

Designing Surveys

Reducing Sources of Error in Data Collection

Contributors: Ronald Czaja & Johnny Blair
Editors: Ronald Czaja & Johnny Blair
Book Title: Designing Surveys
Chapter Title: "Reducing Sources of Error in Data Collection"
Pub. Date: 2005
Access Date: October 15, 2013
Publishing Company: Pine Forge Press
City: A Sage Publications Company
Print ISBN: 9780761927457
Online ISBN: 9781412983877
DOI: <http://dx.doi.org/10.4135/9781412983877.n9>
Print pages: 193-239

This PDF has been generated from SAGE Research Methods. Please note that the pagination of the online version will vary from the pagination of the print book.

<http://dx.doi.org/10.4135/9781412983877.n9>

Reducing Sources of Error in Data Collection

[p. 193 ↓]

This chapter considers how weaknesses in different aspects of data collection can affect the accuracy of our results. There are two main sections: concepts of error and methods to reduce error. The first section introduces the general idea of survey error, discusses the concepts of bias and variance, and provides an overview of unit and item nonresponse. The main measures of survey quality that these core concepts suggest are then summarized. We note the increase in unit nonresponse in recent years and suggest a framework for selecting procedures to address that problem.

This background sets the stage for the second part of the chapter in which we consider methods to reduce the sources of error.¹ Some sources of error are common to both interviewer-administered and self-administered surveys. We will deal with interviewer-administered surveys first, and note issues common to both data collection modes. Last, we take up issues peculiar to self-administered modes, such as conventional mail or Internet data collection.

The Origins of Error

Imagine the perfect sample survey. The survey design and questionnaire satisfy all the research goals. A sampling frame is available that includes accurate information about every population member. The selected sample [p. 194 ↓] precisely mirrors all facets of the population and its myriad subgroups. Each question in the instrument is absolutely clear and captures the dimension of interest exactly. Every person selected for the sample is contacted and immediately agrees to participate in the study. The interviewers conduct the interview flawlessly, and never—by their behavior or even their mere presence—affect respondents' answers. The respondents understand every question exactly as the researcher intended, know all the requested information, and

always answer truthfully and completely. Their responses are faithfully recorded and entered, without error, into a computer file. The resulting data set is a model of validity and reliability.

Except for trivial examples, we cannot find such a paragon. Each step in conducting a survey has the potential to move us away from this ideal, sometimes a little, sometimes a great deal. Just as all the processes and players in our survey can contribute to obtaining accurate information about the target population, so can each reduce that accuracy. We speak of these potential reductions in accuracy as sources of survey error.² Every survey contains survey errors, most of which cannot be totally eliminated within the limits of our resources, and some cannot be eliminated even in principle in a *sample* survey.

The recognition that perfection is unrealistic brings us quickly to some practical questions:

As in all other aspects of the design and conduct of the survey, decisions about how to handle sources of error must balance costs and other resources against the potential harm of not addressing an error source.

Some sources of error are more damaging than others. Before we can assess these sources competently for our particular study, we need to understand sources of survey errors in general. This understanding rests on two concepts: variance and bias. These two error sources, taken together, are used to assess total survey error. For our purposes, only a general understanding of these concepts is necessary, so we will approach them mainly by example.

Variance and Bias

Variance refers to the differences measured in repeated trials of a procedure. This is a useful concept, even though, in most surveys, we do not actually [p. 195 ↓] perform repeated trials. The most common example of variance, already introduced, is that of sampling variance. Recall that if we select a sample of size n and take a measurement on it (i.e., ask each respondent a question such as “How old were you on your last

birthday?”), we produce one sample estimate of a population parameter, in this case, average age.³ Then, if all other aspects of the survey are unchanged, we select a second independent sample of size n , take the same measurement, and produce a second estimate of average age. If this process is continued, we would expect to see variation in our estimate of average age from one sample to another. In other replications of a survey, again holding all procedures constant, we might expect to see random variations in such things as the percentage of respondents who can be contacted for an interview or in the number of refusals to answer a question about personal income. In any particular trial, the magnitude of the variation may be higher or lower than the average across trials.

A similar effect can result from the survey question itself. Imagine asking a sample of respondents, “How many times per month do you go shopping for groceries?” This will produce some sample estimate, say a mean of 3.4 times per month. Now, assume that a week later that same set of respondents is asked the same question again. Many respondents will report the same number as when first asked. But some respondents will report a different number, a bit higher or a bit lower. This may be a result of various factors, including simply thinking about their answers more or less carefully the second time; but the point is that just as we might produce varying estimates from different samples, so we might produce varying estimates from different administrations of a survey question to the same sample. In both instances, if the differences are random we consider them as a source of variance.

By contrast, bias occurs when a measurement tends to be consistently higher or lower than the true population value. In the example just cited, there might be a tendency for some respondents to report that they are older than they actually are, a consequence, for example, of ambiguous wording of the question. Assume we ask simply, “How old are you?” Most respondents will give their current age—that is, their age on their last birthday. But some others may decide, if they are approaching a birthday, that they should report the age they are about to become. It is unlikely that any respondents will report their previous age, even if they just had a birthday. So whatever misreporting occurs is not random, that is, it is not as likely to be in one direction as the other. Misreports are likely to produce higher than actual ages. The resulting sample survey estimate of average age will be higher than the true average. In this case, we would say

that this measurement is [p. 196 ↓] upwardly biased.⁴ Note that this source of error exists in addition to the variance.

More often than not, especially in small-scale research, we do not (or cannot, within our resources) produce empirical measures of the various sources of error in our particular survey, with the major exception of sampling error. Our design decisions concerning the nonsampling error sources are driven by the findings of other surveys and experiments where such measurements have been made. We use the more general (and hopefully robust) findings from such prior methodological research to guide us in rooting out the *likely* sources of error in our study.

In a discussion of the state of survey research as a science, Groves (1987) identified the two approaches to the issue of survey error as *measurement* and *reduction*. That is, there are those “who try to build empirical estimates of survey error and [those] who try to eliminate survey error.” Logically, it would seem that the researcher should be equally concerned with both measurement and reduction and that the researcher's efforts would be directed, based on empirical estimates, to reducing the main sources of error. This is generally not the case (for reasons not discussed here). Nevertheless, the discussion to follow focuses mainly on the reduction of error during data collection, providing guidelines for identifying sources of error and suggesting steps to reduce their effect on the study's results.

Measures of Survey Quality

Survey error arising during data collection can potentially be serious. For example, if nonrespondents differ from respondents, the survey estimates will be biased to some degree. Of course, typically we don't know if respondents differ from nonrespondents on the survey measures because, by definition, we have no data from nonrespondents. In the absence of survey data we often look at indirect indicators. For example, in a general population survey we know from census data the expected distribution of some demographic characteristics: age, sex, race, education, and so forth. If, as is normally the case, we collect some of this information in the survey, we can compare our survey *respondents'* demographic characteristics to the census. Suppose we find

that (a) our respondents underrepresent some demographic groups and (b) some of the substantive survey questions tend to be answered somewhat differently by members of the underrepresented groups compared to nonmembers? Would we be concerned that our estimates may be biased against the under-represented groups? It is important to note that if a group is underrepresented, [p. 197 ↓] then some other group must be overrepresented. Think about why this is true.

The relationship between demographic characteristics and substantive variables, if any, may not be known until after data collection is done and analysis is under way. Our only insurance against these potential biases is a good response rate. The response rate is the percentage of eligible sample units for whom interviews are obtained. This is called the *unit* response rate. We will have more to say about this later, including how it is computed. A similar measure is the *cooperation rate*, which is the percentage of sample members who are interviewed, divided by interviews plus refusals. Consider how these two rates differ.

Ideally, of course, we want to obtain answers from all selected respondents to every questionnaire item. We know that deviations from this objective occur at two levels, the unit, by which we mean a person or household (although it can also be an institution such as a business or school, if that's the survey population), and the item, which is an individual question in our questionnaire. Data that are missing at either the unit or item level can pose potential problems for the quality of our survey estimates. If we fail to obtain any information from some respondents, and for others fail to obtain complete information, our estimates and other analyses may be distorted, sometimes quite seriously.

Unit response rate is the main, and most widely accepted, indicator of survey quality. Of course, respondents who agree to the interview may not answer all the questions. They may refuse to answer particular questions, or inadvertently skip some items, causing *item* nonresponse. While the concerns about item nonresponse are the same as for unit nonresponse, this source of error is usually a less-serious concern. Typically, respondents who agree to the interview answer all, or nearly all, the questions. However, if a survey asks sensitive questions (e.g., sexual behavior or illegal acts), or questions that many respondents simply find too difficult to answer, item nonresponse can become serious. Item nonresponse is usually concentrated in just a few questions.

Both types of response can be affected by interviewer performance. Interviewer behaviors affect respondents' willingness to participate in the survey, and can affect their willingness to answer particular questions.

Interviewers can also impact the reliability of respondent answers. That is, the interviewers themselves can be a source of variability in the survey results, for example by misrecording answers to open-ended items or by being inconsistent in handling respondents' questions or problems. Hence interviewer training and supervision are crucial to effective data collection. [p. 198 ↓] Interviewer performance is seldom measured quantitatively—separately, that is, from the aggregate response rates. Yet their performance is an undeniable potential source of survey variance and bias.

Unit Nonresponse

We are concerned about unit nonresponse because it occurs for reasons that often result in patterns of missing information. For example, suppose that in the crime survey, at the unit level, sample members who are male, or who have less education, or who are elderly living in suburban areas tend to be less likely to cooperate. If such sample members, on average, have different attitudes or experiences than survey cooperators, then our results, which purport to represent the state's *entire* adult population, will be affected. For example, men may be less willing than women to consider sentencing alternatives; people with less education may rate the job police are doing lower than other respondents; or the elderly may be more likely to avoid going certain places because of concerns about crime. Each of these possibilities is speculation, but such patterns are often found in survey results.⁵ To the extent that opinions and behaviors differ by subgroups, their overrepresentation or underrepresentation will affect results.⁶

In the 1970s, as telephone surveys became the predominant means of general population data collection (outside the federal government) Dillman (1978) pointed out the need to examine each step of the survey process in detail, as a contributor to the final success of the survey, particularly response rates. It is useful to return to that advice, taking into account both the new tools and obstacles in conducting surveys today.

If we consider, in sequential order, all the components of survey implementation that may affect the participation decision, we realize that some important factors can easily be overlooked or not given sufficient attention in our planning (Exhibit 9.1). For example, we note that some components that occur before an interviewer ever reaches the respondent may provide information that affect the decision whether or not to participate.

Sometimes a telephone survey is preceded by an advance letter to those sample households with published numbers. As part of random respondent selection, the interviewer will often talk to another household member before reaching the respondent. That person's impressions can affect access to the respondent, the respondent's willingness to participate, or even the respondent's initial understanding of our study. Even if we don't speak to anyone in the household, we may leave a message about the survey on a home recorder. In some studies we may include in the advance letter or leave **[p. 199 ↓]** a message providing a 1–800 number or a Web site URL where information about the survey is available. People who chose not to answer the phone may nonetheless have noticed our phone number or organization name on a caller ID system, perhaps many times.

Each of these steps that precede the interview may affect cooperation and/or data quality. What do we want to do about them? It helps to think this through by putting yourself in the potential respondent's position. This will be a useful exercise. Before proceeding, write a detailed outline of the data collection steps in a telephone, mail, or Web survey. (Exhibit 9.1 below will help you do this.) Then discuss with a colleague what you think could be done at each of these steps to affect a respondent's willingness to participate. One way to do this exercise is for your colleague to “play” the respondent and you “play” the researcher. At each stage of the survey, describe how you plan to carry it out; for example, what information will be in the advance letter, whether or not you will leave a message on answering machines and what it will say, how you will describe the survey to the first person you speak to in the household, and so forth. After describing each step of implementation, your colleague-respondent tells you how he would react to it. Would the planned approach have a positive or negative effect? What questions or doubts might it raise in your colleague-respondent's mind? Each time you get a negative reaction, consider an alternative approach and try that out on the “respondent.” You will find that, without any special expertise, if you and your

colleague try thinking through the process in simply a commonsense manner—*but from the respondents' perspective*—it will produce many ideas, concerns, and insights about how best to conduct data collection.

Some options available to *prevent* unit nonresponse apply to both interviewer and self-administered surveys. These options include designing interesting, logically organized, and nonburdensome questionnaires; using effective devices to provide information about the survey, such as advance letters (or e-mails) and well-crafted introductions or cover letters. Other crucial design factors apply only to one mode or the other.

Exhibit 9.1 Factors Affecting Unit Response

-
- Prior notification about the survey
 - Efforts to reach the respondent
 - Initial contact and respondent selection intermediary (gatekeeper) direct
 - Requesting participation
 - Follow-up efforts
 - Refusal conversion
-

[p. 200 ↓]

For in-person and telephone surveys, it is of critical importance to carefully train the interviewers. In self-administered surveys, respondent instructions must be absolutely clear and easy to follow.

Recent Increases in Nonresponse

Since the mid-1990s several societal and technological factors have affected survey data collection, especially for general population surveys. In some instances, new technologies, such as the World Wide Web and other computer-assisted data collection methods, have created potential opportunities for low-cost data collection; in other instances, technologies such as call blocking have introduced serious difficulties into survey data collection. An increase in the proportion of telephone numbers used solely for Internet access, fax machines and, to a lesser extent so far, cell phones have made sampling household voice telephone numbers more difficult and expensive.

The rise in the volume of telemarketing which many potential respondents find difficult to distinguish from legitimate surveys, is also a serious problem. The practices undertaken by many households to avoid continual bombardment by sales calls have had a detrimental effect on legitimate surveys.

The decline in telephone survey response rates is the most measurable manifestation of these problems. Compared to a decade ago, or even to the last 4 or 5 years, respondents have become more difficult to reach and less willing to participate when contacted. In a review of nonresponse trends in several federally sponsored survey, Groves et al. (2004) show consistent, although not dramatic increases in nonresponse. It is generally agreed that nonfederally sponsored surveys do less well.

So far we do not know how much falling response rates have reduced the quality of surveys and the confidence users should have in their findings. But there is little doubt that continued deterioration in response rates could eventually be very detrimental.

Survey researchers have responded in many ways to halt this trend. While we will not separate new strategies to maintain response rates from procedures that traditionally have been used, it is important to understand that the allocation of survey resources to those aspects of the survey most closely related to obtaining response has generally grown.

Item Nonresponse

Data may be missing randomly; that is, any question is as likely to be missing as any other. Randomly missing data may be a result of, for example, [p. 201 ↓] interviewer mistakes, respondent error (e.g., mistakes in following mail questionnaire skip patterns or instructions), or even because of coding errors. If the amount of randomly missing data is not too large, our results should not be greatly affected. Certainly, if missing data do not exceed a few percent, we are not too concerned about its effects. In such a case, we should not expend many resources to reduce or eliminate the problem of randomly missing data. Unfortunately, when data are missing, most of the omissions are not random.

Even items that are missing as a result of errors by respondents or interviewers are likely to have a pattern. In self-administered surveys, respondents are more likely to skip or not answer questions that are ambiguous, sensitive, or difficult, or that are preceded by an unclear skip instruction. Interviewers may inadvertently encourage nonresponse (or affect answers) to questions that they themselves are uncomfortable asking. We need to be aware of such potential problems during questionnaire design and testing, as well as in interviewer training.

Our Approach: Decisions and Procedures

Next we will review ways in which specific sources of error arising from data collection are addressed. Our approach focuses on two types of decisions, those related to design and those related to procedures for implementing that design. By design, we mean the selection of a data collection method, development of the interviewer training protocol and the data-collection plan (including callback and refusal-conversion efforts). These design decisions, while not irreversible, are, once made, relatively set. In the course of the study, we cannot easily switch from mail to telephone, redesign our training plan, decide to offer cash incentives to participate in the survey, or add a special round of refusal conversion. These decisions, made at the outset, define the main features of data collection—almost always the most expensive stage of the survey—and changing them will usually have serious cost consequences.

We define *procedures* as components of the study's conduct which, although established at the outset of the survey, we can alter or adjust after the study is underway. *Procedures* require ongoing monitoring and microdecisions in the course of the study, such as dealing with problematic interviewer or respondent behaviors and handling problems with gaining cooperation.

To some extent, the distinction between design and procedures is artificial. Still, we think this division will provide a useful framework for separating the main design decisions that must be fixed fairly early from the ongoing microdecisions that occupy most of our attention while conducting survey data collection.

[p. 202 ↓]

As is true of all our decisions about a study, data collection procedural choices are made within the constraint of our budget; each option has some effect on cost, which is interpreted here in a broad sense, encompassing both money and other resources, such as available time of the researcher or unpaid (volunteer) assistants or classmates. Even if we have available to us certain “free” materials and services, they are usually limited, thus requiring decisions about how best to use them. For example, although a university faculty member conducting a survey may not have to pay for a secretary to type the training manual, the secretary's time is not devoted solely to the project. So a choice may have to be made about whether the available time is best used for typing the manual, handling advance letter mailing, or performing other tasks. Finally, sometimes it is necessary to reallocate resources during the survey to address particular problems.

Whether interviews are obtained in person, by telephone, by mail, via the Internet or other means, the data collection process requires routine tasks, such as mailing, dialing sample phone numbers, setting up sample files, and tracking the sample results. These are fairly simple procedures, but they can introduce error. We need to be sure that phone numbers are dialed accurately and that sample tracking accounts correctly for all the sample released for data collection.

Routine processes can be set up to ensure that these largely clerical tasks are done carefully and do not introduce more than trivial error into the survey. These components of the survey are largely record-keeping tasks. Many computerized data-collection tools, such as CATI (computer-assisted telephone interviewing), have utilities that will help with much of this. We will want to think through the steps that must be carried out and for each step, develop a record-keeping form. This mundane procedure will ensure that we do not find ourselves in such situations as omitting followup mailings to some nonrespondents (or wasting resources mailing to sample members we have already interviewed or from whom we have already received questionnaires), not giving each phone number equal call attempts, or neglecting to rework soft refusals,⁷ all of which can contribute to error and wasted resources. However, with moderate attention to such details, these sources of error can be reduced to triviality. The main claim on our resources during data collection will be activities that have the potential to reduce the

more serious survey errors we described above: unit nonresponse, item nonresponse, and interviewer and respondent effects.

Interviewer-Administered Surveys

The respondents' decision to participate in a survey can be affected by a host of factors (see Groves et al., 2004 for a description of some ways to model [p. 203 ↓] these factors). We have little or no control over some of the factors, such as the survey topic, the survey sponsor, a respondent's predisposition to participate in *any* survey, or the social context in which the survey occurs. We try to make the topic appear salient and interesting and we emphasize the survey's importance; but there is only so much we can do without being misleading.

The factors affecting survey response that we can control are the survey administration procedures and, to some extent, interviewer behaviors. We try to design procedures so that each step (see Exhibit 9.6) is implemented to best effect. Yet, procedures such as a good advance letter, a well-written survey introduction, and even monetary incentives can all be undermined by interviewers who are not skillful in their interactions with respondents. A well-conducted survey must be concerned with both data-collection procedures and interviewer effects.

We first turn to data-collection procedures and then consider a number of interviewer performance issues, mainly in telephone studies. Although higher response rates are usually attained in in-person surveys, for cost reasons they are much less common than telephone studies. We include some discussion of in-person surveys because there are exceptions, such as for special populations. In-person data collection also sometimes finds a place in multimode surveys.

It is important that our resources be focused where they will be most effective. In either type of interviewer-administered survey, efforts to address unit nonresponse are often labor intensive and therefore expensive. This is true for both telephone and in-person surveys, though the latter are considerably more costly. In either case, decisions about procedures to enhance cooperation will have an important impact on our limited resources. Our discussion mainly concerns telephone surveys, issues for

any interviewer-administered survey, with additional remarks as appropriate, on points unique to in-person studies.

Administrative procedures and quality control in surveys are greatly aided by computer-assisted data-collection and data-management tools. The acronym CASIC (computer-assisted survey information collection) refers to the wide and expanding array of computer-driven technologies used to collect survey data.⁸ The two major CASIC tools are CATI and computer-assisted personal interviewing (CAPI). Both methods, which have been in use since the late 1970s, allow interviewers to enter survey responses into a computer file while conducting a interview. The main advantage of these systems is quality control. CATI and CAPI systems handle questionnaire skip patterns automatically, reducing interviewer error; limit the range of values that can be entered for some items, for example, requiring two-digit entries for age in a survey of adults; and check some answers against others for internal consistency, for example, alerting the interviewer if someone born in 1970 reports graduating [p. 204 ↓] from high school in 1980. The value of these technologies is that they perform these functions during the interview, rather than at a later, data-cleaning stage, preventing costly recontacts of respondents.

Many of these systems also automatically handle some sample administration and interview scheduling tasks. Most professional and university survey organizations have some type of CATI system; CAPI is, like in-person surveys themselves, far less widespread.

If there is a CASIC system available for our study, then the time and cost to learn how to use the system, including programming the questionnaire into it, must be added to our budget and project plan. For a one-time study, except with the simplest systems, it is probably best to subcontract the programming rather than to try to learn it.

These tools can be very useful to track progress, examine data, and adjust some procedures as data collection progresses. For example, we can more readily make changes in household screening procedures in a computer-assisted environment if we think such changes will make screening more accurate or improve cooperation. Similarly, we can revise the callback plan if we think it can be made more efficient.

We can check the data for patterns of nonresponse. For example, are we having particular difficulties in certain geographic areas (such as central cities of metropolitan areas) or with particular subgroups (such as young males)? By comparing our data to census distributions (in total for in-person studies or just for telephone households), we can get a very good idea of such disproportionate response. If we find such patterns for missing units, we may want to shift our efforts to those underrepresented groups, by either allocating more of the total interviewer hours to these groups or assigning interviewers with particular skills to make the contacts. If, for example, it appears to be more difficult to complete interviews in certain locations, we might increase the proportion of those case assignments given to the more experienced interviewers. If there is a higher refusal rate among men, we might assign those cases to interviewers that are among the better refusal-conversion specialists. Another pattern that can emerge is that particular interviewers are contributing disproportionately to unit nonresponse. If that is the case, retraining or reassignment may be necessary.⁹

It is important to keep in mind the power and flexibility that these CASIC technologies as we discuss the administrative steps in data collection.

Advance Notification

Prior to the start of calling (or of in-person visits) we may want to send an advance letter informing sample households that they have been selected for **[p. 205 ↓]** our survey, about the survey topic and sponsor, and the reason the survey is being conducted. Such a letter should assure confidentiality and provide a phone number (or possibly the URL of a Web site) that respondents can contact for more information about the study. The letter should be brief and, in addition to providing basic information about the survey, explain to the potential respondent why the survey is important.

Returning to the Maryland crime survey, how might we construct an advance letter along these lines? The word “construct” is used intentionally. To form the letter, we want to assemble a set of components, each of which addresses a specific factor that may affect the decision to participate. However, if respondents do not read the letter, it will not serve its purpose. To that end, we strive to keep the advance letter brief and easy

to read. We must include only those points that will most likely affect response, and to express them concisely.

This letter is composed of several brief sections (Exhibit 9.2). Some of the sections simply describe the project, others stress the project's importance, some explain who is sponsoring it and what will be done with the findings, and others state where to get additional information. As an exercise, consider each sentence in the letter one at a time: What is the purpose of the sentence? How important to gaining cooperation do you think the sentence is? Again, try to read it from the perspective of a respondent. If you received this letter, how would you react? Would it help persuade you to participate in the study?

Advance notification is clearly a cost we can choose whether or not to incur. Whether it is worth the cost and time is difficult to know for sure. Also, letters can be sent only to those households with listed telephone numbers. (The vendor from whom you purchase the sample can also provide addresses. If you select your own sample, it can be sent to a vendor who will run in through a commercial database and return listed addresses to you.)

We recommend sending advance letters whenever possible. The letter adds to the legitimacy of the survey and helps to differentiate it from marketing, which is no small issue. It can also serve an additional purpose. For those households with unlisted telephone numbers that raise questions about the survey's legitimacy, the interviewer can offer to send the letter if the potential respondent will provide an address. Having already prepared such a letter will speed up this process.

Reaching the Sampled Respondents

The effort and care we have taken to design and draw a probability sample means little if many of the selected respondents are not contacted [p. 206 ↓] and interviewed. We have noted the potential problems resulting from missing data and touched on the contributions of interviewer and respondent behaviors to this problem. Now we turn to the use of callback procedures. Numerous studies show that repeated callbacks have a major effect on increasing responses from the selected sample regardless of

the data-collection method. Callbacks are the single most-effective strategy in reducing nonresponse.

Exhibit 9.2 Advance Letter

Jones family
2189 Cedar Rd
Adelphi MD 20799

Dear Jones family,

In a few days your household will be called and ask to participate in an important study. The University of Maryland is conducting a survey for the state of Maryland's Summit on Violent Crime.

Your household has been randomly selected for this study. The survey is intended to be representative of all types of households in Maryland, including those that have been touched by crime and those that have not.

All your answers will be kept strictly confidential and the survey results reported only in group form.

The Summit on Violent Crime will use the survey results to help plan crime prevention needs around the state.

Your participation is voluntary, but very important to the success of the study. The interview will take about 10 to 15 minutes. If the interviewer happens to call at an inconvenient time, she will be happy to call back at the time that works best for you.

If you have any questions about this study, please call us toll free at 1-800-314-9090 or visit our web site at www.stopcrime.umd.edu

Thank you in advance for taking the time to help in this important project.

Sincerely,
Robert Ellis, PhD
Department of Research Design

Whether the sampling frame is random-digit dialing (RDD), a list, or some combination, after the first round of calls, the sample is sorted into several groups. These groups include some interviews and identified nonhouseholds that need no further attention. There will also be some early refusals, telephone [p. 207 ↓] numbers whose status (household or not) we have not determined (ring-no-answer numbers), and some households in which the respondent (or household informant) could not be interviewed (at least on the first attempt) because of problems such as difficulty hearing, illness, or languages other than English. In addition, we will have a large number of noncontacts, which include reaching answering machines and finding out that randomly selected respondents are not at home or are busy.

For each of these initial dispositions of the sample numbers, we must decide on a followup callback strategy. How well we do this has a major effect on our success in reaching and interviewing selected respondents and on the costs necessary to do so.

In telephone and in-person surveys, we have some information about the sample units that did not result in an interview on the first call. Even if this information amounts to no more than when the call attempt was made, we can use it to fashion our callback strategy. But we have this information only if interviewers record it carefully.

Interviewers must be trained in the use of forms and procedures for recording, and aggregating, in sample status reports, the result of each call attempt, that is, of each time a sample number is dialed. Regardless of the outcome, the interviewer codes what happened on the call. This kind of coding can be quite elaborate, but we recommend making it no more complex than is absolutely required by the study. The essential call results include completed interview, initial refusal/partial interview, final refusal, noncontact, and other nonrespondent problems. For each of these dispositions the date, time of day, and day of week should also be recorded.

The results of initial calls to each number in the sample determines how we will subsequently “work the sample,” meaning whether and when we will schedule additional call attempts and, to some extent, which interviewers will make the calls. How well we do this has a major effect on both costs and nonsampling errors. Because each call to a sample number has a cost, the objective is to reach a disposition for each number with the minimum effort and to identify eligible sample members and complete interviews with as many of them as possible. Costs are affected because a large part of the total survey budget is for data collection, and calling the sample numbers represents—after the actual conduct of the interviews—the next major portion of data collection costs.

Exhibit 9.3 shows the distribution of initial call results (after one call to each number) and the actual final distribution for the University of Maryland Survey Research Center 1992 State Crime Survey. The first column is the equivalent of our first sample status report. A large number of nonhouseholds are identified after the one call, but more do turn up in the following calls. Initially, we are not sure whether a large group of telephone [p. 208 ↓] numbers are households or not. This category is greatly reduced

(although not to zero) over the subsequent calls. In more than 60% (527 of 835) of the identified households, the selected respondent could not be interviewed on the first call attempt. However, by the end of the study, the noncontact rate is reduced to 5%.

Exhibit 9.3 Maryland Crime Survey: Disposition of the Sample after 1 Call Attempt and after 20 Attempts

	After 1 Call		After 20 Calls	
Total sample	1,816		1,816	
Nonhouseholds	545		702	
Household status unknown	436		53	
Households	835	100%	1,061	100%
Interviews	181	22%	824	78%
Refusals	54	6%	164	15%
Noncontacts	527	63%	49	5%
Problem households (language, problems hearing, illness, etc.)	73	9%	24	2%

Taking each disposition in turn, we consider what types of information we are likely to gather on the first (and subsequent) call attempts and how to use it. Some sample numbers almost immediately drop out because they are not part of the eligible household population: businesses, nonworking numbers, disconnected numbers, government offices, and so forth. However, even some of these numbers may require more than one call.

The second group that is quickly finalized are respondents who are easily contacted and agree immediately to the interview.¹⁰ On average, these are respondents who are home frequently and who quickly comply with the request for an interview. Those at home more often are, as a group, more likely to be elderly, retired, unemployed, and, to a lesser extent, women. While all of these are eligible sample members, clearly they are not a random sample of the population. Using these samples would risk producing very biased population estimates. There are several tempting, but incorrect, procedures that would result in such poor samples. For example, we could draw a very large sample of phone numbers, begin calling and stop when we reached the target number of interviews, leaving the rest of the sample unworked. Or we could make one or two call attempts to each number in a selected sample. Both these approaches would skim off those respondents most available for interview, and both would result in overrepresentation of the demographic groups listed above.

[p. 209 ↓]

Another factor that leads some respondents to immediately agree to do the interview is the survey topic. Respondents who are interested in the survey subject are less likely to put off the interviewer. In the Maryland crime survey, these respondents may be those who have been crime victims themselves or, for other reasons (e.g., because of their profession or the experiences of friends) have strong feelings about crime issues. Again, although they are eligible sample members, a survey that included only these types of respondents would probably provide misleading results. Both the easily reached and the quickly compliant require no further decisions or effort; both also illustrate the nonrandom nature of convenient subgroups of our sample. Now that it is clear that the transition from sampled household to completed interview is not a random process, we must decide how to direct our efforts (resources) to maximize the proportion of the sample that is interviewed and to weed out the ineligible most efficiently.

The sample cases remaining after the first round of calls fall into two categories: households and household status unknown. As we work the sample, these same patterns recur: some respondents are relatively easy to reach and interview; others require greater efforts either to contact or to persuade, or both.

As discussed in Chapter 7 on sampling, we typically begin data collection by releasing a random subsample of numbers for calling. After estimating the interview completion rate (interviews divided by finalized sample) based on this subsample, we release additional random subsamples to attain the total number of interviews we want for the study. When these telephone numbers are released to the interviewers, the call-results pattern is usually similar though not identical.

Number of Contacts. The main factor in successfully contacting a high proportion of selected sample members is simply the number of contact attempts. Surveys that rely on a single contact attempt are likely, in almost all cases, to have serious nonresponse bias. For telephone surveys, the number of attempts typically ranges from 3 to 20, and for mail studies from 2 to 4 (Groves, 1989). We recommend no fewer than five attempts for general population telephone surveys and at least two mailings (each including a questionnaire and a self-addressed, self-stamped return envelope) for

mail studies. If, during data collection, we realize that the planned level of effort is not producing the anticipated response rate, we may need to make some adjustments in our callback plan. Unfortunately, because of cost, we do not have the latitude to make many additional contacts. But adding one or two additional calls for all noncontacted sample members is often useful. This may be especially true if, by examining past call results on each number, we can focus those calls on days and times not yet covered.

[p. 210 ↓]

Answering Machines. In recent years the ownership of home answering machines has greatly increased. Early research (Oldendick & Link, 1994; Triplett, 1994) showed that most of these households can be reached with proper scheduling. Weekend mornings are particularly effective times to contact these households, and they do not appear, once reached, to be any less cooperative than households without answering machines. In the crime survey, after one round of calls, approximately 13% of the identified households were dispositioned as answering machines. By the end of the survey, this number was reduced to approximately 2%. Although this calling strategy is still recommended, there is some evidence that it is becoming less effective. In a large, ongoing immunization survey, conducted for the Centers for Disease Control, independent samples are selected for every calendar quarter. In 2002, the percentage of answering machines at the end of data collection began to rise.

Appointments. It is not uncommon for respondents to telephone surveys to request that they be recontacted at a more convenient time and are willing to set an appointment for the interview.¹¹ It is extremely important that the interviewing effort be organized to ensure that such appointments are kept. Missing them can easily turn a willing respondent into a noncooperator.

Call Scheduling. To properly and cost-effectively work the sample, interviewers must keep an accurate and complete record of the result of each call attempt. This is done through the use of a simple form that the interviewer completes immediately after calling each number. This record allows the interviewing staff manager to look for patterns as to when and when not to attempt additional call attempts for each household. If, for example, a particular household has never answered on weekday late afternoons, it makes sense to shift the next few attempts to later in the evening or to the weekend.

Household informants, once reached, can often provide information about the best time to reach the randomly selected respondent. This information is recorded on the same form.

Using a CATI system that has a call scheduling feature is more efficient than manual methods. However, it will require that some decisions be made and entered into the CATI system. You will need to decide when the first call should be made—day or evening, weekday or weekend—and, if that call is a noninterview, how subsequent callbacks should be scheduled, depending on whether the call result is a ring-no-answer, a refusal, an answering machine, or something else.

In general, for an RDD survey, it is efficient to make the first call during the day because doing so facilitates eliminating businesses and other [p. 211 ↓] nonhousehold numbers. After that, weekday evenings and weekends are when people are most likely to be at home. It is important that callbacks be spread over different days.

Identifying Bad Numbers. We will have difficulty determining the residential status for some telephone numbers even after several call attempts.¹² Because of technical features of the telephone system, a phone number that simply continues to ring each time it is called may or may not be a residence and may or may not be in service. Repeatedly calling a nonresidential number may not only be wasting effort and money but, in the case of the Mitofsky-Waksberg sample design (two-stage RDD), would be costing us the opportunity to replace the number. There are three things we can try: first, schedule a few calls at earlier or later times than usual to try to catch people with, for example, odd work schedules. The few calls at odd times can be useful in bias reduction if the numbers turn out to be residences and an appointment for interview can be set up. Such respondents may be different, both demographically and substantively, than those with more regular schedules. For example, people who work nights may, on the whole, have different responses to some of our questions on crime than do other respondents. Second, when possible, look up the number in a reverse directory.¹³ Third, try the local phone company for information. Unfortunately, the usefulness of this last option varies greatly by the particular company's willingness to help. But it is an inexpensive option to try. Cell phones and computer and fax lines are rapidly increasing. Although most cell phones are usually assigned to different exchanges than land lines,

this is not invariably the case. Also, some respondents may have calls forwarded from a land line to their cell. At this point in the state of technology and sampling frames, all we can do is to train interviewers to be alert for this possibility. Computer and fax lines are more easily recognized and coded, for our purposes, as nonresidential numbers. If a respondent reports that a line is used for voice and computer or fax, classification as residential or not depends on what the respondent says is the *primary* use of the line.

Reworking Refusals. A very important procedure for telephone surveys is refusal conversion. Respondents who refuse to be interviewed initially can often be “converted” on subsequent attempts. While there is little research on this phenomenon, we suspect that conversion is possible partly because the reason for many initial refusals has nothing to do with the survey itself. Keep in mind that while our survey is very important to us, it might mean little to the typical respondent. Some people, for example, caught at a bad[p. 212 ↓] time, will take a moment to set an appointment or just tell the interviewer to try some other time, while other respondents in the same circumstance will refuse, but when recontacted, some will agree to the interview. Professional survey organizations can, using specially trained staff, typically convert 30% to 40% of initial refusers. While the nonprofessional will probably not achieve such rates, a nontrivial number of first refusals can likely still be turned into interviews. In addition to reducing the refusal rate, this procedure can also reduce bias if the initial refusers are generally different in some respects from the initial cooperators.¹⁴

Typically, we will want to let some time pass between the initial refusal and the attempt to convert. If the refusal was a result of a “crisis” in the respondent's household, perhaps in several days it will be resolved. Also, many respondents will not even recall the original contact, so we may not want to even mention it, but, rather, start fresh. The approach we take for trying to convert a particular refusal depends, in part, on what happened the first time. For this reason, it is extremely important for the first interviewer to note the reason, as far as can be determined, for the refusal—the more detail, the better. The record should note the sex of the respondent and indicate, for example, that the person seemed very elderly. Keep in mind that the followup call might reach a different person in the household, if we had not gotten as far as random respondent selection on the first call.¹⁵ Still, it is the circumstances of the first refusal that should be noted. If, for example, a person refused because she was about to leave for work,

it would be foolish to make the conversion attempt at the same time of day as the first call, even though calling at that time increases the chances of reaching the respondent. Similarly, if a respondent refused initially because of a perceived sales purpose, the refusal converter should certainly be prepared to counter this perception quickly if it is raised in the followup call.

Problem Households. In some households, after reaching the randomly selected respondent, we find that the interview cannot be conducted. Some respondents do not speak English at all or not well enough to be interviewed. If we are doing the survey in an area with a large non-English-speaking population (e.g., Miami), we would risk serious coverage bias by not making provisions for interviewing in another language, in this case, Spanish. However, in most small-scale research, we will not have this capability. Nevertheless, these households are still part of the defined population and must be counted as eligible households in our response rate calculations. Similarly, people who are too ill to be interviewed, or who have some physical or mental disability [p. 213 ↓] that prevent them from either hearing, understanding, or answering the questions, are also lost to us.

Minimizing Item Nonresponse

Like unit nonresponse, unanswered questions (item nonresponse) typically do not occur at random.¹⁶ Respondents may be reluctant to answer particular questions (e.g., sensitive items such as whether or not they carry a weapon for protection) or may have difficulty answering others (such as how likely they think it is that in the coming year someone in their household will be a crime victim).

The second class of items that have a higher likelihood of nonresponse are questions that are difficult to answer or require checking records. For example, in a health survey it might be of interest to know how much the respondent spent on prescription drugs in the past year. For many respondents, this will be easy, because the answer is zero. For other respondents, who had many such purchases, this could be very difficult. When faced with such a question, some of those latter respondents will try to answer or hazard a guess; it is quite likely that others will simply say they can't recall. Some items,

like “total household income from all sources,” may be both somewhat sensitive and difficult for certain respondents (e.g., respondents in households with many sources of income).

Similarly, factual questions that request great detail about issues, behavior, or events that are of low salience to the respondent may require more effort than many are willing to expend. While most people will know how many doctor visits they had in the past 6 months, many people will not know how many times they went to a grocery store.

The solution to many of these problems is in questionnaire design. Ask sensitive questions later in the survey, after some rapport has been established between the interviewer and respondent. Preface such questions with reassurances about confidentiality and/or about the need to ask such questions. Provide categories for difficult numerical questions rather than asking for exact figures.

During pretesting note whether particular items seem to have an unusually high proportion of Don't Know or Refused responses. Some item non-response can, of course, result from simple errors in skip patterns, another issue to check carefully prior to and during pretesting.

Other reasons for item nonresponse include poorly written questions. If respondents cannot make sense of the question or have to work to do it, many will not answer it at all. Interviewer behaviors, discussed below, can also affect the item response rate.

Interviewer Effects

[p. 214 ↓]

After we have done all we can by way of instrument design, advance notification, and setting up data collection procedures, gaining cooperation is in the hands of the interviewers. Moreover, interviewers can affect both the willingness of respondents to answer particular questions and the quality of those responses.

Imagine that, for the crime survey, an interviewer contacts a household and selects, at random, an adult who says, “I really don't have time for this. I was just going out. Why

don't you talk to my wife. She keeps up with news and politics more than I do anyway." The interviewer responds with, "OK. She'll probably really like this survey." When the wife comes to the phone, the interviewer starts to read the introduction and the woman interrupts with, "Why is the university doing a poll about crime?" "Well," the interviewer responds, "I'm not too sure, but it probably has something to do with the governor's reelection campaign. You know, to show he's serious about getting criminals off the street." Eventually, the interviewer gets to the question, "In the past year, would you say that the VIOLENT crime situation in the STATE has gotten better, gotten worse, or stayed about the same?" The respondent says, "My daughter tells me she never goes out at night anymore since her next-door neighbor was mugged right outside his house." The interviewer responds with, "So I guess you really think crime's gotten worse."

Each of these interchanges shows how an interviewer can undo the careful design and procedures leading up to and including the interview.¹⁷ First, the random respondent selection procedure is abandoned for convenience; then, the introduction is cut short and a personal comment is interjected that may affect the respondent's reaction to later questions. Finally, instead of probing for a response that matches the answer categories, the interviewer infers the answer from the respondent's comment and puts words in the respondent's mouth. This not-very-farfetched example shows how easily, even naturally, interviewers can affect the quality of the survey.

In-person surveys are subject to more potential interviewer effects than telephone studies. The interviewer's physical characteristics can influence respondent behaviors. For example, if the subject matter of the survey includes racial issues, the race of the interviewer might have an effect. Interviewers' facial expressions and eye contact, which are not pertinent on the phone, may become issues for in-person surveys, particularly those dealing with attitudes or sensitive behaviors. A major difference between the two data-collection modes, as far as interviewers, is the amount of monitoring and supervision that is possible. In-person interviewers, whether conducting [p. 215 ↓] household or special population surveys are much more on their own than interviewers in a centralized telephone facility.

We will spend a good deal of our resources in efforts to control interviewer behaviors through training, monitoring, and, most important of all, showing them how inadvertent,

well-intentioned, actions can be detrimental to the research effort. We must keep in mind that it is the interviewer we depend on to carry out crucial aspects of the survey design. Our job often begins with interviewer selection.

Interviewer Selection

Not only the training but also the selection of interviewers is important. Interviewing is a skill that requires good reading competence, personal interaction ability, and, often, considerable persuasive know-how. We ask interviewers to contact strangers, usually without advance notice, request their cooperation for the interview, explain the purpose of the study, answer their questions about the study or about particular survey items, administer the interview according to instructions, and accurately record the answers. All this must be done while maintaining a professional demeanor, sometimes in trying situations. Not everyone is suited to all the tasks that interviewing requires.

In some instances, the researcher has no role in interviewer selection, for example, when the survey is a class project or a survey organization is hired to collect data. However, even in the case of a class project, it may be that not all the students need to or can properly conduct interviews. As we noted, unlike other aspects of the survey process, interviewing requires both an understanding of what happens in the interview process as well as the skills to carry it out. The best way to find out who has these skills is live practice after some training. One approach is to have everyone participate in the pretest and then decide who should conduct interviews and who should do interview monitoring, data entry, or some other project task.

Whether we are actually recruiting or just making assignments from the interviewers available to us, there are a few guidelines to consider. First, it is useful to have the prospective interviewers go through a structured test. Four areas should be covered: reading and following instructions; gaining respondents' cooperation; reading questions properly; and recording answers accurately. How much skill we can expect in the last two areas depends on how much prior experience, if any, the prospective interviewer has had. But we will find that even novices will differ in how instinctively they react to respondents' reluctance or how naturally, after even a few tries, they can read survey questions. The simplest way to screen possible interviewers is to have a few study-

specific instructions available, along with a draft of the questionnaire. We[p. 216 ↓] should explain to each interviewer the purpose of the study and what we are going to ask them to do. After allowing each interviewer a few minutes to read the relevant instructions and the questionnaire, then a supervisor (or another student) plays the role of respondent, at first going along easily with the request for an interview and answering the questions without incident. On subsequent rounds, the “respondent” varies his or her behavior from this ideal, creating progressively more difficult situations for the interviewer. An observer (or the class as a group) grades the performance. If this test can be set up so that the two parties are in separate rooms actually communicating by telephone, a more realistic assessment of the interviewer's skills, as well as of the interviewer's telephone voice, can be made.¹⁸

Interviewer Training

The training of interviewers should cover both general interviewing skills and the specific requirements of the study at hand (Exhibit 9.4). The amount of general training required depends on the interviewers' prior experience; if possible, it should be conducted by an experienced trainer. The following discussion covers the most basic aspects of training; it supplements a training manual from a professional survey organization and should be followed closely.¹⁹ Survey centers at public universities often will make their general interviewer training manuals available to other (not-for-profit) researchers for free or for a small fee.

The training sessions (and you should plan on multiple sessions) should focus on skill-building practices. At every opportunity, an effective training agenda will emphasize participatory exercises over lectures. We want[p. 217 ↓] constantly to point out behaviors, such as those illustrated at the beginning of this section, that seem natural in conversation and are well-intentioned but are inappropriate in an interview. Rather than simply laying down a series of rules—although we must do that, too—we must show interviewers why such behaviors are detrimental to the project. This point and others are best conveyed through practice and example, not lecture. One aid to doing this is to require that the interviewers read background materials before each training session.

Exhibit 9.4 Key Subjects to Cover in the General Training of Interviewers

-
1. Gaining cooperation
 - Identifying or selecting the correct respondent
 - Explaining the purpose of the survey
 - Persuading reluctant respondents to cooperate
 2. Administration of the interview
 - Getting started
 - Making the transition to the interview
 - Reading questions verbatim
 - Asking nondirective probes
 - Asking all questions and recording answers correctly
 - Following skip instructions
 - Recording answers correctly: open ends
-

Conducting the Interview

A central component of all aspects of interviewer training is role playing, sometimes referred to as mock interviewing. As we have noted, successful survey interviewing requires not so much conceptual understanding (though, of course, a measure of that is essential) as it requires execution of certain skills in gaining respondents' cooperation and properly administering the interview. Such skills are acquired only with practice. The less interviewing experience an interviewer brings to the task, in general, the more practice is necessary. It is also very important to realize that shortcuts on training are false savings. If the interviewers do not practice their skills in a training session, they will, of necessity, practice them with real respondents during the study. The preference is obvious.

In mock interviews, just as in the interviewer-screening process, interviewers take turns playing the role of interviewer and respondent. Each practice round is structured to address a particular set of skills. Practice continues in each area until the interviewer is comfortable, quick, and smooth in handling each situation. Remember that when a real respondent is on the line, there is no time to consult notes, hesitate, or back up and start over. Either an impatient respondent will end the interview or improperly administered questions will result in poor measures.

The exercises should follow the chronological order of interview administration, with at least one exercise to illustrate skill building for each stage of the interview. The number of exercises used and the amount of time devoted to each should be determined by the

nature of the questionnaire, the interviewers' backgrounds, and how quickly they pick up the required skills. Whenever possible, after group instruction, having interviewers practice some of the exercises on the phone will lend realism to the training. The interviewers should know that the final step of the training will be to do a "live" interview with a real respondent. These practice cases should be selected from the study population but not from the actual survey sample. Of course, these final practice cases should be closely monitored.

[p. 218 ↓]

The fundamental instructions to interviewers can be stated simply: Read each question and transition statement *exactly* as written, including all the response categories, without any deviation; ask all the questions, even if you think the answer has already been given; follow the questionnaire's skip patterns; and record answers carefully and, in the case of open-ends, verbatim. These rules are as easy to state as they are to break. This is particularly true of the maxim to read each question verbatim. Anyone who has conducted interviews has run into situations in which it seemed that adding a comment or changing a few words would "clarify" a question or "help" a respondent. The questions must stand on their own, without embellishments of any sort. While we cannot claim that every deviation from the exact wording of the question results in unreliable data, to open the door to invention completely undermines the foundation of uniform data collection. We must strive to develop questions that do not tempt interviewers to such circumvention.

One way to train for this skill is to have each interviewer, in turn, read a question or questions from the survey and have the others critique the reading for verbatim delivery and natural pacing, with proper pauses at punctuation marks and clear pronunciation. The interviewer's tone should be conversational and not sound like a reading from a book. Interviewers often want to skip this practice, feeling that they are literate and can read well enough. It is quickly evident that even well-educated interviewers, on first practice with a new questionnaire, often misread questions, go too fast for many listeners, and make occasional mistakes in pronunciation. These errors seem trivial until there is an irritated or confused respondent on the line. Then these "trivial" errors often lead respondents to hang up or to misunderstand questions. Administering survey questions is not a reading skill but a performance skill.

After the survey introduction and respondent selection, the interview proper begins. We have tried to design the questionnaire so that the first questions ease respondents into the interview; and in most cases the interview will proceed smoothly and uneventfully. But, as we noted in the discussion of behavior coding, there can be some problematic interactions between interviewers and respondents. These have three sources: the questionnaire, the interviewer, and the respondent. We have tried, through the questionnaire design and the training of interviewers to read questions verbatim, to minimize the first two problem sources; we now turn to the third.

Unit Response

The two key training areas are gaining cooperation (unit response) and conducting the interview (including item response). First, we focus on gaining [p. 219 ↓] cooperation. As we have noted, in recent years respondent cooperation in general population surveys has become more difficult to obtain. While the general approach to training interviewers to gain cooperation is largely unchanged, the amount of time spent on gaining cooperation practice exercises is more important and takes a larger share of training time than in the past.

Many respondents will not immediately agree to the interview. Sometimes they want to know more about the survey than what is contained in the interview's introduction. The interviewer needs to be ready to provide this information quickly, concisely, and clearly. A common practice is to have a one page "info sheet" listing the study's sponsor, its purpose, the length of a typical interview, and a thumbnail description of how the results will be used. If, in the course of the pretest or early interviews, other respondent questions or concerns frequently come up, these should be added to the info sheet and circulated to all the interviewers. It is also very useful to provide a phone number respondents can call to verify the legitimacy of the survey.²⁰ A few respondents will want to be reassured about the confidentiality of their answers. But for most reluctant respondents, none of these will be the issue; they will simply not be interested enough to give their time.

The main appeal of a survey is its topic—if the respondent gets to that information. Many refusals in interviewer-administered surveys occur before the topic is mentioned. Survey introductions should be written and interviewers trained to broach the topic—if you think it is an interesting or salient one—as soon as feasible. Advance letters can, of course, help with this.

Topics naturally vary in their appeal to respondents. The crime survey will, in general, be more interesting to a larger number of potential respondents than a study about the public libraries. But even with an interesting, newsworthy topic, many people will not immediately agree to the interview. The interviewer must be prepared for this reaction. In a telephone survey, the interviewer has very little time to gain cooperation. Remember that it is easy for a respondent to hang up.

Two tactics for handling respondents' reluctance are (a) keep the respondent engaged and talking, and (b) address the specific reason the respondent does not want to do the interview. It is very important that the interviewer listen closely to the respondent's reason for not wanting to do the interview, even though many times this will have nothing to do with the survey but simply with bad timing. If the respondent is about to leave for work, is in the middle of watching a ball game, or is dealing with a personal problem at home, the alert interviewer will quickly back off from requesting the interview and try to determine a better time to call back. This approach would seem to be dictated by simple common sense, but it is not unusual for overly aggressive interviewers to push on in such situations, turning reluctance into refusal.

[p. 220 ↓]

Even though the interviewer has read an introduction giving the survey sponsor and topic, some respondents will still suspect it is a disguised sales call or some other solicitation. This is partly because some respondents do not really hear the introduction. Remember that they were not expecting the call; also, many marketing campaigns are disguised as surveys. The interviewer should be ready for this and quickly repeat the sponsor, especially if it is a university or recognizable government agency.

A large number of respondents will simply say that they are not interested or that they don't have time. Although these responses sound different, they often amount to

the same thing, a stock response to end unwanted calls quickly. The main thing the interviewer must try to do is keep the person engaged to avoid the imminent hang up. If the topic is a popular one, the interviewer should try to repeat it, still acknowledging that he or she listened to what the respondent said. For example, “Most people find they like giving their opinions about the crime problem once they get started. It won't take very long.”

Some respondents will say they don't really know much about the issue, particularly if the topic sounds technical, like biotechnology or some aspects of the environment or the economy. The interviewer needs to assure such respondents that the interview is not a test of knowledge.²¹ A response such as, “This is just about your *general* opinions. It doesn't require any special knowledge,” will often suffice.

Finally, in some situations, the reluctance is not from the actual respondent but from a household informant. When we contact a household, any adult can answer the questions used for random respondent selection. If the selected respondent is then someone other than the person on the phone, the interviewer asks to speak to the respondent. At this point, the situation may involve a gatekeeper, that is, a person who answers the phone and does not want to let the interviewer talk to the respondent. The reasons may include some already discussed as well as others. The interviewer strategies are essentially the same, with the added component of expressing the desire to describe the survey briefly to the respondent. If the informant is still reluctant, calling back at another time at least gives a chance to bypass the gatekeeper and reach the respondent directly. Exhibit 9.5 lists suggestions for handling reluctant-respondent situations.

Item Response

Interviewer behaviors can affect people's willingness to respond. For example, a crime victimization survey might include a question about sexual assault. If the interviewer is uncomfortable asking the question, it might [p. 221 ↓] affect how the interviewer asks it, perhaps by reading it very rapidly to get past it and/or lowering his or her voice. Such behaviors will make some respondents more uneasy answering the question than

they would be if the question was asked with the same pace and tone of voice as other questions.

Exhibit 9.5 The Reluctant Respondent: Typical Situations and Some Remedies

Bad Timing

- Don't try to push for the interview.
- Show understanding of the bad timing.
- Use judgment in deciding whether to even ask for a callback time.

Suspicion of Disguised Sales or Solicitation

- Repeat the sponsor and study purpose.
- Assure the respondent that this is not a sales pitch or solicitation; offer an 800 number, if available, or request that the respondent call collect for verification.

No Time/Topic Not of Interest

- If survey topic is a popular one, focus on it; if not, focus on how quickly the interview goes.

Respondent Doesn't Know about Topic

- Focus on the opinion aspect of the survey; downplay any knowledge questions.
- Let the respondent know we're interested in what all people think about the issue.

Gatekeepers

- Ask that the respondent have a chance to hear what the survey is about.
 - Call back to try to avoid the gatekeeper.
-

Poorly trained or supervised interviewers can be even more blatant, with an aside to the respondent such as “you don't have to answer if you don't want to.” There may be instances, as we have noted, when it is important to remind respondents of the confidentiality of their answers or even—at the beginning of a particularly sensitive section of the interview—that if they don't want to answer a particular item, just let the interviewer know. However, we want to control when this happens, lest some interviewers encourage respondents to skip questions they might otherwise answer.

In most surveys, item nonresponse is low. After agreeing to be interviewed, the typical respondent will answer all the questions. But interviewers must be prepared to handle reluctance. The interviewer has two main strategies for [p. 222 ↓] dealing with reluctance. First, let the respondent know why the question is necessary. Some respondents will want to know, for example, why some of the demographic questions (e.g., age, race, or income) are needed. Often, a simple response will suffice: “We want to compare our survey to the U.S. census to be sure it's representative.” Second,

remind respondents that all their answers are completely confidential and will never be linked to their name, address, or telephone number. If, after trying these approaches, the respondent still refuses an item, it is best to skip it and go on. Trying too hard to persuade the respondent to answer a particular question may lead to a breakoff and a lost case. Sometimes, if the item refusal happens early in the interview, it is possible to return to that question later, when the respondent may be more at ease with the interviewer and convinced of the legitimacy of the survey.

A question can be answered but unusable, making it, in effect, a nonresponse. For example, if the answer to an open-ended question (or to an *other specify* option in a closed question) is poorly recorded, the answer may not make sense or otherwise be uncodeable. If the questionnaire has open-ends, interviewers have to have practice recording answers verbatim.

Finally, just as errors in questionnaire skip patterns can cause some respondents to skip questions they should be asked, interviewers can make skip pattern errors when working from paper questionnaires. Again, practice exercises during training will help prepare interviewers to work accurately in the real-time situation of an interview. An important benefit of CASIC systems is to eliminate this type of nonresponse error.

Response Errors

We know that response error can occur when respondents misunderstand questions, cannot recall information, or otherwise have difficulty answering, or even purposely answer falsely. We try to address these problems during instrument design and testing.

The interviewer can also affect some types of response error for better or worse. Interviewers can affect how respondents interpret questions and can sometime influence respondent answers. Most often the effects occur in how interviewers handle some problematic respondent behaviors.

There are three main behaviors of respondents that interviewers must be trained to handle: (a) the respondent does not answer using one of the response categories; (b) instead of answering, the respondent asks a question; and (c) the respondent responds

with a comment about the topic. The interviewer must deal with all of these carefully to avoid affecting the resulting data or having the respondent break off the interview.

[p. 223 ↓]

For each of the problem situations, the interviewer must get the respondent to provide an appropriate answer that can be coded into one of the given categories, yet maintain rapport. Remember that the respondent is volunteering time and trying to be helpful. The interviewer knows that she needs an answer that fits a closed response category and that extraneous remarks cost time and money, but the respondent may view the interaction as something much more like a conversation, with digressions and asides being quite appropriate. It is the interviewer's task to get the respondent on track and keep him there—but tactfully. In doing this, it is paramount that all interviewers handle similar respondent problems in the same way and in a manner that does not affect the respondent's answer choice. It is for these reasons that we instruct interviewers about what sorts of things to expect in the interview and impose strict guidelines on permissible behavior in each situation.

In the mock interviews, “respondents” should take turns deviating from being good respondents while the interviewer practices responding to each situation. After each round, the interviewer's performance should be critiqued by the group.

When interviewers do not receive acceptable responses, their main tool is the probe, a question designed to elicit an acceptable response. Interviewers must learn to recognize which answers are satisfactory and which are not and to use probes that are nondirective, that is, they do not suggest an answer to respondents.

Consider the first crime survey question:

In general, would you say that the crime problem in YOUR NEIGHBORHOOD is very serious, somewhat serious, not very serious, or not at all serious?

If, the respondent's answer is “serious,” the interviewer should probe by repeating *all* the response categories. “Would that be ... (read categories)?” The interviewer should not infer from earlier responses, even to very similar questions, which category to code. Nor should the interviewer's probe make such an inference, as in “So would that be

‘very serious’?” If the categories are such that a partial answer puts the respondent in one subset of response [p. 224 ↓] categories, then the probe can refer to that subset. For example, if the choices are “very satisfied,” “somewhat satisfied,” “somewhat dissatisfied,” and “very dissatisfied,” and the respondent simply says “satisfied,” a proper probe would be “Would that be ‘very satisfied’ or ‘somewhat satisfied’?” A bad probe would be one that did not repeat all the relevant categories. In each case, the respondent must always select the category. This practice should be followed even if the same problem occurs on several questions. Although most respondents will pick up quickly what is wanted, others have to be “trained” by the interviewer to be “good respondents,” and the training must begin at the outset of the interview. Once the respondent is off track, returning is doubly hard.

In training sessions, interviewers should practice suggesting probes they would use in particular situations. It is also useful in practicing probes to note some tempting, but inappropriate, probes. For example, if a respondent answers “serious” to the question above, the interviewer should *not* say, “So, can I put you down as very serious?” Consider an open-ended question such as, “What do you think is the most important problem facing Maryland?” Suppose the respondent answers, “Drugs.” A poor (leading) probe would be “Why do you say ‘drugs’? Is it because people on drugs commit so many crimes?” A better probe would be “Can you explain a little more fully what you mean?”

The second problem type has to do with questions asked by respondents. If the question is off the subject of the interview—for example, the respondent asks the interviewer how she likes doing interviews or what she thinks about a particular issue—the interviewer simply says that interviewers are not supposed to discuss their personal feelings during an interview. If the question is about the survey question, the interviewer must refrain from replying unless there are specific instructions on what to say. One never knows how a comment, even an offhand one, might affect an answer.

In response to the third problem—the respondent makes a comment about the question's topic—the interviewer should refrain from comment and lead the respondent back into the interview. As noted in an earlier example, when, in response to another crime survey question, the respondent mentioned that her daughter had been mugged,

the interviewer should acknowledge hearing what the respondent said but avoid comment. A neutral filler, such as “I see” can serve this purpose.

Exhibit 9.6 provides examples of how to handle common interview problems. The same type of role-playing exercises can be used to practice these as well, mixing up the respondent behaviors, so each interviewer has to think and react quickly. Remember that in all of these situations, a quick—and correct—reaction may mean the difference between an interview and a break-off and refusal; between reliable, unbiased data and bad data.

[p. 225 ↓]

Exhibit 9.6 Conventions for Handling Problematic Respondent Behaviors

Respondent interrupts during the question text with an answer:

- The interviewer should read all of the question. However, if it is part of a series of identically structured items, the interviewer should use her judgment whether to accept the response or finish reading the text.

Respondent interrupts during the response categories with an answer:

- If the question is attitudinal, the interviewer should still read all the categories.
- If the question is factual, the interviewer should accept the response.

Respondent asks what a word or concept means:

- The standard response is to tell the respondent “Whatever it means to you.”
- For particular technical terms, all interviewers may have been provided a uniform definition.
- If the respondent says he or she cannot answer without knowing what is meant by a word or term, the interviewer should code it as a “don’t know.”

Respondent asks for more information or asks about conditions not in the question:

- The interviewer should say, “If your answer is based on just what I’ve told you, would you say . . . ” and then repeat the response categories.
 - If the respondent insists he cannot answer without knowing whether a particular condition applies, or without more information, the interviewer should code it as a “don’t know.”
-

As we have seen, the data collection process is fraught with potential sources of error. The most serious are generally unit nonresponse, respondent error, and interviewer effects. The number of these types of error, of course, is mainly a result of how well the various aspects of the study are designed and implemented. Still, experience has shown that the incidence of such errors is also closely associated with the data-collection method. Exhibit 9.7 is a guide to the typical levels of these sources of error for three data-collection methods.

Quality Control

Several routine procedures are used to track the quality of telephone survey implementation. Each is related to error sources we have noted as important in most telephone studies.

Three procedures are typically used to track interviewers' performance: monitoring, callback verification, and data checking. In monitoring, a small [p. 226 ↓] sample of each interviewer's cases are listened to during the interview, without the knowledge of either the interviewer or the respondent. This procedure, which requires a central telephone facility with appropriate equipment, enables a trained supervisor to evaluate interviewer behaviors on both a general level (e.g., professional interaction with respondents) and on a very specific level (e.g., verbatim reading of all the questions, proper probing of responses when necessary, and proper handling of respondents' questions or difficulties). The percentage of calls monitored depends on available staff time but should be approximately 10% to 20%. More frequent monitoring is advisable at the outset of the study to try to identify any problems as early as possible for new or less-experienced interviewers and for any interviewers who have had difficulty in either gaining respondents' cooperation or conducting the interview. A key aspect of monitoring is immediate feedback—both positive and negative—to interviewers.

Exhibit 9.7 Typical Levels of Error, by Data Collection Method

	<i>Data Collection Method</i>		
	<i>Telephone</i>	<i>Mail</i>	<i>Group Administration</i>
Unit nonresponse	Low	High	Low
Interviewer effects	Medium	N/A	N/A
Respondent errors	Low	Medium-High	Medium-High

Callback verification is sometimes used if monitoring facilities are not available. In this procedure, a sample of each interviewer's cases are recontacted to verify that the interview was done, to check its length, and to ask whether the respondent had any problems or questions and, if so, how the interviewer reacted to them. There is no set rule about how much verification should be done, but a 10% check is frequently

used. Clearly, monitoring is more effective as a quality control procedure and should be considered essential to the conduct of a telephone survey.

For in-person surveys, real-time monitoring cannot be done. Some amount of callback verification is essential. Although interviewer falsification is not a major problem, it does sometimes happen. The temptation to make up an interview (called “curbstoning”) is much greater in the field than in a central telephone facility. CAPI systems, which can record interview length, as well as the day and time of the interview, provide a check and make falsification more difficult. The pay schedule can also affect the tendency to falsify interviews. If an interviewer is paid per completed interview, there is more incentive to falsify than if the interviewer is paid by the hour, regardless of the number of interviews the interviewer completes.

[p. 227 ↓]

Data checking is a procedure in which the collected data are tabulated and inspected during the data-collection period. Data checks can uncover problems in questionnaire administration by interviewers, as well as logic and question-design errors not found earlier. In data checking, frequency distributions are produced for each closed variable, and the verbatim responses to any open-ended items are also generated. These data are examined for such things as skip-pattern errors, patterns of missing items or excessive numbers of missing items, proper use of “other specify” categories (i.e., a check that answers are not being put under “other specify” if they should have been coded into one of the closed categories), consistency between related items, and the clarity of recorded responses to open-ended questions. Data checking is very valuable in spotting problems early enough in the data collection so that, if necessary, corrective actions can be taken.

Self-Administered Surveys: Mail and Internet

Although there is great interest and activity in Internet surveys, whether the data are collected by e-mail or on a Web site, we are still in the early stages of learning how

to conduct them with the rigor and quality we expect in other data collection modes. This state of affairs, in itself, suggests caution in deciding when to use Internet data collection and in selecting survey procedures. A discussion of research on Internet surveys is beyond the scope of this book. Suffice it to note that much current practice follows—whether wisely or not, we cannot yet say—the model of conventional mail surveys. One of the main sources on Internet data collection is by Dillman (2000), whose procedures for mail survey design (Dillman, 1978) have been the standard for more than two decades.

A review of issues and perspectives on Internet surveying can be found in the work of Mick Couper (e.g., 2000). In the following guide to mail and Internet data collection, we will use conventional mail procedures as a foundation and note where Internet practices usually differ. This conservative approach is recommended for the novice researcher.

E-mail Versus Web Data Collection

Internet surveys can be conducted by e-mail or on Web sites. In the early days of Internet surveys, data were often collected via e-mail. The questionnaire was either embedded in an e-mail message or attached to it. The respondent answered the questions and returned the e-mail or attachment. **[p. 228 ↓]** It is still possible to collect data in this manner, but it is not common. We mention it because its modest cost and low reliance on technical skills may, in some instances, make it an attractive alternative.

Three problems led researchers to favor Web site data collection. The wide variety of e-mail systems and of settings within a single system, made it difficult to design a questionnaire that could pass through all systems. For example, many system administrators set size limits on incoming e-mail messages. In those instances, a questionnaire may not get through at all, may be turned into an attachment (which some respondents may be less familiar with handling) or even truncated. Many people have e-mail from one address forwarded to another. If they answer the questionnaire from this second address, the return will not bear the e-mail address the researcher mailed to, making sample tracking difficult.

The second issue is that the available options for designing a questionnaire are typically much more limited in e-mail. If the questionnaire is embedded as an e-mail message, it will simply be text; the responses have to be entered between specified brackets. This increases the chances of respondent error. Skip patterns cannot be handled automatically. Attachments can be executable files that get around some of these problems, but they may be large.

The third issue also has to do with attachments. Because many computer viruses are spread by attachments, many respondents are wary of opening any attachment sent from some person or organization they do not know. This reluctance, of course, will depress response rates. Still, for small surveys with some populations (e.g. students at a university or some membership group) an e-mail survey may be cheap and efficient. We have pointed out some ways that e-mail data collection can contribute to survey error; can you think of some others?

Web surveys are far more common. For this reason, the following discussion about Internet data collection is limited to Web surveys. In these surveys, e-mail is used to contact the respondents, but a URL embedded in the e-mail takes the respondent to a Web site to complete the questionnaire. The Web survey discussion does not mention particular software. Just as with CATI systems, several alternative systems are available. Like other software, the systems change over time or disappear from the market. And just like other software, one has to be wary of bugs, concerned about support, and careful that it is compatible with any other systems it must interact with.

Unit Response

Unit nonresponse is the principal source of nonsampling error in mail surveys, which usually achieve lower response rates than either interviewer-administered general population surveys or surveys of many special [p. 229 ↓] populations. Response rates are typically even lower for Web surveys. For both mail and Web surveys intense followup contacts are essential to obtaining acceptable response rates.

One reason for low response rates in these administration modes is that such surveys are simply easier to decline. An interviewer automatically brings some small amount

of social pressure to at least listen to the appeal to participate. The interviewer can often tailor her appeals to specific respondent reservations. Surveys that don't use interviewers must rely totally on written or graphic material to convince respondents to participate and then provide instructions to complete the questionnaire. It is difficult to know beforehand how well the materials we design will work. Unlike interviewer-administered surveys, conventional pretesting will not provide much information about how the letter and questionnaire influence response. Poor materials will get a low response rate, which is feedback of a sort, but the conventional pretest will provide little information about what to change. As we have noted, the use of focus groups and cognitive methods, in conjunction with conventional testing, will probably be more helpful.

The problem of nonresponse can be greatly reduced if such surveys are limited to special populations, such as members of an organization, employees in an industry or workplace, students at a university, or some other relatively homogeneous group.

Unlike topics in telephone surveys, the topic of a mail or Internet survey can be a major factor in the decision to respond or not. The respondent is told the topic in the cover letter, can guess it from the envelope's return address, and, of course, can flip through the questionnaire. Usually surveys of special populations are done because the topic particularly applies to them; in those cases, making the topic prominent will generally be an advantage.

The option to preview the questionnaire is usually not possible in Web surveys; whether that is an advantage or a detriment depends on the questionnaire. If the instrument is relatively short and appears easy to complete, respondents would notice this and might be more likely to participate. In the absence of the option to preview the questionnaire, it is important to give some indication of its length in the cover letter.

Prior Notification About the Survey

Prior mail or e-mail notification should be given to respondents informing them that a survey is going to arrive soon. This has become a common practice in conventional mail surveys (Dillman, 2000). It can be a useful way to identify bad addresses (via

returns by the post office or e-mails that bounce back as undeliverable). If you have reason to believe that your sampling frame may be somewhat out of date or for other reasons may contain [p. 230 ↓] errors, a prenotification can provide an estimate (not a perfect measure) of frame error. This can tell you how much additional sample you need to send in order to reach the target number of respondents in your questionnaire mailing. Also, if the questionnaire is expensive to print and mail, you will save money by eliminating mailing to some bad addresses.

A prior e-mail notification should be sent in Web surveys. Doing so can identify some bad addresses. A more important purpose is its use as an additional opportunity to let respondents know that a legitimate survey is going to be sent to them. Because of the proliferation of spam and other unwanted e-mails, this is no small issue.

Followup Contacts

After the prenotification, how many additional attempts to obtain an interview are made will depend on available resources. It is essential both in mail and Web surveys that *some* followups be sent, even if the number we recommend is not possible. The final effort should use an alternate approach, as described below, which can be very effective in boosting response.

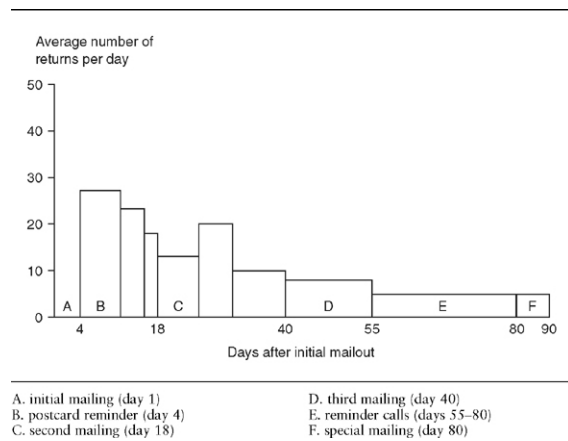
In conventional mail surveys, we follow Dillman's current recommendation of four contacts by first class mail and an additional special contact (Dillman, 2000). These include

This same pattern can be used for Web surveys. Of course, in Web surveys, none of the mailings include a questionnaire. The questionnaire always remains posted on the survey Web site. However, in a conventional mail survey, each mailing (except the reminder postcard) should include a copy of the questionnaire; primary reasons for nonreturns is simple misplacement, loss, or (perhaps inadvertent) discarding of the questionnaire (Dillman, 1978, 2000).

In both mail and Web surveys, mailings should be spaced optimally over the data-collection period. The spacing of the followup mailings depends on the flow of returns

from the prior mailing. Until returns dwindle to a trickle, there is little advantage (or economic sense) in sending out additional mailings. Although the response pattern will differ for Web surveys, where the response is much more rapid, the logic of waiting for returns to decline is the [p. 231 ↓] same. Consequently, we must keep track not only of how many conventional returns we receive but also of how many we receive on each day after the initial mailing and each followup mailing.

Exhibit 9.8 Return Rate in a Mail Survey



Response Patterns

Typically, regardless of the survey population, a few people will respond right away to a mail survey; this is followed by a period of relatively high returns, with about the same number each day. The reminder postcard will help to maintain this flow for a while, but then it will decrease. At some point during this decline, we should send out the second mailing, including a questionnaire. To illustrate this typical pattern, Exhibit 9.8 shows the return rate for a national survey of county officials on the subject of how county priorities are determined. In a Web survey, the response pattern is much more clustered toward the beginning of each mailing; those who intend to respond typically do it almost immediately.

During a mail survey, it is useful to track the number of returns per day (or the percentage of the target sample received) as a way to spot [p. 232 ↓] response-rate problems. Exhibit 9.8 shows that after the initial mailing there is a lag before returns begin to come in. During this waiting period, a postcard reminder is sent to all sample members. There is a large first-wave response, which trails off after a couple of weeks. At that point the second mailing, with a questionnaire, is sent out, with a similar result. Later phone followups to nonrespondents bring in a few more cases, which finally dwindle to the point that we end data collection.

If, after tracking the initial mailing, we conclude that the rate of return is not what we anticipated, what should we do? First, we should determine whether there are any patterns to the nonresponse. If some sample subgroups are less willing to cooperate than others, we may want to shift some resources to them. This may be done by planning an extra mailing to them or allocating them a larger share of the nonrespondent sample to whom we send a special mailing. If monetary incentives are being used, some increased payment to such groups may help (Warriner, Goyder, Gjertsen, Hohner, & McSpurren, 1996).

Reasons for Nonresponse

Second, it is almost always useful to try to determine the reasons for non-response. If telephone numbers are available, a call to some nonrespondents to determine their reasons for not returning the questionnaires may help. If we find in the survey of county officers, for example, that the elected county board members are often not receiving the questionnaire promptly because many are only part-time on the board and spend most of their time at another job, we may want to channel some resources toward telephone calls to obtain these other addresses. Our second mailing would be redirected to the alternate addresses. Remember that in this case, it is the person, not the address, that is the sampling unit.

We might consider these kinds of issues on two other studies. First, we consider a regional survey of recreational boat owners about the problems of trash in public waterways; and second, a study of academic survey organizations about methods they use to develop questionnaires. If we should find that many of the recreational boat

owners are concerned about the confidentiality of their answers to questions about trash disposal from their own boats (because they know some disposal practices are illegal), we may want to tailor the cover letter for the second mailing to address this issue. For the survey of survey organizations, we might find that there are delays because some requested information (e.g., the number of surveys that used particular pretest methods) is not readily available. A subsequent cover letter might acknowledge this likely problem but base its appeal on the need [p. 233 ↓] for the study to have a good response rate because the results will be reported at an industry conference. The organization's respondents, having often been in the same position themselves, may be motivated by this appeal to put in the extra required effort.

The options open to us during a study are limited, but sometimes a small adjustment can have enough of an effect to boost overall response rates or those of particular subgroups to the extent that we do not have to abandon plans for their separate analysis. The key point, with this and other procedural issues, is that if we see that things are not going well, we take what steps are available to us to improve the situation. Things sometimes turn around on their own, but a good researcher does not count on it.

The information to be tracked in a Web survey is the same. Of course, we are dealing with questionnaires completed on-site (not returns), and the technical methods for tracking the count of completed questionnaires is different. The main difference in Web survey tracking is that the tracking system needs to be checked as part of the pretest. It is a mistake to assume that a system—whether “out of the box” or designed specifically for your study—will work exactly as you expect it to work. When main data collection is underway is not the time to discover problems.

Respondent Selection

Random respondent selection (within some sample unit) is rarely used in mail or Web surveys. These modes are not recommended for general population surveys, so selection within a household is not an issue. Most often our frame will be a list with names of eligible respondents. However, in some business or organization surveys, we may have only a title. Even worse, we may have to ask that our questionnaire

go to the (unnamed) person who performs a certain function in the organization (e.g., sets personnel policies), has a certain title (e.g., chief financial officer), or has particular knowledge about the organization (e.g., knows about expenditures for various purposes). The farther away from naming a specific person we are, the greater the likelihood the questionnaire will not reach the correct person and the more difficult it is to obtain a good response rate. Imagine that you are designing a survey of a particular type of business or organization. Think about who would be the person you want to complete your questionnaire and how you would identify or locate that person in the survey.

We lose a lot of control in mail and Web surveys, compared to interviewer-administered studies. For mail and Web surveys, we do not know if an intermediary is the first person to see the mailing or not. The likelihood of this depends heavily on the population surveyed. In business or organization [p. 234 ↓] surveys, conventional mail may well be screened before reaching the target respondent; the higher that person is in the organization, the greater the likelihood this will occur. If an alternate means of contact (e.g., telephone) is affordable, it can be very useful in identifying and persuading intermediaries (i.e., gatekeepers) to pass the questionnaire on to the respondent.

We depend on someone to read and follow the instructions about who should complete the questionnaire. Even if the instructions are well written and are followed, in many instances (especially in surveys of very large organizations), the respondent may delegate the completion of the questionnaire to someone else. What types of response error this leads to depends on the study.

Refusal Conversion

Unfortunately, in mail or Web surveys, the range of tactics used in telephone and in-person surveys is not available. Rarely does a respondent send back a mailed questionnaire saying he refuses, let alone why he refuses. So we cannot separate conscious decisions not to participate from nonreturns for other reasons. Thus, it is difficult if not impossible to tailor our followup efforts to the reasons for nonresponse. We are reduced to using general and multiple appeals. For example, a cover letter accompanying a followup mailing might mention things like the respondents' busy

schedules or the possible misplacing of the first questionnaire (or having problems accessing the Web site). But the reason for the nonreturn may have nothing to do with either. One method we can make use of on mail studies is the special delivery mailing and/or request for address correction. Both may be effective. The request for address correction should be done early to help ensure that subsequent mailings are sent to the correct address. Special delivery or other special methods, because of their cost, have to be focused on fewer cases later in the study. When using these types of mailings, it is important to keep in mind that the mail should be deliverable whether or not the respondent is at home. Having to go to the post office or make other special efforts to get the mailing are more likely to increase sample members' resentment than to increase response rates. (See Dillman, 1978, for a treatment of this issue in general population mail surveys.)

Samples of Nonrespondents

Finally, we want to consider samples of nonrespondents in situations in which unit nonresponse is high. Samples of nonrespondents are routinely used in mail surveys to assess differences between respondents and [p. 235 ↓] nonrespondents, as well as to increase the overall response rate. The method involves selecting a small number of sample members who did not return the mail questionnaire and trying to obtain the interview by another means, usually by telephone. We then compare their characteristics to those of the mail respondents to assess the direction and magnitude of possible bias in the mail sample. While a description of statistical adjustments based on such results is beyond the scope of this book, it is useful simply to know that these "reluctant respondents" are more often of a particular gender or background than the others, or that they tend to be more or less in favor of some proposal referred to in the questionnaire. Nonrespondent sampling is clearly a tactic that has to be planned in advance, as part of the overall design of the survey. But exactly how it is used procedurally can be determined after more is known about the nature of the nonresponse.

Item Nonresponse

Handling missing or ambiguous data can be a major issue in mail surveys. In Web surveys, it is possible to program the questionnaire so that the respondent cannot move forward if the respondent skips a question. This may be a mixed blessing. On the one hand, it certainly reduces item nonresponse; on the other hand, a respondent who feels strongly about skipping a question has the choice of entering false data or refusing to finish the questionnaire. An ethical issue also arises. We typically tell respondents they can skip any question they don't want to answer. We need to think carefully about whether we want to undercut this pledge simply because technology makes it possible.

When a mail questionnaire is received with unanswered items, there are three choices: ignore the items and code as “missing” in the data set; try to determine what the answer “should be” and insert that value into the data set; or recontact respondents to obtain the missing answers.

In making decisions about which option to choose, we must have a sense of which variables are crucial to our study. Obviously, all questions are valuable or we would not have included them, but there are priorities. For example, if the main analysis concerns racial differences, then missing race makes the case much less useful to the study. On the other hand, if differences by race are not central to the study purpose, we would be much less concerned if this item is missing.

If the amount of missing data is both very small (say, less than 3%) and the items are not crucial variables, we want to select the options that use fewest, or no, resources. Either ignore it or see whether the answer can be determined from some combination of other answers in the questionnaire. For example, if gender is missing but elsewhere in the questionnaire the [p. 236 ↓] respondent reported having attended a Catholic high school known to be an all-girl institution, we can then confidently enter “female” for the missing item. Great care must be taken in using this process. Errors in such imputation can make the data worse, not better.

When it appears that data are missing because of a misunderstood “skip” instruction and the item is a key variable, it may make sense to recontact the respondent to obtain

the information. This is especially appealing if the recontact can be done quickly by telephone. If the calls are long-distance, this cost must be factored in. It is useful to have a small amount of resources—funds and schedule time—set aside for such contingencies.

Tracking

The procedures for tracking mail surveys are fairly straightforward (see Mangione, 1995; Dillman, 2000). We have noted the necessity for tracking mail returns by date of return in order to plan future followup efforts. This sort of tracking can be handled simply with a spreadsheet and a simple system of case id numbers. For Web surveys, tracking can be more complex, but depends greatly on the software application being used. Although consideration of alternative Web survey software is beyond the scope of this book, we do caution the first-time Web researcher to determine the capabilities and ease of use of software for both questionnaire administration and for tracking.

Notes

1. The most comprehensive, though somewhat dated, treatment of survey errors is R. Groves's *Survey Errors and Survey Costs* (1989).
2. Note that the term *survey error* is not particularly directed at mistakes per se, such as incorrectly keying in data from a mail questionnaire, but is broad enough to include these as well.
3. In itself, the process of taking a measurement is also subject to error. Such measurement errors are, of course, not restricted to surveys but are found throughout the empirical sciences.
4. Note that this has nothing to do with intending to elicit a false report, that is, to induce people to report their age as younger or to answer some other question in a certain way. As opposed to everyday usage, bias in survey research indicates effect, not intent.

5. This is one reason we standardly collect demographic information as part of the survey. It allows us to compare our sample's demographic distribution to that of all people in the survey area meeting the target population definition.

[p. 237 ↓]

6. There are weighting adjustments that can be made after the data are collected to address this problem to some extent. However, as we will see in Chapter 10, weights bring to the survey analysis their own complications and increases in other error sources.

7. Some respondents who refuse to be interviewed are more adamant than others. The less adamant refusals are termed *soft*.

8. In fact, though only recently introduced, the term has already been extended to include all aspects of survey computerization, such as data transmittal, processing, analysis, and dissemination.

9. In a survey conducted as part of a research class, this would mean moving those individuals to coding, monitoring, or other noninterview work.

10. The response rate is defined as interviews divided by eligible households. The cooperation rate is interviews divided by the sum of interviews, partial interviews, and refusals. At the end of the study, response rate is the main measure of data collection success. During data collection, the cooperation rate is a better indicator of how well interviewing is going. Why is this?

11. This is actually the rule, rather than the exception, for surveys of organizations. In planning that type of survey, allowance should be made for one or two calls simply to set up the appointment with the target respondent.

12. For general population surveys, approximately 80% to 90% of calls should be made between about 6 p.m. and 9 p.m. on weekdays, Saturday and Sunday late mornings and afternoons, and Sunday evenings. Calls at other times are mainly to screen out business numbers and to reach respondents with unusual schedules.

13. Reverse directories are telephone directories that are arranged in numerical order by telephone number, rather than alphabetically. Haines and R. H. Donnelley are the main publishers of these directories.

14. In a study comparing respondents from converted refusals to others from several surveys, Blair and Chun (1992) found that there were some small differences in the number of “don't know” answers and item nonresponses between initial cooperators and those who were converted from refusals.

15. Once a random respondent has been selected in a household, that person remains the target respondent for the survey regardless of what happens on subsequent call attempts.

16. One exception can occur when respondents simply overlook an item in a mail survey. For example, in questions with the instruction Mark All That Apply, respondents select fewer items than when asked to answer the same question Yes or No for each item in the series, in effect producing more item nonresponse (Rasinski, Mingay, & Bradburn, 1994).

17. One should also be aware that these interchanges between interviewer and respondent may potentially affect subsequent respondent behaviors (see Couper, 1997).

18. If this type of training is to be conducted properly, a speakerphone should be placed in one or both rooms.

[p. 238 ↓]

19. The University of Michigan Survey Research Center's *General Interviewing Techniques* (Guensel, Berckmans, & Cannell, 1983) is an excellent source.

20. This should preferably be a 1–800 number if the survey is not local. In any case, it should be a number that is staffed during regular working hours.

21. It is important that surveys in general, and particularly those that might on their surface sound forbidding, not begin with knowledge questions. Remember that all

respondents, but especially reluctant ones, still have the option to break off if the first few questions are difficult or make them uneasy.

<http://dx.doi.org/10.4135/9781412983877.n9>