CHAPTER

1

Sample Surveys in Our Electronic World

Hundreds of times every day someone decides to create a survey. The variety of organizations and individuals who make this decision is enormous, ranging from individual college students to the largest corporations. Community service organizations, nonprofit foundations, educators, voluntary associations, special interest groups, research scientists, and government agencies also all collect needed information by conducting surveys. The topics of these surveys vary greatly, from questions about health, education, employment, and political preferences to inquiries about television viewing, the use of electronic equipment, and interest in buying a new car, among many other things.

The reasons for deciding to conduct a survey are as diverse as the range of survey sponsors and topics. Sometimes, the justification is that the sponsors do not know the opinions or beliefs of those they want to survey. More typically, the sponsor has interests that go much deeper, wanting to know not just how many individuals in a group have a particular attitude, but how that attitude varies with other respondent characteristics that will be asked in the survey, such as across men and women or across different age or socioeconomic groups.

While the need to know something that is unknown drives the decision to conduct most surveys, the uses of survey results are as diverse as those who sponsor them. For example, one of us recently completed a community survey that was used to decide what facilities to include in a new neighborhood park that was about to be developed. University leaders use results from surveys of students to revise their undergraduate and graduate education programs. Public opinion pollsters use results from surveys of likely voters to predict who will win national and local elections. The Federal Reserve uses estimates of the unemployment rate produced monthly in the Current Population Survey to help set economic policy. Data from this same survey are used by individuals and businesses throughout the United States to make investment, hiring, and policy decisions. Market researchers use surveys to provide insights into consumer attitudes and behaviors. Nonprofit groups use surveys to measure attitudes about issues that are important to them and support for possible programs the group might pursue.

Surveys are both large and small. For example, over the course of a year the U.S. Census Bureau asks a few million households to respond to the American Community Survey. Others ask only a few hundred or even fewer individuals to respond. The survey response mode also varies, with some surveys being conducted by a single mode—in-person, web, telephone, or paper—while others provide multiple modes for answering questions. Sometimes respondents are asked to respond only once, while in other surveys a single individual may be asked to answer questions repeatedly over months or years, and surveys may be conducted

in just a few weeks or over several months or years. In some cases people are asked to provide information about themselves or their households, and in other cases they are asked to provide information about a particular business or other organization with which they are affiliated.

Despite this diversity, all surveys still have a lot in common. Each is motivated by the desire to collect information to answer a particular question or solve a particular problem. In some cases the desired information is not available from any other source. In other cases, the information may be available, but it cannot be connected to other important information—such as other characteristics or related attitudes and behaviors—that need to be known in order to solve the problem or answer the question.

In most surveys only some of those in the population of interest are asked to respond. That is, the survey is based on a *sample* rather than being a *census* of every member of the target population. In addition, those who respond are asked questions they are expected to answer by choosing from among predetermined response categories or, occasionally by providing open-ended answers in their own words. These commonalities and the enormous amount of money and effort now spent on surveys point to their importance as a tool for learning about people's characteristics, opinions, and behaviors, and using those results to inform and direct public policy, business decisions, and for many other purposes.

Other nonsurvey means, both quantitative and qualitative, are available to social scientists, marketing professionals, government officials, special interest groups, and others for collecting useful information that will produce insight into the attitudes and behaviors of people and the groups they are a part of. These include unstructured interviews, focus groups, participant observation, content analyses, simulations, small group experiments, and analyses of administrative records or organic data such as birth and death records, sales transactions, records of online searches, social media, and other online behavior. Each of these methods can yield different types of information, and for some questions they are more appropriate than surveys or may be used in combination with surveys to answer the research question or community problem.

The feature of the probability sample survey that distinguishes it from these other methods of investigation is that it can provide a close estimate of the distribution of a characteristic in a population by surveying only some members of that population. If done correctly, it allows one to generalize results with great precision, from a few to the many, making it a very efficient method for learning about people and populations.

The efficiency and importance of the probability sample survey might best be illustrated by considering an alternative way to learn about a population—a census. Every 10 years the U.S. Census Bureau attempts to contact and survey every household in the United States, as required by our Constitution. The resulting information is used to reapportion the U.S. House of Representatives so that each member represents about the same number of U.S. residents. This massive survey, known as the Decennial Census, costs billions of dollars to conduct. A smaller organization that wants to know the opinions of all U.S. residents on a particular issue could hardly afford such an undertaking. But with a probability sample survey, it can learn those opinions for considerably lower costs by selecting only some members of the population to complete the survey.

Even on a smaller scale, few would be able to afford to survey every undergraduate student at a large university in order to assess students' satisfaction in the

education they are receiving. If this were necessary, studies of student satisfaction would seldom, if ever, be done. But probability sample surveys allow us to be much more efficient with our resources by surveying only a sample of students in a way that enables us to generalize to the entire student population.

Whatever the target population or research question, limiting our data collection to a carefully selected sample of the population of interest allows us to concentrate limited resources (e.g., time and money for follow-up communications, data cleaning, and analysis) on fewer individuals, yet obtain results that are only slightly less precise than they would be if every member of the population were surveyed.

Our purpose in this book is to explain how to conduct effective probability sample surveys. We discuss the fundamental requirements that must be met if one wants to generalize results with statistical confidence from the few who are surveyed to the many they are selected to represent. We also describe specific procedures for designing surveys in which one can have high confidence in the results. Regardless of whether your interest in surveys is to understand one of the many national surveys that are conducted for policy purposes or to gain knowledge of how to design your own survey of organization members, college students, customers, or any other population, it is important to understand what it takes to do a good survey and the multiple sources of error that can reduce the accuracy of the survey results—or completely invalidate them.

FOUR CORNERSTONES OF QUALITY SURVEYS

In general, survey error can be thought of as the difference between an estimate that is produced using survey data and the true value of the variables in the population that one hopes to describe. There are four main types of error that surveyors need to try to minimize in order to improve the survey estimates.

1. *Coverage Error* occurs when the list from which sample members are drawn does not accurately represent the population on the characteristic(s) one wants to estimate with the survey data (whether a voter preference, a demographic characteristic, or something else). A high-quality sample survey requires that every member of the population has a known, nonzero probability of being sampled, meaning they have to be accurately represented on the list from which the sample will be drawn. Coverage error is the difference between the estimate produced when the list is inaccurate and what would have been produced with an accurate list.

2. *Sampling Error* is the difference between the estimate produced when only a sample of units on the frame is surveyed and the estimate produced when every unit on the list is surveyed. Sampling error exists anytime we decide to survey only some, rather than all, members of the sample frame.

3. *Nonresponse Error* is the difference between the estimate produced when only some of the sampled units respond compared to when all of them respond. It occurs when those who do not respond are different from those who do respond in a way that influences the estimate.

4. *Measurement Error* is the difference between the estimate produced and the true value because respondents gave inaccurate answers to survey questions. It occurs when respondents are unable or unwilling to provide accurate answers,

which can be due to poor question design, survey mode effects, interviewer and respondent behavior, or data collection mistakes.

We consider reducing the potential for these errors as the four cornerstones of conducting successful sample surveys. Surveyors should attempt to limit each to acceptable levels. None of them can be ignored. As such, each receives detailed attention in the chapters that follow. Because these sources of error are so essential for defining survey quality, we describe each of them here in more detail.

Coverage Error

As we previously mentioned, the strength of a probability sample survey is that it allows us to collect data from only a sample of the population but generalize results to the whole, thus saving considerable time, money, and effort that would be incurred if we had to survey everyone in the population. However, in order to draw a sample, one has to have a sample frame, or a list of members of the target population, and any errors in that list have the potential to introduce coverage error into the final estimates that are produced. If some units from the target population are not included on the sample frame (i.e., undercoverage) *and* they differ from those that are in ways that are important to the survey, the final estimates will contain error.

For example, all other error sources aside, a landline random digit dial telephone survey would likely overestimate the prevalence of higher socioeconomic status because the well-off are more likely than the poor to have landline telephone service (i.e., the well-off are more likely to be on the landline random digit dial sample frame) (Blumberg & Luke, 2013). In fact, one of the challenges now being faced in conducting household telephone surveys is that only about 58% of households still have landlines (Blumberg & Luke, 2013), the traditional source of random digit dialing samples, and those who have them are quite different from those who do not on a number of important characteristics. Using the landline telephone frame alone (without supplementing it with a cell phone frame) for a national household survey would leave out significant portions of the population who are likely to differ in important ways from those included on the frame.

Similarly, conducting a national household survey by Internet would leave out significant portions of the population because, as of May 2013, only 73% of American adults have Internet access in the home (Pew Internet & American Life Project, 2013b). In comparison, an Internet survey of undergraduate students at a university, where all students are required to use the Internet, would likely have little coverage error, provided a list of all students could be obtained. In Chapter 3 we discuss in detail the threat of coverage error, its likely sources, and how to limit it.

Sampling Error

The extent to which the precision of the survey estimates is limited because only some people from the sample frame are selected to do the survey (i.e., sampled) and others are not is known as sampling error. If we have a sample frame with complete coverage (i.e., the list matches the population perfectly), we can say that sampling error is the difference between the estimates produced and the true value because we survey only a sample of the population and not everyone. The power of probability sampling, which is also discussed in detail in Chapter 3, is that

estimates with acceptable levels of precision can usually be made for the population by surveying only a small portion of the people in the population. For example, a researcher can sample only about 100 members of the U.S. general public and, if all 100 respond, achieve estimates with a margin of error of +/−10%. Successfully surveying a sample of 2,000 individuals reduces the margin of error to about +/−2%. Surveying 100 or even 2,000 people rather than the approximately 315 million people in the United States represents an enormous and desirable cost savings, but doing so means that one has to be willing to live with some sampling error in the estimates.

Sampling error is an unavoidable result of obtaining data from only some rather than all members on the sample frame and exists as a part of all sample surveys. For this reason, we describe the importance of reducing survey error to acceptable levels, rather than being able to eliminate it entirely. By contrast, censuses—in which all members on the sampling frame are selected to be surveyed—are not subject to sampling error.

Many novice surveyors find sampling error to be somewhat nonintuitive. They find it difficult to imagine only needing to survey a few hundred or thousand to learn about millions of households or individuals. Yet, during each presidential election in the United States, surveys of between 1,000 and 2,000 likely voters are conducted that correctly estimate (within the limits of sampling error) the votes for each candidate. For example, across polls conducted in the final week of the 2012 campaign, the average error for each candidate was about 2 percentage points. Just as nonintuitive for some beginning surveyors to grasp is that in order to predict the outcome of a local election for a particular state or medium sized U.S. city with perhaps 50,000 voters, nearly as many people need to be surveyed as are needed for predicting a national election.

The exact sampling error is easily calculated mathematically, as described in Chapter 3. However, the ease of making those calculations and the mathematical preciseness of the result leads to overreliance on it as a singular measure of the amount of error in a survey statistic. This tendency should be avoided. Sampling error calculations reflect the completed sample size, that is, only received responses are considered. The larger the number of responses, the greater the reported precision and statistical confidence. But they ignore the possibility for coverage error as well as the fact that many and sometimes most of the invited participants did not respond, which raises the potential for a third source of error, nonresponse error.

Nonresponse Error

Many sponsors think of a survey's response rate (the proportion of sampled individuals that respond to the survey) as the major indicator of survey quality. A major focus of this book is how to obtain high response rates to surveys. However, taken by itself, the response rate is only an indirect indicator of survey quality. The more important response quality indicator is nonresponse error, which occurs when the characteristics of respondents differ from those who chose not to respond in a way that is relevant to the study results. For example, if a survey on environmental attitudes obtained responses mostly from those individuals who have positive attitudes toward the environment and those who have negative attitudes are underrepresented, then that survey's results would be biased because of nonresponse error.

The common mistake sometimes made by novice surveyors is to consider response rate as an adequate indicator of whether nonresponse error exists. Comparisons across many surveys have shown that nonresponse error may occur in surveys with higher as well as lower response rates (Groves & Peytcheva, 2008). For example, in 1989 a study was conducted in Dallas County, Texas, to learn about people's thoughts and behaviors related to acquired immunodeficiency syndrome (AIDS). Sampled individuals were asked to complete a self-administered survey and have a blood sample drawn by a phlebotomist. This study achieved a remarkable 84% response rate: A rate that some might think is a clear indication of high quality. But to ascertain whether there was nonresponse bias, the researchers went back to a random sample of the nonrespondents and were able to get some to participate (some were not asked to give the blood sample at this stage). This effort revealed that the prevalence of human immunodeficiency virus (HIV) risk behaviors like intravenous (IV) drug use and male-to-male sex were underestimated in the original data collection effort. Only 3% of those who initially participated reported engaging in IV drug use compared to 7% of those who participated in the follow-up. Similarly, only about 5% of the initial participants reported engaging in male-to-male sex compared to about 17% of those in the follow-up (Centers for Disease Control and Prevention, 1991). Despite an impressive 84% response rate, the initial estimates were biased because those who responded differed from those who did not respond on characteristics of interest in this study.

While the study just described demonstrates that higher response rates do not guarantee minimal nonresponse error, it is important to recognize that higher response rates do reduce the likelihood of nonresponse error and thus provide greater credibility to surveys' results than do lower response rates. In addition, higher response rates result in larger completed samples, thereby increasing the precision of the estimates in that way. Thus, designing surveys in ways that produce higher response rates can be a helpful tool in reducing nonresponse error.

Response is a function of contact and cooperation. That is, in order to obtain a response, we first have to make contact with sample members and then we have to convince them to cooperate with our request to complete the survey. Using multiple contact attempts and varying the timing, delivery method, and mode of those attempts are a few ways we discuss in this book of increasing the likelihood of making contact with sample members. Respondent-friendly questionnaires, shorter (rather than longer) survey instruments, the use of incentives, follow-up requests that target likely nonrespondents, and switching survey modes are a few of the many features of survey design discussed in this book that are intended to increase the likelihood of sample members cooperating with our request. All of these strategies have the parallel objectives of increasing response while simultaneously reducing nonresponse error. Chapter 2 introduces the discussion of implementation procedures and a theory for guiding those decisions. The majority of this book, from Chapter 4 forward, focuses on many aspects of survey design that can reduce nonresponse as well as measurement error.

Measurement Error

Survey objectives are realized by asking questions to which respondents provide accurate answers. However, in designing a survey that will achieve valid and reliable measurement, one faces a gauntlet of measurement challenges. One of the challenges to asking a good survey question is making sure that it adequately

measures the idea or concept of interest. An example occurred in a survey in which the sponsor wanted to obtain a measurement of household wealth. He had tentatively decided to use household income for the previous year as a measure of wealth until a colleague pointed out that annual income is likely to decrease sharply when a person retires, but wealth typically does not. Similarly, a community survey sponsor proposed using length of time individuals had lived in their current residence as a measure of length of time in the community, but soon discarded the idea because of the likelihood that many people may have moved from one residence to another in the same community. When a question does not measure what it was intended to, as in these cases, it is typically referred to as having specification error (also known as low construct validity). Considerable time and effort can be spent deciding what format of question to use, what type of scale to provide, how to label answer categories, whether to offer a "don't know" option, and any number of other details, but all of that effort is useless if the question does not measure the concept called for by the study objectives.

Once one has selected an acceptable way to measure a specific concept, there are many different ways that accuracy of the estimate may be compromised, resulting in measurement error.

- The substance of the question may encourage a response that, because of perceived societal norms, puts the respondent in a more favorable light to the interviewer and/or survey sponsor. Questions about sex and illegal behaviors are examples.
- The question may be unclear to the respondent because it uses words that are not understood or phrases that are confusing.
- The question structure may encourage certain answers that another structure would not. For example, items that ask respondents to mark all that apply tend to result in fewer selections among later categories than those that ask for an explicit positive or negative answer for each item (i.e., a forced-choice or yes/no format).
- The order in which questions are asked may produce different answers to specific questions than would another order.
- The visual layout of a question may increase the likelihood that certain answers are chosen and others are not, or that some items are overlooked altogether.
- Some types of respondents may be less likely to give accurate answers than others.
- Perception of the expectations of interviewers or the sponsor may also influence answers.
- Interviewer characteristics, such as gender or race, may influence the answers people provide.
- The choice of survey mode may also influence answers to surveys. For example, research has consistently shown that scalar questions are likely to be answered differently in visual versus aural surveys.

These problems can result in two types of measurement error. The first is response bias, in which estimates are systematically shifted one way or the other. Two common examples are underestimating socially undesirable behaviors, like drug use and criminal activity, and overestimating socially desirable behaviors, like volunteering and voting. The second type of measurement error is response

variance, which is akin to the idea of low reliability. That is, if the measurement were taken over and over multiple times, it would produce a different result each time.

A great deal of terminology is often used to indicate why some questions and not others exhibit measurement error, including social desirability, primacy/recency, acquiescence, clarity of figure/ground relationships, the Law of Pragnanz, the norm of evenhandedness, and much more. We mention these many sources of potential measurement differences because writing effective questions requires simultaneously working on many fronts in an effort to reduce measurement problems in surveys to obtain accurate answers to all questions. We discuss this further in Chapters 4, 5, 6, and 7.

Total Survey Error

The need to focus on many design considerations at once sometimes results in ignoring one source of error, a mistake that can have devastating repercussions for a survey. For example, a faculty member concerned with reports of classroom cheating decided to take advantage of the web survey software available in her university and design a survey of students to get their perceptions about whether classroom cheating was happening and to learn what they thought would be appropriate punishment. It was her hope that conducting a probability sample survey of students would produce data she could report to the appropriate university officials to inform new policies for dealing with cheating cases. To avoid the challenge of sending sample members e-mails with individual passwords that would allow only those sampled to respond, she sent generic e-mails and set up the survey website so that anyone who knew about the survey could complete it. She soon learned that the e-mails sent to the carefully selected sample of students had been forwarded to other students and that some students with particularly strong viewpoints had filled out the survey multiple times (i.e., stuffed the ballot box!), which breaks from the requirement for a probability sample that only the people selected for the survey can provide a response and that each person can respond only once. In trying to simplify the administration of this survey, the faculty member ended up making a decision that undermined the probability nature of the sample and discredited the survey's results.

We have also observed situations in which survey designers became excessively concerned over resolving issues with small consequences. Upon learning that a sample of household addresses for a community survey would only reach about 95% of the households in the community, one surveyor became obsessed with how to manually add the missing addresses. To do so would have required tremendous costs and effort, including cross-checking records and potential personal visits to areas in the community to check to see if there were addresses there. In this case, the error from missing 5% of households was likely to be small, and the resources that would be required to fix it were excessive in relation to the likely benefit. It would have been more beneficial to focus on reducing other potential errors.

In another situation this may not be the case. Surveyors designing a national survey that will produce data used to allocate government funds may decide that even though small, the extra precision obtained by enumerating the missing 5% of addresses is worth the extra effort because it will help ensure that federal funds are fairly distributed.

One mistake some survey designers make is to worry most about what error source they know best. The research-based knowledge for dealing with specific

sources of error comes from different academic disciplines. Sampling theory and concepts for defining and understanding coverage effects come principally from statistics. Measurement issues are more likely to be dealt with by the disciplines of psychology and sociology. Nonresponse research draws concepts from all of the disciplines. While understanding of the behavioral reasons for nonresponse as relied heavily on sociological and psychological thinking, potential solutions for such response issues, such as imputing missing responses for individual items or calculating weighting adjustments to mitigate unit nonresponse have been developed primarily by statisticians. Economists, political scientists, and market research professionals have also contributed significantly to the literatures in these areas. Survey error is fundamentally a multidisciplinary problem and nowhere is that more evident than in efforts to reduce multiple sources of survey error. Good survey design requires giving balanced concern to error sources, regardless of one's inclination to focus mostly on what he or she knows best.

This state of affairs has encouraged the development and use of the Total Survey Error (TSE) framework. This term refers to attempting to design surveys in a way that maximizes data accuracy within constraints that cannot be ignored, such as costs and the time available for completing the survey (Biemer & Lyberg, 2003). Reducing total survey error involves careful survey planning, sample selection, questionnaire design, implementation, and data analysis. It is about simultaneously controlling all four sources of error to the extent practical and possible, within the time, cost, and other constraints of the survey. Survey error cannot be completely eliminated, but with diligence to all four types it can be kept to reasonable levels. Our emphasis throughout this book is on how reducing total survey error can be accomplished in large and small surveys alike, including those with generous as well as quite limited budgets.

Often reduction of total survey error focuses on discrete actions that can be taken separately to reduce each type of error, but in other cases a much broader systematic change to the survey design may be undertaken. For many years, the National Household Education Survey conducted by the National Center for Educational Statistics was conducted in a two-step process. Random digit dial telephone surveys (landline numbers only) were used to identify households with children. Then the identified households were surveyed again, also by telephone, to collect detailed information. It became evident early in 2007 that not only were response rates falling dramatically (Montaquila, Brick, Williams, Kim, & Han, 2013), but increasing portions of the nation's children were being raised in homes without landline connections. The proportion of children growing up in cell-only households has continued to increase, and is now over 45% (Blumberg & Luke, 2013). The survey sponsors were concerned about both coverage and nonresponse error and were worried about the costs associated with beginning to call cell phones to reduce the coverage error. A proposal to consider a possible change to address-based sampling using mail methods was met with considerable skepticism. In addition to not being sure it would improve response, changing to mail also meant that questions would need to be asked in different ways, changes that might impact trend lines from data accumulated over many years. But, after extensive testing, it was decided to make the switch based on considerations across multiple types of error.

Making these changes to the National Household Education Survey instead of continuing to try to fix the problems associated with the telephone survey was a major decision that took a lot of guts and hard work. It required extensive institutional change to switch from dealing with telephone to mail, as well as substantial

changes to the survey itself to make it work in a visual rather than aural survey mode. Because this undertaking was so enormous, initial reluctance was only overcome after several years of testing. Ultimately, this testing showed that the new methods were more suitable for the changing survey landscape we now face, and that they were beneficial from a total survey error perspective.

WHAT IS DIFFERENT ABOUT SURVEYING IN THE 2010s?

When the first edition of this book appeared in 1978, personal computers, the Internet, cell phones, and fax machines existed only as ideas that might someday be a part of people's lives. Surveys were limited to landline telephone, mail, and in-person interviews. When the second edition appeared in 2000, the Internet and another intriguing development, telephone Touchtone Data Entry, which eventually evolved into Interactive Voice Response, were added in a single chapter. At this time surveyors were just beginning to consider their possible uses.

Rapid technological development in the past 15 years has changed this situation substantially so that there are now many means for contacting people and asking them to complete surveys. Web and cellular telephone communication have undergone rapid maturation as means of responding to surveys. In addition, voice recognition, prerecorded phone surveys that ask for numerical and/or voice recorded responses, fillable PDFs, smartphones, tablets, and other devices have increasingly been used for data collection. Yet, for many reasons traditional phone, mail, and in-person contacts have not disappeared, and are often being used in combination to maximize the potential of reaching people. In addition, offering multiple ways of responding (e.g., web and mail in the same survey) is common. It is no longer practical to talk about a dominant mode of surveying, as in-person interviews were described in the middle of the 20th century and telephone was referred to from about 1980 to the late 1990s.

The situation faced by surveyors in this decade is in some ways ironic. We can now connect with a huge portion of a survey population in multiple ways; about 98% of U.S. households have either a landline or cellular telephone (Blumberg & Luke, 2013), around 96% have U.S. Postal Service mail delivery (Iannacchione, 2011), and 85% of adults in the United States use the Internet and 73% have Internet access in their homes (Pew Internet & American Life Project, 2013b, 2013c). Individual household access for in-person surveys is harder to estimate because of locked apartment buildings and gated communities that prevent interviewers from gaining access. However, while surveyors now have multiple ways to contact people, their efforts are often thwarted by buffers designed to keep unsolicited messages at bay. Receptionists or guards prevent access to buildings. Answering machines, voice mail, and caller ID technology filter telephone calls. E-mail filters and the ability to preview e-mails without opening them make e-mail survey requests less likely to be seen and answered. Thus, the technology that makes unprecedented and speedy access possible also provides the means of avoiding or ignoring it. In addition, cultural norms have evolved so that control over whether a survey request is received and responded to rests increasingly with the individual to whom the request is being made, and not with the individual making it.

Many years from now when the history of electronic communication is written, it is likely that one of the major themes will be its role in the elimination

of intermediaries. Tasks that once required help—making a bank withdrawal, reserving a room in a hotel or a seat on an airplane, leaving a phone message, and purchasing groceries—can now be done quite well without the assistance of another person. In this environment, why should surveyors expect that positioning an interviewer as a necessary intermediary between the surveyor and respondent remain the most prevalent way of conducting a survey? It should not be surprising that many telephone-only surveys now obtain response rates in the single digits (Keeter, Christian, Dimock, & Gewurz, 2012).

However, the rapid decline of telephone interviewing as a dominant stand-alone way of conducting household and other surveys is occurring for other reasons as well. The shift away from landlines as the predominant method of telephone communication means that the traditional sample frame for random digit dialing that was depended upon to cover the U.S. population no longer covers a considerable portion of households. Combining landline and cell phones poses difficult sampling challenges, some of which occur because many people have both landlines and cell phones, and because landlines tend to be household-based and cell phones tend to belong to individuals. In addition, the portability of cell phone numbers across geographic areas adds to the challenge when one wants to conduct a survey of a specific geographic area like a city or region. Those who keep a cell phone number from another area when they move into the area being surveyed will not appear on the sample frame, and those who kept their local number when they moved out of the area will be erroneously included in the frame. Also, the need to ask all respondents additional questions to establish eligibility is made difficult by the conflicting need to make questionnaires shorter, due to today's culture of people being less willing to reveal information about themselves to a stranger over the telephone.

Many surveyors were optimistic in the late 1990s that as telephone response rates fell, a smooth transition could be made to conducting most surveys over the Internet. This transition has not been as effective as it was envisioned. Not all households have Internet access, and the fact that individuals who do not use the Internet differ sharply (older, less education, and lower incomes) from those who do, makes it difficult to achieve adequate representation for many surveys. Perhaps even more importantly, there are no sample frames for household surveys that allow direct e-mail contact, like traditional random digit dialing for the telephone or address-based lists for mail. Even when e-mail addresses are available (e.g., lists of clients, students, and organization members), contact only by e-mail often produces response rates that are similarly low to those achieved in telephone surveys.

As a result, optimism about the potential for web surveys has more recently given way to puzzlement. Even casual observation in airports, shopping malls, and meetings make it evident that people are increasingly receiving and sending messages on smartphones and a myriad of other electronic devices. Full screen laptops or desktop computers with keyboards are no longer the predominant way that many people connect to the Internet.

While purse and pocket devices provide convenient ways to connect to the Internet, their small screens and input devices make reading and responding to survey requests quite difficult. Obtaining responses to a questionnaire in today's environment often requires getting an electronic survey request successfully through a prescreening on a smartphone (i.e., read but not deleted), and then returned to on a laptop, desktop, or tablet where respondents can more easily view and respond to the survey request. Complicating matters further, as many

young people continue to replace e-mail communication with texts or social networking status updates, it has become harder to reach this group. For these reasons, successfully shifting to electronic communication for all survey requests continues to be very challenging.

Mail surveys have also undergone a significant transformation. Although modern mail survey methods were being developed at the same time that random digit dialing enabled the telephone to become a prominent mode, mail has long been considered a less desirable and lower response rate alternative. This survey mode is also not well suited for the intensive branching that now characterizes many survey questionnaires. But substantial advancements in printing capabilities mean that the personalization and customization of paper surveys and mailing materials have advanced well beyond where they were just a decade ago.

The situation for mail also improved considerably when the U.S. Postal Service began routinely releasing a list of residential addresses of all households receiving delivery of postal mail. Improvements in the proportion of households with city addresses, as opposed to simplified addresses that were somewhat imprecise, now mean that about 95% to 97% of U.S. households are accessible to surveyors by mail (Iannacchione, 2011). At the same time, research has shown that responses to postal surveys have not declined as significantly as responses to telephone surveys (Messer & Dillman, 2011; Rookey, Le, Littlejohn, & Dillman, 2012; Smyth, Dillman, Christian, & O'Neill, 2010).

Ironically, mail has moved from being the lowest response rate mode for many survey designs to now having response rates that are significantly higher than telephone and being competitive with well-financed in-person surveys. It has also shifted from having the poorest coverage for household surveys to having the most comprehensive household sample frame. Mail surveys were also once considered the lowest cost method for conducting surveys but are now a somewhat higher cost method, especially when compared to an e-mail-only contact web survey. That said, mail continues to have its challenges, such as ensuring that the mail is actually delivered to the household and opened by someone in the household and that the person receiving it can read and comprehend it in the language(s) provided.

In sum, single mode surveys, regardless of mode, tend not to be as effective as in years past for many, if not most, survey situations. And increasingly, more than one mode may need to be used to contact and survey different individuals to ensure that various members of the population are represented.

WHY EMPHASIZE MIXED-MODE DATA COLLECTION?

Our emphasis in this book on mixed-mode survey designs stems from our desire to create designs that are most likely to keep the four major sources of error to acceptably low levels while also reducing survey costs. Mixing modes allows us to take advantage of the strengths of certain modes to overcome the weaknesses of others in order to minimize total survey error as much as possible within resource and time constraints. How exactly we mix modes depends heavily on our motivation for mixing them; that is, it depends on what sources of error we are trying to minimize or if we are trying to reduce costs or collect the data quickly.

One goal a surveyor might have is to reduce the costs of their survey. In fact, a recent study of national statistical agency surveys conducted in Europe and the

United States by Luiten (2013) found that reducing costs was the primary reason for the increasing use of mixed-mode designs. A common way to mix modes to reduce costs is to collect as many responses as possible in a cheaper mode before switching to a more expensive mode to try to obtain additional responses. This strategy was used by the U.S. Census Bureau for the 2010 Decennial Census. Paper questionnaires were first mailed to nearly every address in the United States and about 74% of them responded (U.S. Census Bureau, n.d.). Only then were more expensive interviewers sent out to try to obtain responses from households that did not respond by mail. The Census Bureau was able to save considerable money by getting most households to respond by mail and minimizing the number that would need to be visited by in-person interviewers.

However, there are many other reasons that multiple modes of survey response are used. Sometimes the goal is to improve coverage. While it is theoretically possible to contact sampled individuals in many different ways—cell phone, office phone, home phone, home postal delivery, office postal delivery, or through multiple e-mail addresses—it is quite uncommon for our available sampling frames or lists to include all types of contact information for each unit. The lack of available contact information for multiple modes can be due to the inability to match contact information from different frames or because people are unwilling to voluntarily provide multiple types of contact information to organizations requesting it (e.g., some people might provide a phone number, others an e-mail address, and still others a postal mailing address). In this context, developing a sample frame for a single-mode survey often means excluding members of the target population for whom the desired mode of contact is not available, potentially increasing coverage error. Mixing modes is a way to ensure most members of the target population can be included on the sample frame and thus have an opportunity to be sampled.

Sometimes a second or third mode is offered to individuals in hopes they will find an alternative mode particularly appealing or they will be able to respond to it when they are unable to respond by a different mode. An example is that individuals who cannot respond on a computer because of not having developed those skills may be quite comfortable responding by paper or by telephone. Some individuals may not pick up their mail or answer a landline phone but will check their e-mail and answer their cell phone. In cases such as these, using multiple modes can improve response rates and reduce nonresponse error by appealing to different kinds of respondents. In still other instances, one response mode is offered initially, such as web or telephone, and then followed by another (e.g., mail) to improve the speed of response and facilitate quicker processing of results.

Mixing survey modes does not necessarily mean offering people more than one way of completing a survey questionnaire. Different modes can also be used to contact sample members with the survey request even when only one mode is used for collecting responses. Traditionally, people were contacted by the same mode that was also used to complete the survey. However, research has long shown that contacting individuals by mail ahead of a telephone or in-person interview can improve response rates (de Leeuw, Callegaro, Hox, Korendijk, & Lensvelt-Mulders, 2007); similarly, follow-up telephone calls to remind people to respond can sometimes improve response rates for postal surveys.

In fact, in today's survey environment, using multiple survey modes as a means of communication to encourage response in a single mode may be a more powerful way of mixing modes to improve survey response and the quality of those

responses than simply providing an alternative mode for responding to a survey. Several decades of experimentation has consistently shown that sending a token cash incentive of a few dollars with a mail survey request improves response dramatically for that mode (Church, 1993). Recent research has now demonstrated that sending a postal letter with such an incentive and a request to respond over the web improves response over the web more so than with a request to respond to a paper questionnaire (Messer & Dillman, 2011). In these instances, mixing contact modes allows surveyors to incorporate other response-inducing strategies into their surveys.

Perhaps even more important is the potential for creating synergy between contacts via different modes to encourage survey responses. For example, while a postal request containing an incentive can be quite effective at getting people to complete a web survey (Smyth et al., 2010), recent research has shown that following a postal request with an e-mail containing an electronic link to the web survey can improve response rates even more (Millar & Dillman, 2011). Thus, one important area of potential for mixed-mode survey designs is using multiple types of contact information to produce contacts in different modes that work together in synergistic ways to convince sample members to respond.

In the third edition of this book we presented a model proposing four types of mixed-mode surveys:

Type 1: Use one survey mode to encourage response by another mode. For example, use a postal letter to encourage cooperation when an interviewer calls to administer a telephone survey.

Type 2: Use two modes to collect responses from the same respondent. For example, to provide privacy for answering a subset of sensitive questions such as those about sexual behavior or drug use, allow respondents to an in-person interview to answer these questions using a self-administered paper or computer questionnaire.

Type 3: Use two different modes to collect responses from different people in the same survey population. For example, use a telephone survey to obtain responses from individuals who have not responded to a previously sent mail questionnaire.

Type 4: Use two different modes to obtain responses from the same person at different times. A common example is to switch from in-person interviews at time 1 to web follow-ups at time 2, as is sometimes done in longitudinal surveys.

This typology was presented in order to convey how different combinations of contact and response modes may affect costs, coverage, nonresponse, and measurement errors. Whereas Types 1, 3, and 4 are primarily focused on improving coverage and response while controlling costs, Type 2 is primarily focused on improving measurement by reducing social desirability. In addition, Types 3 and 4 have significant implications for measurement error, especially if both aural and visual modes of surveying are used. These risks are likely to be even more serious when attempting to precisely measure change over time as in Type 4.

It is now evident that the mixing of survey modes is likely to be far more complex than suggested by this simple model. Increasingly, modes are being mixed at both the contact and data collection stages. For example, we are aware of a number of surveys that use multiple modes of contact to encourage and facilitate

response in one or more modes of data collection (i.e., Type 1 used in combination with Type 3) in an attempt to maximize response and minimize nonresponse error, improve coverage, or control costs. Examples of mixing modes of contact with and without mixing response modes will be discussed repeatedly in this book.

Although we focus on mixed-mode survey designs, it is important not to ignore single-mode data collection. Often mixed-mode designs are impractical or will not necessarily improve data quality. It is possible, and sometimes most effective, to limit survey contacts and data collection to only one mode. For example, telephone-only preelection surveys will likely continue in the future because of the timeliness with which they can be conducted. In addition, many organizations (i.e., businesses, professional organizations, universities, etc.) that have accurate and complete lists of members' e-mail addresses will likely continue to conduct successful web-only surveys with e-mail contacts. Likewise, contacting households by mail and asking them to complete a paper questionnaire, which will be discussed in this book, has produced response rates and nonresponse error attributes that are as good, or better, than those that can be achieved by mixed-mode designs, and thus will likely continue to be used in the years to come.

In sum, mixed-mode design, from the most simple to the most complex, is about reducing multiple sources of error, with each way of mixing modes having different implications for each source of error. Mixed-mode designs are also justified by the desire for lower costs, achieving greater timeliness of response, and making the response task easier for the recipient of the survey request. These concerns, plus the wide variety and complexity of ways of mixing modes for contact and response, underscore the need to establish criteria for developing specific survey designs.

WHAT IS TAILORED DESIGN AND WHY IS IT NEEDED?

A key premise of this book is that in order to minimize total survey error, surveyors have to customize or tailor their survey designs to their particular situations. This can be illustrated by an experience one of us recently had in a survey design workshop. The workshop participants had just finished a lengthy discussion of topics already discussed in this chapter. One participant responded somewhat impatiently, "You have explained the problems, but you haven't told us how to solve them. The reason I am here is to find out what specific procedures and techniques I should use for my survey in order for it to be a success, whether mixed mode or not."

By asking him to describe his survey problem and then inviting others to share examples of the challenges they were facing, as well by providing additional examples that have come up in other workshops, a list of examples was produced that illustrated the diversity of challenges surveyors face. These included the following:

- An extension service entomologist wanted to survey beekeepers in his state to find out the extent to which they were experiencing winter die-off, and what they were doing to prevent it.
- A university researcher had funding to survey the general public in different parts of the United States in order to understand household water conservation practices. He explained, "I had planned to do a telephone survey

with a 20-minute questionnaire until someone told me I would get a poor response rate."

- A graduate student working on her doctoral dissertation wanted to survey rural and urban people to understand differences in the visual landscapes people preferred for the area in which they lived. "I have to use pictures," she said.
- A federal agency employee wanted to survey a nationally representative sample of home owners in order to better understand effects of the recent recession on their financial well-being.
- Another federal agency employee was concerned with how to find and survey households with children, pointing out that nearly half of the children in the United States are being raised in households without landline telephones.
- An employee of a large corporation wanted to survey consumers about a potential new product and the features they might like or dislike.
- An employee of a large cultural history museum had been asked to develop a way of surveying samples of visitors to measure their satisfaction and collect suggestions for improvement.

Our response to those seeking answers to specific situations such as these is that there is not a simple set of design procedures that if applied to every situation will be most effective in reducing survey error. The populations to be sampled and surveyed, the kinds of questions that need to be asked, the resources available for doing the survey, and other constraints imposed by survey sponsorship differ greatly across individuals and organizations who wish to do surveys. It should be apparent, even from this small list of situations, that the same procedures will not work for all surveys. But how does one go about deciding which procedures to use and not use, and by what criterion does one choose certain methods for collecting data over others? Also, under what conditions should one choose a single survey mode, and under what conditions is it better to use multiple modes?

Tailored design refers to customizing survey procedures for each survey situation based upon knowledge about the topic and sponsor of the survey, the types of people who will be asked to complete the survey, the resources available, and the time frame for reporting results. Tailored design is a strategy that can be applied in the development of all aspects of a survey to reduce total survey error to acceptable levels and motivate all types of sample members to respond within resource and time constraints.

Underlying this general approach are three fundamental considerations. First, tailored design is a scientific approach to conducting sample surveys with a focus on reducing the four sources of survey error—coverage, sampling, nonresponse, and measurement—that may undermine the quality of the information collected. Second, the tailored design method involves developing a set of survey procedures (including the recruitment contacts and the questionnaire) that interact and work together to encourage all sample members to respond to the survey. Thus, it entails giving attention to all aspects of contacting and communicating with people—few, if any, aspects of this process can be ignored when using a tailored design strategy. Finally, tailoring is about developing survey procedures that build positive social exchange and encourage response by taking into consideration elements such as survey sponsorship, the nature of the survey population and variations within it, and the content of the survey questions, among other things.

At first glance, this challenge of tailored design may hardly seem different from that faced for decades by survey researchers. However, the dizzying array of mode possibilities now available, individually and in combination with one another, and each with quite different cost and time implications, adds to the complexity of the situation. In addition, the dramatic changes occurring in the presence or absence of human interaction, trust in the legitimacy of surveys, and changes in people's control over whether and how they can be contacted make what once may have been a more simple survey design situation much more difficult. We utilize tailored design as a means of helping identify which survey procedures are effective and which ones are ineffective within each specific survey context.

We develop our tailored design approach by using an understanding of what causes people to behave in certain ways and not others. Specifically, we use a *social exchange* perspective on human behavior, which suggests that respondent behavior is motivated by the return that behavior is expected to bring, and in fact, usually does bring, from others. It assumes that the likelihood of responding to a questionnaire, and doing so accurately, is greater when the person trusts that the expected rewards for responding to a survey will outweigh the anticipated costs of responding.

Our social exchange approach underlies certain decisions made regarding coverage and sampling (e.g., obtaining sample frame and contact information), heavily influences the way we write questions and construct questionnaires, and determines how we design contacts that will produce the intended representative sample. We explain this social exchange approach in Chapter 2 and discuss how it might be applied to a wide variety of practical survey design situations.

CONCLUSION

The compelling concern that has guided revising this book is that mixed-mode surveys have shifted from being an occasional survey design issue to becoming an enduring concern for many, if not most, survey designers. Even when one decides that a single-mode survey is adequate for her survey needs, consideration of mixed-mode, mixed-device, and/or mixed-communication possibilities often precedes that decision.

Because of this substantial change in the survey landscape, in this edition we have introduced mixed mode front and center in this first chapter, and we treat it as part of the fundamental framework for this book rather than waiting to introduce it until the middle of the book, as was done in the previous edition. It has been presented here as a solution to the inadequacy of individual modes used to recruit sample members to respond and to collect responses.

The mixed-mode framework we have presented focuses the search for high-quality sample survey procedures on finding alternatives for telephone-only, web- and e-mail-only, in-person-only, and mail-only data collection designs. The nature of that approach considers traditional modes as communication mediums in addition to being potential response modes. Tailored design refers to fitting the communication and response modes to the survey topic, population characteristics, and the implementation situation one faces. Using multiple modes in a tailored design framework does not imply a one-size-fits-all approach to surveying. It means getting inside the heads of respondents, to understand what appeals to them and why, and adjusting survey procedures accordingly.

We begin that process with Chapter 2, where we answer the question of why people do and do not respond to sample surveys and provide suggestions for how to increase response rates. In Chapter 3 we focus on issues related to sampling and coverage, or finding and choosing who to survey, for each of the survey modes and for mixed-mode designs. Chapters 4, 5, and 6 are devoted to the topic of designing survey questions and questionnaires. Specifically, in Chapter 4 we cover issues common to all questionnaires; in Chapter 5 we provide guidance for designing specific types of questions; and in Chapter 6 we discuss the differences between aural and visual questionnaires and provide specific guidance for how to design for visual surveys. Chapter 7 is focused on how to order questions in the questionnaire and how to pretest them. These first seven chapters contain information that applies broadly to multiple survey modes.

We then turn to strategies for designing and implementing surveys for specific survey modes: Chapter 8 discusses telephone surveys, Chapter 9 web surveys, and Chapter 10 mail surveys. These chapters will be very useful to readers who are trying to design and carry out single mode surveys but also to those who are using these modes in mixed-mode designs. Chapter 11 then discusses designing questionnaires, contacts, and implementation strategies for mixed-mode surveys, building upon each of the individual mode chapters. Finally, in Chapter 12 we look ahead to how surveyors might respond to technological and societal changes in pursuit of conducting better sample surveys.