberate.

Our concern here is with the effects that interviewing may have on the quality of the data collected in a survey. While errors that result from the use of sampling can be measured and included in survey reports, non-sampling errors are much more difficult to measure, and therefore they can seriously compromise the survey results.

Non-sampling errors can occur in any survey, regardless of the collection method. Moreover, they do not result solely from poor interviewing. For example, some coverage errors may be directly attributable to the use of incomplete frames, and some nonresponse and response errors may be the result of poor questionnaire design. In a mail survey where no follow-up interviewing is done, they may be directly attributable to the questionnaire.

However, poor performance by the interviewers or ineffective interaction with respondents can serious influence the quality of the raw data the interviewers collect, and hence affect the validity of the results.

If the interviewers do not adhere to the respondent rules and follow-up procedures, and do not properly administer the questionnaire, the number of non-sampling errors is likely to be very

large. Many of these errors may be "systematic" errors, which no increase in sample size can reduce or eliminate.

Let's examine the sources of (a) coverage errors, (b) nonresponse errors, (c) response errors, and finally, (d) the main quality control strategies survey researchers have developed to reduce these errors during the interviewing.

(a) Coverage Errors

The main sources of coverage errors in an interview survey are poorly constructed or outdated sampling frames. For example, the interviewers may be given incorrect listings of the households or businesses they are to cover, so some of the units they attempt to contact are unacceptable, non-existent, or otherwise ineligible. These errors cannot be attributed to the interviewers.

In some cases, however, the interviewers may be responsible for coverage errors. They may interview the wrong unit by mistake—because the street number is not clearly marked on the house, for instance. They may even make up the answers to a questionnaire for a hard-to-reach unit, instead of obtaining data from the designated respondent in that unit.[25]

(b) Nonresponse Errors

Nonresponse errors occur, as previously mentioned earlier, when the interviewer gets no data ("total nonresponse") or incomplete data for an item ("item nonresponse") from an eligible sampling unit. Let's look at the sources of these two kinds of nonresponse errors.

(i) Total nonresponse

Total nonresponse occurs when an interviewer does not obtain any data (or less than the minimum amount required to count as a completed interview) from a sample unit that is eligible for an interview.

Frequently, not all sample units assigned to interviewers are eligible for interviewing. In a household survey, for example, units that turn out to be vacant or demolished are ineligible and will not be treated as nonresponse cases. On the other hand, where interviews are not obtained for eligible units because of refusals or inability to contact designated respondents, the units will be counted as nonresponse cases.

It is important that the contract specify in some detail what kinds of units should be defined as ineligible for interview. For example, should households with no English-speaking members be considered ineligible? What about households where all of the eligible respondents are deaf, senile, or otherwise in no condition to be interviewed. These points should be clearly spelled out in the survey contract to avoid later disputes about whether the contractor has achieved the target response rate set in the contract. You will recall that a response rate lower than 75 percent usually is unacceptable for an Agency sponsored survey.

[25] This is called "curb-stoning."

Experienced, well-trained interviewers can do much to minimize the number of nonresponses for eligible units. (See "Locating Respondents" and “Securing Interviews" in section C.) Keep in mind that whatever probability sampling method the contractor uses, every member of the sample has to be accounted for if the statistics are to reflect the target population. Therefore, the interviewers should try to complete interviews with all the units or individuals in the sample assigned to them in accordance with respondent rules and follow-up procedures established for the survey.

(ii) Partial Nonresponse

In addition to total nonresponse, a partial nonresponse can occur. Cases are classified as partial nonresponse if the interviewer fails to obtain acceptable responses to one or more questions but does obtain enough data so the unit need not be counted as a total nonresponse.

The definition of “partial nonresponse” should be included in the contract. This classification is normally assigned to units where responses are missing for any specified questions or more than a certain number of other items.

(iii) Item nonresponse

“Item nonresponse” occurs whenever the interviewer fails to obtain data for a single item on the questionnaire. Either the respondent or the interviewer may be at fault. For example:

- The respondent remains silent or refuses to answer the question;
- The respondent gives an irrelevant answer; or

The interviewer fails to ask one of the questions or skips to the wrong question, which in either case results in a missing reply.

Interviewers are trained to handle the first two kinds of item nonresponse with techniques such as pausing briefly to give the respondent time to answer, using words of encouragement to elicit a reply or a more complete reply, repeating questions, probing adequately, and reading questions exactly as they are worded. (See “Asking Questions” in section C for more information.)

(c) Response Errors

Either the respondent or the interviewer may cause response errors. For example:

- Respondents may give inaccurate replies when they do not understand a question and are reluctant to ask the interviewer to repeat or explain it. Or the respondents simply may not know the answer and, rather than appear uninformed or stupid, will give a false reply. Or respondents may deliberately give inaccurate replies to questions they consider overly sensitive. For example, a 51-year-old man may under-report his age as 47, or overstate his income to impress the interviewer.

- Interviewers may misrecord a respondent's reply (for example, the same respondent truthfully states his age as 51 but the interviewer carelessly records it as 41.) Or interviewers may misread a question, not probe sufficiently when a respondent seems confused or tentative, or skip certain questions altogether in the belief they will be able to fill in the answers themselves later when they edit the questionnaire.

Although it was stated earlier that the respondent or the interviewer causes response errors, the ultimate cause is actually the interaction of the two. Other sources contributing to response errors that are not entirely independent of the interviewing process are: the conditions of the interview, such as the form, content, and wording of the questionnaire; the training and instructions given to the interviewer; and the location of the interview.

To minimize response errors, the interviewers can (a) make an effort to establish a good interaction with the respondent, (b) be faithful to the questionnaire, and (c) maintain an open, neutral position on the questionnaire topics. (See "Asking Questions" and "Recording and Editing the Responses" in section C for details.)

(d) Quality-Control Strategies

Survey researchers have developed numerous quality control strategies to detect and eliminate or reduce non-sampling errors for which interviewers are primarily responsible. The principal strategies used during the data collection phase to control "interviewer effects" are:

Quality-Control Strategies:

(i) Monitoring interviewer completion rates
(ii) Observing interviews
(iii) Screening completed questionnaires
(iv) Validating interviews
(v) Reinterviews

Each of these strategies serves a different purpose. Resources permitting, all five should be used in every Agency sponsored survey where interviewing is the primary collection method. In the work plan, the Agency should require the contractor to specify: (a) the quality-control strategies that will be used, (b) what each strategy is expected to accomplish, (c) how it will be applied and when, and (d) what procedures will be used to make sure it is implemented properly.

Let's look briefly at how the five quality-control strategies listed above typically are used to detect and reduce coverage, nonresponse, and response errors while the interviewing is going on.[26]

[26] Note that in some surveys, quality-evaluation strategies may be used at the end of the survey in an attempt to measure the extent of the non-sampling errors. However, these additional measures are beyond the scope of this Handbook.

(i) Monitoring interviewer completion rates

Often a small proportion of interviewers are responsible for a disproportionate share of nonresponse errors in a survey. To help supervisors track the number of errors each interviewer makes, the interviewers are required to record the specific outcome of each call. For example, to report a (total) nonresponse for any unit, interviewers record exactly why they were unable to secure an interview. If a unit is found to be ineligible for interview, the reason should be given.

Interviewers are usually required to prepare a weekly summary of their work, showing the number of assigned cases in four categories: (1) eligible, interview completed; (2) eligible, nonresponse; (3) ineligible; and (4) pending. Further breakdowns of nonresponse and ineligible cases, by reason, are often required. Alternatively, these reports may be prepared by supervisors or office clerks, based on the questionnaires turned in by the interviewers.

In either case, supervisors to monitor the quality and quantity of each interviewer's work should use these weekly reports. A key indicator of quality is the completion rate—the percent of all eligible cases for which completed interviews are obtained. Another indicator is the proportion of ineligible cases. A high proportion may indicate that interviewers are misclassifying some eligible units. The average number of call-backs per completed case may serve as an indicator of how carefully interviewers are scheduling their calls. Careful review of these and other indicators will allow supervisors to concentrate their attention on interviewers whose work is substandard. (See also the discussion of "Preliminary screening" below.)

(ii) Observing Interviews

Observation of interviews in both face-to-face and telephone surveys is widely used to train and assess interviewers, and to evaluate respondent reactions in pre-test interviews or in exploratory studies.

However, direct observation of face-to-face interviews during the survey itself is relatively uncommon because of the high cost. If resources are available for some direct observation of interviewers in the field, supervisors should observe the work of less experienced interviewers and those with below-average performance, as shown by their activity reports and the failure rates of field screenings of their completed questionnaires (see below). A possible substitute is to ask each interviewer to tape record one or more of their interviews at specified intervals.[27]

Conversely, direct observation of telephone interviews is relatively inexpensive and therefore a valuable tool for controlling all types of nonresponse and response errors. It is widely used to monitor and assess telephone interviewers. Throughout the data collection phase, supervisors can easily monitor the interviewer's side of the conversation, quickly correct deficiencies in the way interviewers ask questions, and make sure they ask all of the questions. Moreover, with the proper equipment and the permission of the respondent, supervisors can monitor both sides of the conversation and give interviewers valuable feedback on how to improve their skills.

[27] This requires the respondent's explicit consent.

The contractor should develop written evaluation criteria for whatever observation techniques are planned. The criteria are needed to guide the supervisors in which aspects of the interviews they need to look at. Supervisors also should be instructed in how to use the results of their observations to help interviewers improve their performance.

(iii) Screening completed questionnaires

An initial "field screening" of the questionnaires turned in by the interviewers is an effective way to detect and correct many types of non-sampling errors. The term "field screening" is more properly applied to face-to-face surveys, but similar procedures are used by supervisors in conventional and mail telephone surveys to control the quality of the interviews.

Questionnaires may be screened by supervisors, or their office assistants, who should look for: (a) missing entries (which may indicate failure to follow skip patterns correctly), (b) inadmissible or questionable entries, (c) unnecessary entries, and (d) illegible entries. The supervisor should record all errors and discuss them with the interviewers.

Field screening may reveal systematic procedural errors by the interviewers, or even faulty instructions or training materials. It is important to detect such systematic errors early in the data collection phase so supervisors can alert the interviewers to their mistakes before they complete too many additional interviews. Once the screening has shown that an interviewer is doing good work, it may not be necessary to review all their completed questionnaires—occasional spot checks may be sufficient.

(iv) Validating interviews

Another important quality-control strategy is for the field staff to verify whether interviewers are actually making all the interviews they claim to have made. Verification is usually accomplished by mailing respondents a card asking (a) if they were interviewed, (b) how long the interview took, (c) if they would be willing to participate again, and (d) if they have any comments or questions about the interview or the interviewer. If a respondent does not return the card within ten days, the supervisor should contact them by phone.

Generally, about 10-30 percent of each interviewer's completed questionnaires should be verified each week. Although professional interviewers rarely forge an interview, if any questionnaire fails the validation test the contractor should verify the entire interviewer's previous work.

(v) Reinterviews

Reinterviews may be an effective method of measuring response errors. They should be done soon after the initial interviews because the respondents' characteristics and availability are likely to change if there is a long interval between the initial interview and the reinterview. Sometimes, an interviewer with similar training and experience will reinterview the original unit; in other cases, supervisors or more experienced interviewers are used. To minimize the burden on the respondents selected for a second interview, usually just a few questions are asked.

The cost of reinterviews is high, however, and the time required to conduct them and process the results—especially if complete reinterviews are done—make them unsuitable as a quick, early strategy for measuring interviewer performance.

Reinterviews sometimes are used to determine whether units that interviewers have termed "ineligible" have been correctly classified. For example, supervisors may reinterview all the housing units in a particular area that interviewers had reported as "vacant." The reinterviews would reveal whether any of these units actually occupied at the time of the survey. Interviewers sometimes are tempted to misclassify occupied housing units where interviews are inconvenient or difficult to obtain as "vacant," thereby eliminating the requirement to obtain interviews for these units.

B. Organizing and staffing field operations

In addition to establishing strategies to assure the quality of the data, in a face-to-face or telephone survey the contractor needs to organize and oversee the work of dozens, perhaps hundreds, of interviewers as well as supervisory and administrative staff.

Although managing the data collection phase of a mail survey is less complex, the contractor must still set up a system to coordinate and control the flow of the questionnaires to and from the respondents. In addition, since mail surveys usually entail some telephone or face-to-face follow-up interviews, staff should be instructed in the proper procedures for these interviews.

This section continues to focus on face-to-face interviews and examines the organizational and administrative tasks a survey contractor typically performs to set up a successful field operation. Many of these tasks also apply to telephone interviewing. The four main tasks are:

Organizing the interviewing:

1. Preparing instructions and training materials
2. Staffing the field operations
3. Training the interviewers
4. Coordinating and controlling the field work

Organizing the "field" operations of a telephone survey is similar to face-to-face surveys in many ways, but less complex. There is no need to set up a far-flung field operation as in a face-to-face survey, for example. Usually the interviewers work in one centralized location, supervised by a few members of the contractor's permanent staff. However, instructions and training materials for the supervisors and interviewers have to be prepared; the interviewers are selected and trained; and a system should be set up to coordinate and control the interviewing activities.

The contractor should fully document these procedures in the work plan well before any of the preparatory tasks are initiated. The sponsoring office should review them at the same time as the quality-assurance procedures, discussed in section A.

1. Preparing Instructions and Training Materials

Once the Agency approves the quality-assurance procedures that will be used to guide the interviewing, the contractor should document them in instructions and training materials for the interviewers, supervisors, and other field staff. How extensive these materials have to be depends largely on the method of collection. Obviously, face-to-face surveys require the greatest number of written materials, and mail surveys require the least.

There are three basic guidance documents prepared for a major face-to-face survey: (a) instructions for the supervisors, (b) an interviewer's manual, and (c) a training guide.

(a) Instructions for Supervisors

It is almost impossible to overemphasize the importance of the field supervisors in controlling the quality of interviewers' work. Yet, all too frequently, written guidance materials for supervisors concentrate on logistic and administrative matters—receipt and shipment of materials, payment and allowances for interviewers, etc. These subjects are important, but they do not deal directly with the supervisor's central responsibility, which is to see that the work is done on schedule and that standards of quality are met.

The instructions to the supervisors should clearly specify:

- The kinds of quality-related problems requiring communication with the central survey staff, and a well-defined procedure for resolving problems that arise;
- The quality-control strategies that will be used to assess the work done by the interviewers, and the supervisor's responsibilities in implementing them and evaluating their effectiveness; and
- The criteria that higher-level field staff or central staff will use to evaluate the supervisor's performance.

(b) Interviewer's Manual

A detailed written instruction manual for the interviewers is essential for every survey. Supervisors will also use this manual in their training and for oversight.

If the contractor has developed a standard training manual covering record-keeping, interviewing techniques, and other features common to all surveys, it may be sufficient to prepare a supplement to their standard manual which will cover only the special features of the Agency's survey, such as:

- How the sample was selected;
- Procedures for locating respondents;

- Respondent rules;
- Follow-up procedures; in particular, how to deal with various nonresponse situations;
- Quality-control strategies to be used;
- The objectives, purpose, and scope of the survey;
- Question-by-question specifications explaining the intent of each question; and
- Any special administrative matters, such as the length of the data collection period, whom to contact in case of problems, and what to do with the completed questionnaires.

(c) Training Guide

A formal training guide for supervisors and others conducting interviewer training sessions is a desirable supplement to the interviewer's manual. The guide should include topics the trainers should cover, the order in which they are to be covered, and practice exercises, quizzes, etc., for each training session.

To supplement the training guide, the contractor may develop other materials such as:

- Test exercises, to be completed at various points in the training;
- Written instructions for "mock" interviews;
- Audio-visual materials such as taped demonstration interviews; and
- Slides and other visual aids showing maps of the sampling areas, questionnaire forms, etc.

2. Staffing the Field Operations

Once the instructions and training materials are ready, the contractor assigns existing staff or recruits new staff to carry out the data collection activities. To complete the fieldwork for a major face-to-face survey–normally several dozen interviewers located in 50-100 sampling points (cities or counties)–several field supervisors and support personnel, staff for overall project supervision, and a full-time central office will be needed. There should be enough supervisors so that they will all have adequate time to monitor the performance of the interviewers assigned to them.

The staff people most directly involved in the field work are (a) the field supervisors and (b) the interviewers themselves. Let's briefly examine their respective responsibilities.

(a) Field Supervisors

Some supervision of the interviewers is essential in every survey to detect poor work and assure

that the fieldwork proceeds smoothly. Sometimes, centrally located supervisors direct the work of a mobile field staff, which moves into the various sampling areas. Some survey research firms prefer a network, of perhaps a dozen supervisors, who work on a regional basis and move with the field staff from area to area. Whether the field supervisors are centrally located or dispersed, they are the main link between the head office and the interviewers in the field.

The contractor should establish some equitable ratio of interviewers (and other field staff) to supervisors. The ratio should be small enough so the supervisors are able to spend sufficient time both in the field and in the regional (or central administrative) unit to regularly review and evaluate the work of the interviewers for whom they are responsible. The appropriate ratio for any specific survey will depend on factors such as the experience of the interviewing staff, the size of the assignment area, the type of transportation and communication facilities available, and the amount of time the supervisors are required to spend on matters not directly related to the survey.

Each field supervisor is responsible for hiring, training, and maintaining a staff of interviewers in the areas assigned to them. They should be in constant communication with interviewers through personal visits, mail, telephone, or e-mail contacts.

The field supervisors, along with a support staff of clerical personnel who usually work in the areas where the interviewing is going on, are responsible for:

- Arranging travel and lodging for staff and interviewers;
- Preparing specific work assignments for the interviewers—areas, times, lists of households—or, in the case of a business survey, coordinating and scheduling interview sessions;
- Logging-in the completed questionnaires and control forms (the interviewers' evaluations, notes, weekly activity reports, etc.);
- Scanning the questionnaires for completeness and accuracy, and forwarding them for editing and coding;
- Regularly evaluating the interviewers' work, using the quality-control strategies discussed in the previous section; and
- Preparing detailed reports on the field activities. These will be used to prepare periodic progress reports for the Agency showing the number of completed or partially completed interviews, the number of refusals, the number of verifications, etc., and the overall response rate.

(b) Interviewers

In any face-to-face or telephone survey, interviewers play a major role in the quality of the responses, and hence in the quality of the results. In some EPA-sponsored surveys, the interviewer is the only link between the contractor's central office staff and the respondents.

Regardless of the size of the survey, the contractor should establish policies and procedures for selecting and training the interviewers and maintaining their morale. A relatively small face-to-face survey of 500 respondents may involve hiring and training as many as 30 interviewers. Keeping interviewer workloads on each survey small will help to (a) keep interviewer travel costs low; (b) minimize the time needed to complete the fieldwork; (c) avoid making the interviewers' job too repetitive and monotonous; and (d) minimize the effects of systematic errors by individual interviewers.

There is a wide range of practices among survey research firms regarding the hiring of interviewers. Most reputable survey research firms maintain a network of skilled interviewers they can call upon. Interviewers usually are recruited on the basis of written applications, followed by a lengthy personal interview and a written test to evaluate the basic clerical skills needed to record, summarize, and edit respondents' answers.

At the end of the project, interviewers generally are rated on their productivity, accuracy, cooperation, and dependability.

Firms typically maintain a file of the names, capabilities, and performance ratings of those who have passed the initial screening. In addition, the file contains detailed information on the interviewers' geographic location, ethnic identification, hours available for work, educational background, special skills, current availability, and performance evaluations on previous surveys.

Before hiring interviewers for a specific project, it is important to make sure that they are able to work at the necessary level during specific hours; are able to get to the interview locations; and are willing to work in the assigned areas.

People become interviewers for many reasons. They are motivated by the flexible working hours, the chance to interact with others, and the opportunity to satisfy their curiosity about a variety of research topics.

While there is no such thing as an "ideal" interviewer—much depends on the nature of the survey—the most sought-after qualities typically are intelligence, dedication, honesty, dependability, attention to detail, a professional attitude (neither overly social nor overly aggressive), and an ability to adapt to a variety of interviewing situations (different types of people, different areas, etc.).

Once interviewers are hired, maintaining morale is vital. Good working conditions, a reasonable schedule of assignments, equitable pay rates, and bonuses for high quality work and difficult assignments all contribute to their efficiency.

3. Training the Interviewers

One of the contractor's most important tasks is to train the interviewers. The contractor should begin training those who will be used for the main survey shortly after the Office of Manage-

ment and Budget approves the clearance request.

No matter how skilled or experienced an interviewer, or how simple the questionnaire, the interviewers need to be:

- Thoroughly instructed in the specific objectives, rules, and procedures of the survey;
- Taught all quality-assurance procedures they will be responsible for, and the procedures for reporting their progress to the supervisor; and
- Taught a standard format for recording respondent replies.

If the interviewers are inexperienced, they should also be instructed in basic interviewing skills (techniques for gaining entry, probing, and so forth), and be taught how to plan and update their calling schedules so as to make the best use of their time and travel.

Survey research firms use a variety of techniques to train or re-train interviewers—interactive lectures, home study programs, practice interviews, and practice in the field. Often a final exam on the field procedures is given as well.

Most face-to-face surveys are complex enough to require interviewers to attend a two-to-five day training conference. These are sometimes held at several different locations around the country. A field supervisor and several professional trainers generally lead the training. Training is guided by the interviewer's manual, the training guide, and various other training aids that the contractor has prepared.

The supervisor should evaluate both the effectiveness of the training sessions, and by rating the trainees' performance in practice exercises, quizzes, and exams of various kinds, the extent to which each interviewer has mastered the essential skills. Interviewers who are clearly incapable of doing work in the field should be eliminated from consideration, re-assigned, or given additional training.

Once the interviewing is in progress, the field staff may provide training for new interviewers or conduct special sessions to reinforce the initial training.

4. Coordinating and Controlling the Fieldwork

In addition to hiring and training interviewers, supervisors, and administrative support staff, the contractor should set up a system to coordinate and control the fieldwork. For most surveys, this means establishing procedures for

- Scheduling and tracking the work of several dozen interviewers for several weeks, or perhaps months.

  Once the contractor has determined how many interviewers will be needed, either the central administrative unit or the field supervisors will prepare a schedule of the units

each interviewer should cover. The assignments are based on the interviewer's availability and experience, and often the special characteristics of the sampling areas that have to be covered, such as the living culture of the neighborhood. For example, although most interviewers are women, if high-crime areas are to be surveyed (particularly at night), male interviewers should be assigned to those areas.

For both economic and administrative reasons, it is necessary to limit the length of the interviewer's assignments. However, from a practical standpoint, the field supervisors should allow the interviewers enough time to cover all their assigned units and to make whatever number of call-backs that were established in the follow-up procedures.

- Controlling the flow of materials to and from the field.

  Once the data collection begins, the pace of the administrative work accelerates rapidly. Unless the contractor establishes close control over the flow of materials to and from the field, chaotic conditions may result. Often a central administrative unit at the contractor's main facility will be given the responsibility of sending instructions and training materials, blank forms and questionnaires, and other necessary supplies to field personnel. This same unit can also receive and screen the questionnaires and other such materials completed in the field. A regional field organization frequently is incorporated into the process. Each unit in the communications chain should maintain accurate records of its own, particularly regarding the response status of each sample unit.

- Resolving problems in the field.

  The contractor needs to develop a system for the field supervisors to report problems encountered in the field to the regional supervisors or the central administrative unit. If the resolution of these problems affects the existing procedures, all staff should immediately be notified of the changes.

C. Conducting the Interviews

It's time to turn now from methodological and organizational concerns, for which the researchers, analysts, and administrators on the contractor's staff are responsible, to the practical aspects of interviewing—the actual conduct of the interviews. Interviewers have four principal tasks in a face-to-face survey:

The interviewer's main tasks:

1. Locating the respondents
2. Gaining respondents' cooperation
3. Asking questions
4. Recording and editing responses

In formal interviews, the interviewer's goal is to obtain full and accurate answers to a fixed set of items and record them on a standardized survey questionnaire. When a structured questionnaire is administered in a uniform way, the researchers and analysts can be reasonably confident that

all the answers are comparable. For this reason, formal interviewing is the norm for statistical surveys. This does not mean that formal interviewing allows no flexibility. The interviewer can explain and probe and adjust the speed of the interview—but within some predetermined limits. Rarely are the interviewers permitted to change the wording or order of the questions, and probing may be allowed only for certain questions.

1. Locating Respondents

In most face-to-face surveys, only about one-third of the interviewer's time is actually spent interviewing. Their most time-consuming pursuit is simply finding the respondents. Studies show that approximately 40 percent of an interviewer's time is spent traveling and locating respondents. The remainder is devoted to clerical and editing tasks. (Note that in a telephone survey, no time is lost in travel and comparatively little is wasted in searching for the respondents. How much of the interviewers' time is spent locating the respondents depends largely on the respondent rules.

In a household survey, usually less than half of the interviewer's initial contacts result in completed interviews—either because no acceptable respondent is home or none of them will agree to be interviewed at the time. Interviewers often have to make several return visits before they secure an interview with an acceptable respondent. If the respondent rules require an interview with one or more specific individuals in the household, a still greater number of callbacks are likely to be necessary. Since the sample units assigned to any one interviewer are often spread over a broad geographic area (a town or county, perhaps), extensive travel—and frustration—is common.

Locating non-household respondents poses somewhat different problems. Physically locating them usually is not difficult. The main problem in business or industrial surveys is finding the people most qualified to answer the questions. Several call-backs may be necessary before the interviewer locates the right people, and is able to schedule interviews with them.

2. Gaining Respondents' Cooperation

Once the interviewer has located a respondent, the next task is to secure an interview. The way interviewers introduce themselves, the identification they carry, what they say about the survey, how they dress and behave, and the courtesy they show to all the people they come in contact with—not simply the respondents—all have a bearing on how successful they are in getting respondents' cooperation. The person the interviewer talks to initially may not be an acceptable respondent, but may be able to provide information on when the desired respondent will be available and ultimately may influence the person's willingness to cooperate.

The interviewer should present a positive, pleasant, relaxed, professional image, and offer the respondent proper credentials—a picture ID showing the name of the survey research firm they represent, possibly a calling card, and other materials that will demonstrate the integrity of the firm and the importance of the research effort.

The interviewer should briefly explain the nature of the study, the purpose of survey research,

and the reasons they want to talk with the respondent. The interviewer also may explain how the data will be used, and who will be permitted access to the data. Explanations about the extent of disclosure of individual responses are especially important to business or industrial respondents, who frequently have strong concerns about revealing trade-sensitive or confidential information.

Most household respondents will agree to be interviewed if approached properly. They do so because they are curious about the subject matter, or about surveys in general, or because they are pleased to have an opportunity to express their views to someone. Sometimes they agree just because it is harder to say "No" than "Yes" to a skillful interviewer.

Some respondents are willing to be interviewed with only a brief explanation of the purpose of the visit; for others it will be necessary to go into some detail. Respondents have various concerns and questions—why they were selected, what good the survey will do, why the person next door isn't being interviewed instead—and the interviewers should give correct and courteous answers.

In no case should an interviewer exert undue pressure to obtain an interview from a reluctant respondent. Responses given reluctantly are likely to be less accurate than those of a more willing respondent**.** Faced with a persistent refusal, it is best to make no further attempts to get an interview. Sometimes a second approach by the supervisor or a more experienced interviewer will succeed in "converting" a refusal to a completed interview.

Respondents may refuse to be interviewed for any number of reasons—they are reluctant to break their daily routine; they have other obligations; they are afraid or suspicious of the interviewer; or they are indifferent or hostile to the Federal government, the subject matter, or research in general. Studies show that the respondent's attitude towards surveys in general, based on their own experience and what they have heard from others, is the overriding factor in their decision to grant or refuse an interview. In addition, the prevalence of "surveys" that are thinly-disguised attempts to sell goods or services can make the interviewer's task all the more difficult; it is important early on to emphasize the research objectives of the survey.

3. Asking Questions

Once the respondent agrees to be interviewed, the interviewer should immediately try to establish a good interaction so the respondent will cooperate in supplying the required data. Ideally, the interviewer will have an opportunity to talk with the respondent in private long enough to complete the questionnaire with no disturbances.

As stated at the beginning of this section, the goal of a formal interview is to obtain full and accurate answers to a fixed set of questions. In addition to reading the questions slowly and deliberately so there is no chance they can be misinterpreted, the interviewer should do whatever is necessary to get satisfactory answers. In fact, an important part of the interviewer's task is to assess the adequacy of the respondent's answers, and if necessary, to take steps to get more information.

When appropriate, the interviewer should:

- Ask the respondent if they would like the question clarified or repeated;
- Provide feedback to indicate that an adequate reply has been given or that something else the respondent said has been noted or understood;
- Clarify aspects of the respondent's task which seem to be problematic or confusing; for example, by confirming the frame of reference of a particular question;
- Check with the respondent to make sure that a particular response was correctly heard or interpreted;
- Motivate the respondent to complete the questionnaire by interjecting a few words of encouragement from time to time; and
- Control the direction and extent of the respondent's replies, by keeping the respondent from digressing or by reading the next question as soon as a satisfactory answer is recorded, for example.

4. Recording and Editing Responses

Although asking questions well is a critical aspect of a formal interview, the information the respondents provide will be lost if it is not recorded accurately and fully. All interviewers should use the same methods and conventions for recording responses and for editing the questionnaire after the interview is over.

Recording answers may seem to be a relatively simple task, but interviewers sometimes make serious errors. The reason is that interviewing is a fairly tiring, repetitive activity, and often a lengthy and complex one as well. In recording replies, interviewers often follow complex skip instructions and coding rules, and at the same time, listen carefully to the respondent so they can be ready to take whatever action is necessary to deal with a vague or inadequate reply. (In computer-assisted interviewing, skip patterns and coding rules are easier to manage.)

To minimize recording errors, interviewers are trained to check the questionnaire for omissions, ambiguities, illegible entries, and clerical errors before concluding the interview and while the respondent is still available. The interviewer should also note where probes were used, and make a few comments on the interview situation. If a tape recorder is used as a backup in a long interview, the interviewer should transcribe and edit any new information onto the questionnaire.

D. Monitoring the Interview Process

As project officer, there are several things you can do, both before and after the fieldwork begins, to foster the collection of high quality data.

Before hiring a contractor, pay particular attention to the following items in the offerors' proposals:

1. The firm's experience in managing surveys where interviews were used to collect a similar volume of data. Selecting a survey research firm with a good track record in conducting surveys of similar size and scope is usually the best guarantee of getting high-quality data from your survey.

2. The proposed interviewing activities. Proposals should include clear-cut plans for: (a) quality assurance; (b) selecting, training, and supervising the interviewers and administrative staff; and (c) organizing and overseeing the interviewing activities. It is strongly recommended that you have a survey expert review these plans, regardless of what primary collection method the contractor plans to use. Even in a mail survey, normally some interviewing is done to follow up nonresponse and response errors.

The quality of the data gathered in a face-to-face survey depends largely on the work done by the interviewers. Inaccuracies, omissions, and biases in the data they collect can be kept to a minimum by good training; rigorous use of the quality-assurance procedures established for the data collection; attentive oversight by the contractor throughout the data collection phase; and close monitoring by the sponsoring office.

Therefore, after the contractor is retained:

1. Have a survey expert review the quality assurance procedures and the procedures for controlling the field operations, as described in the work plan (see sections A and B).

2. Participate in the pilot test. Go along on some of the interviews as an observer. Attend the interviewer debriefing sessions during and following the pilot test. Work with the contractor on revising the interviewing procedures for the survey itself, if necessary. This will expedite any changes in the questionnaire or the interviewing procedures that require Agency approval. Circulate the pilot test report to survey experts, and make sure the contractor takes proper account of all comments and suggestions before any data are collected in the main survey. (See section A of Chapter 3 for more information on pilot tests.)

3. Review drafts of all instructions and training materials the contractor prepares for the interviewers and supervisors. Attend as many interview training sessions as possible. There you can explain the study goals, emphasize the Agency's interest in obtaining high quality data, and answer any questions.

4. Once the data collection begins, make occasional visits to field sites or the facility where the phone interviews are being conducted. If the interviewing is not proceeding according to plan, advise the contracting officer so the Agency can take whatever steps are necessary to correct the problems.

5. Have a survey expert review the contractor's progress reports during the data collection phase to make sure the contractor is (a) maintaining the schedule, (b) achieving the response rates specified in the work plan, and (c) using the quality-control procedures established in the plan.

Bibliography: Chapter 5

Moser, Claus and Graham Kalton, Survey Methods in Social Investigation, Second Edition, New York, Basic Books, 1972.

Judd, Charles M., Louise H. Kidder and Eliot R. Smith. (1991). Research Methods in Social Relations, Fort Worth, Texas, Harcourt-Brace College, 1991.

Chapter 6: Data Processing

In most EPA surveys, the contractor is required to process the "raw" data collected from the sample into usable information. Processing involves a series of manual and computerized operations to reduce responses on the questionnaires to machine-readable form so they can be stored, retrieved, summarized, and analyzed. The desired end-product of these processing operations is a "clean"—virtually error-free—data file, on some magnetic or optical media. The data file is then programmed by the contractor or the Agency to produce a variety of reports, ranging from simple tables summarizing the characteristics of the database to highly sophisticated statistical analyses.

This chapter discusses:

A. The eight fundamental steps in processing survey data; and

B. How to monitor the contractor's data processing activities.

A. Steps in Processing Survey Data

This section examines the eight steps involved in processing the data collected in a typical statistical survey to produce the results for the final report.

Data Processing Procedures:

1. Develop procedures
2. Select and train staff
3. Screen incoming questionnaires
4. Review and edit questionnaires
5. Code open questions
6. Enter data
7. Detect and resolve errors
8. Prepare outputs

The complexity of the steps in any particular survey depends on three factors:

1. The extent of the outputs defined in the analysis plan. The analysis plan, which specifies the preliminary tabulations and the types of analyses to be prepared from the data file, not only influences the design of the questionnaire, the sampling plan, and the data collection procedures, but also guides the processing operations. (See Chapter 1 for more information on the analysis plan.)

2. The size and complexity of the questionnaire. The nature of the questionnaire profoundly influences the processing procedures. If there are many open questions, which require respondents to frame answers in their own words, editing and coding the raw data on the questionnaires will necessarily be more complex. Conversely, if most of the questions offer a fixed range of pre-coded responses, or if a CATI-programmed questionnaire is used, several processing steps may be bypassed.

3. The size of the sample and the complexity of the sampling procedures. These determine how many questionnaires have to be processed and how much weighting and other treatment of the data are needed to produce results for the final survey report. (See Chapter 4.)

Let's turn now to the eight tasks the contractor will typically perform during each of the processing steps listed above.

1. Develop the Processing Procedures

The first step in transforming the raw data collected from the respondents into usable information is to develop a set of procedures for processing the questionnaire data.

The processing procedures are one of the six components of the work plan. The contractor should develop them after major decisions on the questionnaire, the sampling plan, and the analysis plan have been made.

The data processing procedures should specify:

- The specific tasks the contractor will perform after the completed questionnaires arrive at the central processing facility to produce a clean, virtually error-free data file;
- The software, hardware, and personnel to be used for each of these tasks;
- Provisions for training processing personnel in the special procedures developed for the survey;
- The quality control techniques that will be used to minimize errors at each step of the processing;
- A flow chart for the tasks to be completed at each step; and
- A complete listing and schedule of the tabulations and other output reports that will be generated in preparation for the analysis.

The sponsoring office may establish some preliminary specifications for the processing operations during the design phase of the survey, particularly the form and content of the tabulations (or desired outputs.) Once hired, the contractor will have to work with Agency data processing experts, systems analysts, and subject matter specialists to make sure the computerized output reports are clearly defined. This should be done before any computer programs to generate these reports are written. Normally, existing statistical software packages can be modified to accommodate the Agency's tabulation and analysis requirements. However, if the contractor has to develop any new software, sufficient time and resources should be allowed.

Be sure to have appropriate Agency experts review the final processing procedures before giving

the contractor the go-ahead to process any data. If the contractor pre-tests these procedures—usually in a pilot test of the main survey—these experts should also review the adequacy of the preliminary outputs generated from the pilot test data.

2. Select and Train Staff

Most of the people who will be involved in the data processing operations will be permanent members of the contractor's staff with experience in processing survey data. For most surveys the staff also will include a data processing manager, a computer center manager, operations personnel, clerical, coding, and editing personnel, an operational control unit, data entry personnel, systems analysts, and programming personnel.

Usually a supervisor will be assigned to oversee each step of the processing, e.g., the initial screening of the completed questionnaires, the manual edit and coding, the transfer of the data to machine-readable form, the final computer edit and "treatment" of the data, and the preparation of the tabulations.

All processing personnel, especially the editors and coders, should receive formal training in the special procedures developed to screen, edit, and code the survey data. Data entry personnel (if used) also need a short training course. The systems analysts and programmers should also be thoroughly oriented in the informational and analytical objectives of the survey before their work on the project begins.

For most surveys, the contractor will have to prepare instructional and reference materials to train and guide the editors and coders. These materials typically include procedures for coding each open question and for dealing with omissions, inaccuracies, and inconsistencies in the data (item nonresponse). They should be updated throughout the data processing phase.

The actual processing of the data (Steps 3 through 7) begins shortly after the first few batches of completed questionnaires arrive at the processing facility. Appropriate members of the contractor's staff will first check-in and screen the questionnaires (Steps 3 and 4) and code any open questions (Step 5.) Next, other staff will manually key the data (Step 6.) Then comes the final "cleaning" of the data file and the classification and sorting of the data (Step 7.) The last task is the preparation of various tabulations and analyses that summarize and interpret the content of the file, along with the preparation of a report fully documenting the processing procedures (Step 8.)

Note that if computer-assisted interviewing is used as the primary collection method, several steps are bypassed because the respondents' answers are keyed directly during the interviews. Despite the advantages of both CATI and CAPI, they should be used only for large surveys—over 300 respondents, say—because of the high cost of the initial programming. However, a smaller response rate may be cost effective if the survey is very simple.

3. Screen Incoming Questionnaires

Since all members of the sample must be accounted for, strict control of the questionnaires (and

other paperwork generated during the data collection phase) is essential. The contractor should assign a control number to each questionnaire. The number is usually placed on the title page. The purpose of the control number is to permit the processing staff to identify data from each questionnaire at any point in the processing, while maintaining confidentiality.

During this step, clerks at the main processing facility log-in the questionnaires soon after the respondents (in a mail survey) or the field supervisors (in a face-to-face or telephone survey) return them.

4. Review and Edit the Questionnaires

After logging the control numbers, the clerks batch the questionnaires and forward them to an editing and coding supervisor for screening. The amount of screening done at this stage of the survey depends on the method of collection and how much screening was done in the field.

In face-to-face and conventional telephone surveys, questionnaires often receive a preliminary screening by the field supervisors to rectify obvious problems and errors. However, an additional review by the processing staff is almost always done to check for legibility, completeness, and internal consistency. This is especially critical for the first few batches of questionnaires. The hand screening is an effective way of detecting systematic errors the interviewers or other field staff may be making before the interviewing proceeds too far. Any questionnaires containing major problems are generally returned to the field supervisor for action.

Conversely, errors on mail questionnaires are referred to other staff for follow-up to fill in the missing or inconsistent entries—usually via phone interviews—before further processing is done. The purpose of this screening is to isolate questionnaires that:

- Contain omissions and inconsistencies requiring some follow-up (usually in short face-to-face or telephone interviews) before further processing is done;

- Will be counted as "nonresponse" cases because there are too many omissions or illegible answers; or

- Are deemed unacceptable for processing for other reasons, such as being completed for an ineligible unit.

It is essential that you and the contractor fully agree on the precise criteria to be used for the screening operations. Usually, to be considered acceptable for processing, a questionnaire must contain legible and complete responses for all key variables and no more than a specified number of omissions for other items.

The clerks doing the screening may also do a thorough review and edit of the questionnaires, or depending on the complexity of the questionnaires, may forward them to editing or coding specialists.

The purpose of a manual review and edit at this stage of the processing is to catch errors before

the data are converted to machine-readable form. Hand editing is a relatively slow and inefficient way to catch errors, but may be appropriate in a very small survey. A subsequent computer edit (also called "machine edit") involving a more detailed and complete application of the editing rules is vital (see Step 7.) The computer edit also serves to detect and correct human errors introduced during the coding and data entry stages, discussed next.

5. Code Open Questions

Many EPA survey questionnaires include one or more open questions. These questions may generate a large number of different, yet acceptable, responses that must be grouped into a reasonable number of response categories for counting and analysis. This process is called coding.

Codes for open questions often require a lengthy development process. First, the investigators tentatively define a few codes for a set of plausible responses to each open question. The coded response categories are then matched against the answers actually given by respondents in the pretest. Usually, the initial codes have to be redefined to fit the pretest responses, and perhaps tested again. After the first 50 to 100 questionnaires in the final survey are edited and coded, the codes may be further refined. Still further adjustments may be made later if the coders have difficulty fitting existing codes to actual responses on new batches of questionnaires that arrive for processing.

The actual coding of the replies to open items may be done by the interviewers (partially-open questions are coded during the interview); their supervisors (shortly after the interviewers turn in the completed questionnaires); or, most frequently, by experienced coders at the processing facility. Whoever does the coding uses a special coding manual listing the codes defined for each open question.

Quality control of the coding is vital. The work of each coder should be checked periodically for accuracy and consistency with the codes defined in the manual. Processing supervisors normally check 100 percent of each coder's work at the start. Because coding errors tend to decrease as the clerks become more familiar with the subject matter, a random sample—usually 10 percent of the coded questionnaires—is checked after the coders' errors decline to an acceptable level.

To control consistency among the coders, supervisors periodically run tests on a sample of the coded questionnaires and establish a "rate of agreement" for each question. Typically the rate is based on the number of times pairs of experienced coders select the same code for a particular response.

6. Enter Data

The next step in the processing is to transfer the edited and coded data from the questionnaires onto a machine-readable medium. The two most common methods of entering data are to key them manually through on-line terminals or by using some form of optical character recognition (OCR.)

To minimize human error in manual data entry, two different operators key the data from a single questionnaire. Quality control is achieved by a computer-assisted comparison of the data to spot and reconcile any differences. Additional quality control is achieved by programming the data-processing program to identify (and in some cases correct) inadmissible values or codes.

Two more sophisticated method of data entry are optical scanning and a related method called optical character recognition (OCR). In optical scanning, the questionnaire is pre-coded with "bubbles" that are filled in by the respondent (in a mail survey) or by the interviewer (in a face-to-face survey.) Optical character recognition, in addition to recognizing filled-in areas, is able to recognize handwriting; this is becoming increasingly sophisticated, with some systems plausibly claiming accuracy virtually as high as manual keying.

7. Detect and Resolve Errors in the Data File

The next step is to "clean" the data to enhance its quality and facilitate the subsequent tabulation and analyses production. Data cleaning is the process of detecting and resolving inaccuracies and omissions in the data file. Often it is the most complicated and time-consuming step of the processing.

In almost all surveys today, a computer performs the bulk of the work of detecting and resolving data errors. First, an intensive machine edit is performed to identify inaccuracies and omissions, and then various techniques are used to correct or convert unacceptable entries into a form suitable for tabulation and analysis.

Computer Editing

In a computer edit, the first step is to program the computer to check for inconsistent or "impossible" entries, some of which may have been introduced in the previous processing steps. For example, the computer may be programmed to identify errors such as:

1. Inadmissible codes—the code attributed to an item does not correspond with the permissible replies in the coding manual (for example, a code "4" has been entered for an item to which only codes "1" and "2" have been assigned);

2. Out-of-range entries—the amount that has been entered is below or above the permissible values programmed for that item;

3. Omissions—no entry has been made;

4. Inconsistencies—entries for two or more items are not consistent with each other (a respondent is reported to be 14 years old and a physician);

5. Math errors—the total for a list of items should be equal to the sum of the amounts shown for individual items on the list.

The computer may be further programmed to print an error message indicating the nature of the failure, or even to correct certain errors and log them.

Decisions on how much editing should be done by hand and how much by machine depend on many factors. For some surveys, several manual checks, as well as computer runs using special check-and-edit programs, may be necessary to achieve an acceptable error rate. Generally speaking, the more complex the questionnaire, the more difficult it is to develop computer programs for detailed edit-checks; thus considerable manual editing may have to be done. Larger sample sizes tend to make computer editing a more cost effective option.

Error resolution

The computer edit detects errors, but does not resolve them. Several techniques are used to deal with the errors the computer has identified. Survey researchers use several techniques to deal with data omissions and inaccuracies in individual questionnaire items (so-called "item nonresponse".) The principal ones are (1) returning to the original questionnaire to see if errors were made in entering the data, or if it is possible to infer correct responses from other information on the questionnaire; (2) having the computer impute values for missing responses, and (3) creating separate categories to report all missing replies. More specifically:

Consulting Questionnaires

Generally, the most reliable procedure for resolving omissions and inconsistencies in the data file is to consult the questionnaires. Data entry clerks sometimes pick up data from the questionnaires incorrectly. Or, if the respondent has left an answer-space blank, it is sometimes possible to infer the correct answer from other information on the questionnaire. Footnotes or written-in comments also may provide helpful information.

For instance, if respondents fail to state their ages, researchers may be able to infer their correct ages from other information on the questionnaires such as dates of birth or school attendance. Inconsistent responses sometimes can be resolved by considering the whole range of information supplied by a respondent and deciding which of the conflicting entries is most plausible, e.g., from information on the income, education, and marital status of the “14-year old physician” in the example on the previous page, it might be reasonable to assume that the respondent is really 41 years old. However, consulting questionnaires as a means of resolving errors is time-consuming and not always productive.

Imputing Missing Values

Another error-resolution method is to try and compensate for the nonresponse bias by having the computer impute values for the omitted and inconsistent replies. Imputation involves assigning values for missing or unusable responses by drawing on information from other sources such as answers to other items on the same questionnaire, another questionnaire from the same survey, or external sources (administrative records or another survey.) Imputation is similar to the weighting adjustments for total nonresponse, which will be discussed in Step 8.

Imputation generally is a faster and less costly way to resolve errors than consulting questionnaires, but it should be used with discretion. Imputed items should be flagged in the data

file so that tabulations and analyses can be prepared with and without the imputations, if desired. Also, any reports about the survey should indicate the extent of the imputation so that anyone using the data later can distinguish between real and imputed values. The extent to which the contractor intends to impute values for missing or omitted replies should be specified in the data processing procedures submitted with the work plan.

Note that the contractor should aim to get good data from the respondents in the first place, and make data adjustments strictly as a back-up measure. Imputation can be kept to a minimum by instructing interviewers to carefully check the questionnaires immediately after each interview; checking interviewers' work during the data collection phase; and follow-ups of respondents in mail surveys.

Creating Categories for Unreported Responses

If attempts to resolve omissions and inconsistencies in the data file using the above techniques are unsuccessful, the researchers may allow the errors to stand and report them as such in the tabulations. For example, they may report a total for all respondents who provided no valid income data in a new category called "income unknown."

Decisions on whether to impute values for omitted and inconsistent replies, or to add "not reported" categories in the tabulations, depend on a number of circumstances. Using a "not reported" category for such basic characteristics as sex and age may create serious problems in the analysis. Analysts sometimes handle this by imputing values for fundamental demographic variables for which considerable related information is available, and creating "not reported" categories for all others.

8. Prepare the Outputs

The final step in processing survey data is to prepare the tabulations and other outputs called for in the work plan. The contractor's main tasks at this step are to (1) weight the sampled elements to produce the estimates; (2) prepare the preliminary tabulations describing the data base and finalize the analysis plan; (3) apply the procedures described in the sampling plan for calculating the sampling errors; and (4) document the procedures used in preparing the data file.

Weighting the Sampled Elements

The first task in generating the tabulations is to weight the virtually error-free data file prepared in the previous step. Except for simple lists of data items, these preliminary reports summarizing the content of the file should be based on weighted data. Weights (or multipliers) are assigned to survey data for three reasons:

- To account for the probabilities used in selecting the sample

  If all units in the sample have the same probability of being chosen, the survey analysts can obtain valid estimates of some statistics such as proportions, percents, means, and medians without weighting the data. However, to estimate totals, all units are weighted

by the reciprocal of the sampling fraction. For example, if the sampling fraction was 1 in 200, all sample values or totals are multiplied by 200. If the selection probabilities were not the same for all the units, appropriate weights are applied to estimate any statistic. (See section D of Chapter 4 for more information on weighting.)

- To adjust for nonresponse.

  There are two methods of making adjustments for nonresponse:

  (1) One way is to increase the weights applied to individual units that did respond and are similar (based on data available for all the sample units) to those for which no data were obtained. For example, if one sample household in a block did not respond, one of the households in the same block for which data were obtained would be selected at random and given an additional weight of "2."

  (2) The other way is to apply a uniform weight to all the units in the sample or to those in a particular subgroup. For example, in a business survey, if 20 percent of the sample establishments with fewer than 10 employees did not respond, a weight of 1.25 (100 divided by 80) would be applied to all establishments that did respond.

- To apply sophisticated estimation procedures such as ratio or regression estimates.

  These procedures require a determination of relationships between variables or the introduction of independent data from other sources, such as current population estimates.

The overall weights the analysts ultimately assign to the data will reflect the combined effects of these three types of adjustments. Deciding on the sequence and procedures for weighting the data in a particular survey requires a good technical grasp of the sample design and the data processing system. Sampling and data processing experts at the Agency, and on the contractor's staff, should determine the weighting and estimation procedures long before the processing starts. These procedures should be critically reviewed by systems analysts at the Agency before the contractor processes any data collected in the survey.

Preparing the Preliminary Tabulations

After the weighting and estimation procedures are completed, a data file suitable for generating the preliminary tabulations should result.

Using a standard computer software package or software specially designed for the survey,[28] the contractor can then program the data file to generate a set of preliminary tabulations, which

[28] For the long-term viability of the survey data, after the contractor has delivered the final product and received payment, we strongly recommend using standard, "off the shelf" software packages for tabulations and analyses. That way, if EPA or others want to do additional analyses of the survey, they won't be tied to the original contractor's programmers.

normally will include:

- Frequency distributions (sometimes called "marginals") of responses for categorical variables (those based on questions with fixed response categories);
- Some simple cross tabulations;
- Estimated totals, ranges, and means (or medians) for the entire target population and for various subgroups;
- Listings of individual responses for selected items, especially for large sample units; and
- Where applicable, tabulations of key variables showing the number of units for which an item was imputed and how much of the total was imputed.

The preliminary tabulations will give you and the contractor an opportunity to review the database in an organized fashion, and thereby learn its structure and quality before the contractor prepares the final tabulations.

Subject matter specialists should carefully study these preliminary tabulations before the contractor prepares a revised list of the final tabulations to include in the analysis plan. The list should include the computerized output reports (tables and graphs) that will be prepared to fully describe the content of the database.

There is no clear line between the output reports generated at the conclusion of the processing phase and those developed for the analysis. However, the analysis of the database usually goes beyond simple descriptive summaries and explores the underlying relationships among the study variables.

A host of sophisticated analytic techniques may be used to reveal the full informational content of the database.

Usually, the final tabulations include:

- Detailed descriptive statistics (frequency distributions and cross-tabulations);
- Measures of central tendency (means, medians, and modes)
- Measures of variability (standard deviations, ranges); and
- Other analytical statistics such as correlations and regression coefficients.

For each tabulation, the revised analysis plan should specify: (a) the data sources to be used, (b) the variables to be cross-classified, (c) the sub-populations to be included, (d) the statistics to be shown, (e) how the data are to be weighted, (f) the title, subheadings, and footnotes; and (g) the layout. The analysis plan should also include:

- A full description of the methods for quantifying all relevant variables;
- Values of sample weights and all necessary formulas for estimating population means, medians, and variances;
- A list of hypotheses and the tests to be used to evaluate them;
- Descriptions of the variables and respondent groups that may be inter-related, and recommendations for regression and discrimination analyses based on the relationships; and
- Suggested methods for handling problems during the subsequent analysis, such as those that arise from missing data or nonresponse.

You should work with data processing and systems analysts both at the Agency and on the contractor's staff in defining these specifications for the final analysis plan.

Finalizing the Computations of Sampling Errors

The actual calculation of sampling errors (variances) for various estimates should be an integral part of the processing operations.

The estimates of sampling errors serve two purposes:

- They may help evaluate the database. For example, unusually large sampling errors for some items may indicate processing errors; and
- They are essential for determining whether observed relationships are statistically significant or due to random variation introduced by the use of sampling.

As discussed in Chapter 4, sampling errors usually are not calculated for all the statistics produced from the survey. This is generally unnecessary and often costly. The contractor's analysts and sampling specialists should select the items for which sampling error estimates are needed, making sure to include all key statistics and a representative set of other types of statistics that are to be tabulated from the data file. (For more details on calculating sampling errors, see section D of Chapter 4.)

Documenting the Processing Operations

Once the final tabulations are completed, the contractor should create a file documentation manual describing the procedures used to edit, code, and weight the data. The manual should identify the source of each data item on the questionnaire or on other documents used during the data collection phase.

In addition, the contractor should prepare a data dictionary containing the following for each variable:

- **Name**. For greatest compatibility with a variety of computer languages, variable names should contain no more than 8 characters.[29] You are encouraged to develop the data dictionary using one of the commercially available software packages such as SPSS, SAS, or OSIRIS. Although much of the analysis may eventually be done using such Windows-based programs as Microsoft Access or Excel, you are not encouraged to use these programs' flexible naming conventions for variables, because names longer than 8 characters, which may include spaces, are incompatible with many other analysis programs.

- **Description**. Brief description of variable (no more than 20 characters; unlike the variable name, this may contain spaces and special characters);

- **Type**. Numeric or text;

- **Width**. Some measure of width. For numeric variables, the width of the number and the number of decimal places included. For text variables, the number of characters.

- **Codes**. What each possible value means.

- **Location**. If not using one of the commercial packages, the starting column and width of each variable.

If EPA is to analyze the content of the data file, the contractor should submit the documentation manual, the final analysis plan, and whatever other materials Agency analysts will need to study and interpret the data file.

On the other hand, if the contractor is to do the analysis, the documentation manual should be submitted for EPA review and approval along with the final analysis plan before the data are analyzed.

A discussion of data analysis is beyond the scope of this Handbook. To assist you in this regard, a list of excellent sources is provided at the end of this chapter, along with a number of selections offering additional guidance on data processing issues.

B. Monitoring the Processing Activities

Throughout this Handbook OEI has emphasized that EPA's major impact on the successful outcome of a contract survey comes long before the data collection and data processing activities are under way. Achieving a clean data file on which to base the analytic work is largely dependent on the professional, clerical, and management capabilities of the firm the Agency hires to conduct the survey. As in the data collection phase, the sponsoring office has only lim-

[29] The name must begin with a letter. The remaining characters can be any letter, any digit, or the underscore (_) symbol. Variable names should not end with an underscore. Blanks and special characters (for example, !, ?, ', and *) should not be used. Each variable name must be unique; duplication is not allowed. Variable names are not case sensitive—the names NEWVAR, NewVar, and newvar are all considered identical.

ited control over the data processing activities.

Therefore, before the contractor is hired, you should:

- Require the grantors to specify in their proposals:
    - The formal quality-control procedures they intend to use at each step of the processing;
    - How they intend to keep coding and other errors to a minimum; and
    - How they will report production and error rates for each step of the processing.
- Specify the format and any special requirements for the completed data file to ensure compatibility with other EPA data files and otherwise facilitate the analysis.
- Require Agency approval of the key deliverables of the data processing phase (the data file, the tabulations, the estimated sampling errors, and the documentation of the processing procedures.) If the Agency is to do the analysis, you should specify that EPA requires that you approve deliverables before the contract is closed out. If the contractor is to do the analysis, do not let the contractor begin until you have reviewed and approved the above products of the data processing phase.

Other things you can do after the contract has begun to help assure the quality of the data file and the other deliverables are:

- Make sure the questionnaire is designed to facilitate the processing operations.
- Before data for the main survey are collected, carefully review the processing procedures and tabulations specified in the work plan. If necessary, work with the contractor to specify the content and format of the final tabulations. If a pilot test is done, review the procedures and tabulations and make sure the contractor makes any necessary modifications before processing any data from the survey.
- Participate in the development of response codes and procedures for treating nonresponse and "unacceptable" responses.
- Scrutinize all progress reports submitted during the processing to make sure the contractor is (a) adhering to the schedule and budget and (b) following the verification and quality-control procedures specified in the work plan.
- Have Agency statisticians, project personnel, and data processing experts review the preliminary tabulations, the file documentation manual, and the data dictionary. All tables should be reviewed to be sure that (a) they are internally consistent; (b) the estimates appearing in more than one table are consistent with each other; (c) significant changes from comparable data in earlier surveys are adequately explained; and (d) the estimates are "reasonable" based on expectations and data from other sources.

- Finally, if the Agency is to do the analytic work, make sure that all deliverables are in good order before the contract is closed out.

Bibliography: Chapter 6

Data Processing

Appel, Marin V., Robert D. Tortora, and Richard Sigman. "Direct Data Entry Using Touch-Tone and Voice Recognition Technology for the M3 Survey." Bureau of the Census Statistical Research Division Research Report Series, (No. RR-92/01), 1992.

Hoinville, Gerald and Roger Jowell. Survey Research Practices, London, Heinemann Educational Books, 1978. *Chapter 8, "Data Preparation."*

Moser, Claus A. and Graham Kalton. Survey Methods in Social Investigation, Second Edition, New York, Basic Books, 1972. C*hapter 16, "Processing of the Data," and Chapter 17, "Analysis, Interpretation and Presentation."*

U.S. Environmental Protection Agency, Office of Research and Development. Guidance for Quality Assurance Project Plans (QA/G-5) Washington DC, EPA, 1998.

U.S. Environmental Protection Agency, Office of Research and Development. Guidance for Data Quality Assessment: Practical Methods for Data Analysis (QA/G-9), Washington, DC, EPA, 2000.

U.S. Environmental Protection Agency, Office of Environmental Information. Guidance for the Date Quality Objectives Process (QA/G-4), Washington, DC, EPA, 2000.

Warwick, Donald P. and Charles A. Lininger. The Sample Survey: Theory and Practice, New York, McGraw-Hill, 1975. *Chapter 9, "Editing and Coding," and Chapter 10, "Preparation for Analysis."*

Statistical Analysis

Andrews, Frank M. et al. A Guide for Selecting Statistical Techniques for Analyzing Social Science Data, Second Edition, Ann Arbor, Institute for Social Research, University of Michigan, 1981.

Draper, Norman and Harry Smith. Applied Regression Analysis, Third Edition, New York, John Wiley & Sons, 1997.

Gonzalez, Maria Elena et al. "Standards for Discussion and Presentation of Errors in Survey Census Data," Journal of the American Statistical Association, Vol. 70, No. 351, Part II, September 1975.

Hoaglin, David, Frederick Mosteller and John W. Tukey, Editors. Understanding Robust and Exploratory Analysis, New York, Wiley Classics Library, 2000 (reprint of 1983 edition).

Software Packages

Research Triangle Institute. "SUDAAN: Software for the Statistical Analysis of Correlated Data, 2001." http://www.rti.org/sudaan/.

Westat. WesVar 4.0 User's Guide. Rockville, MD, Westat, 2000.

Glossary

BIAS—The difference between the survey estimate, averaged over repeated samples, and the true value. Sampling bias can result from use of a non-probability sample or from errors in the execution of a probability sample design. Non-sampling bias can result from many factors such as use of an incomplete sampling frame (coverage bias), nonresponse in the survey (see NONRESPONSE BIAS), a poorly designed questionnaire, respondent errors, interviewer errors, or processing errors.

BURDEN—In the Paperwork Reduction Act (PRA) of 1995, "burden" is defined as the amount of time required to collect data from the public using a particular data collection instrument (a questionnaire.) The response burden of a particular survey questionnaire is the estimated number of hours each respondent needs to complete the instrument, multiplied by the total number of people to be surveyed. The total number of burden hours for a survey questionnaire is reported to the U.S. Office of Management and Budget (OMB) if data are to be collected from more than nine members of the public. OMB is responsible for overseeing Agency compliance with the Paperwork Reduction Act (PRA) of 1995.

CAPI (computer-assisted personal interviewing)—Face-to-face interview conducted with a laptop computer to assist the interviewer and respondent. Guided by pre-programmed skip patterns, questions are read by the interviewer from the screen, and responses are entered directly into the computer.

CASI (computer-assisted survey information collection)—Signifies the utilization of survey data collection with a computer instead of a simple paper questionnaire. The computer allows a more intricate questionnaire structure that would have been too complex for the paper format.

CATI (computer-assisted telephone interviewing)—A method of telephone interviewing in which a structured questionnaire is programmed into a computer, rather than printed on a form. The interviewer sits before a monitor and asks the questions as they appear on the screen. The interviewer then enters the respondent's replies directly into the computer via a keyboard attached to the terminal.

CLOSED QUESTIONS—Questions offering respondents two or more alternative answers, either explicitly or implicitly, e.g., Yes/No, Male/Female, Strongly Agree/Agree/ Disagree/ Strongly Disagree. When more than two choices are offered, closed questions are sometimes called "multiple choice questions."

CLUSTER SAMPLING—A sample design that deliberately forms geographic groups (clusters) from which the sample is chosen. This is used to reduce travel time and the costs of interviewing, although it increases sample variance. See also STRATIFIED SAMPLING.

CODING—The processing of survey answers into numerical form for entry into a computer. Coding of alternative responses to closed questions (see CLOSED QUESTIONS) can be performed in advance so that no additional coding is required. This is called "precoding." If some items are precoded or keyed directly (numerical amounts), then coding refers only to the

coding of open questions (see FIELD CODING).

COEFFICIENT OF VARIATION—The variance of the statistic divided by the statistic itself, usually expressed as a percentage. This is a common measure of the variability of a statistic.

COVER LETTER—A letter sent or given to the sampled person or entity, explaining the purpose of the survey, and asking for cooperation. It should be convincing without being "leading" (unintentionally indicating how the sponsor wants the respondent to answer, thereby causing biased answers.)

DATA DICTIONARY—see "DICTIONARY"

DEBRIEFING—A meeting of interviewers, supervisors, research analysts, etc., immediately after a pretest or during the early stages of the data collection phase of the main survey. Debriefings alert project personnel to problems with the questionnaire, so they can be corrected before the rest of the interviews are done.

DEMOGRAPHIC CHARACTERISTICS—The basic variables used by survey researchers to classify population groups; for example, sex, age, marital status, race, ethnic origin, education, income, occupation, religion, and residence.

DEPENDENT/INDEPENDENT/INTERDEPENDENT VARIABLES—Dependent variables are the behaviors or attitudes whose variance the researchers are attempting to explain. Independent variables are those variables used to explain the variance in the dependent variables. Variables such as "occupation" or "income" may be dependent or independent, depending on the purposes of the research and the model used. In more complex models, variables may be interdependent; that is, variable A affects variable B while, simultaneously, variable B affects variable A.

DIARIES—Written records kept by respondents to keep track of events that may be difficult to recall accurately later. Diary-keepers are requested to make entries immediately after an event occurs. Sometimes they are compensated with money or gifts for their efforts.

DICTIONARY—A list of survey variables, usually in computerized form. Includes variable name, type (alphabetic or numeric), size (number of characters for alphabetic variables; size and number of decimal places for numeric variables), and description of each possible code. Also called "data dictionary."

FACE-TO-FACE INTERVIEWS—One of the traditional interviewing methods used to collect statistical data. In face-to-face interviewing, a trained interviewer poses questions in the presence of the respondent. See also: TELEPHONE INTERVIEWS, MAIL SURVEYS, and CAPI.

FIELD CODING—The coding of responses to open questions by the interviewer during the interview. When this technique is used, the questionnaire includes a set of pre-printed, coded replies. Instead of writing down the respondent's answer verbatim, the interviewer checks the pre-printed reply that most nearly matches the respondent's reply.

FIELD TEST—See PRETEST and PILOT TEST.

FOCUS GROUP—An exploratory interviewing technique involving small, informal group discussions "focused" on selected topics of concern to the researchers. The discussions are led by a moderator knowledgeable about the subject matter. The participants are selected from the target population or a specific subgroup of the target population.

FRAME—The source or sources from which the survey sample is drawn. The sampling frame may consist of one or more lists of individuals or organizations, but it also may be a set of city blocks, a set of telephone exchanges, etc. Also called LIST.

IMPUTATION—The process of replacing missing or unusable information with usable data from other sources such as responses to other items on the same questionnaire, another questionnaire from the same survey, or external sources (another survey or administrative record.) The use of imputation techniques is rapidly expanding in scope and sophistication due to advances in computer technology.

INTERVIEWER INSTRUCTIONS/DIRECTIONS—Instructions to interviewers regarding which questions to ask or skip, how to enter responses, and when to probe (see PROBES). Interviewer instructions are printed on the questionnaire but not read to respondents.

LIST—See FRAME.

LOADED QUESTION—A question worded in a way that increases the likelihood of a particular kind of response. Loaded questions may legitimately be used to overcome respondent reluctance to report sensitive information. Poorly written questions using "loaded" words or expressions may inadvertently produce biased responses.

MAIL SURVEY—A survey conducted by mailing a questionnaire and cover letter to the sample. For non-respondents, usually supplemented by TELEPHONE or FACE-TO-FACE interviewing.

MULTIPLE-CHOICE QUESTIONS—See CLOSED QUESTIONS.

NONRESPONSE BIAS— nonresponse bias results when units who do not respond to the survey differ significantly from those who do respond. It can also result from nonresponse to individual items on the questionnaire.

OPEN (OR OPEN-ENDED) QUESTIONS—Questions allowing respondents to answer in their own words. The open format encourages respondents to express themselves in language that is comfortable to them. Some open questions are coded during the interview using a fixed set of response categories (see FIELD CODING). Questions that should be answered as a written-in number (age or income, for example) are also considered open-ended.

PILOT TEST—A small field-test replicating the field procedures proposed for the main survey. Usually a purposive sample of 10 to 50 members of the target population is used for the test. A

pilot test is more elaborate than a pretest (see PRETEST) in that the proposed collection procedures, as well as the questionnaire, are tested. Its purpose is to alert the researchers to any operational difficulties not anticipated during the planning and pretesting stage. (Note that some researchers use “pretest” and “pilot test” synonymously.)

PRECODING—See CODING.

PRETEST—A small field test of the questionnaire proposed for the main survey. Usually a purposive sample drawn from various subgroups of the target population is used. Pretests are vital for all Agency-sponsored surveys involving new topics or populations. (Also, see PILOT TEST.)

PROBABILITY SAMPLE—A sample drawn in such a way that each unit (person, household, organization, etc.) in the target population (see TARGET POPULATION) has a known, non-zero probability of being included in the sample. This method of selecting the survey respondents permits statistically valid inferences about the population that the sample is designed to represent.

PROBES—Questions or statements used by the interviewer to obtain additional information from the respondent when the initial answer appears incomplete. Examples of probes are: "How do you mean?" "In what way?" or "Could you explain that a little?"

QUESTIONNAIRE—The complete data collection instrument used by an interviewer or respondent during a survey. The questionnaire includes not only the questions and spaces for the answers, but also interviewer or respondent instructions and an introduction. The questionnaire usually is printed, but non-paper versions can be used on computer monitors (see CAPI, CATI).

RANDOM DIGIT DIALING (RDD)—A method used to select samples for telephone surveys by random selection of telephone numbers within working exchanges. This method permits coverage of both listed and unlisted telephone numbers.

RANDOM SAMPLE/NON-RANDOM SAMPLE—In practice, the term “random sample” is often used loosely to mean any kind of probability sample. “Simple random sample” is a technical term for a sample in which each unit in the population has the same probability of selection and in which all possible samples of a given size are equally likely to be selected. The term “purposive sample” is used to mean any sort of non-probability sample such as a quota sample, a convenience sample, or a judgment sample.

RECORDS—Documents used to reduce memory error on factual questions. Memory errors are unintentional errors in respondent reports caused by forgetting or incorrectly recalling events or details of events. Examples of records are bills, checkbook records, cancelled checks, and inventory accounts.

RESPONSE BURDEN—See BURDEN.

RESPONSE EFFECTS—Variations in the quality of data resulting from the process used to

transmit information from the respondent to the interviewer (where applicable) and ultimately to the data user. The principal sources of variation in quality are the interviewer's performance, the respondent's performance, and the nature of the data requirements and collection methods established by the survey designers.

SAMPLING—Selection of some of the units (a sample) from a population (see TARGET POPULATION) to obtain information that that can be used to characterize or describe the whole population. See PROBABILITY SAMPLE.

SCALE QUESTION—A multiple-choice question that asks respondents to rate a particular quality in themselves or some other person or thing. For example, they may be asked whether they agree or disagree with a statement of opinion, about the frequency of a type of behavior, or whether they like or dislike a certain product. Some scales are entirely verbal (sometimes referred to as "fully-anchored scales"—e.g., Excellent/Very Good/ Fair/Poor.

SELF-ADMINISTERED QUESTIONNAIRE—A questionnaire requiring respondents to read and answer the questions themselves. Self-administered mail questionnaires are one of the traditional methods of collecting survey data. Note that a questionnaire can be considered to be self-administered even if an interviewer is present to hand it out, collect it, and clarify questions, as long as the respondent is primarily responsible for reading the questions and answering them.

SKIP INSTRUCTIONS—Directions on the questionnaire that show the person completing the form which question to ask or answer next, based on the answer to the previous question. Skip instructions make it possible to use a single questionnaire for many different types of respondents because they need answer only those items that are relevant. Also known as "skip patterns."

SOCIAL DESIRABILITY/SOCIAL UNDESIRABILITY—This refers to the perception by respondents that the answer to a question will enhance or hurt their image in the eyes of the interviewer. Examples of socially desirable behavior are voting, being well informed, and fulfilling moral and social responsibilities. Examples of socially undesirable behavior include alcohol and drug abuse, deviant sexual practices, and traffic violations.

STANDARD ERROR—The square root of the VARIANCE. Also known as the "standard deviation."

STATISTIC—A summary measure derived from sample data. "Statistics" (plural), in everyday language, refers to a collection of numerical data. "Statistics" (singular) is an academic discipline concerned with methods of converting numerical data into information useful for scientific research, business decision-making, and other similar purposes.

STRATIFIED SAMPLING—A sample design that draws samples from specific groups (strata) of individuals, thereby assuring representation from each of the groups. This decreases sample variance. It is often used in conjunction with CLUSTER SAMPLING.

STRUCTURED/UNSTRUCTURED QUESTIONNAIRES—Structured questionnaires specify

the wording of the questions or items and the order in which they are asked. They are used for all statistical surveys, regardless of whether the questionnaire is administered by interviewers (in person or by telephone) or by the respondents themselves. Unstructured questionnaires are essentially topic outlines in which the wording and order of the questions are left to the interviewer's discretion. Unstructured survey questionnaires are used primarily in exploratory research for in-depth individual interviews or focus-group studies.

SENSITIVE QUESTIONS—These are questions that are likely to make respondents feel uneasy or threatened and to which they may be reluctant to respond. They include questions about socially desirable and socially undesirable activities (see SOCIAL DESIRABILITY/SOCIAL UNDESIRABILITY). For businesses, sensitive questions include those covering information that they may not want to reveal to their competitors or to government regulatory authorities.

TARGET POPULATION—The complete set of people, households, organizations, businesses, or other units that is of interest and from which the samples for pretests and the main survey are drawn. Also known as UNIVERSE.

TELEPHONE INTERVIEWS—One of the major methods of collecting statistical data. Data are obtained using a structured telephone interview. As in face-to-face interviewing, the interviewer both asks the questions and records the responses. A relatively recent innovation in telephone interviewing is computer-assisted telephone interviewing. (See CATI.)

UNIVERSE—See TARGET POPULATION.

VALIDATION—The process of recontacting respondents to determine whether an interview was actually conducted. In a broader sense, "validation" also refers to the process of obtaining data from other sources to measure the accuracy of respondent reports. Validation may be at either the individual or group level. Examples include the use of financial or medical records to check on reports of assets or health care expenditures. Unless public records are used, validation of individual responses usually requires the consent of both the respondent and the custodian of the records.

VARIABILITY or VARIANCE—For estimates based on samples, variance refers to differences between estimates from repeated samples selected from the same population using the same selection procedures. In a population, it is the average squared distance between the mean and each item. For statistical definitions of variance, see any statistics textbook. See also COEFFICIENT OF VARIATION and STANDARD ERROR.

VARIABLES—See DEPENDENT/INDEPENDENT/INTERDEPENDENT VARIABLES.