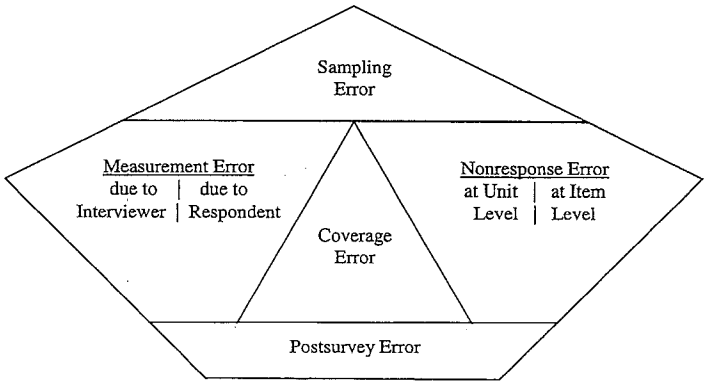
The Total Survey Error Approach

A GUIDE TO THE NEW SCIENCE OF SURVEY RESEARCH

Herbert F. Weisberg

THE UNIVERSITY OF CHICAGO PRESS • CHICAGO AND LONDON

The Total Survey Error Approach



Types of Survey Error

PART I

Survey Error Theory

1

Scientific Survey Research
THE DEVELOPMENT OF A DISCIPLINE

Around 1939 George Gallup's American Institute of Public Opinion published a pamphlet entitled *The New Science of Public Opinion Measurement.* That was a heady era in the survey world, with researchers having learned how to measure attitudes and how to sample the mass public. As the pamphlet indicated, this new science would use the "sampling referendum" (as it termed what we now call the sample survey), based on understanding the sample size required for confidence in results and on devising methods for obtaining representative cross-sections of respondents. Yet in retrospect it is clear that survey research was still in its infancy back then. Both the statistical underpinnings of survey methods and the underlying social psychology were primitive. It was appropriate for Gallup's shop to exult in the early successes of their survey approach, but much more research was needed before it would be time to call this a science. Today, more than sixty years later, it finally is time to proclaim *the new science of survey research.* In the last few decades, survey research has accumulated a solid basis in both statistical methods and social psychology. The "total survey error" approach now provides a paradigm by which to understand and study this new science.[1]

The purpose of this book is to explain the survey research process in the context of this new approach to the field. The first three chapters establish the background by reviewing the development of the survey field, outlining and extending the total survey error approach, and providing a common understanding of the different survey modes. The middle of the book follows the total survey error approach's emphasis on the different sources

of survey errors, discussing the different elements of surveys through a focus on the sources of error. The final chapters turn to broader and summary issues.

The notion of a scientific survey will be explained carefully in chapter 2, but it is useful to give a quick definition here. Surveys have people answer questions so as to develop quantitative descriptions; usually a sample of people are interviewed, with the intention of generalizing the findings to a larger population.

The Development of Modern Survey Research

It is appropriate to begin with a brief recounting of the development of survey research as a scientific field. This section reviews the early institutionalization of survey research, later changes due to evolving survey technology, and the recent professionalization of the field.

The Early Institutionalization of Survey Research

Polling can be dated back to censuses taken several millennia ago and to straw polls in early U.S. elections, but surveys did not become commonplace until the twentieth century. Jean Converse (1987) identifies three more proximate sets of ancestors: social reformers, science, and business. Social reformers around 1900 sought to document the extent of poverty in cities in England (Booth 1889; Rowntree 1901) and the United States (Residents of Hull House 1895; DuBois 1899) through "social surveys" (Bulmer, Bales, and Sklar 1991) and to detail living patterns in American rural areas through the Country Life movement (Bailey 1906; Bowers 1974). Sampling techniques were also developed during the first half of the twentieth century, particularly for agricultural research. A second independent source was the world of social science, resulting from the work in the 1920s on attitude measurement in psychology and sociology (Thurstone 1928). This led to some early academic political surveys (e.g., Gosnell 1927; Rice 1928).

Converse argues, however, that the third source was the most direct ancestor of modern survey research: the world of business as represented by market research and opinion polling in journalism. Consumer research developed in the decade of the 1910s in several large industries. Such researchers as George Gallup, Elmo Roper, and Archibald Crossley pioneered in straw-vote journalism with their correct prediction that Franklin Delano Roosevelt would win reelection in 1936.[2] This led to innovation in sampling procedures, attention to interviewing techniques, and developing formats for questions. By the 1935–40 period, linkages were being established between those doing applied research for business and government and academic survey researchers. This was typified by the founding of the *Public Opinion Quarterly* journal in 1937 and the establishment of survey organizations in the government and at universities. Converse documents the early history of three influential survey organizations: Rensis Likert's Division of Program Surveys in the Department of Agriculture; Hadley Cantril's Office of Public Opinion Research at Princeton University, where Gallup was headquartered; and Paul Lazarsfeld's Office of Radio Research (to develop measures of listenership so that radio stations could charge for ads) at Columbia University, which became the Bureau of Applied Social Research.

What is notable is how many of the basic understandings that underlie survey research were in place by 1940. It was recognized that people could be studied and could give useful information and that different people would give different information so that it was important to study multiple people. Statisticians realized that they could generalize from a sample without studying the entire population, and they had developed probability sampling methods to permit generalizations (Wright 2001). Psychologists learned that attitudes could be measured. Panel surveys in which the same respondents were reinterviewed over time were devised. Each of these understandings is taken for granted today, but each was an important advance for survey research.

But the events that were most important for the development of the survey field occurred just after this time. Many social scientists went to Washington, D.C., during World War II to work for the federal government in the war effort. That led to significant development of polling technology, as evidenced in the publication of the major work *The American Soldier* (Stouffer et al. 1949), based on surveys of soldiers to deal with wartime morale and issues that might affect their return to civilian society. After the war, the social scientists themselves went back to academia, often in teams, such as Likert's Department of Agriculture team, which went to Ann Arbor to found their Institute of Social Research (ISR) with its important Survey Research Center. Lazarsfeld's Bureau of Applied Research at Columbia University did several single community surveys in the 1940s as well as contributing to the intellectual understanding of survey research and data analysis. The National Opinion Research Center (NORC), a group that had been founded just before the war at the University of Denver, moved to the University of Chicago in 1947 and did national face-to-face interviews for social scientists. Thus by 1950 several large

academic survey organizations were in place, and on the commercial side of polling, George Gallup and Elmo Roper, among others, had their own national survey operations.

The late 1940s was a period of building the institutions of survey research. This is best exemplified by the founding of a professional organization, the American Association for Public Opinion Research (AAPOR), in 1947 after an initial meeting in 1946, with *Public Opinion Quarterly* becoming its official journal in 1948. The *International Journal of Opinion and Attitude Research* was also started in 1947. The Survey Research Center at the University of Michigan began its summer training program in survey research methods in 1948, attracting students from across the United States and around the world. In the 1950s, textbooks on survey research methods were written (Parten 1950; Hyman 1955), the range of topics studied through surveys was expanded, and survey research spread to more countries.

The federal government also had a major involvement in the development of survey research through its Census Bureau. Originally commissioned to conduct only a single census every ten years, the Census Bureau became increasingly involved in the conduct of surveys and the study of survey accuracy in order to assess the precision of its own census efforts. Additionally, the federal government's need to generate statistics about such diverse topics as employment, crime, and health as well as a desire to be able to evaluate the results of government programs led to the development of excellent survey research capabilities in other federal government agencies. Indeed, some of the best work on methodological issues involving surveys is still done in the federal statistical agencies, including the Bureau of the Census, the Bureau of Labor Statistics, and the National Center for Health Statistics.

Both fact-based and attitude surveys became common by the 1950s. Fact-based surveys, which include most surveys undertaken directly by the federal government, generally focus on past behavioral experiences, such as asking people about their medical history or asking businesses about their sales in the past month. Attitude surveys ask people how they feel toward particular objects, from political candidates in election surveys to household products in marketing research. Attitudes themselves are the positive or negative orientations toward these objects. Along with attitudes, these surveys typically measure preferences between objects, based on comparisons of attitudes toward objects, such as whether the person prefers the Republican or Democratic candidate or prefers one brand of toothpaste or another. Attitude surveys also often include questions about beliefs—the respondents' opinions about the objective state of the world. These are measured in attitude surveys because people's attitudes will depend on what they think is true of the world. Thus public views about the health of the economy are likely to affect voting plans as well as purchasing plans.

The first surveys were used mainly for simple reporting of the results of individual questions in order to assess the frequency of particular facts or attitudes, often with comparisons of different subgroups of the population. By the 1960s repeated surveys became common, so that trends over time in both facts and attitudes could be traced. Examples of large-scale continuing surveys include the National Election Studies in political science, the General Social Survey in sociology, and the National Longitudinal Study of Labor Market Experience. Social science researchers increasingly used these surveys to construct causal explanations of behaviors and attitudes by examining the relationships between those dependent variables and predictor variables that are appropriate in the relevant fields of study.

Looking back at this period, Jean Converse (1987) argues that the early survey researchers made significant contributions in addition to founding survey organizations. She calls particular attention to the development of theory, especially attitude theory, to the creation of exemplars of quality research, and to progress in methods, including sampling, interviewing, and question wording. Yet she concludes that even by 1960 survey research was not itself a discipline: "while it was no garden-variety tool, it was still something less than a discipline" (390–91).

The Impact of Evolving Survey Technology

While the survey research field had begun to broaden its scope by the 1960s, it growth was being limited by technology. The understanding at the time was that quality surveys could be obtained only through face-to-face interviewing in dwelling units, and the expense of that mode of polling restricted its expansion potential.

By the mid-1970s the field was rethinking the optimal way to conduct surveys. Whereas telephone surveys were originally frowned upon because many households did not have phones, telephones had finally become nearly universal in the United States by the 1970s, with a coverage rate nationally of about 90%. Furthermore, the costs of telephoning across the nation began to fall around the same time that the gasoline crisis of that era increased the transportation costs involved in personal interviewing. Procedures for sampling telephone numbers were developed

and refined. The shift from room-size computers to desktop computers in the 1980s made it possible for even small telephone survey shops to own their own computers to process survey results. By the 1980s telephone interviewing had become the predominant mode of conducting surveys in the United States.

The shift to telephone technology permitted widespread growth in the field of survey research. It became possible to conduct a national survey from just a small room equipped with a dozen telephones. The simultaneous development of computer-assisted telephone-interviewing (CATI) technology further facilitated this switch and provided the added benefit of being able to generate results virtually instantaneously. Led by Berkeley's Survey Research Center, most academic survey centers and the established commercial pollsters switched to phone interviewing. The institutionalization of the survey field continued as more universities established survey research centers, additional commercial polling companies entered the field, and several important research firms were founded in the Washington, D.C., area to consult for federal agencies that used survey data. The survey research industry blossomed and experienced robust growth after the nation emerged from the recessions of the early 1990s, with the federal and state governments commissioning many more surveys.

Further technological changes occurred in the 1990s, continuing into the first decade of the 2000s. Just as telephone survey methods were being refined, changes in telephone usage patterns began to make phone interviews more problematic. The chances of getting through to a residential phone when calling a phone number decreased with the popularity of cellular phones and with people putting answering machines, fax machines, and computer modems on their telephone lines. Thus it may be that the high point of telephone surveys is already in the past. Meanwhile, the development of the Internet permits new ways of contacting respondents through e-mail and through Web surveys. The different modes of survey research will be discussed in more detail in chapters 3 and 12.

Professionalization of the Field

The survey field has professionalized as it has institutionalized and evolved. One aspect of its professionalization has been the development of professional standards. This includes disclosure standards (such as what information should be provided in press releases on surveys, as discussed in chapter 13), technical standards (particularly standard ways of computing response rates, as will be described in chapter 8), and ethics codes that are incorporated in statements of professional conduct (see chapter 14).

Several professional associations have been formed. AAPOR brings together survey researchers from academia, government, and commercial firms. The Survey Research Methods Section of the American Statistical Association initially focused on sampling methods but subsequently broadened to cover the full range of survey-related methods. The Council of American Survey Research Organizations (CASRO) is a trade association for commercial survey firms, the Council for Marketing and Opinion Research (CMOR) emphasizes market research, and the National Council of Public Polls (NCPP) focuses on media reporting of polls.

Another aspect of professionalization is that survey research has become truly international. The World Association for Public Opinion Research (WAPOR) has been in existence since the late 1940s, while the International Association for Survey Statisticians was formed in 1973. Excellent research on the survey process is being conducted at universities and government statistical agencies in a wide variety of countries. There are frequent international conferences on various aspects of survey research, yielding major compendia of research reports, such as *Survey Nonresponse* (Groves et al. 2002), and *Cross-Cultural Survey Methods* (Harkness, van de Vijver, and Mohler 2003). The internationalization of the field is reflected in the articles published in journals such as *Public Opinion Quarterly*, *Survey Methodology* (published by Statistics Canada), the *Journal of Official Statistics* (published by Statistics Sweden), the *Survey Statistician* (published by the International Association for Survey Statisticians), and the *International Journal of Public Opinion Research* (published in England for WAPOR).

The professionalization of survey research is also evident in the establishment in the 1990s of several university-based degree programs. The Joint Program in Survey Methodology of the Universities of Michigan and Maryland in conjunction with Westat is a major training program for federal government employees in the Washington, D.C., area. Additionally, certification, master's, and/or doctoral programs in survey research were instituted at the Universities of Cincinnati, Connecticut, Maryland, Michigan, Nebraska, and Ohio State University. If survey research was not yet a separate discipline in 1960, it was approaching that status by 2000.

One distinguishing feature of the survey research field is that it is inherently multidisciplinary. It is a social science field since it deals with human behavior, but it is based on sampling procedures developed in

statistics. Many of the most important insights come from psychology, as it is essential to understand how people react to survey questions as well as how people form their attitudes. The scientific study of communication is also relevant since interviews are a communication process. Many important developments in the field also come from political science because of the importance of election polls and from market research because of extensive polling of consumers.

Approaches to Survey Research

As the institutionalization, technology, and professionalization of survey research have evolved, approaches to the field have also changed. The original approach to the field was based on the lore that accumulated from early surveys, assembling the experiences of survey researchers into lists of recommended do's and don'ts. The procedures that seemed to work best became the accepted practices in the field. Some procedures were based on the social psychology of the time, such as structuring the relationship between a survey interviewer and respondent on the basis of the relationship between a psychological therapist and patient, but most of this early work was atheoretical.

The survey field gradually began to emphasize *standardization*. The more each interview proceeded in the same way, the more scientific the polling seemed to be. Interviewers were instructed to proceed in a standardized fashion. The best questions were thought to be those that were the standards in the field. This approach was still atheoretical, but it seemed more systematic than just codifying experiences of survey researchers.

The survey research field underwent *scientization* in the 1990s, applying theories from other fields to the survey research field. In particular, the cognitive revolution in psychology has led to new understanding of the survey response process (see chapter 5), and developments in statistics have provided new imputation techniques for dealing with missing data (chapter 7).

As part of this scientization, a new approach to survey research began to predominate by the 1990s: the "total survey error approach." It was always recognized that there are several types of error involved in a survey, but the greatest emphasis was on the type that is easiest to estimate statistically: sampling error—the error that occurs in a survey when one surveys a sample of the population rather than the entire

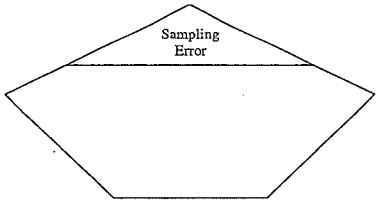


Figure 1.1 Sampling error as the tip of the iceberg

population. However, sampling error is just the tip of the iceberg in surveys (see fig. 1.1).

Early depictions often represented the error in surveys using a right triangle, with sampling error as one leg of the triangle, nonsampling error as the other leg, and the hypotenuse as the total survey error (fig. 1.2). This representation reminded people that the total survey error is much greater than just sampling error.[3] In practice, however, the emphasis in the field was mainly on sampling error because there are elegant mathematical formulas to estimate its magnitude, while nonsampling error was seen as too difficult to estimate. Furthermore, sampling error could be lowered if the sample size was increased, whereas the other sources of error were seen as more difficult to minimize.

The "total survey error" phrase first appeared in a book title involving health surveys (Andersen, Kasper, Frankel, and associates 1979). Applying ideas developed by Hansen, Hurwitz, and Madow (1953) and Kish (1965), the approach took off with Robert Groves's book *Survey Errors and Survey Costs* (1989). The Groves book was the first systematic treatment of all the different considerations involved in this approach, detailing the several sources of survey error and analyzing the costs involved in minimizing each. The total survey error approach has subsequently dominated research in the field, becoming the prevailing paradigm for understanding surveys. Chapter 2 will elaborate on this approach, including extensions of its usual formulation.

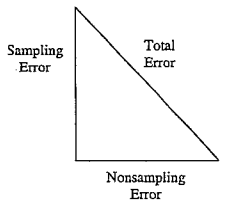


Figure 1.2 Sampling error as one component of total error

It must be admitted, however, that the recent explosion of research on surveys has led to reappraisals of several longstanding assumptions. As an example, a review by Jon Krosnick (1999) challenges much of the conventional wisdom about survey methodology. Additionally, there are some other frameworks that have been proposed for surveys, particularly Dillman's (2000) "tailored design method" for non–interviewer assisted surveys such as mail and Internet surveys and self-administered questionnaires.[4] Thus the move to a new paradigm is not fully smooth, reminding us that paradigms in science are fragile and the development of a paradigm is not necessarily decisive.

Survey Limitations

Survey research has become an important means of data collection that is useful for assessing the frequency of attitudes and human behaviors as well as measuring changes in those attitudes and behaviors over time. The underlying assumption is that people will respond candidly, sharing their views and experiences with researchers. Survey research will be important as long as governments, businesses, and social scientists value these measures. Yet as with any research approach, it is also necessary to recognize the limitations of surveys. There are several sources of error in surveys to consider, as will be detailed throughout this book. Furthermore, there are cost, time, and ethical issues that impinge on survey research (see chapter 2).

The most fundamental challenge would be widespread public disillusionment with surveys, leading to diminished cooperation and lack of faith in the results. George Gallup justified public opinion polling by emphasizing its usefulness in a democracy, allowing officials to know what the public is thinking. However, as will be seen in chapter 8, the rate of cooperation with surveys has already decreased considerably. There always has been some public suspicion of surveys, especially when results differ from the person's own views. Surveying becomes even more problematic when public figures encourage people to lie to pollsters on their vote intentions, though few real people are likely to go to that extreme. An even more negative argument against polls would be that the public does not know enough and is not intelligent enough for its views to be given credence. Gallup instead accepted Charles-Maurice de Talleyrand's assertion that "the only thing wiser than anybody is everybody," because democracy requires acceptance of the primacy of the public.

These limitations do not diminish as survey research becomes more scientific. Indeed, many people are skeptical of claims of the scientific status of surveys. Still, it is important for survey researchers to keep developing its scientific basis. The development of the total survey error approach is best seen as a stepping-stone in the continuing progress of the field.

2

Survey Error

THE ESTABLISHMENT OF A PARADIGM

The total survey error approach is based on analyzing the several different sources of error in surveys and considering how to minimize them in the context of such practical constraints as available money. This chapter introduces the total survey error approach by summarizing the different sources of error that will be treated in more detail in later chapters and by explaining the statistical implications of error. Additionally, this chapter extends previous statements of the total survey error approach by generalizing the notion of survey constraints to include time and ethics considerations and by recognizing the limitations imposed by survey-related effects.

The Theory of Survey Error

The total survey error approach is based on the fact that error can occur at every stage of a survey. Table 2.1 outlines these stages and indicates the chapter in this book in which each is discussed. A preliminary step is to decide on one's research objectives and whether a survey is the best way to meet them. As part of that decision, it is necessary to select a target population, such as residents of a particular city or senior citizens across the country.

The next step is to decide how to administer the survey: whether to do a face-to-face survey, a phone survey, a mail survey, an Internet survey, or some other type. This is known as the survey *mode*. Along with that decision is the choice of a survey *design*: whether it is a one-time *cross-section*

Table 2.1 Steps of the Survey Process

Stage	Chapter
Decide research objectives	
Determine target population	
Choose survey mode and design	3, 12
Choose sampling frame	9
Select sampling method	10
Write questions	5
Pretest questionnaire	6
Recruit respondents	8
Ask questions	4
Process data	11
Analyze results	

survey or a *longitudinal* survey that involves repeated interviews of the same people to look at changes in their attitudes and/or behavior over time.

Then comes deciding on how to sample respondents. Part of this decision involves moving from the target population to a realistic *sampling frame*—the set of units from which the sample will be drawn. A sampling frame for senior citizens, for example, might consist of single- and multiple-family dwelling units but would necessarily exclude hospitals, thereby missing part of the target population. Another part of this decision is the choice of sampling method—whether people will be selected randomly or some other sampling technique will be used.

The questions to be asked must be written and assembled into a questionnaire (sometimes termed the *survey instrument*). It is best to "pretest" the questionnaire by trying it out on some people before the actual interviewing so as to make sure that it works as intended.

The actual interviewing comes next. That includes recruiting actual respondents and asking them the questions. Afterward, the data must be processed and the results analyzed.

While these steps are listed in order in Table 2.1, the process is more fluid than has been described so far. Often, for example, researchers decide on the questions they want to ask people before deciding on a sampling method. But the basic steps are the same, regardless of their order.

Types of Error

While sample surveys draw responses from individual people, the analysis is usually based on aggregating the results to obtain statistics for the sample, statistics that hopefully estimate the corresponding population values

(known as population parameters). The survey may be designed to measure some abstract constructs (such as abortion attitudes). It does so by a process of measurement, asking specific questions intended to measure the construct. The ultimate intention is to infer population values from the sample results.

The term *error* is usually thought of as a synonym for *mistake*, but in the survey context it refers to the difference between an obtained values and the true value, usually the true value for the larger population of interest. Several sources of error are distinguished in the total survey error approach.

The first three types of error are related to selecting respondents for a survey. The best known is *sampling error*. As stated in chapter 1, sampling error is the error that occurs when a sample of the population rather than the entire population is surveyed. There can be systematic bias when nonprobability sampling is employed, whereas probability sampling has the advantage of permitting mathematical computation of the sampling error.

Another type of respondent-selection error is *coverage error*, the error that occurs when the list from which the sample is taken (known as the *sampling frame*) does not correspond to the population of interest. This is the type of error that concerned researchers before the 1970s, when they avoided phone surveys: telephone coverage was not universal, so samples obtained through phoning people would not have covered the full adult population.

The other category of error related to respondent selection is *unit nonresponse error*. This error occurs when the designated respondent does not participate in the survey, thereby limiting how representative the actual respondents are of the population of interest. This form of nonresponse error calls attention to a key question: Are high response rates essential in surveys?

Three additional types of survey error involve the accuracy of responses. First is *item nonresponse error*—the error that occurs when the respondent participates but skips some questions. "Don't know" responses are an important item nonresponse problem, making it necessary for researchers to decide how to deal with missing item data in the data analysis.

Response accuracy issues also arise because of *measurement error*, the error that occurs when the measure obtained is not an accurate measure of what was to be measured. *Measurement error due to the respondent* involves whether the respondent gave an accurate answer to the question, which is often really a matter of how well the researcher worded the survey question. *Measurement error due to the interviewer* occurs when effects associated with the interviewers lead to inaccurate measures.

The final type of error is *postsurvey error*, the error that occurs in processing and analyzing survey data. Some accounts of the total survey error approach include postsurvey error, though others do not since postsurvey error is really not error from the survey process itself.

Another survey administration issue has to do with *mode effects*, the effects that the choice between face-to-face interviewing, phone surveys, mail questionnaires, and other survey modes has on the results that are obtained. The final survey administration issue involves *comparability effects*, the differences between survey results obtained by different survey organizations or in different nations or at different points in time.

Figure 2.1 takes the iceberg depiction of survey error and breaks it into three tiers in accord with the above discussion. Sampling error is still shown as the tip of the iceberg, with other respondent-selection issues also near the top of the iceberg. Response accuracy issues are further below the surface. Survey administration issues are shown in the bottom tier, detachable from the rest, as postsurvey error, mode effects, and house effects are usually not included in discussions of survey error. The relative sizes of the different areas in figure 2.1 should not be taken literally, though it is true that two of the larger areas, those for sampling error and measurement error due to respondents, both depict major topics.

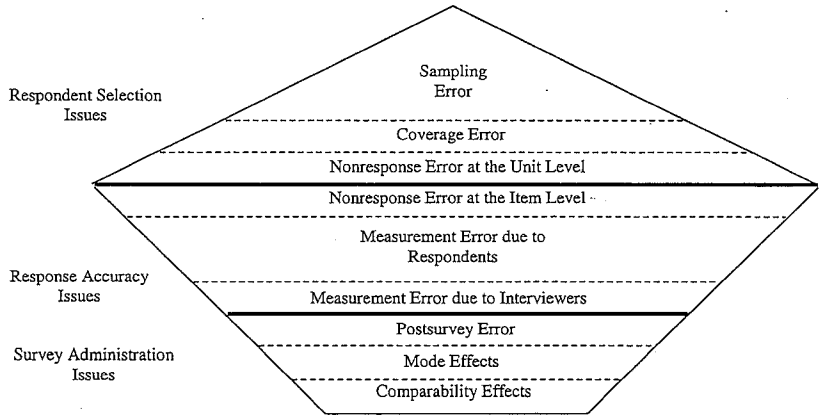


Figure 2.1 Types of survey error

The chapters that follow discuss these several potential sources of error in detail. These are called *potential sources of error* because they are not necessarily errors. For example, survey nonresponse does not lead to error if the people who do not participate in the survey are similar to those who do participate. Yet we generally cannot tell whether they are similar, since we lack data on the people who do not participate; this makes it important to understand nonparticipation as a potential source of error.

Scientific Sample Surveys

Since this book is dealing with surveys, it is appropriate to consider what exactly is a scientific survey. The American Association for Public Opinion Research (AAPOR) attempted in 2001–2 to define a "scientific sample survey" as a means of helping the media and, through them, the public to distinguish between high-quality surveys and ones that should not be taken seriously. As could be expected, there was not full agreement on the initial draft, so in the end AAPOR did not adopt an official definition. Still, it is useful to examine how the AAPOR committee of distinguished survey researchers initially attempted to define the term (AAPOR 2001, 476–77).

First, they emphasized that surveys and polls provide a means to study people's "attitudes, behaviors, and characteristics." Their focus was on sample surveys, not censuses or other surveys that interview the entire population of interest.

Their list of four characteristics of a scientific sample survey or poll emphasizes the points listed in the previous section: coverage, sampling, nonresponse, and measurement.

1. Coverage: A scientific survey "samples members of the defined population in a way such that each member has a known nonzero probability of selection. Unless this criterion is adhered to, there exists no scientific basis for attempting to generalize results beyond those individuals who completed the survey."
2. Sampling: A scientific survey "collects data from a sufficient number of sampled units in the population to allow conclusions to be drawn about the prevalence of the characteristic in the entire study population with desired precision (for example, + or – 5%) at a stated level of confidence (e.g. 95%)."
3. Nonresponse: A scientific survey "uses reasonable tested methods to reduce and account for unit and item nonresponse error (differences between characteristics of respondents and nonrespondents) by employing appropriate procedures for increasing unit and item response rates and/or making appropriate statistical adjustments."
4. Measurement: A scientific survey "uses reasonable tested methods to reduce and account for errors of measurement that may arise from question wording, the order of questions and categories, the behavior of interviewers and of respondents, data entry, and the mode of administration of the survey."

At the same time, the AAPOR preliminary statement admits that "in practice all sample surveys fall short of perfection in meeting one or more of the above criteria." Because of that lack of perfection, its definition emphasizes the importance of "disclosure of the exact procedures by which a survey is conducted, including the sample size and sampling methods, coverage characteristics, the questionnaire itself, data collection methods, response rates, and information about the characteristics of respondents as well as nonrespondents, [since this] makes it possible for others to evaluate likely survey errors and reach an informed judgment on the resultant quality of the findings."

Finally, the AAPOR committee gave a few examples of what leads to a survey's "failing to meet the test of being scientific." Letting volunteers fill out a questionnaire rather than using random selection is nonscientific, for generalization beyond the people who filled out the questionnaire is not warranted. A low response rate can make the survey unscientific: "In some surveys, a relatively small percentage of people selected in a sample actually complete a questionnaire, and the respondents differ significantly from non-respondents on a characteristic of relevance to the survey. If this fact is ignored in the reporting of results, the survey fails to meet an important criterion of being scientific." The final example is that a survey is not scientific if the question's wording produces biased answers.

This definition of a scientific sample survey has some ambiguities, which shows the complexity of defining a scientific survey. Take the matter of response rates. The nonresponse paragraph seems to indicate that a survey is scientific so long as it tries to minimize and/or adjust for nonresponse, but the later paragraph on nonscientific surveys seems to say that nonresponse can invalidate a survey. This inconsistency may properly reflect the nature of response rate as a problem—it can render a survey nonscientific but does not necessarily do so—but it also shows that there always will be ambiguities in working through what makes a survey scientific.

The Statistical Impact of Error

While it would be best if all of the types of error described above could be eliminated, that is not possible. The goals instead are to keep them at minimal levels and to measure them so that their statistical impact can be assessed. Survey error affects the statistical analysis of survey data. Statisticians make two distinctions between types of error that have important implications for statistical analysis.

The first key distinction is between systematic and random error. *Systematic error* is usually associated with a patterned error in the measurement. This patterned error is also known as *bias*. An example of systematic bias would be measuring the average age from a survey if older people were less willing to be interviewed. Another example is when people who are dissatisfied with a product are more likely to choose to participate in a customer satisfaction survey, rendering estimates of the extent of dissatisfaction wrong. Thus systematic error directly affects the average value of the variable (its *mean*), depressing it in these two examples. In statistical terminology, the sample mean is a biased estimator of the population mean in the presence of systematic error.[1] The *validity* of a measure is whether it measures the construct of interest. Variables that are measured with systematic bias are not valid. If the systematic error is just a constant, the variance of the variable will still be correct. Also, adding a constant to a variable does not affect correlations or regression coefficients, so those statistics can still be correct, depending on the nature of the systematic error.

By contrast, *random error* is error that is not systematic. For example, if people do not think hard enough about a question to answer it correctly, some may err on one side and others may err in the opposite direction, but without a systematic tendency in either direction. Random error would have a mean of zero, so random error does not affect the mean of a variable. However, random error does increase the variance of a variable.[2] The *reliability* of a measure has to do with the amount of random error it contains. The reliability of a measure can be thought of as the extent to which repeated applications would obtain the same result. An unreliable measure has a large amount of response variance. Random measurement error reduces the reliability of a measure, as when asking a question in different tones of voice yields different answers. Because of the increased variance, random error directly affects correlations with other variables, which are reduced in magnitude (an effect known as *attenuation*). Regression coefficients are similarly attenuated by random error in the independent variable, and hypothesis tests have less statistical power (making it harder to obtain significant results) than if measures were perfectly reliable.

One way to put systematic and random error together is to think in terms of the *mean squared error* on a variable—the average of the squared differences between survey respondents' values on the variable and the true population mean. The mean squared error is the sum of the squared bias and the observed variance, respectively reflecting validity and reliability concerns (Biemer and Lyberg 2003, 56). This formulation is of only theoretical interest, since of course the true population mean is unknown, but it usefully shows how validity and reliability concerns fit together in survey measurement.

A second key distinction is between uncorrelated and correlated measurement error. The discussion of systematic and random error above was based on *uncorrelated error*. Uncorrelated error occurs when the errors for different individuals are unrelated, as when an interviewer happens to ask a question in one tone one time and in a different tone of voice another time, with no particular pattern as to when each tone is used. This sounds benign, but, as seen in the above paragraphs, even uncorrelated random measurement error does affect statistics. One solution (Fuller 1991) would be to reinterview some respondents to obtain estimates of the reliability of measures in order to adjust the statistics, but researchers generally would prefer spending money to interview more respondents rather than to take more interviews with the same respondents.

Correlated error occurs when the errors for different respondents are related. If an interviewer systematically asks questions in a particular wrong way, for example, all of the interviews that person conducts will be slanted in the same way. Correlated error can occur when interviewers take multiple interviews and when coders code open-ended questions (those that respondents answer in their own words, rather than choosing between response categories that are part of the question) in multiple interviews. As shown in table 2.2, correlated error is more serious than uncorrelated error because it directly multiplies the variance of the variable, which makes it harder to get significant results. The mathematics of correlated error will be discussed in chapter 4 as part of the treatment of measurement error due to interviewers.

Survey-Related Effects

While the standard total survey error approach naturally emphasizes survey errors, it is important to extend this approach to take explicit account

Table 2.2 The Effects of Error on Variance and Estimates

	Random Error	Systematic Error
Uncorrelated Errors	Unbiased estimates; increased random variance	Biased estimates
Correlated Errors	Unbiased estimates; variance multiplied	Biased estimates

of *survey-related effects*. These effects are recognized in the current literature, but they have not been previously incorporated into the total survey error approach. They should, however, be considered explicitly, because they limit the conclusions that can be drawn from survey evidence even if they do not constitute errors.

As an example, one set of survey-related effects is question-related. No single question wording is "correct," but, as will be seen in chapter 5, different question wordings can yield somewhat different results. Question order is known to affect results, but question order effects cannot be eliminated, since it is not possible to have every question come first in a survey. Similarly the order in which response options are offered affects answers; the *recency effect*, for example, occurs when phone respondents tend to choose the last answer on a list because they remember it better than those read to them earlier. What distinguishes these various effects is that there is no way to minimize them. Different ways to structure the question yield somewhat different results, but that does not mean that a particular question wording is more valid than any other wording. One way to think about this is that attitudes are always context dependent, so it would be chimerical to try to eliminate context. It is better to conceptualize true answers as a range and then view alternative question wordings (including question orders and response-option orderings) as permitting researchers to examine that range. According to this perspective, it is better to experiment with multiple question wordings than to seek a single optimal wording.

As already mentioned above, there are also mode effects associated with the choice between face-to-face, phone, mail, Internet, and other modes of administering surveys, and there are comparability effects involved with comparisons between surveys. Similarly, the term "interviewer effects" is sometimes used rather than "interviewer error," since some of the effects associated with interviewers are not mistakes but are simply the inevitable results of interviewer gender, race, or tone of voice—factors that cannot be totally eliminated so long as interviewers are used. Again, such effects cannot be eliminated; instead it is important to understand such effects and study how much they can affect survey results.

The statistical effects of survey-related effects vary considerably. They can have either random or systematic effects and can be uncorrelated or correlated. Fortunately, it is possible to study these effects, such as by testing the effects of different question orderings or different response-option orders.

Survey Error and Survey Constraints

Survey researchers want to minimize error in their work, but the degree to which error can be reduced is limited by practical constraints. Groves's (1989) classical exposition of the total survey error approach focuses on the balance between survey errors and one important constraint: costs. However, it is important to generalize this treatment by recognizing that time requirements and ethics also impose important constraints.

The *cost* side is relevant in every research endeavor, but survey research involves particularly large costs, especially when face-to-face or telephone interviewing is used. Some survey costs are fixed, but most are variable, depending on the number of interviews and the average length of the interviews. Survey organizations typically generate cost proposals to potential clients based on those considerations, taking the number of interviews and average interview length into account and adding a term to cover fixed costs. For example, if it is estimated that the interviewer will, on average, complete one interview per hour for a study, and if 1,200 interviews are to be taken, that multiplies out to 1,200 interview hours. If the cost per interview hour were $15 (covering administrative costs as well as interviewer wages), variable costs would multiply out to $18,000. Additionally, the fixed costs associated with taking on the project must be added on to cover the development of the questionnaire, programming it if the survey is computer assisted, and data management. Survey cost estimates also take into account the number of callbacks requested, the amount of screening required (screening is expensive since it means more calls are required to obtain a completed interview), and how complicated the sampling is.

In practice, the cost side is more complicated because of the competitive nature of the survey field. Since researchers always want to get the most they can with their limited funds, they typically will seek bids from different survey organizations. As a result, survey organizations must price their product at a competitive level to win survey contracts. Of course, there are also nonprice bases for such competition, as when survey organizations emphasize the quality of their research and argue that

high-quality research is more expensive. For example, it is common for university-based telephone survey operations to justify their charges on the basis of making many callback phone calls to find an intended respondent at home, rather than just going on to the next phone number as some commercial survey firms do.

Time is also an important consideration in survey research. Surveys are often conducted on a tight time schedule. Sometimes this is due to external constraints that make the information less valuable if it is not available by a particular time. Preelection polling, for example, must by definition end before the election actually occurs. Also, some attitudes and facts will change over time, so it is necessary to measure them in a timely fashion. When commercial airlines hire polling operations to phone recent customers and ask about their satisfaction, that information must be gathered soon enough after a flight that the passengers remember it and before they have been on subsequent flights. However, there are also drawbacks to instantaneous polling. Some survey organizations conduct tracking surveys every evening to look for changes in attitudes, but that research design limits the results since it overrepresents the types of people who are at home on particular evenings.

A third relevant constraint on survey research involves *ethics*. As explained in chapter 14, there are a number of ethical guidelines for research that affect surveys, and it is important that researchers think through some of these ethical issues before conducting surveys. Indeed, preclearance is necessary to conduct human subjects research at universities and other organizations that receive funds from the federal government. Ethical considerations can be considered to be a constraint on research, as the rules restrict certain types of surveys. Surveys on children, for example, are tightly regulated. The review boards that consider proposals for human subjects research may not approve questions that would cause stress for respondents. Having to justify these types of questions to a review panel has the side benefit of discouraging such questions unless they are essential to the survey. Thus ethical considerations can constrain survey activities just as survey costs and time needs do.

The Total Survey Error Approach

The total survey error approach is often discussed in terms of a trade-off between survey errors and costs. However, a broader statement of this approach emphasizes the several different types of survey error, inevitable survey-related effects, and the full set of cost, time, and ethical considerations that constrain research. Research always involves compromises, and here the compromise is between trying to minimize survey error and recognizing that survey constraints may require accepting more error, while also taking into account the impact of survey-related effects.

This is not meant as an argument for tolerating large amounts of error in surveys but rather as an argument for finding the best survey design within the researcher's financial and time constraints as well as applicable ethical factors. For example, many specialists would view face-to-face interviews as having less error than most other types of surveys because of their higher response rates and the greater rapport that interviewers can achieve with respondents, but they are also much more costly and take much longer to complete. At the opposite end of the continuum, Internet polls can be fairly inexpensive and can give fast results, even if they have lower response rates and interviewers cannot probe "don't know" responses. What is important is to understand the trade-offs involved in the choices, since it is not possible to maximize all benefits at once.

The total survey error approach is used in many ways in the survey research field. Some survey organizations use it to structure their discussions with clients, making them aware of the different sources of error and working with them to achieve a research design that deals with those error sources within their budget. A survey can be thought of as no better than its weakest link (Fowler 2002, 8), which makes it important that clients understand the extent to which each source of error affects their survey. The natural inclination is to maximize the sample size and/or the response rate, but careful consideration of the different sources of error may suggest that it is more important to take steps to work on coverage or other errors. There are inevitably trade-offs in research, and the total survey error approach helps make those trade-offs more explicit.

Some survey organizations also use the total survey error approach in training interviewers and their supervisors, making sure that all employees understand how their performance affects the quality of the product delivered to clients. The total survey error approach is relevant after the data are collected as well. In particular, reports on the survey should discuss the relevant sources of errors, detailing the decisions in the data collection stage that affect interpretation of the results.

Additionally, the total survey error approach is used to structure some academic training programs in the field, including the Joint Program in Survey Methodology (JPSM) of the University of Maryland and the University of Michigan in collaboration with Westat.

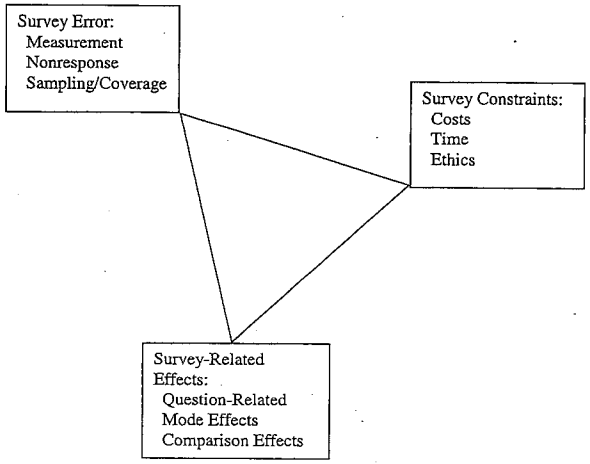


Figure 2.2 The survey research triangle: Survey error, constraints, and effects

There is, of course, never a perfect paradigm for understanding everything in a field, and so there are other perspectives that can be used in approaching survey research as a field of study. However, the total survey error approach is certainly the dominant approach today, and part of its advantage is that some alternative approaches can be treated as part of its perspective. For example, one recent aspect of the scientific approach to surveys has been the application of increased understanding of the human thought process (the so-called cognitive revolution in psychology) to understanding the survey research process. Although this could be considered an alternative perspective, it can also be subsumed under the total survey error approach, as part of increasing our understanding of the errors in the measurement process associated with the respondent.

Finally, the total survey error approach provides a useful way to structure this book. It reminds us of the importance of identifying, discussing, measuring, and seeking ways to minimize all the factors that lead to errors in surveys. It emphasizes the need to balance error minimization against survey costs, time, and ethical considerations. Additionally, it recognizes that there are a number of survey-related effects that will always affect survey results. Figure 2.2 summarizes the main elements of these three factors, while implying the difficulty of balancing between them. Chapters 4 through 14 will focus on the various sources of survey error, survey constraints, and some survey effects.

3

Survey Modes

RESPONSES TO EMERGING TECHNOLOGIES

There are several different modes through which surveys can be administered: face-to-face, telephone, mail, Internet, self-administered, and so on. Before the sources of survey error are examined in detail, it is useful to give preliminary consideration to these different modes of survey administration. Two basic distinctions between survey modes will be developed in this chapter, after which some related research approaches are discussed. Finally, computer-assisted survey information collection and Internet surveys will be described in more detail.

Survey Modes

A variety of different survey modes have been used over the years, with much of the change reflecting changes in available technology. As will be seen, the different modes vary in their costs and time requirements, factors that must be considered when one is choosing an interviewing mode.

Two Important Distinctions

There are two important distinctions to make about survey modes (see table 3.1). One is how personal they are in terms of the extent of interviewer involvement: whether the survey is *self-administered* or *interviewer-administered.* Face-to-face and telephone surveys are generally interviewer administered. Face-to-face interviews are usually in the respondents' natural surroundings, as when interviewers go to people's homes or offices. Skillful interviewers can usually get a high response rate and can conduct