5

Survey Research

Scott Keeter

Overview

The survey is the most commonly used social research method for collecting data from individuals in a population of interest. Surveys are conducted by nearly every government in the world, by academic researchers, by nonprofit and nongovernmental organizations, by journalists, and by corporations and businesses. Much of what we know about the economic, physical, and social characteristics of the populations of the world is derived from survey data.

Perhaps the best-known survey in the United States is the U.S. census, mandated by the Constitution to gather information about the size of the population through an "actual enumeration" of every person every 10 years. The word *census* means a collection of data from all units in a population. But most surveys are not censuses. Rather, they entail some type of sampling in order to save time and money in the data collection process.

Surveys can be conducted in many different ways. An interviewer may ask the questions, either in person or by telephone, or the respondent may complete the survey by reading the questions in a paper questionnaire or on a computer screen and providing written or electronic responses. And there are other possibilities as well, including prerecorded questions administered to respondents, who answer via the touch-tone pad on their phones or on a computer screen. Some surveys use more than one data collection technique.

As with all social research, surveys raise important and difficult ethical questions as well. Good research takes ethical concerns seriously and attempts to minimize the time and effort required of respondents, to protect their privacy, and to ensure that the survey does no harm to those who participate.

This chapter will review the principles of good survey research in an effort to provide some of the knowledge needed to design and conduct

effective surveys. Even if you do not conduct your own survey, an understanding of good survey practice will enable you to assess the quality of survey evidence used in others' research.

The discussion here will be organized around four critical aspects of the survey process. The first is the overall *research design,* which encompasses elements of the other three aspects. Key decisions in the overall research design include the question of whether to do a survey at all or to use some other type of data collection method, the choice of the population to study, the timing of the data collection, the choice of whether data collection is to be done at one point or at multiple points in time, and the nature and level of resources to be made available to the study.

The second aspect is *measurement,* or how the key variables of interest will actually be measured. In practical terms, what questions will be asked of the respondents? The third aspect of survey research is *sampling.* Unless a census is undertaken, choices must be made regarding who in the population will be selected for interviewing. How samples are drawn and implemented determines how well the data will represent the population of interest. The fourth aspect is *survey administration.* Choices must be made regarding the mode of data collection and how it will be implemented. For example, if the telephone is used to collect data, when will calls be made and how many attempts will be made to reach each sampled telephone number?

Although surveys are usually thought of as a type of *quantitative* research, they stand at an intersection between the qualitative and the quantitative, and the best surveys make use of key principles and methods from the major forms of qualitative work. Surveys can make a strong claim for reliability because of the relatively large number of cases in a typical survey (at least by comparison with the typical qualitative study), but reliable evidence is an insufficient base on which to make important decisions; the evidence must also be valid, and qualitative research is fundamentally concerned with validity.

Ultimately, the goal of a successful survey is to describe the population with a minimum of error. At each stage of the process, we must consider how to reduce error so as to maximize *both* reliability and validity. Achieving greater reliability and validity in surveys requires attention to a range of details in the design and implementation of a survey. With a finite amount of time and money, we are constantly faced with choices about how to allocate resources most effectively; greater effort placed on one aspect of the survey may mean less effort can be made in another aspect. Do we spend money enlarging the sample size (which would reduce sampling error and improve reliability), or do we spend the money on achieving a higher rate of response with a given sample? Do we spend more time testing our questionnaire before putting it into the field, or do we spend the time making sure that our sampling frame is clean and accurate? These are the trade-offs inherent in the complex process of survey research. The goal of this chapter is to help you make

educated choices about your own survey and to be an informed consumer of surveys conducted by others.

Survey Research Design

After settling on a research topic, the first big choice in a project is how to collect the data. Is a survey the best way to obtain the needed data? Most of the major alternatives to a survey are covered in other chapters in this book. Surveys are complicated and costly, both for the researcher and for the respondents because of the time they require to complete. Some subjects simply are not amenable to survey research because respondents do not have the knowledge necessary to respond in a useful way or may not be willing to answer candidly.

Archives of Survey Data

Inter-university Consortium for Political and Social Research, University of Michigan. http://www.icpsr.umich.edu/

Roper Center for Public Opinion Research, University of Connecticut. http://ropercenter.uconn.edu

Central Archive for Empirical Social Research, University Cologne (Zentralarchiv für Empirische Sozialforschung, Cologne) http://www.gesis.org/en/za/index.htm

Pew Research Center for The People & The Press, Washington, D.C. http://people-press.org

U.S. Census Bureau, U.S. Dept. of Commerce. http://www.census.gov/

One consideration is that survey data regarding your topic may already exist and could be subjected to *secondary analysis* to answer the key research questions. An astounding quantity of excellent survey data can be found in several important *data archives* around the world (see box). Additionally, many researchers who have collected survey data for a particular purpose are willing to share the data after they have mined it for their needs. But often these surveys are not donated to archives, and finding them will require a little detective work. Even if you do decide to go forward with your own survey project, looking at other surveys on your topic can be extremely helpful. For example, U.S. census data can give you a good idea about the incidence of certain groups in the population, a concern you might have if you are targeting a relatively small

minority of the population. Census data can also tell you where you are likely to find certain kinds of people, which might help you be more efficient in your sampling. And a review of other surveys on your topic can help you find survey measures that may be appropriate, along with data with which to make comparisons across populations or over time.

If you decide to conduct a survey, several important decisions loom. Many of these will be discussed in greater detail in the sections on measurement, sampling, and survey administration below. Assuming that you have a clear goal in mind for your research, some overarching questions include the following:

- How much money (or other comparable resource) do you have available for the survey?
- How much time do you have to complete it?
- Given the resources you have available—specifically, money and time—what is the appropriate scope of the study?
- How much data do you need to collect from each respondent? How many respondents do you need to interview?
- Is it essential or even feasible to use a probability sample? If not, can an acceptable non-probability sample be created?
- What mode of data collection should be employed? When is the best time to conduct the survey?
- Will this study be a one-shot data collection, or is it a longitudinal study with data collections at multiple points in time?

Advertising Slogan for a Survey Research Firm

We offer . . .

- Quality
- Speed
- Low Cost

Pick two.

Answering these questions, even if only tentatively, will help to set the general contours for the research. Answering them may illuminate basic tensions and tradeoffs among the elements. In almost every kind of research, there is an inherent tradeoff among quality, speed, and cost. Because the money available is almost always limited to some degree, decisions must be made about which aspects of the survey would benefit the most from greater investment and which would harm the study the least if corners were cut. One of the clearest examples of this tradeoff can be found in the choice of a sample size. The marginal cost of additional data collection is usually very high, so it is worth considering how the

money that would otherwise be used for an expanded sample could be better used. If we settle upon the smallest possible sample that will satisfy the research needs, we might be able to increase the size of the questionnaire and measure more concepts of interest, or obtain better measures of concepts by using multiple measures. We might do more extensive pre-testing of the questionnaire to ensure that respondents understand the questions and can answer them. Or we might even use the savings to obtain a higher response rate and reduce the chance that nonresponse will introduce a serious bias into our results.

THE ELEMENT OF TIME

One key decision about your research design does not fit neatly into the other three aspects of the survey process but is logically prior to them—the timing and number of surveys. For many, if not most, research projects, it may not matter a great deal when the data are collected, and this decision is made largely by default. The data collection occurs whenever the researcher can assemble the necessary resources and logistical wherewithal to execute the study. But often the timing of the study is critical to the validity of the inferences we want to draw from it. Or, our population of interest may be more accessible to us, or more amenable to cooperating, at some times rather than others. Researchers in the United States often try to avoid conducting important surveys of the general public during the last two weeks of August or during the Christmas holidays, both traditionally times of vacations and travel. Surveys of schoolchildren conducted during the last week of the school year may not receive the careful attention from the students we would otherwise hope for. And, of course, researchers should respect the privacy of respondents by not conducting telephone surveys late in the evening or early in the morning. In general, researchers should consider the personal habits and schedules of those in the population and adjust the timing of the surveys accordingly.

Perhaps a more important element of timing pertains to knowledge, perceptions, or experiences we wish to measure with our survey. In general, the sooner we interview someone about an experience they recently have had, the more reliable their recall of the experience. At the same time, the full import of events may not be felt until later. Thus, the choice of when to conduct a survey depends on the research goals. Surveys about public response to the terrorist attacks in the United States on September 11, 2001, conducted within a few days of the event yielded different results from surveys conducted months later. If we are interested in assessing the long-term impact of the event, the latter surveys may be more valid because they will reflect the thinking of the public after considering the issue, discussing it with friends and family, and absorbing news coverage and commentary. But if we want to know exactly what people felt at the time, how they reacted emotionally, and how they followed news coverage of it, surveys conducted near the event provide better evidence.

Time is an important basis for comparison as well. Many surveys are conducted to measure change in opinions or experiences over time. For some phenomena, we try to draw inferences about change using a single survey by comparing people who are older and younger, using the assumption that the difference between them is an indication of the effect of aging, or changing financial or life cycle circumstances. Or, we compare people who have been married for different lengths of time, in an effort to gauge whether people and relationships change as time passes. Surveys used in this fashion are sometimes called *cross-sectional surveys.* But this is a relatively weak way to assess change; people in a single survey may differ for many other reasons besides their ages or how long they have been married.

In order to make better inferences about change, we need surveys taken at more than one point in time. The simplest design for a *longitudinal* study would use two surveys of the same population at different time points. Inferences about change over time are much stronger with this kind of design because we can make comparisons between similar people in the two surveys (e.g., lower income African Americans or older divorced women) and thus control for other possible reasons for differences between them. If similar groups in the two surveys are different, we can draw the inference that something happened in the interim to produce that change. If the amount of time between surveys is very large, normal aging and movement through the life cycle are also a plausible explanation for the changes we observe. One technique for examining this type of process is *cohort analysis,* in which we compare people from the same birth cohorts—that is, people who were born in the same year or set of years—at the two time points. A simple example will illustrate.

One of the more interesting attitudinal changes observed in the United States during the past three decades is the decline of interpersonal trust. According to standard measures that have been asked of survey respondents since the 1970s, the percentage of people who indicate that they believe others will take advantage of them has increased. In 1973, the large national General Social Survey (http://www.norc.uchicago.edu/projects/gensoc.asp) found 34% saying people would try to take advantage of them. By 2002, 40% felt this way. What has happened to social trust in the United States? Three general mechanisms could be at work, separately or together.

- *Aging effects:* Have people become less trusting as they grew older through a process of growing skepticism that came with greater maturity?
- *Period effects:* Did the social climate in the United States deteriorate so much that people lost faith in others as they experienced greater conflict, crime (and the depiction of crime on television), and disorder?
- *Cohort effects:* Did change occur because successive generations of young people came of age socialized to regard others with lower levels of trust, replacing older, more trusting cohorts who were dying out?

Table 5.1 Example of Cohort Analysis

Percent Who Say People Will Try to Take Advantage of You				
Birth Year	*1973–1982*	*1983–1992*	*1993–2002*	*Total*
< 1915	28	26	22	26
1915–1924	31	28	29	29
1925–1934	28	33	33	31
1935–1944	33	33	33	33
1945–1954	40	36	36	37
1955–1964	48	43	40	42
1965–1974		51	49	49
1975–			54	54

Source: General Social Survey.

Although there are many complexities to cohort analysis, and it is nearly impossible to definitively isolate the relative importance of each of the three mechanisms, we can gain considerable insight into what's going on by comparing successive cohorts across time. People aged 20–29 in 1973 were aged 30–39 in 1983. Even though the same individuals were not interviewed again, those who are ages 30–39 in the 1983 survey are a sample of the same people who were 20–29 in 1973. In 1993, this group would be 40–49. As the table shows, the evidence is very strong in support of a cohort effect. Each new generation of young people is less trusting than the one that came before it, but once into adulthood, the cohorts do not become less trusting as time passes. The overall level of trust goes down because of generational replacement; older cohorts, who are more trusting, leave and are replaced by younger ones, who are less trusting (see Table 5.1).

Measurement

During the 1980s, in the early days of survey research about the growing problem of AIDS in the United States, researchers were testing questions for a survey on sexual behavior. At one university survey center, a middle-aged woman was taking part in an in-depth interview that would help clarify how

people understood the language and terminology to be used in the final survey questionnaire. A critical question was about sexual orientation. She was asked, "Are you heterosexual, homosexual, or bisexual?"

The woman thought about the question for a moment, and then replied, "Well, it's just me and my husband, so I guess we're bisexual." The researchers then knew that they needed to work on the question a little more before putting it into the field.

Creating and executing a successful survey entails a skillful combination of art and science. In no part of the survey process is the need for art greater than in measurement: the creation of the questions that respondents will be asked. This is not to say that measurement is uninformed by science; to the contrary, our ability to write clear, specific, comprehensible questions that people are able and willing to answer has been greatly enhanced by psychological theory and research as well as by rigorous experimentation (see Tourangeau, Rips, & Rasinski, 2001). Yet the fundamental task of communicating clearly remains a quintessential application of human sensitivity and judgment. Fortunately, even though few of us will have the good fortune to always write the perfect question, there are many habits we can develop that will help us improve our skills to write good ones.

The first principle of good question writing is to know what the heck you want to learn from respondents. While this may sound obvious, it is sometimes harder to put in practice than it seems. A typical research project has a clear idea of the general subject of interest and of many important elements of the subject. But surveys tend to be quite detailed in content, and the process of writing questions often founders because of a lack of clarity about what information is actually needed from respondents.

Several important points to be considered in this section are the following:

- Always ask yourself these questions before writing survey questions: Will respondents have the necessary *knowledge* and *cognitive skill* to understand the questions and to answer them? Will respondents be *willing* to answer the questions? Will they be willing to answer them *honestly?*
- Seek as much input and criticism about your draft questions and questionnaire as you can afford.
- Always allow adequate time and resources to test your questionnaire with typical respondents before putting the survey into the field.
- Put as much effort into planning the sequencing of questionnaire items as you put into the construction of the individual items themselves.

There are important differences between measures of *subjective* and *objective* phenomena. Attitudes and opinions fall into the category of subjective phenomena, while events, demographic characteristics, and physical conditions are objective phenomena. Both are important in most surveys, but each presents different challenges to good measurement.

Consider the audience for your research and, more generally, the purpose of the survey when writing specific questions that deal with the heart of your project. There is often a trade-off to be made between optimal measurement for statistical analysis versus ideal measurement for face validity and interpretability of the questions. For example, if you are building a statistical model to explain the extent of commitment to an ideology or a negotiating position and plan to use multiple regression analysis, you might choose to measure popularity through a "feeling thermometer" that gauges warm and cold feelings on a scale of 0 to 100, or through a set of familiar 10-point scales. The use of these quasi-continuous variables may produce greater variance in the dependent variable and distributions that better conform to the assumptions of regression and thus perform better in the statistical analysis. But if you are going to rely largely on cross-tabular analysis and are writing for a nonspecialist audience, a measure that uses commonly understood words and modifiers such as "strongly committed" or "somewhat committed" would be preferable. Nonspecialists will find it easier to understand that "60% of Muslims compared with only 35% of Christians are strongly committed" than "the position's mean thermometer rating among Muslims is 60 while it is only 35 among Christians."

OPEN AND CLOSED QUESTIONS

One of the most fundamental choices in drafting survey questions is whether to ask respondents to choose among alternative answers or to simply ask the question and let them respond in their own words. Most survey questions, especially for nonspecialist populations, are *close ended,* meaning that respondents are provided with a set of answer categories. For example:

Here are four problems facing our nation today. Please tell me which one is the most serious threat to our nation [READ RESPONSE OPTIONS]:

(1) terrorism
(2) poverty
(3) racial conflict
(4) AIDS and other diseases
(9) Don't know/No answer [DO NOT READ]

The alternative to a close-ended question is an *open-ended* question, which does not provide alternatives to the respondent:

What is the most serious problem facing our nation today?

(1) Answer given [RECORD VERBATIM]
(9) Don't know/No answer [DO NOT READ]

Both types of questions can provide valuable information, but there are distinct advantages and disadvantages to each. In exploratory research, we may not know a great deal about how respondents think about an issue or problem, and there is a danger that our close-ended questions may not include the proper response options or that we may impose a framework on them that is inappropriate to the way they perceive the issue. In this case, an open-ended question allows us to hear respondents talk about a problem in their own words, to see the range of ideas they bring to the question, and to better understand what they believe our question is asking. For the "most important problem" question, an open-ended question might show that something other than the four problems we list (the economy, for example) is actually more commonly mentioned. One way to put this advantage is to say that sometimes open-ended questions can have better validity than close-ended questions.

But there are disadvantages as well. As is often the case in research, greater validity of one sort often comes at the expense of reliability. Respondents vary greatly in their willingness and ability to express themselves verbally, which means that one respondent will give much more complete and insightful answers than another one—even when their underlying opinions may be the same. Because we would then code the two respondents differently (even though their basic opinion was the same), our measure would be unreliable.

Open-ended questions also impose a considerable burden on respondents. They require substantial cognitive effort, and too many of them may lead some respondents to refuse to continue the interview. They are also a burden for the researchers because effective use of the responses requires a lot of work to categorize and code them for analysis. The coding process itself is susceptible to inconsistencies. A large survey may require more than one person to code, and there are often differences between coders in how certain responses are categorized. This problem is another source of unreliability in the measurement process.

Yet open-ended questions are valuable if used strategically. For pilot or exploratory studies with relatively small samples, where the goal is to learn enough about the population to design a larger, more rigorous survey questionnaire, open-ended questions can be very helpful in gauging how respondents perceive an issue or problem. For the main data collection, open-ended questions as a follow-up to close-ended questions (e.g., "Why do you feel this way?") and for certain kinds of topics—the "most important problem" item, for example—can still be useful and can give the survey additional credibility with an audience because respondents were not "steered" toward a particular subset of issues. An important principle to keep in mind is that open-ended questions should be placed early in the questionnaire, so that later content does not suggest answers to respondents or remind them of issues that might not be especially salient to them. If the "most important problem" question follows a

lengthy series of items about crime, you can bet that crime will be mentioned far more often than if it had been placed earlier in the survey. That said, surveys should *not* begin with open-ended questions because respondents find them somewhat intimidating. It is better to "warm up" the respondent with some relatively easy items before asking them to discuss an issue or problem in their own words.

The open-end format is also appropriate when the information sought is concrete and likely to be readily retrievable by the respondent. Examples include the following: "What was your age as of your last birthday?" "In what country were you born?" or "How many times in the past year did you consult with a mediator?" In each of these instances, you can record the responses and then code the answers into categories at a later date. Capturing the exact information gives you the option to categorize it in many different ways—including leaving it uncategorized—according to the purposes of your analysis. One caveat about this is that some kinds of concrete information may be too difficult to remember with the specificity necessary to give a single answer. For example, if we ask, "How many times in the past year did you and your spouse discuss money and family finances?" we may find that few people are able to provide a specific response. We would have better luck by determining whether they *ever* discuss money and finances and how often they typically do it within a shorter time frame, such as 1 month.

We often want to have the benefits of both the closed- and open-ended formats at the same time. Questions can be written that provide response options but also allow for the capture of responses from people who volunteer an alternative to the offered choices. Sometimes we use the open-ended format but *precode* a set of expected responses to make it easier for the interviewer to capture the answers and keep the interview flowing. For example, a list of important problems can be provided to the interviewer but not read to the respondents; if a respondent mentions the economy, the interviewer simply checks that category and moves on. If the respondent does not mention any of the problems on the pre-coded list, the interviewer records the answer verbatim.

As this discussion suggests, the best close-ended questions are ones that offer response options that accurately reflect the way respondents actually think about the issue or topic at hand. At a minimum, the options offered must be *exhaustive* and *mutually exclusive*. That is, the list provided should cover all of the possible answers, or at least include a provision for accepting "other" responses that are not included in the alternatives. (The alternatives are exhaustive of the possibilities.) And it should be clear to the respondent and to the interviewer that a logical answer can fit into only one of the possible alternatives offered. (The alternatives are mutually exclusive.)

The list of alternatives in a close-ended question should be relatively short—ideally, three or four—unless the list is likely to be familiar to the respondents.

Even with a short list, there is evidence that items appearing first or last in the list of response options may be favored because they are easier for respondents to remember (see McClendon, 1991). For this reason, it is often useful to rotate the order in which response options are presented to respondents.

ATTITUDES AND NON-ATTITUDES

In the 1980s, the Gallup Organization, one of the world's preeminent survey organizations, conducted a series of international surveys tapping public opinion on world political issues. The questionnaire for Germany had been finalized and transmitted to the German field house that would collect the data. When the survey was conducted and the data transmitted back to the United States, the Gallup president, Andrew Kohut, immediately saw that there were no "no opinion" responses recorded. He contacted the German field house and asked how they had obtained results with no missing data. "The questionnaire you sent us did not include a 'no opinion' category. So we made them answer," was the response. Normally, Kohut would personally conduct a final review of all questionnaires, but was unable to take a last look at this one, and assumed that the final version of the questionnaire would be reviewed by a project manager. Unfortunately, that review did not occur. No one caught the fact that the "no opinion" category was not explicitly included on the questionnaire. The German survey organization implemented the instrument exactly as written.

This affair illustrates two important points about survey research. One is that details matter, and attention to detail is critical in the success of such projects. Review by multiple participants can help ensure that errors do not slip by. The second point is that people will answer questions if asked to do so, even if they don't really have opinions on the subject at hand. We can think about the stereotype of the obedient German and laugh at the notion that *every* respondent would answer every question when prodded to do so, but the vast majority of respondents in any culture will provide a response if the interviewer insists. Thus, it is important to make sure that respondents are capable of answering the questions we write and that our interviewing procedures do not put undue pressure on people to offer an opinion if they have none (Smith, 1984).

The more practical question is whether to explicitly offer a "no opinion"—a "don't know (DK)"—response in the question itself or to simply leave it implied that a non-response is an acceptable alternative. Not surprisingly, extensive experimentation indicates that more people will choose "no opinion" when it is offered than when it is not—anywhere from an eighth to a third, depending on the topic (Converse & Presser, 1986). DK responses remove otherwise useful respondents from your multivariate analyses and reduce your sample size. This is because these responses are usually coded as missing data in such statistical procedures as regression and factor analysis. Studies show that

the overall distribution of opinion (e.g., the ratio of approval versus disapproval) remains the same even when DK is not offered, and thus the substantive "bottom line" is not harmed. On the other side is the indisputable fact that many of the people who answer in the absence of an explicit DK are simply inventing a response on the spot. This response may not be entirely invalid because it appears that people use their values and whatever knowledge is available at the time to figure out a position on the question—and thus often come up with a response not unlike that of their better-informed or more opinionated brethren. But encouraging answers from those who have thought less about the issue undoubtedly introduces more random error into the data, depressing correlations and making it more difficult to detect causal relationships.

Another related question is whether to provide an explicit middle-of-the-road option to respondents when measuring opinion that falls along a continuum. Consider, for example, a survey of residents in a nation undergoing post-war reconstruction. Many decisions about a new political system may need to be made, and surveyors could ask respondents whether they approve or disapprove of each proposal. Should they be offered the option to say they "neither approve nor disapprove" of each one? Assuming that the topics being asked about are not entirely outside of the experience or knowledge of the population being surveyed, the best approach would be to omit the middle option and capture the extent of certainty or uncertainty through a measure of intensity. The "DK" response is available for respondents who simply cannot or do not want to offer an opinion. The measure of intensity can be built into the question or asked as a follow-up to the initial item.

Do you approve or disapprove of the provisional authority's proposal to shift ownership of the country's electricity generating plants to the private sector over the next 12 months?

If approve: Do you completely approve or mostly approve?
If disapprove: Do you completely disapprove or mostly disapprove?

(1) Completely approve
(2) Mostly approve
(3) Mostly disapprove
(4) Completely disapprove
(9) Don't know/refused

THE ACQUIESCENCE PROBLEM

Although the approve/disapprove format is useful when measuring judgments about leaders or specific proposals or options for future action, we must be on guard against a potential bias that results from a tendency of less educated or less informed respondents to agree with assertions or proposals

offered to them in a survey. Experiments have revealed that the less educated—compared with the better educated—are more apt to acquiesce to the suggestions of the researchers, to agree rather than disagree to a statement presented in an agree/disagree format (see Schuman & Presser, 1996). Such a bias can result in drawing incorrect inferences; for example, one could mistakenly believe that certain proposals are more popular among poorer people than among wealthier ones. And by inflating support or agreement among a portion of the sample, the acquiescence phenomenon may inflate overall levels of apparent support. Although we do not find acquiescence bias in every instance, it is a common enough problem that we should use the forced-choice format when respondents are being asked to consider alternative proposals or values.

Acquiescence Bias

Agree/Disagree Format (October 1999)

The best way to ensure peace is through military strength (54% agree, 42% disagree)

Forced-Choice Format (July–September 1999)

The best way to ensure peace is through military strength (33%)

OR

Diplomacy is the best way to ensure peace (53%)

Source: U.S. surveys by Pew Research Center.

To illustrate what difference this format can make, and to highlight the difficult nature of trade-offs faced in survey research, consider the example of the Pew Research Center's longstanding series of questions tracking American values. These agree/disagree items have been asked of large national samples since 1987. In 1994, the center decided to add forced-choice versions of many of these core questions to its inventory of standard questions for use in its voter typologies. In the example shown in the box, far more people agreed with the assertion that "the best way to peace is through military strength" than were willing to choose this option when the alternative of diplomacy was available to them. (See also Druckman, 1970, for a discussion of the bias on reversed attitude scaled items.)

Because of the value of the time series using the agree/disagree format, the center has continued to use them. Even though they may not provide the most valid measure of the precise level of opinion, they remain valuable for tracing *changes* in opinion over time.

SOCIAL DESIRABILITY

The human tendency to present oneself in the best light (even if only to oneself!) constitutes a challenge to accurate survey measurement. Some of the content of contemporary surveys touches on sensitive topics about which people may not be motivated to reveal the full truth about their actions. When surveys have evoked inaccurate responses to questions about sensitive matters, we say they have been affected by *social desirability bias* (Druckman, 1970; Groves, 1989). Social desirability can lead respondents to understate behaviors they may be ashamed of or to overstate behaviors that are considered praiseworthy. Respondents may fail to report interpersonal conflict in the family or violent behavior on their part. They are likely to underestimate the amount of alcohol they consume and may falsely deny using illegal drugs. There is evidence that people overstate the extent of their charitable giving, churchgoing, and even the regularity with which they wash their hands after using the toilet. These biases do not affect all individuals in the same way. Men may overstate the number of sexual partners they have had, whereas women may understate the number. (For a provocative discussion of this issue, see Lewontin, 1995.)

Respondents are more apt to tell the truth if they believe that their responses will be kept confidential and that they will not be judged by the interviewer. This means that it is important to establish a high level of trust with respondents. Sponsorship of the survey by an organization known to and trusted by the respondent is helpful. Professional, serious, empathetic interviewers are critical. Highly sensitive questions should placed later in the interview, after the respondent has had a chance to become comfortable with the interview and to judge the seriousness of the research by reflecting on the other content of the questionnaire. It is also helpful to remind the respondent of the importance of thorough and complete answers.

Unfortunately, no magic tricks exist to solve the problem of social desirability bias, but a number of techniques have been employed and found useful in certain circumstances. When asking about socially desirable behaviors such as voting or churchgoing, use a preface that "normalizes" behavior that does not conform to the ideal. For example, a typical voter registration question in the United States might be prefaced by the following: "These days, many people are so busy they can't find time to register to vote, or move around so often they don't get a chance to reregister. . . Are you *now* registered to vote in your precinct or election district, or haven't you been able to register so far?"

A similar technique first asks if the respondent has "ever" engaged in a behavior (such as volunteering to help others in the community), then asks about activity within the past 12 months (or other shorter time frame). The "ever" question allows people to provide a socially desirable answer, reducing the pressure to overstate how recently they engaged in the behavior.

When asking for an estimate of the frequency with which an individual engages in socially undesirable behaviors (such as binge drinking), offer categories of response that include relatively high levels of the behavior. Respondents take their cues about what is normal from the alternatives offered. Consider the question, "On how many days in the past month did you have five or more drinks of alcohol?" This could be asked in open-ended format, allowing the respondent to provide a number. But given the tendency for some people to underestimate this behavior, it is better to offer a set of categories. Yet if the categories include a relatively low range (e.g., 0, 1–2, 3–4, 5 or more), people may gravitate to the lower end of the range. A better range might be 0, 1–2, 3–5, 6–9, 10 or more.

For certain kinds of highly sensitive questions, special methods can be employed. For questions about drug use or sexual behavior, researchers have had success in gaining better answers by allowing respondents to take part of the survey by sitting at a computer keyboard while listening to a recorded interviewer ask the questions through headphones. This method is known as audio-CASI (computer-assisted self-interviewing; see Tourangeau et al., 2001). Another method is the *list experiment*, in which respondents are randomly divided into two groups, each of which is read a short list of behaviors and asked how many—not which ones—they have engaged in. One group is given a list of four items (including common and non-sensitive items such as shopping today for an article of clothing), whereas the other gets the identical list plus an additional item—the item of interest to the researchers (e.g., struck a spouse or loved one in anger). The difference in the mean number of items is the proportion of respondents who engage in the behavior. Although the technique can yield a valid estimate of the incidence of particularly sensitive behaviors for the overall sample and for subgroups (e.g., men versus women), it cannot identify individuals who engage in the behavior.

ASSEMBLING QUESTIONNAIRES

Let us assume that you have constructed excellent questions for each of the concepts of interest in your survey. You now face the task of putting them together into a coherent and attractive questionnaire for administration. In many respects, this aspect of the measurement process is just as important as the design of the questions themselves. Reasonably good survey questions of concepts on which respondents have opinions or readily retrievable experiences can work well, even if flawed. But how questions are embedded within the context of other questions can have a substantial impact on results–even greater than changes in the wording or format of individual questions themselves. And how questionnaires look and sound to respondents can have a big impact on their willingness to undertake the survey or to finish it.

A detailed discussion of questionnaire composition and layout is beyond our scope here, but fortunately a fine resource exists in Dillman (2000). Most of the principles discussed are applicable to personal and telephone interviews as well.

A few key issues in questionnaire construction are worth reviewing. First, pay attention to the first question in the survey. This is particularly important in telephone interviews, where the first question arrives before a good rapport with the respondent can be established. Even in a self-administered survey where the respondents can "see beyond" the first question, we should be concerned about its impact on their willingness to continue with the interview. In general, the first question should be something that everyone in the sample can answer easily, but it should not be so dull or mundane as to suggest that the survey is going to be boring. Many political surveys begin with a question about presidential approval, or a judgment about the state of the nation. Others may begin with a question asking respondents to rate their community or state as a place to live. Never begin a survey with a sensitive demographic question such as family income—and do not begin with innocuous demographics either, because they are boring. If you need to ask demographics at the beginning of the survey in order to find respondents who meet a specific criterion for the sample, consider offering one or two substantive questions first, even if the responses to these are simply going to be "thrown away."

Demographic questions should be placed at the end of the survey unless they are needed to route respondents to certain questions (e.g., Catholics to questions about their opinion of how the church has handled the sexual abuse scandal among priests; mediators to questions about the way they implement their third-party role in a particular conflict). Some respondents resent being asked about age, race, and especially income, and many people find the full range of demographic questions to be somewhat intrusive.

Perhaps the most difficult choices in questionnaire construction arise over the sequencing of the substantive questions. Despite our best efforts to write clear questions, the context in which they are asked is often very important in how they are interpreted by respondents. The dilemma of choosing an appropriate context is a classic issue of validity. Is it better to measure approval of the mediator at the beginning of the survey, before asking about a variety of issues that bear upon judgments of his or her performance? Or is it better to ask a series of questions about different issues to get the respondent immersed into the political soup and then to ask about the job a particular mediator is doing? Arguably, the latter approach would yield a more thoughtful, considered response, akin to what we might achieve after a discussion with a friend over dinner or in a bar. We could even say that this is a more valid measure of opinion. But a question about a mediator's performance asked at the beginning of the survey will certainly yield a more reliable measure over time. Unless we

precede the question with the exact same series of questions in successive surveys—questions about approach, about empathy, about advice given, about personal qualities, and so forth—we cannot be confident that the context was the same from one survey to the next.

Sampling

Sampling is a critical aspect of survey research, but the subject is relevant to nearly all scientific inquiry and not just to surveys. The principles that underlie survey sampling are the same as those at work in fields such as agriculture, business, accounting, biology, and the law. A solid understanding of sampling can be useful in all sorts of human endeavors, including CA&R.

Our lives entail the taking of one sample after another. The people we meet, the food we eat, the places in which we live and visit, and the ideas we encounter represent small samples of all of the people, food, places, and ideas that could be experienced. At a gut level, we understand this, and also realize that our samples of reality are *biased* in important ways. The people we know are a lot more like us socially, financially, and culturally, than a true cross section of the world's population. Similarly, the ideas we encounter are a combination of things available to us in the physical, social, educational, and media environment. From these settings and sources, we exert some control over what we see and hear, but we recognize the limitations on the breadth of what is available to us.

Similarly, survey research typically gathers data from a sample of a population of interest. We use samples for many reasons: We do not have time to gather data from everyone; we do not have the money to pay for data collection from the entire population; we do not want to ask everyone in the population to take the time to give us information, either out of respect for their limited resources of time or, sometimes, because we do not want the entire population exposed to our work.

The critical question for sampling is how well the sample we obtain will resemble the population in which we are interested. The sample is a *model of the population.* As with any model, it should be as similar to the population as is possible. Achieving this can be very difficult and costly, and we rarely achieve perfection. Fortunately, the laws of probability are our good friend in this endeavor.

If we set out consciously to gather a sample of Americans that represents the general public in every possible respect, we might assemble a list of all of the attributes we want to ensure are properly reflected in the sample. This list could include state of residence, type of housing, race and ethnicity, sex, and age. Or it might include parental and marital status, experience with drugs and

alcohol, automobile ownership, preference in pets, psychological dispositions to anger and violence, and physical fitness. We could continue to list characteristics, including anything of specific relevance to our research project, and would eventually collapse in exhaustion. It seems apparent that a conscious effort to make a sample representative is bound to fail, simply because we cannot anticipate every attribute that may be relevant. It is also bound to fail because we have no way to characterize people according to these attributes until we have actually interviewed them, so these cannot be used as a basis of selecting people in the first place. Even if we did know the attributes, we would still have to have a way to select a few people from among the many who shared attributes we wanted represented. This process might inevitably introduce biases, in which we picked people who were the easiest to find, or most willing to be interviewed, or most attractive.

The best solution is a selection method that is essentially blind to people's attributes, that ignores such characteristics in favor of a neutral selection process that allows each person in a population to have an equal—or more properly, a *known*—chance to be selected, *regardless of their attributes.* This general approach is known as *probability sampling,* also often called *random sampling.* For some types of research probability, sampling may simply be impossible or too costly to conduct. In such cases, we may have no alternative to *non-probability sampling.* Although non-probability samples are much less desirable, when properly designed they can yield useful information and are widely used in certain situations—especially those relevant to conflict. Examples include respondents in war zones, dialogue workshop participants, and peacekeepers returning from missions.

PROBABILITY SAMPLING

The great virtue of probability sampling is that it permits us to compute an estimate of the accuracy of the sample. Without probability sampling, there is no way to gauge the degree to which the sample actually matches the population. But with it, we can state the likelihood that our results are within a specific margin of the true value for the population. Such estimates of precision are critical for decision making under conditions of uncertainty where the cost of action (or inaction) may be very high. For example, manufacturers often sample products coming off an assembly line to assess the quality of the manufacturing process. If occasional defects are discovered, it is important to be able to estimate exactly how common they are because stopping production and addressing the problem may be extremely costly. Probability sampling of items on the line can allow the decision makers to make a precise estimate of how common the problem is within a given range. In other words, they can estimate how many total units will be defective out of the total number

produced. With this information, a decision can made as to whether stopping the line and fixing the problem is worth the cost.

Probability sampling entails several steps.

The first step is to decide how precise the estimates need to be—that is, how much uncertainty about the findings can be tolerated. This decision affects the size of the sample needed, and can also affect decisions about how the sample will actually be drawn.

The second step is to specify exactly what population you wish to represent. This sounds obvious, but in practice it can be difficult. If you want to represent "the general public of the United States," does this include people who do not speak English? (If so, you will need to make provisions to have your questionnaire translated into many languages and to hire interviewers who are fluent in those languages.) Does it include people without telephones? (If so, you cannot collect all of your data with a telephone survey.)

The third step is to assemble a list of the population, also called a *list frame* or just *sampling frame*. Sometimes a list is simply unavailable, and often a list will have serious problems. Take one commonly used list: the telephone directory. People in the population will be missing from the list (e.g., people with unlisted numbers do not appear in the telephone directory), others will appear more than once (e.g., people with more than one telephone number may appear in the directory more than once), and still others appear in the list but are no longer in the population (e.g., people who moved away will still be in the book until an updated directory is printed). For professional populations, such as the Society for Professionals in Dispute Resolution (SPIDR), printed directories or lists of members may be relatively complete and up-to-date, providing an excellent list frame. This list was used by Birkhoff (2001) to study mediators' perspectives on power. She mailed a survey to each of the 500 SPIDR members categorized as trained mediators; because these members defined the population of professional dispute resolution mediators, a sampling frame was not used. Although only 184 mediators returned the survey (about 37%), this "sample" was shown to represent the population on various demographic variables. In general, the completeness of list frames varies considerably from population to population.

The final step is to actually draw the sample and attempt to contact the sampled individuals. There are many methods for doing this, which are discussed below.

How Large Should the Sample Be?

Because one of the great benefits of sampling relative to taking a census is the savings in time and money that a sample provides, deciding how big the sample should be has implications for both the accuracy of the study and the costs—with accuracy going up along with the costs. In general, a sample should be large enough to provide an acceptable degree of precision not only

for estimates about the population as a whole, but also for important subgroups that are represented in the sample (for example, racial minorities or the elderly). The precision of a sample is affected by three factors, which for most situations can be arrayed in order of importance, as follows:

- The size of the sample (very important)
- The variability of the population (important)
- The fraction of the population taken into the sample (usually not important)

Simply put, the precision of a probability sample increases as the size of the sample increases. So bigger is better, but as we shall see, there is a law of diminishing returns that makes it relatively straightforward to choose an optimal sample size because additional precision beyond a certain point can be very costly. But size is not the only factor; populations that have greater variation in the variables of interest to the study yield less precise estimates with a given sample size than do populations that are more homogeneous. And although it is usually of no consequence except in cases of very small populations, the proportion of the population actually included in the sample can affect the precision of the results. As logic would suggest, taking a bigger percentage of the population into the sample improves the precision, but—contrary to what common sense would tell us—for relatively large populations (say, several thousand and up), the sampling fraction has almost no impact on precision for samples of the size we are likely to be dealing with.

To explore these practical matters in greater detail, we need to take a side journey into the theory underlying probability sampling. Even though we are typically drawing only one sample from the population for our study, the easiest way to grasp the basics of sampling is to think about *many* samples drawn from the same population and not just one sample. The big question is how similar or different successive samples are likely to be. Ideally, we would like to know that repeated samples of the same size from the same population will yield results that are nearly identical every time, and that this result is close to what we would get if took a census and not just a sample. If so, it means that any given sample is likely to be a good model of the population. If samples vary a lot—if they are very different from each other even though we are using the same techniques in drawing them—then we will have less confidence that any given, single, sample we draw is going to be accurate.

Fortunately, we have control over how much samples are going to vary. The biggest factor under our control is the size of the sample. Consider the simple example of coin flipping. Coins have two sides: heads and tails; thus, the population value for heads or tails is .50. Our experience with flipping a coin tells us that occasionally we will get several heads in a row. But over a longer series of flips, the number of heads and tails will even out. We would not be

surprised to get 7 heads (70%) in a sample of 10 flips, but how likely is it that we would get 70 heads in 100 flips? Unlikely. Or 700 in 1,000 flips? Very, very unlikely. Indeed, we can specify *exactly* how unlikely for each of these outcomes because the results of random samples are themselves distributed according to the normal curve, with a mean and a standard deviation just like any statistic. This distribution has a special name: *the sampling distribution,* which is the hypothetical distribution of the results of all of the possible samples of a given size drawn from the population. According to sampling theory, the mean for any given statistic plotted from all of the possible samples we could draw from a population will equal the population mean (in the same way that lots of coin flips eventually establish the fact that the coin has two faces that are equally likely to turn up), and the variability of this statistic is affected by—you guessed it—three things: the size of the samples we have been drawing, the variability of the variable we are estimating with the samples, and the sampling fraction. (See Chapter 4 for a discussion of the sampling distribution and the coin flipping example in the context of experimentation.)

A more thorough discussion of how to compute the standard error is beyond the scope of this chapter, but here are some useful tools that both illustrate the principles described above and provide a basis for making decisions about sample size. First, Figure 5.1 plots the relationship between sample size and sampling error of a proportion for samples ranging from 10 cases all the way up to 4,000 cases. This graph "cheats" by changing the scale along the *x*-axis (twice), but this manipulation helps to show visually why most survey samples tend to fall within a narrow range of sizes.

Take a moment to orient yourself with Figure 5.1. The margin of error is plotted on the *y*-axis, and the sample size is plotted on the *x*-axis. A given point on the line tells us the expected error margin for a sample of a given size; for example, the expected error for a sample of 100 cases is plus or minus 10 points. As you can see, the amount of error drops rapidly as samples increase through the range of 10 to about 50, then the curve flattens out. When the *x*-axis scale changes from tens to hundreds beginning at 100, another dip occurs, with expected error declining from 10 points down to about 7 points. So adding 100 cases to your sample of 100 improves the precision of the sample by 3 points, about the same as adding 10 cases to a sample of 20 or 30. After the drop from 100 to 200 cases, the curve again flattens out, with each additional increment of 100 cases improving the accuracy of the sample, but by smaller and smaller amounts. Between 1,000 and 2,000 cases, we see another decline in the margin of error (slightly less than 1 point), but after that, even adding an extra 1,000 cases makes only a marginal difference.

The practical implication of the relationship depicted here is that samples need not be much larger than about 1,000 to provide a reasonably precise estimate of the population. Adding more cases can be very expensive

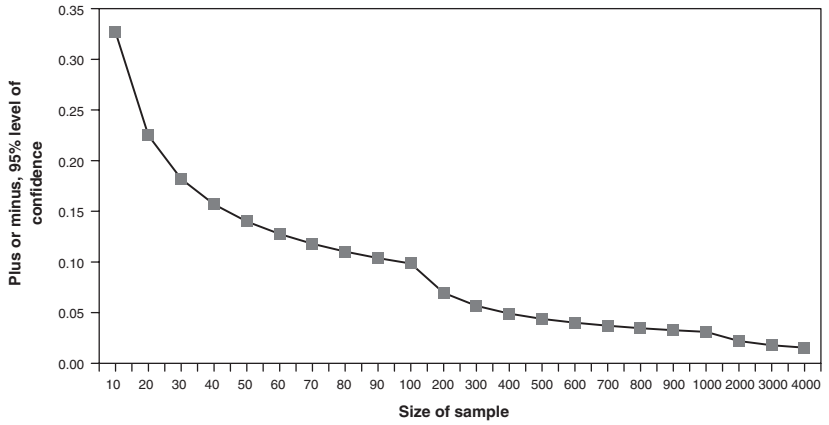

Figure 5.1 Sampling Size and Sampling Error

but does not yield much in the way of additional precision. At the other end of the chart, it is clear that additional cases with small samples can make a lot of difference in the sampling error. One caveat—mentioned earlier—is that we are ordinarily not interested only in the overall sample estimates for the population as a whole. We are often interested in subgroups in the population as well, such as foreign-born residents. If such a group represents 10% of the population, we can anticipate obtaining interviews with 100 of them (assuming they are just as accessible and amenable to being interviewed as native residents). Thus, although the overall margin of error for the sample of 1,000 will be plus or minus approximately 3 percentage points, the margin for the foreign-born residents will be plus or minus 10 percentage points. If we need a more precise estimate for this group, we will have to obtain a larger sample of the overall population, or we will have to employ some other method to increase the number of foreign-born residents in the survey (for example, by drawing a larger portion of our sample in cities than in small towns and rural areas).

Figure 5.2 shows how the amount of variation in the characteristic we are trying to measure can affect the amount of sampling error. The top line shows the sampling error for a survey question in which the responses divide evenly: 50% give one answer, 50% give another, for example, whether the respondent is male or female. This is the highest level of variation that a proportion can achieve. The bottom line on the graph shows the sampling error for a question in which 5% of respondents give one answer, and 95% give a different answer, for example, that there is or is not an extremist living in the household. This variable is almost a constant. Nearly everyone gives the same answer; only a few

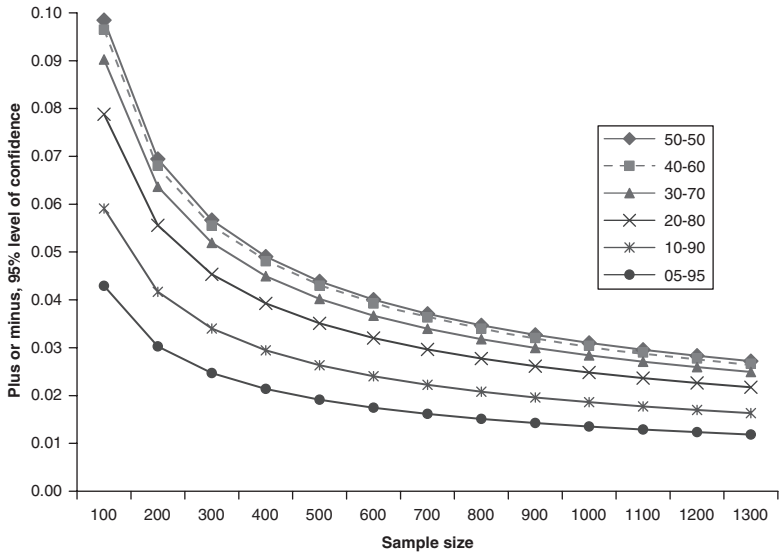

Figure 5.2 Effect of the Variability of the Variable of Interest on Sampling Error

people report living with someone prone to extreme violence. As Figure 5.2 shows, the sampling error is much smaller for a question with little variability than for one that varies a lot.

Figure 5.3 illustrates how little difference the sampling fraction makes. Only when the sample is a significant portion of the population, for example, 50%, can we see a clear improvement in the sampling error. As the chart shows, a sample of 500 taken from a population of 1,000 will have a sampling error of plus or minus approximately 3 percentage points. The same size sample from a population of 5,000—a sampling fraction of 1/10—has a sampling error of plus or minus 4.2 percentage points.

Types of Probability Samples

There are many different ways to create a probability sample, and the choice of approach depends on several factors: whether a good list frame is available, the mode of the data collection (telephone, mail, Internet, personal interviews), whether or not additional interviews with certain kinds of people are needed, and so forth.

The somewhat misleadingly named *simple random sampling (SRS)* requires a list frame, wherein the elements are numbered and then selected

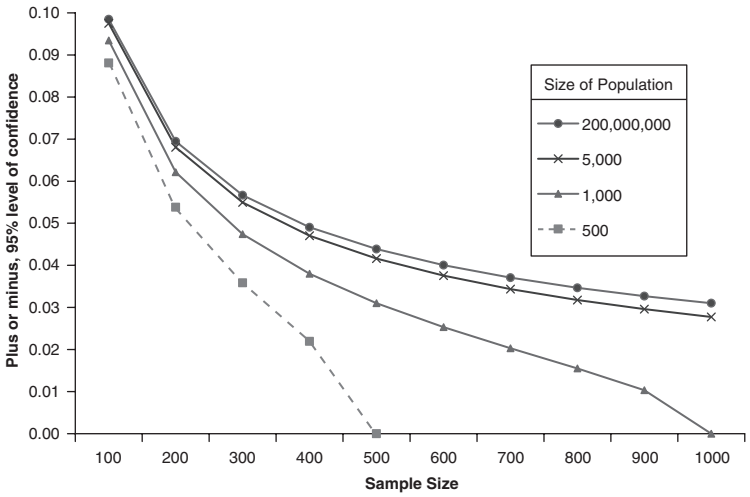

Figure 5.3 Sample Size and the Sampling Fraction

randomly using a computer process or a random-number table. This method is akin to selecting ping pong balls with lottery numbers, or drawing names or numbers out of a hat, though neither of these common approaches may be truly random. In practice, SRS is often complicated to implement because it requires a list and can be cumbersome if the list is large. A more common method is *systematic sampling,* which entails the selection of every *n*th element of a list. For example, if we need a sample of 100 college students taken from a telephone directory containing 10,000 names, we could simply choose a random number between 1 and 100, count from the beginning of the directory to the student whose name fell at that point in the list, and then select every 100th name in the directory after that. Systematic sampling is also used in situations where no list is available, such as in intercept surveys of voters leaving polling places (also known as *exit polls*). Researchers will estimate the expected number of voters at a precinct during the day (e.g., 200) and then divide that figure by the number of interviews they need at the location (e.g., 10) to determine the *sampling interval* needed (200/10 = 20). The interviewer will then attempt to conduct an interview with every 20th voter.

Often our list frame will contain information about the people in the population, information that can be used to improve the accuracy of the sample or to help us target certain groups of special interest. This information makes it possible to sort people into groups or strata, from which random samples can be drawn. This is called *stratified random sampling.* For example,

a list of college students may include class standing. With this information, we could ensure that the sample we draw matches the population on this characteristic: the correct percentage of freshmen, sophomores, and so on. Or if we were particularly interested in freshmen and wanted to make sure we had enough interviews with them to provide good estimates of their views and experiences (i.e., a low margin of error), we could oversample them, taking a higher percentage of freshmen into our sample than would occur by chance.

Lest stratified sampling sound like the hypothetical purposive sampling we described (and rejected) at the beginning of this section, they are not the same thing. Sorting the population into strata and then sampling randomly from within the strata cannot damage the representativeness of the sample; in fact, it can actually improve the statistical accuracy of the sample because it removes one source of variation between the population and the sample. The sample is now guaranteed to match the characteristic on which we stratify. Nor does over-sampling necessarily damage the sample. However, it is necessary to statistically adjust the results when oversampling has occurred.

Recall that the strict definition of a probability sample is one in which every object in the population has a *known* chance of selection, not necessarily an *equal* chance. Consider a study of racial conflict and cooperation in an American university. We hypothesize that freshmen are apt to experience higher levels of conflict but also more experiences with cooperation because they are more likely to be placed in living situations with a diverse population. Let us say that we decide to take twice as many freshmen into our sample as there should be, based on their actual percentage in the student body. (For example, if freshmen are 25% of the student body, we might decide to make our sample 50% freshmen because we want to make sure we have enough interviews with them to do detailed analysis.) When we use our survey findings to report on the opinions and experiences of the entire student population, our results will be biased in favor of the freshman experience because there are twice as many freshmen in the sample as is appropriate. For example, we might estimate rates of racial conflict that are much too high for the overall student population. But we can use *weighting* to adjust the results so that freshmen interviews account for only 25% of the total—in other words, count each freshman interview as half of an interview. Computer programs for survey data analysis have the capability to make these adjustments; all we need to do is tell the computer what weight to assign to cases for a given type of respondent. The bottom line is that stratified sampling—even with oversampling—is still probability sampling.

Most of the examples of samples discussed thus far entail a single stage: We have a list and choose people from the list. But many samples entail *multistage sampling*. Multistage samples are commonly used when no list frame exists or it would be impractical to construct one for the whole population.

Objects in the population may belong to categories or groups that can be sampled at the first stage. At the second or a subsequent stage, it may be more feasible to construct a list or to use some type of systematic sampling. A couple of examples will illustrate the concept.

Exit polls of voters are multistage samples. At the first stage, a sample of precincts from all of the available precincts in a jurisdiction is drawn. Because some precincts are larger than others, it may be necessary to give some precincts a better chance of being selected than others. If ping pong balls were labeled with precinct names and used to conduct the random selection, we could give bigger precincts a higher chance of selection by including more ping pong balls with their names. We can also stratify precincts according to past voting history to make sure that a politically diverse and representative mix of precincts is included in the final sample.

When we have a sample of precincts, the second-stage sampling occurs when interviewers go to the precincts on election day and conduct a systematic sampling of voters. The resulting sample meets the fundamental test of probability sampling: Every object (every voter) in the population had a chance to be included in the survey (because any precinct *could* have been selected at the first stage, and any voter in selected precincts could have been intercepted by an interviewer); moreover, we can compute the probability that any given voter would be selected if we know how many voters live in each of the precincts in the population and what fraction of voters we sampled within each of the selected precincts.

Another kind of multistage sample can be very expensive to implement but provides the best opportunity to achieve a truly representative sample of an entire population. *Area probability sampling* is employed in face-to-face surveys for which no list frame of the population is available and there is a desire to reach people regardless of whether they have a telephone or not. At the first stage, geographic areas (such as counties) are enumerated and sampled, usually taking into account the size of the population living in them (as with exit polling, jurisdictions with larger populations may have a larger chance of being selected). At a second stage, the selected counties might be subdivided further, for example, by city blocks, and a sample of these divisions taken. At a third stage, all of the housing units in the selected divisions will be listed and a sample of housing units drawn. At a fourth stage, the residents of each sampled housing unit will be enumerated, and a sample of one adult in each household will be selected for an interview. Area probability sampling is employed to obtain probability samples in most developing nations, because many, if not most, households lack telephone service. Although travel costs can be high, low labor costs make it more feasible to use personal interviewers. The presence of an interviewer is useful because literacy levels are often low.

Virtually all multistage sampling involves *cluster sampling* as well, in which more than one respondent is interviewed within the second or subsequent

stages of the sample. This is usually a practical necessity, as it is much more efficient to obtain multiple interviews at a single voting precinct than it would be to send an interviewer to a precinct and interview only one voter. Travel costs may be very great for area probability samples, so many households within a sampled community may be chosen for interviews.

Clustering in multistage samples does have a slight negative impact on the precision of the samples, relative to a non-clustered sample with the same number of interviews. This occurs because objects within a cluster (e.g., children within a selected classroom, voters in a selected precinct) are likely to be more similar to each other than would be a set of randomly selected children or voters from all of the possible classrooms or voting precincts in the population. Put another way, cluster samples will be slightly more internally homogeneous than non-clustered samples. But cluster samples are still probability samples because we are able to specify the probability of selection for any individual in the population.

Practical Sampling

Often a list for a population of interest will be available, though sometimes it may require a good bit of detective work to locate and obtain. Privacy concerns usually make it difficult to obtain lists of the general public, such as college student directories, driver's license records, or auto registration lists. But lists of professionals and other elite populations are often available from state or local governments or from nonprofit organizations. Most states have regulatory agencies that license or certify professionals (such as nurses, lawyers, EMTs, mediators, and the like), and these lists may be relatively up-to-date and available for purchase. States or localities may also have lists of businesses or employers, and "yellow page" directories can also provide a reasonably complete list of certain kinds of businesses. Professional and trade associations may have mailing lists of members for sale, and voter registration lists are available for purchase in most states.

Several companies exist for the purpose of creating and selling survey samples. These companies design *random digit dialing* samples for telephone interviewing, but they also may have specialized samples of businesses, professionals, individuals with particular consumer or lifestyle interests, and the like. Many specialized samples do not qualify as probability samples, but some do. Two respected companies in the sampling business are Survey Sampling International of Fairfield, Connecticut (http://www.surveysampling.com), and Genesys Sampling Systems of Fort Washington, Pennsylvania (http://www.m-s-g.com/genesys/genesyshme.htm).

Sometimes, however, no commercial sample or obvious source for a list frame exists, and a certain amount of ingenuity is needed to construct one or

come up with an acceptable alternative approach to creating a probability sample. Multistage sampling is often an option but can introduce a good deal of complexity into the statistical analysis. Most university departments of mathematics or statistics have specialists who can help with this type of problem, but the planning necessary to create a multistage sample is best done by someone who is familiar with the population of interest and the topic at hand—typically that would be you, the researcher!

Consider the prospect of interviewing homeless people about their experiences with violence on the streets. Believe it or not, obtaining probability samples of homeless people is possible. But the researcher's knowledge of the habits of the homeless—where they tend to stay, when they move to shelters and when they do not, and so on—is essential in creating a practical plan. The parts of a city known to be occupied by the homeless can be divided into sectors and a sample of sectors chosen. A specific day for the survey can be designated and interviewers dispatched to the sampled locations to search for the homeless. Every homeless person encountered would be counted and a sample of them selected systematically for interviewing. All shelters that are open at the time of the survey—or a sample of them—would also be visited, an enumeration made of the occupants, and a sample of them interviewed. Although difficult to plan and implement, such a sample design could yield a probability sample of homeless people because all locations in the city likely to have homeless residents had a chance of inclusion. The more accurate the researcher's familiarity with the population and its habits, the more successful the study would be.

NON-PROBABILITY SAMPLING

The discussion of sampling will conclude with a few words about non-probability samples. For some purposes in survey research, it may not be essential or even feasible to have a probability sample. But as with everything in life, some non-probability samples are better than others.

In general, exploratory research may be an appropriate setting for non-probability sampling. Typically, there is an interest in gathering in-depth data from a relatively small number of cases, and although such research is relevant only to the extent that it can be generalized to the population of interest (otherwise, why do it?), there would be no effort to generalize the findings to a broader population using numerical data and estimates of sampling error. In-depth interviews with people to explore key concepts for a survey are often conducted with individuals chosen purposefully to represent a range of perspectives. Focus groups conducted with small numbers of participants to help in crafting language for survey questions are usually conducted with non-probability samples of individuals who are located near the facility in which

the group will meet. Focus groups are widely used in CA&R research to provide data difficult to collect by other means. Often the participants are experts or individuals who have firsthand experience in situations of interest. They can provide estimates of the internal stability of a regime and the potential for a coup, the activities of terrorist organizations, evidence about progress toward democratic institutions, and judgments about the success of peacekeeping missions.

More systematic research is also conducted with non-probability samples. A great deal of consumer behavior research uses *quota sampling,* often through intercept surveys in shopping malls. Interviewers are instructed to intercept shoppers more or less randomly who pass by a certain point in a mall and invite them to participate. The interviewers are asked to interview a specified number of people who fit each of a set of different categories (e.g., by age, race, gender). Quotas for each type of person are established (based upon judgments about the population of interest, which may be the consumers of a particular product), and the interviewer tries to fill the quotas by selectively intercepting individuals who appear to match the categories. Unlike probability sampling, quota sampling usually allows the interviewer to use his or her discretion in approaching and intercepting individuals for interviews. Although quota samples do not provide any basis for estimating sampling error—or indeed any kind of gauge of representativeness—they are relied upon for billions of dollars of business and marketing decisions worldwide, to get views from selected individuals living in war zones regarding the prospects for peace, and are often accepted in courts of law as credible evidence in legal disputes.

Data Collection

A useful survey entails not only developing a good questionnaire and a valid sampling plan, but success in actually interviewing the sampled individuals. One of the first decisions in a survey design is how the data will be collected. The choices of a data collection method and its implementation are key elements in the survey process. The four principal modes of data collection are as follows:

Personal interviewing, in which an interviewer interacts face-to-face with a respondent. Sometimes this interaction may also involve the use of a written questionnaire or a computer, with the respondent completing certain parts of the interview on paper or on the computer.

Telephone interviewing, in which an interviewer interacts over the telephone with a respondent. There are also telephone surveys using recorded interviewers that ask the respondent to answer by keying responses on the

telephone touch-tone keypad, but we do not consider these to be true telephone interviews. They are more akin to self-administered surveys.

Printed self-administered interviewing, in which respondents complete a printed questionnaire that has been provided to them through the mail, by personal delivery (such as a drop-off at home), by an employer, or in a group setting.

Internet interviewing, in which respondents complete a survey on a Web site or in an e-mail message.

Often the decision regarding survey mode is driven almost entirely by monetary considerations. Graduate students conducting large-scale surveys without grant funding may find it impossible to use personal interviewing or even telephone interviewing because of the costs involved. Even well-funded entities find that they cannot afford personal interviewing for most projects.

But money is not the only consideration relevant to the choice of survey mode. This decision usually entails the balancing of a number of considerations and trade-offs. The population of interest may be more amenable to one type of mode than another. (For example, people with low levels of education may lack the skills to handle printed questionnaires; some elite populations may be inaccessible by telephone.) Certain kinds of survey content may require a specific mode (the need for detailed information about particular negotiations would point to printed questionnaires that can be completed after records are consulted). Some populations cannot be reached by certain modes. (About 5% of U.S. households do not have home telephones; perhaps one quarter do not have access to the Internet; an increasing number of young people have a cell phone but no land line.)

ADMINISTRATION OF THE SAMPLE

A sample is only as good as its implementation. That is, a good sampling plan, executed with a good sampling frame, may yield a terrible set of interviews if administered in such as way as to result in serious biases. For example, if a mail survey of the general public in a community obtains a high rate of response from upper-middle-class residents but almost no response from lower-income residents, we would say that it obtained a biased sample. The results are certainly not going to be representative of the larger community. The infamous *Literary Digest* straw poll of presidential voting intentions in 1936 is a good example of the problem. The sample for the poll was itself biased, in that it was drawn from lists of automobile owners, households with telephones, and magazine subscribers. In the midst of the Great Depression, this sample was certainly skewed toward upper-income citizens.

But the sample itself was not the most important reason the poll predicted that President Franklin Roosevelt would be defeated for reelection. The

problem was that voters who disliked Roosevelt were more motivated to respond than were those who planned to vote for him. The poll drew about 2 million responses—a huge number by the standard of any contemporary poll—but a very low rate of response. Thus, a bias toward the Republican candidate was introduced, and the poll failed to predict the outcome of the election. If the *Literary Digest* had prodded non-respondents to participate, this bias might have been avoided. The magazine went out of business shortly after publishing these results.

We frequently encounter online polls at Web sites we visit. These polls invite us to register our opinions, and it is entertaining to take these polls and compare our results to the totals registered on the site. But these polls cannot be generalized to any known population and are essentially meaningless as gauges of public sentiment. Because any visitor to the site (and *only* visitors) can take the poll—and can take it numerous times—such polls are subject to serious biases. America Online posted online polls asking people what should happen to President Clinton after his affair with a White House intern was revealed. A majority said Clinton should resign, and the site tallied over 100,000 "votes." But valid public opinion polls representing the American public at the same time showed big majorities opposing impeachment or resignation. Those who visited the AOL site and were motivated to take the poll were more hostile toward the president than those who did not do this, so the poll was biased.

Much has been made about declining response rates in the United States and elsewhere. It is true that response rates are lower now than even 6 years ago. Well-designed telephone surveys by major media organizations typically obtain response rates in the range of 25% to 40%. And concern about response rates in any given survey is not misplaced. But implementation of a good probability sample in such as way as to minimize the possibility of bias in the obtained responses is more important than achieving some kind of mythic target response rate.

Personal Interviewing

For many purposes, personal interviews are ideal. A skillful interviewer approaching a potential respondent in person is more likely to solicit cooperation than a telephone interviewer or an approach via mail or e-mail. The rapport that a good interviewer can establish with a respondent makes it possible for personal interviews to be much longer than is possible in other modes (1-to 1½-hour interviews are common), and in-person interviews can yield significant benefits in terms of item non-response (people are more willing to answer questions) and in respondent comprehension of the questions. Respondents appear to take personal interviews more seriously than other kinds of interviews and to put more effort into providing the requested information.

Personal interviewing is very costly, chiefly because of the travel time required. Unlike telephone interviewing, where many interviewers are working in a common location and are able to efficiently conduct a large number of interviews, personal interviewers must travel to the sampled locations, find the respondents at home, and secure their cooperation. Not surprisingly, much of their time is wasted in travel and unsuccessful efforts to contact respondents. This is similar to the problem of subjects not appearing for a scheduled session of an experiment.

Mail Surveys

Self-administered questionnaires delivered by mail probably constitute the most common type of survey in the United States. Despite their reputation as a "method of last resort," the mail survey can achieve high response rates and in many situations is the method of choice (see Fox, Crask, & Kim, 1988). Certainly mail surveys can be inexpensive to mount and are popular with students, nonprofit organizations, and others who want to conduct an affordable survey.

Successful mail surveys require a significant degree of skill and attention to detail, however. The first element is a good questionnaire, discussed earlier. The self-administered questionnaire must be easy to comprehend and complete ("user-friendly") and not so long or intimidating as to turn off respondents.

Equally important is how the survey is presented and what effort is made to convince the respondent of the importance of the research. Donald Dillman is the acknowledged expert of self-administered surveys. In the 1970s, he assembled a set of principles for good survey practice under the rubric of the "Total Design Method," which has evolved into the "Tailored Design Method" (see Dillman, 2000). Underlying his approach is the notion of the social exchange, in which the researcher communicates to the respondent the seriousness of the enterprise and the social benefits and rewards he or she would receive in exchange for the effort involved in responding to the survey.

In practical terms, this means that mail surveys must appear professional and personal to the respondent. Essential elements of the design would include

- First-class postage—stamps preferred—on the envelope delivering the survey to the respondent. Self-addressed stamped envelope included for return of the questionnaire
- Personalization of the address (no letters to "occupant")
- A personalized cover letter to the respondent, explaining the value and purpose of the study, how the data will be used, who is conducting it, and how to get further information about it. The letter is signed by the researcher (preferably with ball point pen)

- Questionnaire is in booklet format, with an attractive cover, no questions on the inside cover, plenty of margin space, relatively large type, room for comments on the back cover
- Carefully timed follow-up mailings, including a reminder postcard 1 week after the initial mailing, a second questionnaire and new cover letter to non-respondents 2 weeks after that, and additional mailings a few weeks later (perhaps using express mail or overnight delivery to convey the seriousness of the enterprise)

Fortunately, modern computer technology, desktop publishing, and mail-merge programs make it possible for almost anyone to generate professional looking surveys, cover letters, and envelopes. It is easier than ever to mount a high-quality mail survey. Depending on the population, properly designed mail surveys can achieve response rates comparable to telephone surveys and, for certain populations, may be the only way to gain access.

Telephone Surveys

The telephone remains the most common mode of survey administration for most large-scale surveys in industrialized nations. In addition to the United States, it is feasible to conduct national population surveys by telephone in Great Britain, Germany, France, Japan, the Netherlands, all of the Scandinavian countries, Belgium, Italy, and Spain. The telephone is much less feasible as a data collection tool in the less-developed countries that are often the focus of research in CA&R. In these countries, the personal interview is often the preferred mode of data collection.

Telephone administration has an advantage over other modes in terms of speed and quality control. A large national telephone survey can feasibly be conducted in 2 to 3 nights, and acceptable samples may be gathered in a single night if certain caveats are accepted. Telephone interviewers typically work in a centralized facility where their work can be monitored to ensure that all questions are being asked in a standardized way with appropriate probes and follow-ups. Productivity in these facilities is usually very high.

Long surveys may be problematic on the telephone, though some telephone surveys extend to 40 minutes or longer. Additionally, respondents with hearing problems or limited formal education may find telephone surveys challenging because of the absence of visual and emotional cues that help people understand the content of interpersonal communication.

The greatest concern about telephone surveys today pertains to declining rates of response. Many factors are contributing to this problem. The public encounters a growing number of telemarketers and solicitations for non-profit organizations. Concerns about crime such as identity theft, credit card fraud, or burglary lead people to be less willing to give out information about

themselves on the telephone. Growing numbers of households have call screening devices such as answering machines, caller ID, and call blocking. The result is that response rates have declined over the past three decades, with perhaps an acceleration of the decline in the past 6 years as new technologies for call screening have become available.

We have a great deal of evidence that declining response rates have not yet damaged the ability of telephone surveys to provide valid data about the U.S. public. Telephone surveys conducted at the time of national elections still yield accurate forecasts of how the public will actually vote in the elections. Experiments comparing well-designed surveys with higher and lower rates of response find little difference in the substantive conclusions they draw (see Keeter, Miller, Kohut, Groves, & Presser, 2000).

Internet Surveys

Online surveys are a welcome innovation in the survey world. The Internet makes it possible for anyone with access worldwide to receive a solicitation and complete a questionnaire online in very short order. As with telephone or computer-assisted personal interviewing (CAPI), online questionnaires can be dynamic and contingent, skipping and branching according to the responses given. Unlike the telephone survey, they can provide the respondent with visual materials.

It is not yet clear whether data quality with online surveys will be better or worse than in other modes. Online surveys are like all self-administered surveys in that we can expect higher levels of item non-response—there is no interviewer present to urge respondents to answer. But there is also evidence that respondents take the interview more seriously and expend more cognitive energy in online surveys than in telephone surveys.

The chief concern about online surveys is how to get the sample to complete the survey. In some situations, this is relatively easy. Employee surveys are often conducted online. Employees are contacted at work by e-mail or through a mailing and asked to go to a Web site and complete the survey. If Web access is provided at the work site, everyone will be able to access the survey, and we can expect high rates of response. Even outside the work setting, Internet surveys of professionals who are interested in the survey tend to yield high response rates. College students are also apt to complete online surveys. But if respondents are not intrinsically motivated to respond, it may be essential to reach them through mail, telephone, or personal contact to urge them to go to the Web site and complete the interview. Because of spam and the very impersonal nature of e-mail, it has tended to be a less effective recruitment tool for online surveys. (For an example of an online survey of negotiation experiences, see Druckman, Ramberg, & Harris, 2002.)

Multimode Surveys

Sometimes the most successful surveys involve more than one mode of contact and data collection. Some surveys provide respondents with a choice of how to respond, for example, to be interviewed by telephone or to go to a Web site and complete the questionnaire. Multimode surveys present special challenges and can be more costly than single-mode studies, and there is often a question about the comparability of data collected in different modes. (For example, we expect somewhat greater item non-response in self-administered surveys.) But for certain populations such as professionals, offering a choice of mode may help you achieve a better response rate. And for populations that may be missed entirely by one mode—for example, people without telephones, people in less-developed nations where telephone coverage is incomplete—a survey that needs to represent the entire population may need to consider using multiple modes.

SUMMARY OF ADVANTAGES AND DISADVANTAGES OF DIFFERENT SURVEY MODES

A summary of important considerations in choosing a survey mode is presented in Table 5.2. Cost is almost always a critical concern and may override other considerations, but many of the other criteria shown in the table are important to the choice of mode as well.

PRE-TESTING SURVEYS

The final topic we discuss in planning a survey is not at all the least important—the *pre-test,* or preliminary test. Compared with many other kinds of research, survey research uniquely benefits from real-world testing of questionnaires and survey administration techniques. Regardless of how carefully we try to craft questionnaires to capture how people feel and think about a topic, we can never anticipate all of the ways in which our words can be misunderstood, misconstrued, or simply offend or bore a respondent. Nor can we fully anticipate how responsive a population will be to our request for cooperation, unless we are replicating a well-known methodology such as an RDD (random digit dialing) telephone survey.

For these reasons, preliminary testing of the questionnaire (and, where necessary, the data collection process itself) is essential to success in the survey process. Pre-tests typically consist of administering the draft questionnaire to a small sample of respondents drawn from the same population and in the same way as is planned for the real study. A small part of the RDD sample is used for the pre-test, and 15–25 interviews may be conducted. Often the pre-test is administered by the most experienced interviewers on the staff, because they can do a good job with an unfamiliar questionnaire and can provide useful feedback on questions that do not work or on problematic sequencing of questions.

Table 5.2 Survey Modes

	Personal Interviewing	*Mail Surveys*	*Telephone Surveys*	*Internet Surveys*
Cost	Very high	Low to moderate	Moderate to high	Low to moderate
Accessibility to population of interest	Very high	High	High, except for non-telephone households	Moderate; 25% of individuals in the U.S. have no Internet access
Length of data collection	Lengthy	Lengthy	Fast	Moderate to lengthy, depending on respondent motivation
Infrastructure and staffing needed to administer	Substantial	Minimal	Substantial	Minimal to moderate
Level of questionnaire complexity	High	Moderate (depending on population)	Moderate	Moderate
Maximum length of survey	Long	Moderate	Moderate to long	Moderate
Response rates	Moderate to high	Low to moderate	Low to moderate	Very low to moderate
Respondent cognitive skills needed	Low	High	Low to moderate	High
Ease of use of open-ended questions	Easy	Difficult	Easy to moderate	Difficult
Use of graphics or other visual aids	Yes, including video and animation	Yes, static graphics only	No	Yes, including video and animation
Item nonresponse	Low	Moderate to high	Low	Moderate to high
Control over order of items answered	Complete	Low	Complete	Depends on implementation
Ability to control who responds	High	Low	High	Low

In almost all surveys, we create draft questionnaires that are simply too long, due to the "nice to know" problem ("it would nice to know if. . . ."). The pre-test can tell us how long the survey actually runs and helps us focus on how to pare it down to a manageable length.

Perhaps the most important value of pre-tests is to identify questions that respondents do not understand, or struggle with, or do not try to answer accurately. The principal investigators should listen to the pre-tests—most interviewing facilities have the capability to allow the unobtrusive monitoring of the telephone calls—and the talk with the supervisor and the interviewers about the experience of administering the new questionnaire. Experienced interviewers often have good suggestions about how to reword questions, and simply hearing respondents as they wrestle with questions often yields insights about how to reframe a query.

Another useful role for pre-tests is to assess a data collection method. With RDD surveys of the general public using organizations that have conducted many such studies, we can make reliable estimates of response rates. But when conducting surveys of specialized populations—either by telephone, mail, or in person—we often do not know what to expect. This uncertainty is problematic, as we may not know important parameters, such as how many questionnaires to mail out in order to achieve a certain sample size. If time permits, a pre-test of the survey mode with the selected population can provide insight about the amenability of the population to the survey. If we get an especially low response rate from the test, we can consider alterations to the design such as including a monetary incentive, using a prenotification letter prior to calling or sending the survey, or other similar techniques.

Pre-testing is a broad topic that encompasses even more complex methods than we have described here. But any study will benefit from even a simple pretest, and often more than one attempt is useful.

Ethics

Ethical concerns in research are discussed in Chapter 1, but a few points are worth repeating in the context of surveys. The first is that your survey should do no harm. Most surveys are unlikely to harm anyone, but certain topics—especially in CA&R—are highly sensitive and can cause distress. Respondents in such surveys should understand that they are free to skip questions that are upsetting to them. What is more important, researchers must consider whether the questions being asked pose any risk to the respondents. For example, public health researchers may want to know whether a respondent has been the victim of domestic violence, but such questions asked in a telephone survey may pose a risk to the respondent. An abusive (and suspicious) spouse or partner might be listening to the interview.

A second concern is that responses must be kept confidential. This means that identifying information that could connect a respondent with an interview must be kept hidden and, as soon as possible, destroyed. Identifying information (such as a telephone number) may be needed so that interviews can be verified, but after a reasonable time period, these links should be deleted. Most surveys are subject to subpoenas by courts of law, but if the information linking the surveys with specific respondents has been destroyed, the subpoena cannot result in the identification of individual respondents and the disclosure of confidential responses.

A third ethical concern about surveys is that they take the time of respondents on behalf of your research, and this contribution of time should be respected and not wasted. A 10-minute survey with 1,000 respondents represents the equivalent of a month of full-time work (8-hour days times 5 days per week) contributed by volunteers to your research. Although these individuals may derive psychological benefits from participating in the research, they have nevertheless contributed a finite and valuable resource to it. It is our obligation to use this time wisely and carefully.

Conclusions

For professionals in CA&R, the ability to describe and analyze the experiences and perspectives of larger populations—those who have lived through wars and other violent conflicts, mediators in all types of settings, people who are currently dealing with conflict in their work lives, and so forth—is often essential to achieving their goals. The sample survey is a primary tool for providing this description and analysis.

Surveys can provide credible evidence when they draw from representative samples of the population of interest and when they measure concepts of interest in a valid and reliable way. This chapter has focused on how these objectives can be met. An important theme has been the trade-offs we must make throughout the survey process. Especially in the field of conflict analysis and resolution, conditions are often not optimal for survey research. Probability sampling may be difficult or impossible. Political considerations or restrictions may limit the questions we can ask, or we may be asking about topics that are highly sensitive to respondents. This does not mean that surveys should not be conducted, but we must be aware of the problems and adapt our methods to meet these challenges.

The sample survey is a critical tool for giving voice to the silent, for empowering the powerless, for offsetting the advantage that the articulate and well connected have in the world of politics and public affairs. The difficulties of conducting surveys notwithstanding, it is a critical tool for ensuring that all sides in a conflict can be heard and considered.

Discussion Questions

In this chapter, we have covered many aspects of survey research, including issues of design, questionnaire construction, sampling, and administration. You should review this material before embarking on a survey. The following discussion questions are intended to help with the review process. They cover the various parts of the chapter.

1. For what types of research projects are surveys an appropriate mode of data collection? When would surveys be inappropriate?
2. Hundreds of thousands of survey questions exist in archives, providing valuable comparisons with other populations and other times in history. What considerations should be used to decide between a question from an archive and one that might be tailored for a specific research situation?
3. Discuss the practical implications of the humorous advertising slogan that tells potential survey researchers to choose two options among quality, speed, and price.
4. For what topics in conflict analysis and resolution would open-ended survey questions be most useful? For each topic you suggest, draft one or two close-ended questions that would address the same subject matter.
5. Describe how you would go about creating a sample of mediators for a survey. What lists, if any, could you use? What proportion of the population of mediators would these lists cover? What kinds of mediators would be omitted from such a sample?
6. If you were conducting a survey of the general public about experiences with conflict in daily life, how would you draw the sample? Would an RDD sample suffice, or should certain groups in the population be over-sampled? If so, how would you do this?
7. Probability samples offer the ability to estimate the accuracy of a survey. For what kinds of research in the field of conflict resolution is this ability important for the credibility of the results? How does a researcher resolve the tradeoffs between the higher costs of probability samples and the benefits of greater reliability of the results?
8. For surveys of U.S. peacekeepers recently returned from conflict regions, what survey mode would be most appropriate? What mode is most appropriate for people experiencing serious conflict in their work lives?
9. Describe the steps you would take to ensure that your survey would not exacerbate the anxiety felt by refugees from a conflict region.
10. A government official demands an explanation of why you are spending thousands of dollars on survey research, drawn from funds intended to assist people in war-torn regions of a nation of interest. Aside from your specific research interests, what general justification for survey research in conflict regions can you offer?