Sampling

How well a sample represents a population depends on the sample frame, the sample size, and the specific design of selection procedures. If probability sampling procedures are used, the precision of sample estimates can be calculated. This chapter describes various sampling procedures and their effects on the representativeness and precision of sample estimates. Two of the most common ways of sampling populations, area probability and random-digit-dialing samples, are described in some detail.

There are occasions when the goal of information gathering is not to generate statistics about a population but to describe a set of people in a more general way. Journalists, people developing products, political leaders, and others sometimes just want a sense of people's feelings without great concern about numerical precision. Researchers do pilot studies to measure the range of ideas or opinions that people have or the way that variables seem to hang together. For these purposes, people who are readily available (friends, coworkers) or people who volunteer (magazine survey respondents, people who call talk shows) may be useful. Not every effort to gather information requires a strict probability sample survey. For the majority of occasions when surveys are undertaken, however, the goal is to develop statistics about a population. This chapter is about sampling when the goal is to produce numbers that can be subjected appropriately to the variety of statistical techniques available to social scientists. Although many of the same general principles apply to any sampling problem, the chapter focuses on sampling people.

The way to evaluate a sample is not by the results, the characteristics of the sample, but by examining the process by which it was selected. There are three key aspects of sample selection:

1. The sample frame is the set of people that has a chance to be selected, given the sampling approach that is chosen. Statistically speaking, a sample only can be representative of the population included in the sample frame. One design issue is how well the sample frame corresponds to the population a researcher wants to describe.

2. Probability sampling procedures must be used to designate individual units for inclusion in a sample. Each person must have a known chance of
selection set by the sampling procedure. If researcher discretion or respondent characteristics such as respondent availability or initiative affect the chances of selection, there is no statistical basis for evaluating how well or how poorly the sample represents the population; commonly used approaches to calculating confidence intervals around sample estimates are not applicable.

3. The details of the sample design, its size and the specific procedures used for selecting units, will influence the precision of sample estimates, that is, how closely a sample is likely to approximate the characteristics of the whole population.

These details of the sampling process, along with the rate at which information actually is obtained from those selected, constitute the facts needed to evaluate a survey sample.

Response rates are discussed in Chapter 4, which also includes a brief discussion of quota sampling, a common modification of probability sampling that yields nonprobability samples. In this chapter, sampling frames and probability sampling procedures are discussed. Several of the most common practical strategies for sampling people are described. Interested readers will find much more information on sampling in Kish (1965), Sudman (1976), Kalton (1983), Groves (1989), Henry (1990), and Lohr (1998). Researchers planning to carry out a survey almost always would be well advised to obtain the help of a sampling statistician. This chapter, however, is intended to familiarize readers with the issues to which they should attend, and that they will likely encounter, when evaluating the sampling done for a survey.

THE SAMPLE FRAME

Any sample selection procedure will give some individuals a chance to be included in the sample while excluding others. Those people who have a chance of being included among those selected constitute the sample frame. The first step in evaluating the quality of a sample is to define the sample frame. Most sampling schemes fall into three general classes:

1. Sampling is done from a more or less complete list of individuals in the population to be studied.

2. Sampling is done from a set of people who go somewhere or do something that enables them to be sampled (e.g., patients who received medical care from a physician, or people who attended a meeting). In these cases, there

is not an advance list from which sampling occurs; the creation of the list and the process of sampling may occur simultaneously.

3. Sampling is done in two or more stages, with the first stage involving sampling something other than the individuals finally to be selected. In one or more steps, these primary units are sampled, and eventually a list of individuals (or other sampling units) is created, from which a final sample selection is made. One of the most common such sampling schemes is to select housing units, with no prior information about who lives in them, as a first stage of selecting a sample of people living in those housing units. These multistage procedures will be described in more detail later in this chapter.

There are three characteristics of a sample frame that a researcher should evaluate:

1. **Comprehensiveness**, that is, how completely it covers the target population.

2. Whether or not a person’s probability of selection can be calculated.

3. **Efficiency**, or the rate at which members of the target population can be found among those in the frame.

**Comprehensiveness.** A sample can only be representative of the sample frame, that is, the population that actually had a chance to be selected. Most sampling approaches leave out at least a few people from the population the researcher wants to study. For example, household-based samples exclude people who live in group quarters such as dormitories, prisons, and nursing homes, as well as those who are homeless. Available general lists, such as those of people with driver’s licenses, registered voters, and homeowners, are even more exclusive. Although they cover large segments of some populations, they also omit major segments with distinctive characteristics. As a specific example, published telephone directories omit those without landline telephones, those who have requested that their numbers not be published, and those who have been assigned a telephone number since the most recent directory was published. In some central cities, such exclusions amount to almost 50% of all households. In such cities, a sample drawn from a telephone directory would be representative of only about half the population, and the half that is represented could easily be expected to differ in many ways from the half that is not.

A growing threat to telephone surveys is the increase of cell phone use. Most telephone surveys have depended on sampling telephone numbers that can be linked to households. Households that are not served by any “landline”
are excluded using the techniques mos: often used to draw samples for telephone surveys. Those households which are served only by cell phones are therefore left out of such samples (Blumberg, Lake, & Cynannon, 2006).

E-mail addresses provide another good example. There are some populations, such as those in business or school settings, that have virtually universal access to e-mail, and more or less complete lists of the addresses of these populations are likely to be available. On the other hand, as an approach to sampling households in the general population, sampling those with e-mail addresses leaves out many people and produces a sample that is very different from the population as a whole in many important ways. Moreover, there is not currently a way to create a good list of all or even most of those who have e-mail addresses.

Two recent innovations, spurred by the desire to conduct survey via the Internet, deserve mentioning. First, large numbers of people have been recruited via the Internet for survey participation. These people fill out initial baseline questionnaires covering a large number of characteristics. The answers to these questions can then be used to “create” a sample from the total pool of volunteers that roughly matches those of the whole population a researcher wants to study. When such a “sample” is surveyed, the results may or may not yield accurate information about the whole population. Obviously, no one is included in such a sample who does not use the Internet and is not interested in volunteering to be in the surveys. Often, the same people participate in numerous surveys, thereby further raising questions about how well the respondents typify the general population (Couper, 2007).

In an effort to address some of these concerns, another approach is to carry out a telephone survey to recruit a pool of volunteers for Internet surveys. Those without access to computers may be given a computer to use. Even with those efforts, the “sample frame” consists of that subset of the population that lives in a household with telephone service and agrees to be part of a pool for research studies. From a statistical perspective, statistics based on samples from that pool do not necessarily apply to the balance of the population. Rather, in both of the examples above, those responding to a survey can only be said to be representative of the populations that volunteered or agreed to be on these lists (Couper, 2007). The extent to which they are like the rest of the population must be evaluated independently of the sampling process.

A key part of evaluating any sampling scheme is determining the percentage of the population one wants to describe that has a chance of being selected and the extent to which those excluded are distinctive. Very often a researcher must make a choice between an easier or less expensive way of sampling a population that leaves out some people and a more expensive strategy that is also more comprehensive. If a researcher is considering sampling from a list, it is particularly important to evaluate the list to find out in detail how it was compiled, how and when additions and deletions were made, and the number and characteristics of people likely to be left off the list.

**Probability of selection.** Is it possible to calculate the probability of selection of each person sampled? A procedure that samples records of visits to a doctor over a year will give individuals who visit the doctor numerous times a higher chance of selection than those who see the doctor only once. It is not necessary that a sampling scheme give every member of the sampling frame the same chance of selection, as would be the case if each individual appeared once and only once on a list. It is essential, however, that the researcher be able to find out the probability of selection for each individual selected. This may be done at the time of sample selection by examination of the list. It also may be possible to find out the probability of selection at the time of data collection.

In the above example of sampling patients by sampling doctor visits, if the researcher asks selected patients the number of visits to the physician they had in a year or if the researcher could have access to selected patients’ medical records, it would be possible to adjust the data at the time of analysis to take into account the different chances of selection. If it is not possible to know the probability of selection of each selected individual, however, it is not possible to estimate accurately the relationship between the sample statistics and the population from which it was drawn.

"Quota samples," discussed near the end of Chapter 4, are another common example of using procedures for which the probability of selection cannot be calculated.

**Efficiency.** In some cases, sampling frames include units that are not members of the target population the researcher wants to sample. Assuming that eligible persons can be identified at the point of data collection, being too comprehensive is not a problem. Hence a perfectly appropriate way to sample elderly people living in households is to draw a sample of all households, find out if there are elderly persons living in selected households, then exclude those households with no elderly residents. Random-digit dialing samples select telephone numbers (many of which are not in use) as a way of sampling housing units with telephones. The only question about such designs is whether or not they are cost effective.

Because the ability to generalize from a sample is limited by the sample frame, when reporting results the researcher must tell readers who was or was not given a chance to be selected and, to the extent that it is known, how those omitted were distinctive.
SELECTING A ONE-STAGE SAMPLE

Once a researcher has made a decision about a sample frame or approach to getting a sample, the next question is specifically how to select the individual units to be included. In the next few sections, the various ways that samplers typically draw samples are discussed.

Simple Random Sampling

Simple random sampling is, in a sense, the prototype of population sampling. The most basic ways of calculating statistics about samples assume that a simple random sample was drawn. Simple random sampling approximates drawing a sample out of a hat: Members of a population are selected one at a time, independent of one another and without replacement; once a unit is selected, it has no further chance to be selected.

Operationally, drawing a simple random sample requires a numbered list of the population. For simplicity, assume that each person in the population appears once and only once. If there were 8,500 people on a list, and the goal was to select a simple random sample of 100, the procedure would be straightforward. People on the list would be numbered from 1 to 8,500. Then a computer, a table of random numbers, or some other generator of random numbers would be used to produce 100 different numbers within the same range. The individuals corresponding to the 100 numbers chosen would constitute a simple random sample of that population of 8,500. If the list is in a computerized data file, randomizing the ordering of the list, then choosing the first 100 people on the reordered list, would produce an equivalent result.

Systematic Samples

Unless a list is short, has all units prenumbered, or is computerized so that it can be numbered easily, drawing a simple random sample as described above can be laborious. In such situations, there is a way to use a variation called systematic sampling that will have precision equivalent to a simple random sample and can be mechanically easier to create. Moreover, the benefits of stratification (discussed in the next section) can be accomplished more easily through systematic sampling.

When drawing a systematic sample from a list, the researcher first determines the number of entries on the list and the number of elements from the list that are to be selected. Dividing the latter by the former will produce a fraction. Thus, if there are 8,500 people on a list and a sample of 100 is required, 100/8,500 of the list (i.e., 1 out of every 85 persons) is to be included in the sample. In order to select a systematic sample, a start point is designated by choosing a random number that falls within the sampling interval, in this example, any number from 1 to 85. The randomized start ensures that it is a chance selection process. Starting with the person in the randomly selected position, the researcher proceeds to take every 85th person on the list.

Most statistics books warn against systematic samples if a list is ordered by some characteristic, or has a recurring pattern, that will differentially affect the sample depending on the random start. As an extreme example, if members of a male-female couples club were listed with the male partner always listed first, any even number interval would produce a systematic sample that consisted of only one gender even though the club as a whole is evenly divided by gender. It definitely is important to examine a potential sample frame from the perspective of whether or not there is any reason to think that the sample resulting from one random start will be systematically different from those resulting from other starts in ways that will affect the survey results. In practice, most lists or sample frames do not pose any problems for systematic sampling. When they do, by either reordering the lists or adjusting the selection intervals, it almost always is possible to design a systematic sampling strategy that is at least equivalent to a simple random sample.

Stratified Samples

When a simple random sample is drawn, each new selection is independent, unaffected by any selections that came before. As a result of this process, any of the characteristics of the sample may, by chance, differ somewhat from the population from which it is drawn. Generally, little is known about the characteristics of individual population members before data collection. It is not uncommon, however, for at least a few characteristics of a population to be identifiable at the time of sampling. When that is the case, there is the possibility of structuring the sampling process to reduce the normal sampling variation, thereby producing a sample that is more likely to look like the total population than a simple random sample. The process by which this is done is called stratification.

For example, suppose one had a list of college students. The list is arranged alphabetically. Members of different classes are mixed throughout the list. If the list identifies the particular class to which a student belongs, it would be possible to rearrange the list to put freshmen first, then sophomores, then juniors, and finally seniors, with all classes grouped together. If the sampling
design calls for selecting a sample of 1 in 10 of the members on the list, the rearrangement would ensure that exactly 1/10 of the freshmen were selected, 1/10 of the sophomores, and so forth. On the other hand, if either a simple random sample or a systematic sample was selected from the original alphabetical list, the proportion of the sample in the freshman year would be subject to normal sampling variability and could be slightly higher or lower than was the case for the population. Stratifying in advance ensures that the sample will have exactly the same proportions in each class as the whole population.

Consider the task of estimating the average age of the student body. The class in which a student is a member almost certainly is correlated with age. Although there still will be some variability in sample estimates because of the sampling procedure, structuring the representation of classes in the sampling frame will constrain the extent to which the average age of the sample will differ by chance from the population as a whole.

Almost all samples of populations of geographic areas are stratified by some regional variable so that they will be distributed in the same way as the population as a whole. National samples typically are stratified by region of the country and also by urban, suburban, and rural locations. Stratification only increases the precision of estimates of variables that are related to the stratification variables. Because some degree of stratification is relatively simple to accomplish, however, and because it never hurts the precision of sample estimates (as long as the probability of selection is the same across all strata), it usually is a desirable feature of a sample design.

**Different Probabilities of Selection**

Sometimes stratification is used as a first step to vary the rates of selection of various population subgroups. When probabilities of selection are constant across strata, a group that constitutes 10% of a population will constitute about 10% of a selected sample. If a researcher wanted a sample of at least 100 from a population subgroup that constituted 10% of the population, a simple random sampling approach would require an overall sample of 1,000. Moreover, if the researcher decided to increase the sample size of that subgroup to 150, this would entail taking an additional 500 sample members into the sample, bringing the total to 1,500, so that 10% of the sample would equal 150.

Obviously, there are occasions when increasing a sample in this way is not very cost effective. In the latter example, if the researcher is satisfied with the size of the samples of other groups, the design adds 450 unwanted interviews in order to add 50 interviews that are wanted. In some cases, therefore, an appropriate design is to select some subgroup at a higher rate than the rest of the population.

**Sampling**

As an example, suppose that a researcher wished to compare male and female students, with a minimum of 200 male respondents, at a particular college where only 20% of the students are male. Thus a sample of 500 students would include 100 male students. If male students could be identified in advance, however, one could select male students at twice the rate at which female students were selected. In this way, rather than adding 500 interviews to increase the sample by 100 males, an additional 100 interviews over the basic sample of 500 would produce a total of about 200 interviews with males. Thus, when making male-female comparisons, one would have the precision provided by samples of 200 male respondents and 400 female respondents. To combine these samples, the researcher would have to give male respondents a weight of half that given to females to compensate for the fact that they were sampled at twice the rate of the rest of the population. (See Chapter 10 for more details about weighting.)

Even if individual members of a subgroup of interest cannot be identified with certainty in advance of sampling, sometimes the basic approach outlined above can be applied. For instance, it is most unusual to have a list of housing units that identifies the race of occupants in advance of contact. It is not uncommon, however, for Asian families to be more concentrated in some neighborhood areas than others. In that instance, a researcher may be able to sample households in areas that are predominantly Asian at a higher average rate to increase the number of Asian respondents. Again, when any group is given a chance of selection different from other members of the population, appropriate compensatory weighting is required in order to generate accurate population statistics for the combined or total sample.

A third approach is to adjust the chance of selection based on information gathered after making contact with potential respondents. Going back to the college student survey, if student gender could not be ascertained in advance, the researchers could select an initial sample of 1,000 students, have interviewers ascertain the gender of each student, then have them conduct a complete interview with all selected male students (200) but only half of the female students they identified (400). The result would be exactly the same as with the approach described above.

Finally, one other technical reason for using different probabilities of selection by stratum should be mentioned. If what is being measured is much more variable in one group than in another, it may help the precision of the resulting overall estimate to oversample the group with the high level of variability. Groves (1989) provides a good description of the rationale and how to assess the efficiency of such designs.
MULTISTAGE SAMPLING

When there is no adequate list of the individuals in a population and no way to get at the population directly, multistage sampling provides a useful approach.

In the absence of a direct sampling source, a strategy is needed for linking population members to some kind of grouping that can be sampled. These groupings can be sampled as a first stage. Lists then are made of individual members of selected groups, with possibly a further selection from the created list at the second (or later) stage of sampling. In sampling terminology, the groupings in the last stage of a sample design are usually referred to as “clusters.” The following section illustrates the general strategy for multistage sampling by describing its use in three of the most common types of situations in which lists of all individuals in the target population are not available.

Sampling Students From Schools

If one wanted to draw a sample of all students enrolled in the public schools of a particular city, it would not be surprising to find that there was not a single complete list of such individuals. There is, however, a sample frame that enables one to get at and include all the students in the desired population: namely, the list of all the public schools in that city. Because every individual in the study population can be attached to one and only one of those units, a perfectly acceptable sample of students can be selected using a two-stage strategy: first selecting schools (i.e., the clusters) and then selecting students from within those schools.

Assume the following data:
- There are 20,000 students in a city with 40 schools
- Desired sample = 2,000 = 1/10 of students

Four different designs or approaches to sampling are presented below. Each would yield a probability sample of 2,000 students.

The four approaches listed all yield samples of 2,000; all give each student in the city an equal (1 in 10) chance of selection. The difference is that from top to bottom, the designs are increasingly less expensive; lists have to be collected from fewer schools, and fewer schools need to be visited. At the same time, the precision of each sample is likely to decline as fewer schools are sampled and more students are sampled per school. The effect of this and other multistage designs on the precision of sample estimates is discussed in more detail in a later section of this chapter.

<table>
<thead>
<tr>
<th>Overall Probability of Selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability of Selection at Stage 1 (schools)</td>
</tr>
<tr>
<td>(a) Select all schools, list all students, and select 1/10 students in each school</td>
</tr>
<tr>
<td>(b) Select 1/2 of the schools, then select 1/5 of all students in them</td>
</tr>
<tr>
<td>(c) Select 1/5 of the schools, then select 1/2 of all students in them</td>
</tr>
<tr>
<td>(d) Select 1/10 schools, then collect information about all students in them</td>
</tr>
</tbody>
</table>

Area Probability Sampling

Area probability sampling is one of the most generally useful multistage strategies because of its wide applicability. It can be used to sample any population that can be defined geographically, for example, the people living in a neighborhood, a city, a state, or a country. The basic approach is to divide the total target land area into exhaustive, mutually exclusive subareas with identifiable boundaries. These subareas are the clusters. A sample of subareas is drawn. A list then is made of housing units in selected subareas, and a sample of listed units is drawn. As a final stage, all people in selected housing units may be included in the sample, or they may be listed and sampled as well.

This approach will work for jungles, deserts, sparsely populated rural areas, or downtown areas in central cities. The specific steps to drawing such a sample can be very complicated. The basic principles, however, can be illustrated by describing how one could sample the population of a city using city blocks as the primary subarea units to be selected at the first stage of sampling.
Assume the following data:
A city consists of 400 blocks
20,000 housing units are located on these blocks
Desired sample = 2,000 housing units = 1/10 of all housing units

Given this information, a sample of households could be selected using a strategy parallel to the above selection of students. In the first stage of sampling, blocks (i.e., the clusters) are selected. During the second stage, all housing units on selected blocks are listed and a sample is selected from the lists. Two approaches to selecting housing units are as follows:

\[
\text{Probability of Selection at Stage 1 (blocks)} \times \text{Probability of Selection at Stage 2 (housing units in selected blocks)} = \text{Overall Probability of Selection}
\]

(a) Select 80 blocks (1/5), then take 1/2 of units on those blocks
\[
1/5 \times 1/2 = 1/10
\]

(b) Select 40 blocks (1/10), then take all units on those blocks
\[
1/10 \times 1 = 1/10
\]

Parallel to the school example, the first approach, involving more blocks, is more expensive than the second; it also is likely to produce more precise sample estimates for a sample of a given size.

None of the above sample schemes takes into account the size of the Stage 1 groupings (i.e., the size of the blocks or schools). Big schools and big blocks are selected at the same rates as small ones. If a fixed fraction of each selected group is to be taken at the last stage, there will be more interviews taken from selected big schools or big blocks than from small ones; the size of the samples (cluster sizes) taken at Stage 2 will be very divergent.

If there is information available about the size of the Stage 1 groups, it is usually good to use it. Sample designs tend to provide more precise estimates if the number of units selected at the final step of selection is approximately equal in all clusters. Other advantages of such designs are that sampling errors are easier to calculate and the total size of the sample is more predictable. To produce equal-sized clusters, Stage 1 units should be sampled proportionate to their size.

The following example shows how blocks could be sampled proportionate to their size as the first stage of an area probability approach to sampling housing units (apartments or single family houses). The same approach could be applied to the school example above, treating schools in a way analogous to blocks in this process.

1. Decide how many housing units are to be selected at the last stage of sampling—the average cluster size. Let us choose 10, for example.
2. Make an estimate of the number of housing units in each Stage 1 unit (block).
3. Order the blocks so that geographically adjacent or otherwise similar blocks are contiguous. This effectively stratifies the sampling to improve the samples, as discussed above.
4. Create an estimated cumulative count across all blocks of housing units. A table like the one below will result.

<table>
<thead>
<tr>
<th>Block Number</th>
<th>Estimated Housing Units</th>
<th>Cumulative Housing Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>43</td>
<td>43</td>
</tr>
<tr>
<td>2</td>
<td>87</td>
<td>130</td>
</tr>
<tr>
<td>3</td>
<td>99</td>
<td>229</td>
</tr>
<tr>
<td>4</td>
<td>27</td>
<td>256</td>
</tr>
<tr>
<td>5</td>
<td>15</td>
<td>271</td>
</tr>
</tbody>
</table>

Hits (Random Start = 70; Interval = 100 HUs)

Determine the interval between clusters. If we want to select 1 in 10 housing units and a cluster of about 10 on each selected block, we need an interval of 100 housing units between clusters. Put another way, instead of taking 1 house at an interval of every 10 houses, we take 10 houses at an interval of every 100 houses; the rate is the same, but the pattern is "clustered."

After first choosing a random number from 1 to 100 (the interval in the example) as a starting point, we proceed systematically through the cumulative count, designating the primary units (or blocks) hit in this first stage of selection. In the example, the random start chosen (70) missed block 1 (though 43 times in 100 it would have been hit); the 70th housing unit was in block 2; the 170th housing unit was in block 3; and the 270th housing unit was located in block 5.

A list then is made of the housing units on the selected blocks (2, 3, and 5), usually by sending a person to visit the blocks. The next step is to select housing units from those lists. If we were sure the estimates of the sizes of blocks were accurate, we could simply select 10 housing units from each selected block, using
either simple random or systematic sampling; a systematic sample would usually be best because it would distribute the chosen units around the block.

It is common for estimates of the size of stage 1 units such as blocks to be somewhat in error. We can correct for such errors by calculating the rate at which housing units are to be selected from blocks as:

\[
\text{Rate of HU selection on block} = \frac{\text{Avg. cluster size}}{\text{Estimated HUs on block}} = \frac{10}{87} = 0.1\text{.}
\]

(On block 2)

In our example, we would take 1 per 8.7 housing units on block 2, 1 per 9.9 housing units on block 3, and 1 per 1.5 housing units on block 5. If a block is bigger than expected (e.g., because of new construction), more than 10 housing units will be drawn; if it is smaller than expected (e.g., because of demolition), fewer than 10 housing units will be drawn. If it is exactly what we expected (e.g., 87 housing units on block 2), we take 10 housing units (87\( \times \) 0.1 = 10). In this way, the procedure is self-correcting for errors in initial estimates of block size, while maintaining the same chance of selection for housing units on all blocks. No matter the estimated or actual size of the block, the chance of any housing unit being selected is 1 in 10.

The area probability sample approach can be used to sample any geographically defined population. Although the steps are more complicated as the area gets bigger, the approach is the same. The key steps to remember are the following:

- All areas must be given some chance of selection. Combine areas where no units are expected with adjacent areas to ensure a chance of selection; new construction may have occurred or estimates may be wrong.

- The probability of selecting a block (or other land area) times the probability of selecting a housing unit from a selected block should be constant across all blocks.

Finally, even careful field listers will miss some housing units. Therefore, it is good practice to include checks for missed units at the time of data collection.

Random-Digit Dialing

Random-digit dialing (RDD) provides an alternative way to draw a sample of housing units in order to sample the people in those households. Suppose the 20,000 housing units in the above example are covered by six telephone exchanges. One could draw a probability sample of 10% of the housing units that have telephones as follows:

1. There is a total of 60,000 possible telephone numbers in those 6 exchanges (10,000 per exchange). Select 6,000 of those numbers (i.e., 10%), drawing 1,000 randomly generated, four-digit numbers per exchange.

2. Dial all 6,000 numbers. Not all the numbers will be household numbers; in fact, many of the numbers will not be working, will be disconnected or temporarily not in service, or will be businesses. Because 10% of all possible telephone numbers that could serve the area have been called, about 10% of all the households with telephones in that area will be reached by calling the sample of numbers.

This is the basic random-digit-dialing approach to sampling. The obvious disadvantage of this approach is the large number of unfruitful calls. Nationally, fewer than 25% of possible numbers are associated with residential housing units; the rate is about 30% in urban areas and about 10% in rural areas. Waksberg (1978) developed a method of taking advantage of the fact that telephone numbers are assigned in groups. Each group of 100 telephone numbers is defined by a three-digit area code, a three-digit exchange, and two additional numbers (area code-123-45XX). By carrying out an initial screening of numbers by calling one random number in a sample of groups, then calling additional random numbers only within the groups of numbers where a residential number was found, the rate of hitting housing units can be raised to more than 50%. In this design, the groups of 100 telephone numbers are the clusters.

In recent years, most survey organizations have begun using a list-assisted approach to RDD. With the advancement of computer technology, companies can compile computerized versions of telephone listings. These computerized phone books are updated every 3 months. Once all these books are in a computer file, a search can yield all clusters (area code-123-45XX) that have at least one published residential telephone number. These companies can then produce a sample frame of all possible telephone numbers in clusters that have at least one published residential telephone number. Sampling can now be carried out using this sample frame. This approach has two distinct advantages. The first is that the initial screening of telephone numbers required by the Waksberg method is no longer needed. The construction of the sample frame has already accomplished this. The second advantage is that the sample selected using this frame is no longer clustered. By using all clusters that contain residential telephone numbers as a sample frame, a simple or systematic random sample of telephone numbers can be drawn. This new approach to
RDD is more cost effective and efficient than its predecessors were. A limitation is that telephone numbers in clusters that have no listed residential numbers have no chance of selection. Brick, Waksberg, Kulip, and Stare (1995) have estimated that, on average, about 4% of households with telephone service in the United States are left out. Lepkowski (1988) provides a good summary of the various ways to sample telephone numbers in order to sample households.

The accumulation of lists of individuals and their characteristics has made possible some other efficiencies for telephone surveys. One comparatively simple advance is that reverse telephone directories can be used to tie addresses to some telephone numbers. One of the downsides of RDD is that households do not receive advance notice that an interviewer will be calling. Lists make it possible to sort selected numbers into groups (or strata) based on whether or not there is a known residential address associated with a number. Those for whom there is a known address can be sent an advance letter.

More elaborately, if there are lists of people who have known characteristics that are targeted for a survey—an age group, those living in a particular geographic area, people who gave to a particular charity—a stratum can be made of telephone numbers likely to connect to households that are being targeted. Numbers in the other strata or for which information is not available may be sampled at lower rates, thereby giving all households a known chance of selection but increasing the efficiency of the data collection by concentrating more effort on households likely to yield eligible respondents. Note that if the probabilities of selection are not the same for all respondents, weighting must be used at the analysis stage, as described in Chapter 10.

There are several additional issues to note about the random-digit-dialing approach to sampling. First, its value depends on the fact that most households have telephone service. Nationally, only about 5% of the households lack household service, but in some areas, particularly central cities or rural areas, the rate of omission may be greater than that. The growing use of individual cell phones has also posed a growing problem for RDD sampling. Most current RDD sampling focuses only on household service and avoids exchanges devoted to cell phone use. It is possible to sample from both kinds of services, but the complexity of sampling, data collection, and post-survey weighting are greatly increased if cell phone numbers are included in the sample frames (Brick, Dipko, Presser, Tucker, & Yuan, 2006; Lavrakas, Shutless, Steeh, & Fienberg, 2007).

To give one example of the complexity: RDD sampling uses area codes to target populations in defined geographic areas. However, cell phone numbers are much less tied to where people actually live. A survey based on cell phone area codes will reach some people who live outside the targeted geographic area and, worse, will omit those who live in the area but whose cell phones have distant area codes.

Like any particular sampling approach, RDD is not the best design for all surveys. Additional pros and cons will be discussed in Chapter 5. The introduction of RDD as one sampling option, however, has made a major contribution to expanding survey research capabilities in the last 30 years. With the longer-term impact of cell phones and response rate challenges (discussed in Chapter 5), the future use of RDD sampling remains to be seen.

**Respondent Selection**

Both area probability samples and RDD designate a sample of housing units. There is then the further question of who in the household should be interviewed.

The best decision depends on what kind of information is being gathered. In some studies, the information is being gathered about the household and about all the people in the household. If the information is commonly known and easy to report, perhaps any adult who is home can answer the questions. If the information is more specialized, the researcher may want to interview the household member who is most knowledgeable. For example, in the National Health Interview Survey, the person who "knows the most about the health of the family" is to be the respondent for questions that cover all family members.

There are, however, many things that an individual can report only for himself or herself. Researchers almost universally feel that no individual can report feelings, opinions, or knowledge for some other person. There are also many behaviors or experiences (e.g., what people eat or drink, what they have bought, what they have seen, or what they have been told) that usually can only be reported accurately by self-reporters.

When a study includes variables for which only self-reporting is appropriate, the sampling process must go beyond selecting households to sampling specific individuals within those households. One approach is to interview every eligible person in a household. (So there is no sampling at that stage.) Because of homogeneity within households, however, as well as concerns about one respondent influencing a later respondent's answers, it is more common to designate a single respondent per household. Obviously, choosing the person who happens to answer the phone or the door would be a nonprobabilistic and potentially biased way of selecting individuals; interviewer discretion, respondent discretion, and availability (which is related to working status, lifestyle, and age) would all affect who turned out to be the respondent. The
The key principle of probability sampling is that selection is carried out through some chance or random procedure that designates specific people. The procedure for generating a probability selection of respondents within households involves three steps:

1. Ascertain how many people living in a household are eligible to be respondents (e.g., how many are 18 or older).
2. Number these in a consistent way in all households (e.g., order by decreasing age).
3. Have a procedure that objectively designates one person to be the respondent.

Kish (1949) created a detailed procedure for designating respondents using a set of randomized tables that still is used today. When interviewing is computer assisted, it is easy to have the computer select one of the eligible household members. The critical features of the procedure are that no discretion be involved and that all eligible people in selected households have a known (and nonzero) probability of selection. Groves and Lyberg (1988) review several strategies for simplifying respondent selection procedures.

One of the concerns about respondent selection procedures is that the initial interaction upon first contacting someone is critical to enlisting cooperation. If the respondent selection procedure is too cumbersome or feels intrusive, it may adversely affect the rate of response. Thus, there have been various efforts to find streamlined ways to sample adults in selected households.

One popular method is the “last birthday” method. The household contact is asked to identify the adult who last had a birthday, and that person is the designated respondent. In principle, this should be an unbiased way to select a respondent. In practice, it depends on the initial contact having information about all household members’ birthdays.

Another relatively new approach keys selection to the person the interviewer first talks with. First, the number of eligible people in the household is determined. If there are two or more eligible, a randomized algorithm chooses either the initial informant at the appropriate rate or chooses among the “other” eligible adults (if there is more than one) (Rizzo, Brick, & Park, 2004).

However the respondent is chosen, when only one person is interviewed in a household, a differential rate of selection is introduced. If an adult lives in a one-adult household, he or she obviously will be the respondent if the household is selected. In contrast, an adult living in a three-adult household only will be the respondent one third of the time. Whenever an identifiable group is selected at a different rate from others, weights are needed so that oversampled people are not overrepresented in the sample statistics. In the example earlier in this chapter, when male students were selected at twice the rate of female students, their responses were weighted by one half so that their weighted proportion of the sample would be the same as in the population. The same general approach applies when one respondent is chosen from households with varying numbers of eligible people.

The simplest way to adjust for the effect of selecting one respondent per household is to weight each response by the number of eligible people in that household. Hence, if there are three adults, the weight is three; if there are two eligible adults, the weight is two; and if there is only one eligible adult, the weight is one. If a weighting scheme is correct, the probability of selection times the weight is the same for all respondents. (See Chapter 10.)

MAKING ESTIMATES FROM SAMPLES AND SAMPLING ERRORS

The sampling strategies presented above were chosen because they are among the most commonly used and they illustrate the major sampling design options. A probability sampling scheme eventually will designate a specific set of households or individuals without researcher or respondent discretion. The basic tools available to the researcher are simple random and systematic sampling, which are modified by stratification, unequal rates of selection, and clustering. The choice of a sampling strategy rests in part on feasibility and costs; it also involves the precision of sample estimates. A major reason for using probability sampling methods is to permit use of a variety of statistical tools to estimate the precision of sample estimates. In this section, the calculation of such estimates and how they are affected by features of the sample design are discussed.

Researchers usually have no interest in the characteristics of a sample per se. The reason for collecting data about a sample is to reach conclusions about an entire population. The statistical and design issues in this chapter are considered in the context of how much confidence one can have that the characteristics of a sample accurately describe the population as a whole.

As described in Chapter 2, a way to think about sampling error is to think of the distribution of means one might get if many samples were drawn from the same population with the same procedure. Although some sources of error in surveys are biasing and produce systematically distorted figures, sampling error is a random (and hence not a systematically biasing) result of sampling. When probability procedures are used to select a sample, it is possible to calculate how much sample estimates will vary by chance because of sampling.

If an infinite number of samples are drawn, the sample estimates of descriptive statistics (e.g., means) will form a normal distribution around the true
population value. The larger the size of the sample and the less the variance of what is being measured, the more tightly the sample estimates will bunch around the true population value, and the more accurate a sample-based estimate usually will be. This variation around the true value, stemming from the fact that by chance samples may differ from the population as a whole, is called "sampling error." Estimating the limits of the confidence one can have in a sample estimate, given normal chance sampling variability, is one important part of evaluating figures derived from surveys.

The design of sample selection (specifically, whether it involves stratification, clustering, or unequal probabilities of selection) affects the estimates of sampling error for a sample of a given size. The usual approach to describing sampling errors, however, is to calculate what they would be for a simple random sample, and then to calculate the effects of deviations from a simple random sampling design. Hence, the calculation of sampling errors for simple random samples is described first.

**Sampling Errors for Simple Random Samples**

This is not a textbook on sampling statistics. Estimating the amount of error one can expect from a particular sample design, however, is a basic part of the survey design process. Moreover, researchers routinely provide readers with guidelines regarding error attributable to sampling, guidelines that both the knowledgeable reader and the user of survey research data should and understand. To this end, a sense of how sampling error is calculated is a necessary part of understanding the total survey process.

Although the same logic applies to all statistics calculated from a sample, the most common sample survey estimates are means or averages. The statistic most often used to describe sampling error is called the standard error (of a mean). *It is the standard deviation of the distribution of sample estimates of means that would be formed if an infinite number of samples of a given size were drawn.* When the value of a standard error has been estimated, one can say that 67% of the means of samples of a given size and design will fall within the range of ±1 standard error of the true population mean; 95% of such samples will fall within the range of ±2 standard errors. The latter figure (±2 standard errors) often is reported as the "confidence interval" around a sample estimate.

The estimation of the standard error of a mean is calculated from the variance and the size of the sample from which it was estimated:

\[
SE = \sqrt{\frac{\text{Var}}{n}}
\]

**SE** = standard error of a mean

**Var** = the variance (the sum of the squared deviations from the sample mean over \(n\))

\(n\) = size of the sample

The most common kind of mean calculated from a sample survey is probably the percentage of a sample that has a certain characteristic or gives a certain response. It may be useful to show how a percentage is the mean of a two-value distribution.

A mean is an average. It is calculated as the sum of the values divided by the number of cases: \(\sum x / n\). Now suppose there are only two values, 0 (no) and 1 (yes). There are 50 cases in a sample; 20 say "yes" when asked if they are married, and the rest say "no." If there are 20 "yes" and 30 "no" responses, calculate the mean as

\[
\frac{\sum X}{n} = \frac{(20 \times 1 + 30 \times 0)}{50} = \frac{20}{50} = 0.40
\]

A percentage statement, such as 40% of respondents are married, is just a statement about the mean of a 1/0 distribution; the mean is 0.40. The calculation of standard errors of percentage is facilitated by the fact that the variance of a percentage can be calculated readily as \(p \times (1 - p)\), where \(p\) = percentage having a characteristic (e.g., the 40% married in the above example) and \((1 - p)\) is the percentage who lack the characteristic (e.g., the 60% not married).

We have already seen that the standard error of a mean is as follows:

\[
SE = \sqrt{\frac{\text{Var}}{n}}
\]

Because \(p(1 - p)\) is the variance of a percentage,

\[
SE = \sqrt{\frac{p(1-p)}{n}}
\]
is the standard error of a percentage. In the previous example, with 40% of a sample of 50 persons being married, the standard error of that estimate would be as follows:

\[
SE = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.40 \times 0.60}{50}} = \sqrt{\frac{0.24}{50}} = 0.07
\]

Thus we would estimate that the probability is .67 (i.e., ±1 standard error from the sample mean) that the true population figure (the percentage of the whole population that is married) is between .33 and .47 (.40 ± .07). We are 95% confident that the true population figure lies within two standard errors of our sample mean, that is, between .26 and .54 (.40 ± .14).

Table 3.1 is a generalized table of sampling errors for samples of various sizes and for various percentages, provided that samples were selected as simple random samples. Each number in the table represents two standard errors of a percentage. Given knowledge (or an estimate) of the percentage of a sample that gives a particular answer, the table gives 95% confidence intervals for various sample sizes. In the example above, with 50 cases yielding a sample estimate of 40% married, the table reports a confidence interval near .14, as we calculated. If a sample of about 100 cases produced an estimate that 20% were married, the table says we can be 95% sure that the true figure is 20% ± 8 percentage points (i.e., 12% to 28%).

Several points about the table are worth noting. First, it can be seen that increasingly large samples always reduce sampling errors. Second, it also can be seen that adding a given number of cases to a sample reduces sampling error a great deal more when the sample is small than when it is comparatively large. For example, adding 50 cases to a sample of 50 produces a quite noticeable reduction in sampling error. Adding 50 cases to a sample of 500, however, produces a virtually unnoticeable improvement in the overall precision of sample estimates.

Third, it can be seen that the absolute size of the sampling error is greatest around percentages of .5 and decreases as the percentage of a sample having a characteristic approaches either zero or 100%. We have seen that standard errors are related directly to variances. The variance \( p(1-p) \) is smaller as the percentages get further from .5. When \( p = 0.5 \), \( (0.5 \times 0.5) = 0.25 \). When \( p = 0.2 \), \( (0.2 \times 0.8) = 0.16 \).

Fourth, Table 3.1 and the equations on which it is based apply to samples drawn with simple random sampling procedures. Most samples of general populations are not simple random samples. The extent to which the particular sample design will affect calculations of sampling error varies from design to design and for different variables in the same survey. More often than not,

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>5/95</th>
<th>10/90</th>
<th>20/80</th>
<th>30/70</th>
<th>50/50</th>
</tr>
</thead>
<tbody>
<tr>
<td>35</td>
<td>7</td>
<td>10</td>
<td>14</td>
<td>15</td>
<td>17</td>
</tr>
<tr>
<td>50</td>
<td>6</td>
<td>8</td>
<td>11</td>
<td>13</td>
<td>14</td>
</tr>
<tr>
<td>75</td>
<td>5</td>
<td>7</td>
<td>9</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>100</td>
<td>4</td>
<td>6</td>
<td>8</td>
<td>9</td>
<td>10</td>
</tr>
<tr>
<td>200</td>
<td>3</td>
<td>4</td>
<td>6</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>300</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>500</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>1,000</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>1,500</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

NOTE: Chances are 95 in 100 that the real population figure lies in the range defined by ± number indicated in table, given the percentage of sample reporting the characteristic and the number of sample cases on which the percentage is based.

This table describes variability attributable to sampling. Errors resulting from nonresponse or reporting errors are not reflected in this table. In addition, this table assumes a simple random sample. Estimates may be subject to more variability than this table indicates because of the sample design or the influence of interviewers on the answers they obtained; stratification might reduce the sampling errors below those indicated here.

Table 3.1 will constitute an underestimate of the sampling error for a general population sample.

Finally, it should be emphasized that the variability reflected in Table 3.1 describes potential for error that comes from the fact of sampling rather than collecting information about every individual in a population. The calculations do not include estimates of error from any other aspects of the survey process.

Effects of Other Sample Design Features

The preceding discussion describes the calculation of sampling errors for simple random samples. Estimates of sampling errors will be affected by different sampling procedures. Systematic sampling should produce sampling
errors equivalent to simple random samples if there is no stratification. Stratified samples can produce sampling errors that are lower than those associated with simple random samples of the same size for variables that differ (on average) by stratum, if rates of selection are constant across strata.

Unequal rates of selection (selecting subgroups in the population at different rates) are designed to increase the precision of estimates for oversampled subgroups, thus

(a) they generally will produce sampling errors for the whole sample that are higher than those associated with simple random samples of the same size, for variables that differ by stratum, except

(b) when oversampling is targeted at strata that have higher than average variances for some variable, the overall sampling errors for those variables will be lower than for a simple random sample of the same size.

Clustering will tend to produce sampling errors that are higher than those associated with simple random samples of the same size for variables that are more homogeneous within clusters than in the population as a whole. Also, the larger the size of the cluster at the last stage, the larger the impact on sampling errors will usually be.

It often is not easy to anticipate the effects of design features on the precision of estimates. Design effects differ from study to study and for different variables in the same survey. To illustrate, suppose every house on various selected blocks was the same with respect to type of construction and whether or not it was occupied by the owner. Once one respondent on a block reports he is a home owner, the additional interviews on that block would yield absolutely no new information about the rate of home ownership in the population as a whole. For that reason, whether the researcher took one interview per block or 20 interviews per block, the reliability of that estimate would be exactly the same, basically proportionate to the number of blocks from which any interviews at all were taken. At the other extreme, the height of adults is likely to vary as much within a block as it does throughout a city. If the respondents on a block are as heterogeneous as the population as a whole, clustering does not decrease the precision of estimates of height from a sample of a given size. Thus, one has to look at the nature of the clusters or strata and what estimates are to be made in order to evaluate the likely effect of clustering on sampling errors.

The effects of the sample design on sampling errors often are unappreciated. It is not uncommon to see reports of confidence intervals that assume simple random sampling when the design was clustered. It also is not a simple matter to anticipate the size of design effects beforehand. As noted, the effects of the sample design on sampling errors are different for every variable; their calculation is particularly complicated when a sample design has several deviations from simple random sampling, such as both clustering and stratification. Because the ability to calculate sampling errors is one of the principal strengths of the survey method, it is important that a statistician be involved in a survey with a complex sample design to ensure that sampling errors are calculated and reported appropriately. The problem of appropriately taking into account design features when estimating sampling errors has been greatly simplified by the fact that several available analysis packages will do those adjustments. (See Chapter 10.)

Finally, the appropriateness of any sample design feature can be evaluated only in the context of the overall survey objectives. Clustered designs are likely to save money both in sampling (listing) and in data collection. Moreover, it is common to find many variables for which clustering does not inflate the sampling errors very much. Oversampling one or more groups often is a cost-effective design. As with most issues discussed in this book, the important point is for a researcher to be aware of the potential costs and benefits of the options and to weigh them in the context of all the design options and the main purposes of the survey.

HOW BIG SHOULD A SAMPLE BE?

Of the many issues involved in sample design, one of the most common questions posed to a survey methodologist is how big a survey sample should be. Before providing an approach to answering this question, perhaps it is appropriate to discuss three common but inappropriate ways of answering it.

One common misconception is that the adequacy of a sample depends heavily on the fraction of the population included in that sample—that somehow 1%, or 5%, or some other percentage of a population will make a sample credible. The estimates of sampling errors discussed above do not take into account the fraction of a population included in a sample. The sampling error estimates from the preceding equations and from Table 3.1 can be reduced by multiplying them by the value \((1 - f)\), where \(f\) is the fraction of the population included in a sample.

When one is sampling 10% or more of a population, this adjustment can have a discernible effect on sampling error estimates. The vast majority of survey samples, however, involve very small fractions of populations. In such instances, small increments in the fraction of the population included in a sample will have no effect on the ability of a researcher to generalize from a sample to a population.
The converse of this principle also should be noted. The size of the population from which a sample of a particular size is drawn has virtually no impact on how well that sample is likely to describe the population. A sample of 150 people will describe a population of 15,000 or 15 million with virtually the same degree of accuracy, assuming that all other aspects of the sample design and sampling procedures are the same. Compared to the total sample size and other design features such as clustering, the impact of the fraction of a population sampled on sampling errors is typically trivial. It is most unusual for it to be an important consideration when deciding on a sample size.

A second inappropriate approach to deciding on sample size is somewhat easier to understand. Some people have been exposed to so-called standard survey studies, and from these they have derived a typical or appropriate sample size. Thus some people will say that good national survey samples generally are 1,500, or that good community samples are 500. Of course, it is not foolish to look at what other competent researchers have considered to be adequate sample sizes of a particular population. The sample size decision, however, like most other design decisions, must be made on a case-by-case basis, with the researchers considering the variety of goals to be achieved by a particular study and taking into account numerous other aspects of the research design.

A third wrong approach to deciding on sample size is the most important one to address, for it can be found in many statistical textbooks. The approach goes like this: A researcher should decide how much margin of error he or she can tolerate or how much precision is required of estimates. Once one knows the need for precision, one simply uses a table such as Table 3.1, or appropriate variations thereon, to calculate the sample size needed to achieve the desired level of precision.

In some theoretical sense, there is nothing wrong with this approach. In practice, however, it provides little help to most researchers trying to design real studies. First, it is unusual to base a sample size decision on the need for precision of a single estimate. Most survey studies are designed to make numerous estimates, and the needed precision for these estimates is likely to vary.

In addition, it is unusual for a researcher to be able to specify a desired level of precision in more than the most general way. It is only the exception, rather than the common situation, when a specific acceptable margin for error can be specified in advance. Even in the latter case, the above approach implies that sampling error is the only or main source of error in a survey estimate. When a required level of precision from a sample survey is specified, it generally ignores the fact that there will be error from sources other than sampling. In such cases, the calculation of precision based on sampling error alone is an unrealistic oversimplification. Moreover, given fixed resources, increasing the sample size may even decrease precision by reducing resources devoted to response rates, question design, or the quality of data collection.

Estimates of sampling error, which are related to sample size, do play a role in analyses of how big a sample should be. This role, however, is complicated.

The first prerequisite for determining a sample size is an analysis plan. The key component of that analysis plan usually is not an estimate of confidence intervals for the overall sample, but rather an outline of the subgroups within the total population for which separate estimates are required, together with some estimates of the fraction of the population that will fall into those subgroups. Typically, the design process moves quickly to identifying the smaller groups within the population for which figures are needed. The researcher then estimates how large a sample will be required in order to provide a minimally adequate sample of these small subgroups. Most sample size decisions do not focus on estimates for the total population; rather, they are concentrated on the minimum sample sizes that can be tolerated for the smallest subgroups of importance.

The process then turns to Table 3.1, not at the high end but at the low end of the sample size continuum. Are 50 observations adequate? If one studies Table 3.1, it can be seen that precision increases rather steadily up to sample sizes of 150 to 200. After that point, there is a much more modest gain to increasing sample size.

Like most decisions relating to research design, there is seldom a definitive answer about how large a sample should be for any given study. There are many ways to increase the reliability of survey estimates. Increasing sample size is one of them. Even if one cannot say that there is a single right answer, however, it can be said that there are three approaches to deciding on sample size that are inadequate. Specifying a fraction of the population to be included in the sample is never the right way to decide on a sample size. Sampling errors primarily depend on sample size, not on the proportion of the population in a sample. Saying that a particular sample size is the usual or typical approach to studying a population also is virtually always the wrong approach. An analysis plan that addresses the study's goals is the critical first step. Finally, it is very rare that calculating a desired confidence interval for one variable for an entire population is the determining calculation in how big a sample should be.

**Sampling error as a component of total survey error**

The sampling process can affect the quality of survey estimates in three different ways:

- If the sample frame excludes some people whom we want to describe, sample estimates will be biased to the extent that those omitted differ from those included.
If the sampling process is not probabilistic, the relationship between the sample and those sampled is problematic. One can argue for the credibility of a sample on grounds other than the sampling process; however, there is no statistical basis for saying a sample is representative of the sampled population unless the sampling process gives each person selected a known probability of selection.

The size and design of a probability sample, together with the distribution of what is being estimated, determine the size of the sampling errors, that is, the chance variations that occur because of collecting data about only a sample of a population.

Often sampling errors are presented in ways that imply they are the only source of unreliability in survey estimates. For surveys that use large samples, other sources of error are likely to be more important. A main theme of this book is that nonsampling errors warrant as much attention as sampling errors. Also, it is not uncommon to see sampling errors reported that assume simple random sampling procedures when the sample design involved clusters, or even when it was not a probability sample at all. In these ways, ironically, estimates of sampling errors can mislead readers about the precision or accuracy of sample estimates.

Sampling and analyzing data from a sample can be fairly straightforward if a good list is used as a sampling frame, if a simple random or systematic sampling scheme is used, and if all respondents are selected at the same rate. With such a design, Table 3.1 and the equations on which it is based will provide good estimates of sampling errors. Even with such straightforward designs, however, researchers need to consider all sources of error, including the sample frame, nonresponse, and response errors (all discussed in subsequent chapters) when evaluating the precision of survey estimates. Moreover, when there are doubts about the best way to sample, or when there are deviations from simple random sampling, it is virtually essential to involve a sampling specialist both to design an appropriate sampling plan and to analyze results properly from a complex sample design.

EXERCISES

1. In order to grasp the meaning of sampling error, repeated systematic samples of the same size (with different random starts) can be drawn from the same list (e.g., a telephone directory). The proportions of those samples having some characteristic (e.g., a business listing) taken together will form a distribution. That distribution will have a standard deviation that is about one half the entry in Table 3.1 for samples of the sizes drawn. It is also valuable to calculate several of the entries in Table 3.1 (i.e., for various sample sizes and proportions) to help understand how the numbers were derived.

2. What percentage of adults in the United States would you estimate:
   a. Have driver’s licenses?
   b. Have listed telephone numbers?
   c. Are registered to vote?
   d. Have a personal e-mail address (not through their work)?

3. What are some likely differences between those who would be in those sample frames and those who would not?

4. Compared with simple random samples, do the following tend to increase, decrease, or have no effect on sampling errors?
   a. Clustering?
   b. Stratifying?
   c. Using a systematic sampling approach?

Further Readings

Nonresponse

Implementing a Sample Design

Failure to collect data from a high percentage of those selected to be in a sample is a major potential source of survey error. Approaches to contacting respondents and enlisting cooperation for mail, telephone, Internet, and personal interview surveys are discussed. The biases associated with nonresponse are described, as are the disadvantages of strategies such as quota samples to avoid the effort required to obtain high response rates.

The idea of a probability sample of people is that every individual in the target population (or at least the sample frame) has a known chance to have data collected about him or her. A sampling procedure will designate a specific set of individuals (or units of some kind); the quality of sample data depends on how well those from whom data actually are collected reflect the total population with respect to the variables the survey is designed to describe. The procedures used to collect data are as important as the sample selection process in determining how well data from a sample describe a population.

Of course, the accuracy of any particular estimate from a survey depends on who provides an answer to a particular question. In every survey, there are some respondents who do not answer every question. Although nonresponse to individual questions is usually low, occasionally it can be high and can have a real effect on estimates. The focus of this chapter, however, is on those people who do not provide any data at all. (See Chapter 10 for further discussion of item nonresponse.)

There are three categories of those selected to be in a sample who do not actually provide data:

- those whom the data collection procedures do not reach, thereby not giving them a chance to answer questions
- those asked to provide data who refuse to do so
- those asked to provide data who are unable to perform the task required of them (e.g., people who are too ill to be interviewed, who do not speak the researcher’s language, or whose reading and writing skills preclude their filling out self-administered questionnaires)
The procedures that a researcher decides to use can have a major influence on the percentage of a sample that actually provides information (i.e., the response rate) and the extent to which nonrespondents introduce bias into sample data. For the most part, the likely effect of nonresponse is to bias samples, that is, to make them systematically different from the population from which they were drawn, thereby producing potentially biased estimates. In this chapter, the effect of nonresponse on survey estimates and procedures for reducing nonresponse are discussed.

**Calculating Response Rates**

The response rate is a basic parameter for evaluating a data collection effort. It is simply the number of people who complete the survey divided by the number of people (or units) sampled. The denominator includes all people in the study population who were selected but did not respond for whatever reason: refusals, language problems, illness, or lack of availability.

Sometimes a sample design will involve screening to find members of a population to be studied. Screened units that do not include people in the study population do not enter the response rate calculation. Hence vacant houses, telephone numbers that are not working or that do not serve residential units, and households where no eligible person resides (e.g., households in which no elderly people live when one is drawing a sample of elderly persons) are omitted in calculating response rates. If there are some units for which information needed to determine eligibility is not obtained, however, the response rate is uncertain. The best approach in this situation is to calculate the response rates using conservative and liberal assumptions about the rate of eligibility of unscreened units and report the range, together with a best estimate.

Response rates usually are reported as the percentage of a selected sample from whom data were collected. A further calculation can sometimes be made of the fraction of the population represented in the sample. If the sample frame did not omit anyone in the study population, the response rate is the same as the percentage of the population represented in the sample. If only 95% of the population has a telephone, however, the best estimate of the percentage of the population represented in a sample is .95 times the response rate for a telephone survey.

It is important to know the details of the way response rates are calculated. Differences in the way they are calculated can make comparisons difficult or inappropriate. For example, some organizations report a "completion rate," the percentage of households contacted at which an interview was completed.

**Nonresponse**

Such numbers will always be higher than the response rate outlined above, which includes selected uncontacted units in the denominator. A publication available from the American Association for Public Opinion Research (AAPOR) provides an excellent discussion of response rate calculations and how to report them (AAPOR, 2006).

**Bias Associated With Nonresponse**

The effect of nonresponse on survey estimates depends on the percentage not responding and the extent to which those not responding are biased—that is, systematically different from the whole population. If most of those selected provide data, sample estimates will be very good even if the nonrespondents are distinctive. For example, when the U.S. Bureau of the Census carries out the National Health Interview Survey, it is successful in completing interviews in over 90% of selected households. It is easy to show that even if the nonresponding 10% is very distinctive, the resulting samples are still very similar to the population as a whole.

The experience of the Bureau of the Census is extreme in the positive direction. At the other extreme, one occasionally will see reports of mail or Internet surveys in which 5% to 20% of the selected sample responded. In such instances, the final sample may have little relationship to the original sampling process; those responding essentially are self-selected. It is very unlikely that such procedures will provide any credible statistics about the characteristics of the population as a whole.

Response rates for most survey research projects lie somewhere between these two extremes. Response rates generally are higher in rural areas than they are in central cities. It also is easier to collect information from any responsible adult in a household than to obtain an interview with a specific designated respondent. Some subjects (e.g., health) may interest more people than other topics (e.g., economic behavior or public opinions). Moreover, survey organizations differ considerably in the extent to which they devote time and money to improving response rates.

There is no agreed-upon standard for a minimum acceptable response rate. The Office of Management and Budget of the federal government, which reviews surveys done under contract to the government, generally asks that procedures be likely to yield a response rate in excess of 80% and requires a nonresponse analysis if a survey does not meet this standard (OMB, 2006). In the United States, academic survey organizations sometimes are able to achieve response rates for designated adults in the 70% range for in-person
surveys with general household samples. The General Social Survey conducted by the National Opinion Research Center is an example. Rates of response for surveys of central city samples or using random-digit dialed telephone samples are likely to be lower—often much lower.

The nature of bias associated with nonresponse differs somewhat among mail, telephone, and personal interview procedures. One generalization that seems to hold up for most mail surveys is that people who have a particular interest in the subject matter or the research itself are more likely to return mail questionnaires than those who are less interested (Groves et al., 2006). This means that mail surveys with low response rates may be biased significantly in ways that are related directly to the purposes of the research (e.g., Fillion, 1975; Heberlein & Baumgartner, 1978; Jobber, 1984). For example, one study of those who did not return a mail questionnaire about health care experience, but who were later interviewed by phone, indicated that mail nonrespondents were younger, healthier, used fewer health services, and were more likely to be male than the mail respondents (Fowler et al., 2002). Gallagher, Fowler, and Stringfellow (2005) reported on another study in which health records were available on nonrespondents, enabling them to conclude that the early mail respondents were significantly different from later respondents in several ways highly relevant to the survey.

Another example of significant bias from low response to mail questionnaires is the oft-cited Literary Digest presidential poll in 1936, which managed to predict a victory for Alf Landon in an election that Franklin Roosevelt won by a political landslide. The story is told that a sample was drawn from telephone books, and Republicans (those in Landon’s party) were more likely to have telephones in 1936. In addition, however, the Literary Digest survey in 1936 was a mail survey. Its failure also was one of nonresponse; only a minority of those asked to return questionnaires did so. As is typical of mail surveys, those who wanted the undervote to win, the Landon supporters, were particularly likely to want to express their views (Bryson, 1976; Converse, 1987).

Availability is a more important source of nonresponse for telephone and personal surveys than for mail surveys. It is obvious that if a data collection effort is carried out between 9:00 a.m. and 5:00 p.m. on Mondays through Fridays, the people who will be available to be interviewed will be distinctive. Of course, most survey organizations emphasize contacting household on evenings and weekends. Nonetheless, those most likely to be found at home are stay-at-home parents, unemployed persons, and retired people. They will tend not to have busy volunteer and social lives. They are more likely to be parents of small children. Large households are more likely to have someone at home than households with only one or two members.

Accessibility of a different kind also produces biases associated with nonresponse. National surveys using personal interview procedures almost always have lower response rates in central cities than in suburbs and rural areas. There are three main reasons for this. First, the rate of hard-to-find single individuals is higher in central cities. Second, an increasing fraction of individuals in central cities live in apartment buildings that have significant security features that make it hard for interviewers to gain direct access to residents. Third, there are more areas in central cities where visits at night are uncomfortable for interviewers; hence, they may not give difficult-to-find people as a good a chance to be found at home.

The continuing increase in central city populations and other broad social changes (more single-person households, fewer families with children, more women in the labor force) have made the achievement of high response rates harder during the past 20 years for both in-person and telephone surveys. Telephone survey response rates may also suffer from increased use of caller ID and the rise of cell phone use—both of which may have decreased the rates at which people answer their landline phones and can be exposed to the interviewers’ introductions. De Leeuw and de Heer (2002) report trends for decline in response rates internationally, as well as in the United States. The relative roles of noncontact and refusals in nonresponse vary from country to country.

There is some evidence that telephone procedures may reduce the differential response rate between central cities and rural areas because it is possible to give more thorough coverage to urban households, to make contact with people in high-security buildings, and to make a large number of efforts to find single people at home. On the other hand, less-educated people seem less willing to be interviewed in a random-digit telephone procedure, as are those over 65 years of age. These biases are found less often in personal interview surveys. Groves and Kahn (1979), Cannell, Groves, Magilavy, Mathiowetz, and Miller (1987), Groves and Lyberg (1988), and Groves and Couper (1998) present good evaluations of the nonresponse biases in telephone surveys.

Finally, there is bias associated with people who are unable to be interviewed or to fill out a form. These persons usually compose a small fraction of a general population. Leaving out people who are in a hospital, however, may be a very important omission when trying to estimate health care utilization or expenditures. There are also neighborhood areas or groups where omitting people who do not speak English would be a significant factor. If special steps are not taken to collect data from a particular group, the sample estimates apply to a more restricted population: the population that actually had a chance to answer the questions or provide data, given the data collection procedures implemented.

Although there tend to be demographic differences between respondents and nonrespondents for interviewer-administered surveys, particularly for random-digit dialing based telephone surveys, the effect of nonresponse on survey estimates is less clear. Keeter, Miller, Kohut, Groves, and Presser (2000) report
a careful comparison of results of two telephone surveys: one yielded a 36% response rate, the other a 60% response rate. With respect to the political and social attitudes that were the topics of the surveys, there were very few statistically significant differences between the two survey results. The researchers reported a replication of this study with similar results (Keeter, Kennedy, Dimock, Best, & Craighill, 2006).

However, these studies should not lead researchers to believe that nonresponse is not an important source of error. Groves (2006) reports on an analysis of response rates and nonresponse error. He studied over 200 estimates based on 30 surveys that used various modes of data collection. He found considerable evidence of error due to nonresponse. However, he reached two other important conclusions:

1. The response rate for a survey was not a very good predictor of nonresponse error. The correlation he reports between the two is .33.
2. One main reason for the comparatively low association is the variability within surveys in the amount of nonresponse error for different variables.

Thus, for any given survey, some estimates may be affected a lot by nonresponse while others are affected very little.

The key issue is the extent to which nonresponse is related to the estimates the survey is designed to make. Groves, Presser, and Dipko (2004) report experimental results that show that people with roles relevant to a survey topic (such as new parents asked to do a survey on child care) are more likely than average to respond. However, they also found that such identifiable role groups often were small and, hence, might not have a measurable effect on survey estimates. It seems reasonable and is consistent with existing evidence, that respondent interest in the subject may play a bigger role in response to mail surveys than when an interviewer enlists cooperation.

Altogether, we have clear evidence that nonresponse can affect survey estimates, but we usually lack the information to reliably predict when, and how much, nonresponse will or will not affect survey estimates. Moreover, the effect of nonresponse on one variable can be very different than for others in the same survey.

Table 4.1 presents an example to help thinking about the potential effects of nonresponse on results. Suppose a sample of 100 is drawn, and 90 respond (response rate of 90%). Of those 90, 45 say "yes" to some question; the other 45 say "no." There are 10 people (the nonrespondents) whose views we do not know. If they all were "yeses," the true figure for the population would be 55% "yes." Thus, given:

Response rate = 90%
50% of respondents say "yes"

<table>
<thead>
<tr>
<th>Responding* (Answers Known)**</th>
<th>Nonrespondents (Answers Unknown)</th>
<th>Total Sample (Possible Range of Answers if Everyone Responded)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>45</td>
<td>0 to 10</td>
</tr>
<tr>
<td>No</td>
<td>45</td>
<td>0 to 10</td>
</tr>
<tr>
<td>Total</td>
<td>90</td>
<td>10</td>
</tr>
</tbody>
</table>

*If response rate was 90%.
**If 50% of respondents answered "yes".

Table 4.2 works out this logic for a range of response rates. It can be seen that when response rates are low, the potential for error due to nonresponse is very large.

<table>
<thead>
<tr>
<th>When Response Rate Is:</th>
<th>90%</th>
<th>70%</th>
<th>50%</th>
<th>30%</th>
<th>10%</th>
</tr>
</thead>
<tbody>
<tr>
<td>True value if everyone in sample responded could range from:</td>
<td>45%–55%</td>
<td>35%–65%</td>
<td>25%–75%</td>
<td>15%–85%</td>
<td>5%–95%</td>
</tr>
</tbody>
</table>

SOURCE: Adapted from table developed by Jack Elison and Mitchell D. Elison. Personal communication.
It may be instructive to compare the potential effects of nonresponse bias, such as those presented in Table 4.2, with the effects of sampling error presented in the preceding chapter (see Table 3.1). One usually does not know how biased nonresponse is, but it is seldom a good assumption that nonresponse is unbiased. Efforts to ensure that response rates reach a reasonable level and to avoid procedures that systematically produce differences between respondents and nonrespondents are important ways to build confidence in the accuracy of survey estimates.

**REDUCING NONRESPONSE**

**Telephone or Personal Interview Surveys**

Just as one can always say that a larger sample will be more reliable than a small sample, all other things being equal, one also can say that a survey with a higher response rate probably will produce a better and less biased sample than one that has more nonresponse. At the very least, given that nonresponse error is usually unmeasured, one can say that higher response rates increase the credibility of a survey’s results. As with any design decision, a researcher must choose how much effort to invest in reducing nonresponse.

Two different problems must be addressed in order to achieve a high rate of response for telephone and personal surveys: gaining access to the selected individuals and enlisting their cooperation.

To reduce nonresponse resulting from lack of availability

- Make numerous calls, concentrating on evenings and weekends. The number needed depends on the setting. Six calls per household are probably a minimum in urban areas. For phone studies, more calls can be made cheaply, and many organizations use 10 as a minimum. Often, many more calls than 10 are made.
- Have interviewers with flexible schedules who can make appointments at any time that is convenient to respondents.

To enlist cooperation

- If possible, send an informative advance letter. It reassures some respondents, and interviewers feel more confident as well.
- Effectively and accurately present the purposes of the project. Make sure respondents know their help is important and how it will be useful.

- Make sure that respondents will not be threatened by the task or the uses to which the data will be put.
- Have effective interviewers. Make sure they know that the response rate is important. Identify interviewers who are having response rate problems quickly and either retrain or do not continue to use interviewers who are not effective.
- Teach interviewers to listen to respondents and tailor the interaction to the individual rather than following a largely scripted approach to introducing the survey.

Striking the right balance between persistence and responsiveness to reluctant respondents is not easy. Participation in surveys is voluntary, but reluctance to participate is often based on ignorance. Groves and Couper (1998) present useful analyses of the reasons for respondent reluctance to be interviewed.

It is appropriate to ask respondents to be informed about the purposes of a survey before refusing. Most surveys serve a useful purpose from some vantage point. When a person refuses, the resulting data will be less accurate. Interviewers should be required to make a good effort to get respondents to the point that they know what they are being asked to do and why. It also is reasonable routinely to ask people who initially refuse to reconsider. A significant percentage of refusals result from contacting the respondent at the wrong time rather than a fundamental unwillingness to be interviewed. Between one quarter and one third of people who initially refuse will agree to be interviewed when asked again at a later time.

The interview process is generally a positive experience for respondents. If a survey research project is being conducted by a responsible group, responses will be kept strictly confidential. Survey researchers routinely respect confidentiality with the same zeal as psychiatrists and journalists protect their sources. Many survey research projects are serving some reasonable cause to which the majority of people would be willing to contribute. If the interviewer is willing to arrange an interview at the respondent’s convenience, pressures for time should not be extraordinary for most respondents. Finally, most respondents report that being interviewed is pleasurable. People like to have an opportunity to talk about themselves to a good listener.

RDD poses particularly significant challenges for achieving good response rates. When list-assisted approaches are used to select the sample, letters can be sent to households for which addresses are known. However, even when that is done, the burden is on the interviewer to engage the person who answers the telephone. The first few sentences are critical to
the questionnaire is returned. Should the respondent letter be signed in blue ink? Is a real stamp better than a postage-paid envelope?

Generally speaking, almost anything that makes a mail questionnaire look more professional, more personalized, or more attractive will have some positive effect on response rates. Tending to such details probably is worthwhile in the aggregate; Dillman (2007, 2008) reviews these issues well. It probably is also important to make the instrument easy to complete. More details about design are provided in Chapter 6, but there are three points worth mentioning here.

1. The layout should be clear so it is easy to see how to proceed.
2. The questions should be attractively spaced, easy to read, and uncluttered.
3. The response tasks should be easy to do. Do not ask respondents to provide written answers, except at their option. The response tasks should be to check a box, circle a number, or some other equally simple task.


Several reviews of published studies indicate that prepayment to respondents of several dollars increases mail responses rates (Fox, Crask, & Kim, 1988; James & Bolstein, 1990). Almost all studies showed a benefit, even surveys of doctors (Berry & Kanouse, 1987). However, it is only advance payment that has been shown to affect responses. Delaying payments, offers to give money to charities, and enrollments in drawings have not been shown to increase response rates (Warriner, Goyder, Jersten, Hohner, & McSpurren, 1996). Finally, there is no question that the most important difference between good mail surveys and poor mail surveys is the extent to which researchers make repeated contact with nonrespondents. A reasonable sequence of events, such as that outlined by Dillman (2007), might include the following:

1. About 10 days after the initial mailing, mail all nonrespondents a reminder card, emphasizing the importance of the study and of a high rate of response.
2. About 10 days after the postcard is mailed, mail the remaining nonrespondents a letter again emphasizing the importance of a high rate of return and including another questionnaire for those who threw the first one away.
3. If the response rate is still not satisfactory, probably the best next step is to call nonrespondents on the telephone. If phone numbers are not available or
if the expense of telephone calls seems too great, additional persuasion letters, night telegraph letters, overnight delivery, or other follow-up procedures that stand out and seem important have been shown to be helpful.

With telephone follow-ups, of course, response rates will be higher than without them. If the researcher is persistent, and if it is a reasonably well-conceived and well-designed study, acceptable response rates can be obtained by mail. There are many examples of mail surveys achieving response rates as high or higher than other modes (e.g., Fowler, Roman, & Di, 1998).

If a researcher is going to recontact nonrespondents, the researcher must know who has not returned a questionnaire. The process need not be circuitous or complex. A simple identifying number can be written on the questionnaire or on the return envelope. It is good practice to tell people in the covering letter what the number is for.

Occasionally, a researcher may want to reassure respondents that they will not be identified. There is a simple alternative strategy that works very well and still permits follow-up. The respondent is sent a questionnaire that has no identifier on it. Attached to the questionnaire is a separate postcard that has a respondent identifier, as follows:

Dear Researcher, I am sending this postcard at the same time that I am putting my completed questionnaire in the mail. Since my questionnaire is completely anonymous, this postcard will tell you that you need not send me a further reminder to return the questionnaire.

This procedure maintains the respondent’s anonymity, at the same time telling the researcher when someone has completed the questionnaire. Some might think that respondents simply would send back the postcard in order to avoid further reminders, but this seldom happens. The numbers of postcards and questionnaires returned almost always come out to be about the same. Sieber (1992) discusses a wider range of ways to collect data anonymously.

Internet Surveys

Although surveys on the Internet are comparatively new, so there is not the same body of experience that exists for mail and interview surveys, the dynamics and challenges seem likely to closely parallel those for mail surveys. The problem is to induce people to respond without the intervention of an interviewer.

There are two main ways that the Internet is used for surveys: respondents are asked to answer questions by replying to an e-mail questionnaire or they are asked to go to a Web site where a survey form is waiting to be completed. For a variety of reasons, using e-mail as a data collection approach is not a very good idea (see Chapter 5), and the use of the Internet for surveys is primarily about getting respondents to visit a survey site and fill out a questionnaire.

The most common way to ask people to complete an Internet survey is to send an e-mail invitation. As one might expect, this can produce widely varying results. Dillman (2007) reports a survey of faculty that obtained a response rate near 60%, nearly the same rate as they obtained in a parallel mail version of the same survey. Kaplovitz, Hadlock, and Levine (2004) report a similar experiment with students that obtained an Internet response rate that was half as high as the Dillman study and in which the postal mail got a significantly higher response rate than the e-mail strategy.

Both of these experiments were with populations that have virtually universal access to e-mail and used it routinely—and the survey request was identifiable as coming from an institution of which respondents were members. When survey requests come from less known or unknown sources and go to people who vary widely in how and how much they use the Internet, results are predictably variable. Sometimes, virtually no one responds.

The same kinds of steps that have been found to be helpful to postal surveys are likely to help enlist cooperative for Internet surveys: identifiable sponsors, well-designed instruments, financial incentives, and repeated contacts, including trying mail or phone requests for those who do not respond to an initial e-mail request. Dillman (2007) presents a thoughtful discussion of approaches to improving response rates.

Multimode Surveys

One of the best ways to minimize survey nonresponse is to use more than one mode to collect data. The key issues, as noted, are access, motivation, and cost. Mixing modes can enable researchers to reach people who are inaccessible via a single mode. It also can allow them to collect data from less intrinsically motivated sample members. For example, one attractive protocol is to use e-mail or mail for the first phase of data collection, followed by telephone interviews with nonrespondents. Combining telephone and in-person interviews is another effective design.

A critical issue in multimode surveys is the comparability of data across modes. Answers to some questions are affected by the mode of data collection; others are
not. In order to combine data collected using different modes, it is important that the data are comparable. Those issues are discussed more in Chapter 5.

TWO OTHER APPROACHES TO REDUCING NONRESPONSE ERROR

Proxy Respondents

Many surveys routinely collect data from one household respondent about other household members. If a respondent is unable or unwilling to be interviewed, asking another household member to report for the designated respondent is one option. Studies of the quality of such proxy data usually, but not always, indicate that proxy reporting is not as good as self-reporting for most topics. Moreover, few researchers would accept proxy reports of subjective states such as feelings, knowledge, or opinions. For factual information, however, if there is a knowledgeable proxy available, using proxy respondents can be an effective way to reduce error resulting from nonresponse. Groves (1989) reviews the inconclusive literature on the quality of data provided by proxy respondents.

Surveying Nonrespondents

Suppose a mail survey was done, and 60% of those sampled responded. The researcher thinks many nonrespondents would respond to a request to give a telephone or personal interview, but lacks the funds to try those procedures for all nonrespondents. An option is to draw a sample of nonrespondents to be contacted with the more expensive methods.

Two different uses can be made of such efforts.

First, the data from this second round of data collection can be used to estimate the direction and amount of bias in the initial sample. Of course, the second round also will have nonresponse, so it may produce data that do not fully represent all nonrespondents. However, subject to that limitation, the data may be used to improve the statistical adjustments (discussed in Chapter 10). With this perspective, the researcher might ask only a subset of the key questions, rather than all the questions, as one way of increasing the chances that a reluctant respondent might agree to cooperate. Some research organizations use significant financial incentives in these efforts as well.

Second, if the new round of data collection replicates questions in the initial survey, the results can be added to the initial sample data set. To do that, the data need to be weighted to adjust for the fact that only a sample of nonrespondents received the follow-up treatment.

If half the nonrespondents are followed up, then the respondents from this phase of data collection should be weighted by a factor of two when they are combined with the initial data. Moreover, the results can be appropriately reported as reflecting an adjusted percentage of the sample population (an adjusted response rate), calculated as follows:

\[
\text{Adjusted Response Rate} = \frac{\text{Phase I Responses} + 2 \times \text{Responses From Phase II}}{\text{Original Eligible Sample}}
\]

NONPROBABILITY (OR MODIFIED PROBABILITY) SAMPLES

The discussion in this chapter so far has assumed a probability sample design whereby respondents are designated by some objective procedure. The researcher's problem is to collect data about those designated. However, as we have been discussing, often it is difficult and expensive to get responses from a high percentage of a probability sample, particularly a general population sample that on average has no particular reason to be invested in a particular survey. It is therefore understandable that investigators have explored alternative approaches to easing the data collection burden.

Substituting Respondents

The disagreement among survey researchers about the importance of probability sampling is intriguing. The federal government generally will not fund survey research efforts designed to make estimates of population characteristics that are not based on probability sample techniques. Most academic survey organizations and many nonprofit research organizations have a similar approach to sampling. At the same time, almost all of the major public opinion polling groups, political polling groups, and market research organizations rely heavily on nonprobability sampling methods (see Converse, 1987, for a discussion of the historical roots of this difference).

The heart of probability sampling strategies is that the inclusion of someone in a sample is based on a predetermined procedure that sets a rate of selection for defined population subgroups. Beyond that, neither respondent characteristics nor interviewer discretion influence the likelihood that a person will be
in a sample. Although nonprobability modifications of sampling procedures vary, they all share the property that, at the last stage, interviewer discretion and/or respondent characteristics are part of the sample design and affect the likelihood of being included in a sample. The two most common procedures are described below.

For a personal interview study involving nonprobabilistic sampling, the researcher might draw blocks in much the same way that a sampler would draw blocks for an area probability sample. The difference would be that once a block was selected, the interviewer would be instructed to visit that block and complete some fixed number of interviews with people who reside on the block. There would be no specific listing of housing units on the block. One approach is to give the interviewer freedom to call on any home located on that block; the interviewer would not make callbacks to find people not at home on the initial visit.

A similar strategy is used for telephone surveys. Within a particular exchange, or cluster of numbers within an exchange, a target is set for completing a certain number of interviews. If there is no answer or no available respondent at the time the interviewer calls, another number is called from within the same cluster until the desired number of interviews is obtained. The first stage of sampling, if it is carried out as indicated above, distributes the sample around a geographic area more or less in the way that the population is distributed. At the point of household and respondent selection, however, there are three very clear kinds of biases that can be introduced.

In the personal interview strategy, but not the telephone strategy, interviewers can make a choice about which houses to visit. It turns out that interviewers will visit more attractive houses rather than less attractive houses, and first-floor apartments rather than second- and third-floor apartments. Interviewers also prefer housing units without dogs. Other factors that influence choices made by individual interviewers can be left to the reader’s imagination.

Some research organizations attempt to restrict interviewer discretion by providing instructions about where on the block to start and asking interviewers not to skip housing units. Without an advance listing of units on the block, however, it is virtually impossible to evaluate whether or not an interviewer followed these instructions. Moreover, if there is an advance listing of units on the street, one portion of the cost savings of this approach is eliminated.

In addition to the potential to substitute the households at which interviews are taken, there is great value in taking an interview with any household adult who is available and willing to be interviewed. Probability sampling requires specifying a respondent within a household by some objective method; in contrast, the substitution strategies permit interviews with any member of a contacted household who is willing to be interviewed.

One of the most obvious potentially biasing features of the nonprobability methods is the effect of availability. If one is not going to call back to the housing units where no one was home, or call back for household members who were not at home when the interviewer made contact, people who spend more time at home have a higher chance of being selected than those who routinely are not at home.

Uncontrolled sampling in this way produces some obvious sample biases. The most common approach to increasing the quality of the samples is to introduce quotas for obvious biases. Thus an interviewer may be required to interview half males and half females from any particular block or telephone cluster. Occasionally, some additional constraints will be set, such as the expected racial composition or the number of old or young adults. It is important, though, not to put too many constraints on the quotas, or interviewers will have to spend a great deal of time calling or wandering around blocks looking for eligible respondents.

The final bias inherent in allowing substitution has to do with the enlistment of cooperation. In the event that a respondent says he or she is busy or that it is not a good time to be interviewed, the interviewer has no incentive to enlist cooperation. If a project is not effectively presented, a significant fraction of the population will not be interested in helping. Letting people refuse easily without strenuous effort to present the study to them not only will bias a sample against the busy people, it also will bias it against the people who have less prior knowledge or less intrinsic interest in research and/or in the particular subject matter being studied.

Sudman (1967, 1976) argues that there is nonresponse in all surveys, even those in which every effort is made to contact nonrespondents. Once it has been learned that an individual will not cooperate or cannot be reached after several calls, he suggests that substituting a respondent from the same household or block actually may improve the quality of estimates. He argues that having a neighbor in the sample may be better than having neither the designated respondent nor his or her neighbor in the sample. When careful control is exercised over the interviewers’ discretion, however, as Sudman advocates, the savings in reduced callbacks are offset largely by increased supervisory costs.

Nonprobability sampling methods produce cost savings for personal interview surveys (less so for telephone surveys). One other critical advantage of these methods, particularly for political surveys, is that they make it possible to conduct surveys overnight or in a few days. There are many people for whom many callbacks over several days are needed to catch them at home. Quick surveys obviously have to rely mainly on response from people who are more available.

When a quota sample is effectively implemented, the resulting samples often look rather similar to probability sample data, to the extent that they can.
be compared. Even so, two facts should be kept in mind. First, because the key to saving money is to make no callbacks, only about a third of the population has a chance to be in most nonprobability sample polls (i.e., the population that is at home on a typical first call). A sample that only gives a third of the population a chance to be selected, a third of the population with known distinctive characteristics, has great potential to be atypical in ways that will affect sample statistics.

Robinson (1989) provides an excellent example of how distorted a presumably well-done nonprobability sample can be. He compared results from two surveys, one based on a probability sample, the other on a nonprobability sample, both aimed at estimating interest in the arts and attendance at art-related events. The nonprobability sample survey greatly overestimated the level of interest in the arts.

In addition to potential bias, another downside of quota samples is that the assumptions of probability theory and sampling error, which routinely are touted as describing the reliability of nonprobability samples, do not apply. If there are substitutes, the sample is not a probability sample, though it may be spread around the population in a reasonably realistic way.

There are times when nonprobability samples are useful. Henry (1990) describes the various kinds of nonprobability samples and when they might be appropriate. If a researcher decides to use a nonprobability sample, however, readers should be told how the sample was drawn, the fact that it likely is biased in the direction of availability and willingness to be interviewed, and that the normal assumptions for calculating sampling errors do not apply. Such warnings to readers are not common. In many cases, nonprobability samples are misrepresented seriously, and that constitutes a serious problem for the credibility of social science research.

**NONRESPONSE AS A SOURCE OF ERROR**

Nonresponse is a problematic, important source of survey error. Table 4.2 demonstrates the great potential of nonresponse to affect results. Yet, although we can calculate a rate of response, we usually do not know the effect of nonresponse on data. The Keeter et al. (2006) study illustrates a survey with a comparatively low response rate producing results that are very similar to one with a much higher response rate, but the Groves (2006) analysis reminds us not to be complacent.

The key problem is that we usually lack good data about when nonresponse is and is not likely to be biased with respect to the content of the survey. Certainly one unintended positive effect of the increasing concerns about response rates is to increase pressure on researchers to collect data about nonresponse bias when they can. However, it is hard to do. In the absence of such data, perhaps the strongest argument for efforts to maximize response rates is credibility. When response rates are high, there is only a small potential for error due to nonresponse to be important. When response rates are low, there is great potential for important error; critics of the survey results have a reasonable basis on which to say the data are not credible.

**EXERCISE**

If a sample of housing units is selected as the first stage of sampling adults aged 18 or older, the response rate is the number of completed interviews divided by the number of individuals in the study population designated by the sampling procedure to be in the sample. Would you include or exclude the following groups from the denominator when calculating the response rate? (Why?)

- Vacant housing units
- Those who were away on vacation
- Those temporarily in the hospital
- Those who refused to be interviewed
- Housing units in which all residents were under 18
- Those who could not speak the researcher’s language
- Those who others in the households said were mentally ill or too confused to be interviewed
- Those who were never at home when the interviewer called
- Those who were away at college

Define the population to which your sample statistics (and your response rate) apply.

**Further Readings**


Methods of Data Collection

The choice of data collection mode, mail, telephone, the Internet, personal interview, or group administration, is related directly to the sample frame, research topic, characteristics of the sample, and available staff and facilities; it has implications for response rates, question form, and survey costs. Computers can be used in the data collection process via all of these modes. An inventory of the various considerations in and consequences of the choice of a data collection mode is presented in this chapter.

One of the most far-reaching decisions a researcher must make is the way in which the data will be collected. Should an interviewer ask the questions and record the answers, or should the survey be self-administered? If an interviewer is to be used, there is the further decision about whether the interview will take place in person or over the telephone. If the respondent is to read and answer questions without an interviewer, there are choices about how to present the questionnaire to the respondents. In some cases, questionnaires are handed to respondents, in groups or individually, and returned immediately. In household surveys, questionnaires can be dropped off at a home or mailed and returned in a similar fashion. For those with Internet access, questions can be imbedded in e-mails or respondents can be asked to go to a Web site to answer questions.

Computers can be added to the mix in various ways. Of course, interviewers routinely use computers when doing telephone and in-person interviews. The more interesting variations have respondents entering answers directly into a computer. Surveys over the Internet involve no interviewer at all. Data can be collected in places where respondents go, such as doctors’ offices, by having respondents answer questions into computers. Finally, over the telephone, a computerized voice can ask questions that are answered by using the touch-tone numbers.

Although the majority of surveys utilize a single data collection method, it is not uncommon for combinations of methods to be used. For example, personal interview surveys sometimes have series of questions that respondents answer by filling out a self-administered form or entering answers directly into a laptop computer. Computerized data collection has also been combined with telephone interviews. To reduce nonresponse, people who fail to return mail questionnaires sometimes are contacted by an interviewer on the phone or in
person. Surveys using e-mail are often supplemented by mail survey protocols for those who lack an e-mail address or do not respond to e-mail. Respondents whom personal interviewers are unable to find at home or who have moved out of an area may be interviewed by telephone or asked to complete a self-administered form. Finally, some household surveys utilize telephone interviews for people at addresses for which telephone numbers can be obtained, but use a personal interviewer in households for which no telephone number can be found.

There are conditions under which each of these approaches to data collection is the best. In this chapter, the goal is to discuss the bases on which to choose among the various data collection approaches.

**MAJOR ISSUES IN CHOOSING A STRATEGY**

**Sampling**

The way a researcher plans to draw a sample is related to the best way to collect data. Certain kinds of sampling approaches make it easy or difficult to use one or another data collection strategy. If one is sampling from a list, the information on the list matters. Obviously, if a list lacks either good mailing addresses, good e-mail addresses, or good telephone numbers, trying to collect data by the corresponding mode is complicated. RDD greatly improved the potential of telephone data collection strategies by giving every household with a telephone a chance of being selected. Assuming one is willing to omit those without landline telephone service, it is perhaps the least expensive way to draw a general household sample.

Of course, it is possible to use random-digit-dialing strategies simply to sample and make initial contact with households, followed by data collection through the use of some other mode. Once a household has been reached, one can ask for a postal or e-mail address to permit an electronic or paper questionnaire to be sent or an interviewer to visit. Such designs are particularly useful when one is looking for a rare population, because both the sampling and the screening via telephone are comparatively less expensive than doing the same task with a personal interviewer. The difficulty with such strategies may lie in the rate of cooperation at the time of data collection.

When the basis of the sample is a set of addresses, either from a list or from an area probability sample, telephone, personal interview, and mail procedures all may be feasible. Obviously, if one has a good address, one can send an interviewer. In addition, it is possible to obtain telephone numbers for many addresses using commercial or Internet services that match names and addresses to published numbers and Directory Assistance.

Such approaches will not produce telephone numbers for all addresses. Therefore, some other mode of data collection will be needed as a supplement. It is often possible to interview a majority by telephone, however, thereby realizing the potential advantages of that mode. If one samples from a good list of addresses, a mail survey also is feasible. In an urban area with many multiunit dwellings, though, it is important that there be apartment unit designations (or household names) as well as a street address. Without an apartment unit or name, a mailing to a multiunit structure will go undelivered or will not reach the right unit. The problem is equally troublesome in rural areas where people may receive their mail in post office boxes. If a rural sample frame does not include names or addresses at people’s houses, a mail survey is out of the question.

Sample listings that include e-mail addresses obviously open that avenue of data collection. While e-mail is not yet a reasonable option for general population surveys, there are many populations (employees, students, members of professional organizations) for which e-mail addresses are nearly universal and are easily available. In those cases, using the Internet as the main, or at least one, data collection mode may be a good idea.

A final sampling issue to consider is designating a respondent. If the sample frame is a list of individuals, any procedure, including mail, is feasible. Many surveys, however, entail designating a specific respondent at the time of data collection. If a questionnaire is mailed to a household or organization, the researcher has little control over who actually completes the questionnaire. Therefore, the involvement of an interviewer is a critical aid if respondent designation is an issue.

**Type of Population**

The computer skills of the population, their reading and writing skills, and their motivation to cooperate are salient considerations in choosing a mode of data collection. Self-administered approaches to data collection place more of a burden on the reading and writing skills of the respondent than do interviewer procedures, and computer use and skills are added if one is considering an Internet approach. Respondents who are not very well educated, whose reading and writing skills in English are less than facile (but who can speak English), people who do not see well, people who do not use computers very much, and people who are somewhat ill or tire easily all will find an interviewer-administered survey easier than filling out a self-administered form.
Another problem for self-administered approaches is getting people to return a completed questionnaire. If no interviewer is involved, the intrinsic motivation of the respondent is likely to be critical to who responds; people who are particularly interested in the research problem tend to be most likely to respond (Fowler et al., 2002; Heberlein & Baumgartner, 1978; Jobber, 1984). In this context, if one is collecting data from a population that is highly literate and that, on the average, is likely to be highly interested in the research, procedures such as mail or e-mail become more attractive. At the other extreme, if one is dealing with a population in which reading and writing skills are likely to be low and/or where the average level of interest and motivation is estimated to be low, interviewer-administered data collection strategies are likely to be preferable.

Ease of contact is another consideration. Interviewer-administered strategies depend on an interviewer being able to contact the respondent and make arrangements for data collection. One great advantage of the self-administered strategies is that if the contact information is correct, the questions will get to the respondents. Moreover, busy people can respond at any time that is convenient for them. On the other hand, if a survey is work related, busy people in work settings are used to talking on the telephone and to making appointments; they often have people who will schedule appointments for them. Making an appointment for a telephone interview can be the best way to collect data from such people.

**Question Form**

Generally speaking, if one is going to have a self-administered questionnaire, one must reconcile oneself to closed questions, that is, questions that can be answered by simply checking a box or circling the proper response from a set provided by the researcher. In part, that is because ease of response is a priority to maximize returns. Second, self-administered open answers often do not produce useful data. With no interviewer present to probe incomplete answers for clarity and for meeting consistent question objectives, the answers will not be comparable across respondents, and they will be difficult to code. If such answers are useful at all, it usually is when they are treated as anecdotal material, not as measures.

Although open-ended questions usually require an interviewer, there are also some instances when closed questions can be handled better by self-administered procedures. One very good example is when a researcher wants to ask a large number of items that are in a similar form. Having an interviewer read long lists of similar items can be awkward and tedious. On such occasions, a good strategy may be to put the questions in a self-administered form either in a questionnaire or on the computer. Such an approach also provides a welcome change of pace for an interview.

Self-administered procedures also have an advantage when question response categories are numerous or complex. In a personal interview, it is common to hand the respondent a card listing responses to help respondents keep all the possible answers in mind. Telephone surveys, however, require some adjustments. Three approaches are used. First, researchers simply may limit response scales for telephone interviews. Some have argued that four categories is a comfortable maximum on the telephone; for many telephone surveys, two- or three-category responses predominate.

Second, a longer list can be handled if a telephone interviewer reads the categories slowly, then reads them again, with the respondent choosing the correct category when the interviewer gets to it. It has not been demonstrated, however, that answers obtained in this way are identical to those given to a visual list of categories. For some kinds of questions, the answers are affected by the order in which responses are read (Bishop, Hippler, Schwarz, & Strack, 1988; Schuman & Presser, 1981).

Third, researchers can break down complex questions into a set of two or more simpler questions. For example, it is common to ask people to report their income in more than four or five categories. A nine-category question can be asked in two phases as follows: Would you say that your total family income is less than $30,000, between $30,000 and $60,000, or more than $60,000? Then, depending on the answer, the interviewer proceeds to ask another three-alternative question, such as: Well, would you say that your total family income was less than $40,000, $40,000 to $50,000, or more than $50,000? These variations do work to make questions answerable, but sometimes the question form itself affects the answers (Groves, 1989).

There are question forms, including those with complex descriptions of situations or events and those requiring pictures or other visual cues, that cannot be adapted to the telephone. If such measurement is a critical part of a survey, some form other than the telephone probably is needed. Researchers have shown, though, that they can adapt the majority of survey questions to telephone use. If an instrument is to be used in both interviewer and self-administered modes, it is wise to design the interview version first. It usually requires fewer changes to adapt an interview schedule to self-administration than vice versa.

Computer-based modes provide a number of advantages that cannot be replicated with paper and pencils. For example, rules for which questions to ask that are contingent on the answers to more than one question are nearly impossible to follow without computer assistance. Sometimes it is desirable to randomize the order of questions or response alternatives—an easy task with a computer, almost impossible without computer assistance.
Finally, when respondents are working directly with a computer, much more complicated material can be included in a survey. Thus, pictures, audio material, and combinations thereof can be presented to respondents as part of the survey process.

**Question Content**

Many studies have compared the results of different data collection strategies (e.g., Cannell et al., 1987; Groves & Kahn, 1979; Hochstirn, 1967; Mangione, Hingson, & Barret, 1982). Good summaries of results are by de Leeuw and van der Zouwen (1988) and Dillman (2007). For most survey questions studied, the aggregate distributions obtained by personal interview, telephone interview, and self-administered procedures have been very similar.

Researchers have argued persuasively that one or another of the strategies should have an advantage when dealing with sensitive topics. Self-administered procedures are thought to be best because the respondent does not have to admit directly to an interviewer a socially undesirable or negatively valued characteristic or behavior. Others have argued that telephone procedures lend an air of impersonality to the interview process that should help people report negative events or behaviors. Moreover, random-digit dialing at least provides the option of having a virtually anonymous survey procedure, because the interviewer need not know the name or location of the respondent. Still others argue that personal interviewers are the best way to ask sensitive questions, because interviewers have an opportunity to build rapport and to establish the kind of trust that is needed for respondents to report potentially sensitive information.

Though all of these arguments sound plausible, the data clearly indicate that sensitive information is more frequently, and almost certainly more accurately, reported in self-administered modes than when interviewers ask the questions. Both self-administered paper forms and computer-assisted self-administration have been shown to produce the same results in comparison to interviewer-administered protocols (Aquilino, 1994; Dillman & Tonnai, 1991; Tourangeau & Smith, 1998; Turner et al., 1998). Moreover, these results apply to very sensitive material (such as illegal drug use and sexual behavior) and more subtle issues related to self-image, such as reported health status or the prevalence of "problems" after prostate surgery (Fowler, Roman, & Di, 1998; McHorney, Kosinski, & Ware, 1994). If potentially sensitive answers are an important focus of a survey, finding a way to get those answers without interviewer involvement will almost certainly improve the estimates.

When comparing telephone and in-person interviews in this respect, the data are not as clear (de Leeuw & van der Zouwen, 1988). However, it is probably most common for a telephone procedure to show differences in the direction suggesting a social desirability bias in the answers, compared with personal interviews. One of the most striking such differences was found by Mangione and colleagues (1982) in the rate at which people admitted having past drinking problems. Hochstirn (1967), Henson, Roth, and Cannell (1977), Aquilino (1994), and Fowler et al. (1998) found consistent results.

An entirely different aspect of question content that may affect the mode of data collection is the difficulty of the reporting task. In some surveys, researchers want to ask about events or behaviors that are difficult to report with accuracy because they extend over a period of time or are quite detailed. In such cases, reporting accuracy may benefit from a chance to consult records or to discuss the questions with other family members. The standard interview is a quick question-and-answer process that provides little such opportunity; this is especially true for telephone interviews. Self-administered procedures provide more time for thought, for checking records, and for consulting with other family members.

When great detail about small events is desired, such as information about what people eat or how they spend their money, or what television programs they watched, the very best approach is to have respondents keep a diary. Although there are limits to how many people will keep a diary, and often respondents are paid for doing so, it is a good way to get details. There are computerized alternatives to a paper diary, including having people periodically call an 800 number or give information over the Internet, that may be practical and can achieve the same data quality.

Overall, when samples are comparable, researchers have found that many survey estimates are unaffected by the mode of data collection. Unless some of the issues outlined above are very central to the research project, it is likely that the decision about how to collect data should be made on grounds other than the interaction between the subject matter of the questionnaire and the mode of data collection. Nonetheless, attention to the way question form or content might interact with mode of data collection to affect the results is an important part of the survey design process.

**Response Rates**

The rate of response is likely to be much more salient in the selection of a data collection procedure than other considerations. Obviously, one of the great strengths of group-administered surveys, when they are feasible, is the high rate of response. Generally speaking when students in classrooms or workers at job settings are asked to complete questionnaires, the rate of response is near 100%. The limits on response usually stem from absenteeism or scheduling (shifts, days off).
The response rate for mail or e-mail surveys depends critically on the population and the survey's purpose. A survey of prostate surgery patients covered by Medicare achieved an 82% return by mail before telephone contacts brought the return rate over 90% (Fowler et al., 1998). In contrast, there are mail surveys done that achieve returns from fewer than 20% of those surveyed.

There is no doubt that the problem of nonresponse is central to the use of mail surveys. As noted in the previous chapter, if one simply mails questionnaires to a general population sample without appropriate follow-up procedures, the rate of return is likely to be less than 50% (Heberlein & Baumgartner, 1978). If extensive and appropriate follow-up procedures are utilized if the project is otherwise well designed and executed, response rates often can be obtained for mail surveys that are similar to rates obtained using other modes (e.g., Dillman, 2007, Fowler et al., 1998).

The effectiveness of telephone strategies in producing high response rates depends in part on the sampling scheme. One way of utilizing the telephone for surveys is to replicate personal interview procedures. If one has a list of addresses as well as telephone numbers, an advance letter can be mailed introducing the study and explaining its purposes. After that, an interviewer can call on the telephone and ask for cooperation. Under these circumstances, telephone and personal response rates do not differ significantly. This is particularly true if interviewers can offer the option to nonrespondents of being interviewed in person (Groves, 1989; Hochstim, 1967, Mangione et al., 1982).

The procedures outlined above, however, are representative of only a minority of telephone surveys. Much more common are studies that link the telephone with random-digit-dialing sampling. One distinctive characteristic of RDD is that advance notice is not usually sent, though some organizations use reverse-address directories and send letters when an address is matched to a selected number. For a variety of reasons, response rates to random-digit-dialing telephone surveys have been falling over the past decade. That phenomenon, plus the rise of cell phone use, have made the representativeness of random-digit-dialing samples of increasing concern.

In conclusion, for list samples when advance letters are possible, and telephone coverage is nearly universal, there is little difference between telephone and personal interview procedures with respect to response. Moreover, when researchers want to reinterview people previously interviewed to get further information, the rate of response via the telephone is not different from that obtained by personal interviewers. For broader populations, it appears that one of the costs of random-digit-dialing telephone surveys is that the rate of response of selected households is lower, often much lower, than would be obtained by a personal interview survey. When a 5% to 20% reduction in response rates is multiplied by the rate at which people without phones, or landlines, also are omitted from such samples, the differential in the rate of response is not trivial. This is a disadvantage that researchers must be prepared to accept, or to work very hard to avoid, when they choose a random-digit-dialing approach.

**Costs**

The great appeal of mail and telephone survey procedures is that they usually cost less per return than personal interviews. Least expensive of all, of course, are surveys using the Internet. Survey costs depend on a multitude of factors. Some of the more salient factors are the amount of professional time required to design the questionnaire, time to program and test the computer-assisted program, the questionnaire length, the geographic dispersion of the sample, the availability and interest of the sample, and the availability of trained staff.

The perceived costs for a mail survey can be misleading. The cost of postage, clerical time for mailing, and of printing questionnaires turns out not to be trivial. Moreover, if there are telephone reminder calls, the expense gets higher.

Another key to the comparison is the telephone charges that are involved. Telephone use costs also will affect the personal-telephone cost comparison, but personal household interviews almost always will cost more per interview than telephone interviews with the same sample. Necessarily, the wages and expenses for an interviewer to visit a house and make contact with a respondent will exceed those for telephoning.

The comparative costs of mail and telephone modes will also depend on the population. If it is a highly motivated sample that readily returns surveys, mail costs will be lower than telephone costs. However, in a more typical case, to achieve similar rates of response, mail and phone modes may be fairly similar.

Although the choice between mail and telephone surveys offers can be made on grounds unrelated to cost, cost usually has to play a role in choosing a personal interview procedure. Yet there are many cases in which the strengths of the personal interview procedure make it the best choice to achieve a given set of research objectives.

Finally, of course, if a survey can be done over the Internet, the costs per return potentially are the lowest of all. Depending on the kind of data collection, there may be some initial investment in software. Other than that, the main data collection cost is the staff time to design the survey instrument and test it. The key issues to the comparisons with other modes, of course, are whether or not an appropriate sample can be contacted via e-mail and the rate of response that can be achieved.
Available Facilities

The facilities and staff available should be considered in choosing a data collection mode. The development of an interviewing staff is costly and difficult. Attrition rates are generally high for newly trained interviewers. Many new interviewers are not very good at enlisting cooperation of respondents, producing high refusal rates at the start. In addition, people who are good at training and supervising interviewers are not easy to find. Thus one very practical consideration for anyone thinking about doing an interviewer-conducted survey is the ability to execute a professional data collection effort. If one has access to an ongoing survey operation or if staff members have experience in interviewer supervision and training, interviewer studies become more feasible. If not, self-administered surveys have a real advantage.

Length of Data Collection

The time involved in data collection varies by mode. Mail surveys usually take 2 months to complete. A normal sequence involves mailing the questionnaires, waiting for a while, doing some more mailing, some more waiting, and some eventual telephone or in-person follow-up. Of course, the Internet eliminates the time waiting for delivery, but it still usually involves repeated contacts and reminders. At the other extreme, it is quite feasible to do telephone surveys in a few days. The very quickest surveys pay a cost in nonresponse, because some people cannot be reached during any short period. Telephone surveys routinely, however, can be done more quickly than mail or personal household interview surveys of comparable size.

The length of time required to do a personal household interview survey defies generalization, simply because it depends so much on the sample size and the availability of staff. It is safe to say, though, that it is only a very unusual circumstance in which the data collection period for personal interviewing would not be greater than for a comparable telephone survey.

Computer-Assisted Data Collection

The traditional survey has relied on paper-and-pencil procedures, in the hands of either interviewers or respondents. In the past decade, however, computers have replaced paper and pencils: Questions to be asked pop up on a computer screen to be read by interviewers or respondents, and answers are recorded by keying codes into the computer. The principal advantage of computer-assisted data collection is having answers instantaneously in machine-readable form.

For some surveys, there also are advantages for data collection:

- The computer can follow complex question skip patterns that are difficult in a paper-and-pencil version.

- Information from previous questions, or even from previous interviews, can be taken into account in question wording or the sequence of questions asked.

- If inconsistent data are given, the computer can identify the inconsistency, and it can be reconciled at the point of data collection.

Offsetting these advantages, considerable lead time is needed to make sure the computer-assisted data collection is error free, and as is discussed in somewhat more detail in Chapter 9, researchers lose the ability to check or exercise any quality control over the data entry process. So, like most decisions about survey design, the value of having computer-assisted data collection varies with the requirements of each individual project. Couper and colleagues (1998) present the most recent studies related to computer-assisted data collection.

Computer assistance is used most commonly in connection with telephone surveys. In fact, to some people, computer-assisted telephone interviewing (CATI) is virtually synonymous with a telephone survey. To date, there is no documentation that the quality of data from telephone surveys is affected by whether or not data collection is computer assisted, except for reduction in missing data (Catlin & Ingram, 1988). Most of the documented advantages and disadvantages are practical: the ease of managing question form and order, the speed of data entry, sample management, and the potential to provide interviewers with "help" in the form of instructions or definitions as needed. The disadvantages include the need for error-free programs, the difficulty of interviewers going back to make corrections, and the risk of computer systems going down. In addition, although questions in narrative form can be handled by having interviewers type in verbatim answers, computer-assisted data collection further increases pressure to ask only fixed-response questions.

Most computer-assisted interviewing is done from a centralized telephone facility. With lightweight portable computers, however, personal household interviewers also carry out computer-assisted data collection. In addition, in selected settings, such as physicians' offices, computers have been used to collect data from people in a self-administered way: Respondents sit down at a computer, read questions on a screen, and respond by entering answers without benefit of an interviewer. Computers with touch screens or mouse capabilities are particularly suitable for such data collection efforts. The
advantages of computer-assisted personal interviewing (CAPI) are about the same as those for CATI: ease of question management and rapid compilation of data. At the end of an interviewing day, an interviewer can transmit data to a central office over telephone lines.

Although computer assistance for self-administered surveys is still evolving, there are some additional interesting potentials that are likely to be realized. For example, computers make it possible to present information and stimuli in forms other than words (e.g., pictures).

Computers have the potential to adjust the language of the questions to the language of the respondent, as well as to read questions out loud for those who have difficulty reading. The ability of computers to alter the choice or sequence of questions to fit previous answers is a particular strength in self-administration, where complex skip instructions are difficult for respondents. Call-in computers can ask questions and record answers via touch-tone data entry, offering an alternative to the Internet to let respondents provide data at any time they choose. Finally, respondents appear to be more comfortable keying sensitive information into computers than providing the same information to an interviewer.

**SUMMARY COMPARISON OF METHODS**

The preceding discussion is not exhaustive, but it does cover most of the major considerations. The choice of a data collection mode is a complex one that involves many aspects of the survey research process. A summary of the strengths and weaknesses of the main approaches to collecting data follows.

*Potential advantages of personal interviewing:*

- Some sample designs can be implemented best by personal interview (e.g., area probability samples).
- Personal interview procedures are probably the most effective way of enlisting cooperation for most populations.
- Advantages of interviewer administration such as answering respondent questions, probing for adequate answers, accurately following complex instructions or sequences are realized.
- Multimethod data collection including observations, visual cues, and self-administered sections, on paper forms or into a computer, are feasible.
- Rapport and confidence building are possible (including any written reassurances that may be needed for reporting very sensitive material).
- Probably, longer survey instruments are possible in person than by any other mode.

*Potential disadvantages of personal interviewing:*

- It is likely to be more costly than the alternatives.
- A trained staff of interviewers that is geographically near the sample is needed.
- The total data collection period is likely to be longer than telephone procedures.
- Some samples (those in high-rise buildings or high-crime areas, elites, employees, students) may be more accessible by some other mode.

*Potential advantages of telephone interviewing:*

- Unit costs are usually lower than for personal interviews.
- Random-digit dialing (RDD) can be used to sample general populations.
- It provides better access to certain populations, especially compared to personal interviews.
- Data collection periods are usually shorter than for alternatives.
- The advantages of interviewer administration (in contrast with mail or Internet surveys) can be realized.
- Interviewer staffing and management is easier than for personal interviews: A smaller staff is needed, it is not necessary to be near the sample, and supervision and quality control potentially are better.
- There is likely to be a better response rate from a list sample than from a mail sample.

*Potential disadvantages of telephone studies:*

- There may be sampling limitations, especially as a result of omitting those without landline telephones or the inability to contact those on a list if a correct phone number cannot be found.
- Nonresponse associated with random-digit-dialing sampling is higher than for personal interviews.
- Questionnaire or measurement constraints are associated with the telephone, including limits on response alternatives, use of visual aids, and interviewer observations.
- Telephone is possibly less appropriate for personal or sensitive questions.

*Potential advantages of self-administered (versus interviewer-administered) data collections:*

- Presenting questions requiring visual aids is possible (in contrast to with telephone interviews).
- Asking questions with long or complex response categories is also feasible.
• Asking batteries of similar questions may be more acceptable to respondents.
• The fact that the respondent does not have to share answers with an interviewer makes collection of sensitive data likely more valid.

**Potential disadvantages of self-administration:**

• Especially careful questionnaire design is needed.
• Open questions usually are not useful.
• Good reading and writing skills are needed by respondents.
• The interviewer is not present to exercise quality control with respect to answering all questions, meeting question objectives, or the quality of answers provided.
• It is difficult to control who answers the questions.

Self-administered surveys can be done by mail, via group administration, or in households. Each approach has strengths and potential weaknesses.

**Advantages of group administration:**

• Cooperation rates are generally high.
• It provides a chance to explain the study and answer questions about questionnaire (in contrast to mail surveys).
• Generally, the unit costs are low.

The main disadvantage is that only a small number of surveys can use samples that can be gotten together in a group.

**Advantages of mail procedures:**

• Unit costs are relatively low.
• They can be accomplished with minimal staff and facilities.
• Mail provides access to widely dispersed samples and samples that for other reasons are difficult to reach by telephone or in person.
• Respondents have time to give thoughtful answers, to look up records, or to consult with others.

**Disadvantages of mail procedures:**

• Mail contact may not be an effective way to enlist cooperation (depending on the group to be studied and the topic).
• There are various disadvantages of not having interviewer involved in data collections.
• Good mailing addresses for the sample are needed.

**Advantages of dropping off (and later picking up) questionnaires at households:**

• The interviewer can explain the study, answer questions, and designate a household respondent, in contrast to mail.
• Response rates tend to be like those of personal interview studies.
• There is more opportunity for respondents to give thoughtful answers and consult records or other family members than in personal or telephone interview surveys.
• Trained interviewing staff is not required.

**Disadvantages of dropping off questionnaires:**

• This procedure costs about as much as personal interviews.
• A field staff is required (albeit perhaps a less thoroughly trained one than would be needed for personal interviews).

**Potential advantages of Internet surveys:**

• The unit cost of data collection is low.
• There is potential for high speed of returns.
• All the advantages of a self-administered instrument can be realized.
• All the advantages of a computer-assisted instrument can be realized.
• Like mail surveys, they provide time for thoughtful answers, checking records, or consulting with others.

**Potential disadvantages of Internet surveys:**

• Samples are limited to Internet users.
• Good addresses are needed.
• There is the challenge of enlisting cooperation (depending on the sampled groups and the topic).
• The various disadvantages of not having interviewer involved can affect the data collection.

Finally, when considering options, researchers also should consider combinations of modes. As noted, answers to many questions are not affected by mode of data collection. Combinations of personal, telephone, mail, and Internet procedures may offer the cost savings associated with the less expensive modes without the sampling or nonresponse prices they sometimes entail. Dillman (2007) and de Leeuw, Dillman, and Hox (2008) discuss some such combinations.
CONCLUSION

It is clear that the choice of mode is a complex decision and depends very much on the particular study situation. All of the above strategies are the best choice for some studies. It is appropriate, however, to note that the trend definitely has changed with respect to surveys of general, household-based samples. Thirty-five years ago, a researcher would have assumed that a personal interview survey was the method of choice for most studies. The burden of proof was on the person who would argue that another method could produce data that were as satisfactory. Because of the cost advantages, in the last two decades of the 20th century, a researcher had to address directly the question of why an interviewer-administered survey could not be carried out by telephone. However, because of the widespread concerns about nonresponse to telephone surveys, the trend has changed again. While telephone surveys with random-digit-dialing samples are still widely done, there is a serious ongoing search for alternative ways to do general population surveys.

The role of self-administered techniques has grown in the past decade for two reasons. First, of course, the development of the Internet has opened a brand new approach to data collection. Researchers are eager to explore the Internet’s potential. Second, considerable research evidence has demonstrated that self-administered procedures, particularly those that are computer assisted, can collect better data about sensitive topics than can interviewers. Those findings, along with expanding research into areas such as drug use and risky sexual behavior, have led to increased interest in integrating those strategies into data collection protocols.

Finally, it should be clear that the total survey design approach is critical when making a decision regarding mode of data collection. A smaller sample of personal interviews may produce a more useful data set than a larger sample of telephone interviews for the same price. A good sense of methodological goals and thoughtful consideration of all the design issues affecting cost and data quality are necessary before an appropriate decision can be made about how to collect survey data.

EXERCISE

Disregarding monetary costs, describe a survey research problem for which a mail survey would probably be the best choice, and explain why it would be better than the alternatives. Do the same for a random-digit telephone survey, for a personal interview household survey, and a survey using the Internet for data collection.

Further Readings