Mass Communication Research and Theory

EDITED BY

Guido H. Stempel III
Ohio University

David H. Weaver
Indiana University

G. Cleveland Wilhoit
Indiana University

PUB 2003

Boston • New York • San Francisco
Mexico City • Montreal • Toronto • London • Madrid • Munich • Paris
Hong Kong • Singapore • Tokyo • Cape Town • Sydney

13

Survey Research

Pamela J. Shoemaker

Maxwell E. McCombs

Media professionals spend a lot of time dealing with the products of survey research. Journalists write public opinion poll stories to tell Americans who's going to win the next election or to report what we think about the latest issue. Advertisers assure us that "9 out of every 10 dentists surveyed" support their products. And media management's marketing strategies are based on what surveys say the audience wants. Surveys have even invaded the Internet.

A *survey* is a study that collects information by asking people questions. The information collected—the data—is generally numerical and suitable for statistical analysis. Although the United States Constitution specified that a survey of Americans be conducted every ten years in order to ensure proper representation in the House of Representatives,[1] this census of all U.S. households is not what we commonly think of as a survey. The vast majority of survey research projects are *sample surveys* in which data are collected from a subset of individuals in the population. Inferences about the larger population are made from the information gathered from those people in the sample. U.S. public opinion polls commonly include survey interviews with 1,000 to 1,200 people, even though the adult population of the United States exceeds that by millions.

Surveys have proven to be an effective way to assess people's voting intentions, to explain their votes, to predict use of products, and to assess changes in opinions. In fact, the effectiveness of the telephone survey has resulted in the public being bombarded with polls. Unfortunately, many polls are thinly disguised attempts to sell products and services, thus making survey research more difficult for the legitimate pollster.

Surveys of Americans' actual or intended votes have been conducted for 200 years, with one of the earliest polls on record being conducted after the election for Massachusetts governor in 1787.[2] Supporters of the incumbent Governor James Bowdoin conducted a poll to explain why John Hancock won the race. The Bowdoin poll showed that Hancock's support came primarily from "labourers, servants, and so on," whereas Bowdoin's base was among "merchants and traders."

Three days later a Hancock supporter revealed his own poll, a recategorization of the vote that he called the "authentic breakdown," which showed that Bowdoin's support was actually among "speculators in publick securities," whereas Hancock's vote came primarily from "merchants, tradesmen, and other worthy members of society" and from "friends to the revolution." The Hancock poll also revealed that the one wizard voting in the election supported the losing candidate![3]

Although such canvassing of voters became fairly common in the 1800s, surveys of people's opinions on issues of the day were rare. One of the first U.S. opinion polls was conducted in the 1880s by the Iowa Labor Statistics Bureau. As part of its survey on labor and economic conditions, the bureau asked farmers' attitudes on the liquor issue and immigration.[4]

The advent of modern random sampling techniques in the early 1900s made surveys more practical, since a smaller number of interviews meant savings in time and expense,[5] but surveys using random samples also proved to be more accurate. In the 1936 U.S. presidential election, pollsters George Gallup, Elmo Roper, and Archibald Crossley used the expertise of statisticians, psychologists, and market researchers to correctly predict Franklin D. Roosevelt's victory over Alf Landon, whereas the less methodologically sound *Literary Digest* magazine poll of 10 million people picked Landon by a wide margin.[6]

Gallup and other pollsters enjoyed popularity until they predicted incorrectly that Thomas Dewey would beat Harry Truman in 1948. Confidence in polling collapsed until in the 1960s, when secret polls captured the imagination of politicians.[7] John Kennedy used polls in his 1960 election campaign, thus dignifying them for use by others, and Jimmy Carter was the first U.S. president to have access to continuous opinion polls.[8]

Much of the interviewing in opinion polls now is being done by telephone, rather than in face-to-face interviews, because of the rising costs and lower response rates face-to-face interviewing has yielded in recent years.[9] About 98 percent of U.S. households are reachable by telephone.[10] Internet polling is also becoming increasingly popular.

Survey Designs and Applications

One of the first decisions to be made in planning a survey involves the *design*—who will be interviewed, over what period, and how many times? There are three basic survey designs: the **cross-sectional** survey, the **panel study**, and the **trend study**.

The decision about whom to interview is largely a theoretical one driven by the hypotheses that are to be tested, but it is not a trivial decision. For example, in an election survey, we could interview adults, or only adult citizens, or only registered voters, or only people who actually intend to vote. The population we select for study will affect the data we collect and the conclusions we draw.

The period of time during which interviews are conducted—generally called being "in the field"—is also crucial. Some polls are in the field for less than one day, meaning that interviews have been conducted only with those individuals who are home at the instant that the interviewer calls. Such a sample of respondents may underrepresent busy individuals. Other polls are conducted over several months, making the results vulnerable to events occurring while the study is in the field. For example, a poll asking about people's

attitudes toward AIDS could yield two very different data sets if a cure for AIDS is revealed in the middle of the field period. The optimal field period is long enough to make several attempts to complete an interview in each sampled household, but short enough to minimize threats to validity due to intervening events.

In a *cross-sectional* survey, the respondents are interviewed only once, and the data collected provide a snapshot of the population at the time the field work (the interviewing) is done. Cross-sectional surveys are common, because they are far more economical than longitudinal studies. Cross-sectional surveys can describe the characteristics of the sample (e.g., respondents' attitudes or behaviors) at a given point in time, but such one-shot studies cannot address questions about change (e.g., is a change in a person's attitude followed by a change in his or her behavior?). To study changes among the elements in a sample (and, by inference, in the population from which the sample is drawn) requires that data be collected at more than one point in time—a *longitudinal* study. The panel study and the trend study both involve longitudinal data collection.

The *panel study* involves interviews with the same individuals at more than one point in time. The panel design permits the comparison of a person's responses at each time point, thereby allowing us to assess the extent and direction of any changes that may have occurred *within that individual.* For example, we might compare changes in an individual's media use with changes in support for a political candidate.

The *trend study* is also used to study change, but change in the population as a whole rather than changes in individuals. In a trend study, a new sample of individuals is drawn from the population at each point in time—the trend study is composed of several cross-sectional studies that are compared over time. This permits the researcher to evaluate change in the population over time (e.g., changes in support for two presidential candidates); however, because such changes (e.g., in vote intention) are attributes of individuals, the trend design limits our ability to explain change. Because different individuals are interviewed at each point in time, we do not know why the observed changes have occurred, and we cannot easily separate changes in individuals over time from possible differences between the samples drawn at each time point.

The panel is a superior design for explaining change, but data from a panel study may be flawed by *panel mortality*—the loss of individuals from the panel between each interviewing wave. Individuals leave panels for many reasons: They decline to participate; they move during the study or change telephone numbers and are now unreachable; they go on vacation or business trips or become ill; or they may even die during the study. Panel mortality is generally not the result of a random process, and therefore the individuals who leave a panel generally have characteristics in common. Their failure to participate in the entire study may make the panel unrepresentative of the population, thereby limiting the inferences we can make about the population.

Attempts to limit panel mortality can make the panel study more expensive than the trend design, but the panel remains the more valid way of assessing changes in individuals. The trend design is useful in the *secondary analysis* of preexisting data sets. Many survey organizations (such as the University of Michigan Institute for Social Research and the National Opinion Research Center) archive survey data and make data sets available to others. Secondary analysis is "the extraction of knowledge on topics other than those which were the focus of the original surveys."[11] When the same questions are asked in

successive surveys, the secondary analyst can perform a trend analysis of several data sets. For example, McCombs used the well-known Michigan national election surveys to trace increasing political participation by blacks through their use of newspapers and television for political information during the 1952–1964 civil rights era.[12]

A fourth design is not recommended—the *pseudopanel*. If a researcher asks her respondents to report their past and current attitudes or behaviors, she may decide to treat her data as if they represent different time points. Such a design is vulnerable to many problems, however, not the least of which is the human being's ability or willingness to remember and report what he or she thought or did in the past. The pseudopanel design is subject to many potential errors and should be avoided.

Using Survey Data to Establish Causal Relationships

Establishing a causal relationship requires: (1) showing covariation between the presumed cause and the presumed effect; (2) establishing appropriate time order, that is, that the cause precedes the effect and not the reverse; (3) ruling out alternative explanations for the observed relationship; and (4) controlling error variance, that is, keeping errors to a minimum and explaining as much of the independent variable as possible.

The cross-sectional design—representing the bulk of all survey projects—can address only three of the four criteria necessary for establishing a causal relationship: covariation, alternative explanations, and error. Covariation can be shown between two or more variables through the use of statistics that test the strength of the relationship between two variables (e.g., Pearson's correlation coefficient). But, as the saying goes, "correlation is not causation"—showing that changes in one variable (e.g., exposure to a daily newspaper) are associated with changes in another (e.g., levels of political interest) is only one step, but a necessary one, toward proving causality.

Survey researchers have become adept at *multivariate statistical techniques* that help rule out alternative explanations for a relationship by statistically controlling for the effects of other variables on the observed relationship. The use of statistical controls, however, is only as good as the researcher's ability to identify and measure possible alternative explanations. It is far easier to rule out alternative explanations in a randomized experiment, where the experimenter randomly assigns subjects to treatment groups, thereby controlling for a vast array of unidentified and unmeasured variables. In a survey, respondents select themselves into high or low political interest groups, thus making differences in these groups related to many causes.

Time order can be established in survey research through the use of longitudinal designs, with the panel's ability to study changes within individuals making it far superior to the trend or cross-sectional design in establishing causality. In a panel study, we can look at whether, for example, individuals first change their level of newspaper reading and then their level of political interest, or vice versa. If we observe that newspaper reading and political interest change simultaneously, then we have evidence of a third variable affecting both reading and interest or a methodological problem with the time intervals between our panel waves. The waves of interviewing need to be frequent enough to "catch" changes in variables that may occur rapidly one after the other.

Controlling error variance is a major challenge in all research, and, to the extent that our studies are flawed, our data will not permit the establishment of valid causal relationships. Later in this chapter, we will discuss the types of errors that must be considered in survey research.

Types of Surveys

Survey interviews are either *self-administered* or conducted by an *interviewer*. Self-administered questionnaires are either distributed in group settings or by mail (or, occasionally, through home delivery), whereas interviewer-conducted surveys can be done by telephone or face to face.

Self-Administered Questionnaires

This popular survey method involves giving a potential respondent a questionnaire, usually by mail. The respondent is asked to complete the questionnaire and return it to the study director. The major advantage of the self-administered questionnaire is a savings of time and expense when compared to telephone and face-to-face surveys. Self-administered questionnaires also avoid biases due to interviewers, ensure standardized presentation of questions, give respondents more privacy (important for sensitive questions), and may increase the validity of responses that require the respondent to check information or to think about his or her answer. Their major weakness is that response rates are generally lower than in telephone or face-to-face studies, thus increasing threats to validity due to the lack of representativeness for the sample.

Mail Surveys. Intensive follow-ups in **mail surveys** are often necessary to increase the response rate above 50 percent, thus increasing study time and expense. A variety of techniques has been used to increase the percentage of returns in a mail survey.[13] Follow-up contacts by mail have been shown to increase return by 20 percent for one follow-up and about 12 and 10 percent for second and third follow-ups, respectively. Supplementing telephone follow-ups with mail follow-ups can increase the return by another 15 to 30 percent. The role of monetary inducements has not been clearly established.

Some questionnaires seem to get better returns than others. Questionnaires that are particularly salient to the respondent can yield a 30 to 40 percent higher response than nonsalient questionnaires. Those dealing with sensitive topics may get lower response rates. The length of the questionnaire was not related to response rate in one experiment, where length varied from one to twenty-two pages. A regular stamped return envelope yields higher responses than a business reply envelope. In addition, the best response rates have been associated with questionnaires printed in booklet format and with an attractive cover page. Among other factors influencing response rate is the type of population being surveyed. Response rate is positively related to the education level of the respondents; nonreaders are automatically excluded from responding. Altruistic appeals and an institutional sponsor can also be helpful in increasing responses.

Internet Surveys. An increasing number of self-administered questionnaires are being delivered via email or through web pages on the Internet. The decision of whether to include the questionnaire as part of an email message is related to the amount of bandwidth it takes—how long it is. Very long questionnaires may irritate potential respondents as a breach of "netiquette."[14] It is probably more reasonable to send an email message as a "cover letter," with an accompanying link to a web page where the questionnaire can be accessed. Other Internet questionnaires begin with a link from another web page, such as a browser that is used by many people. Either way, **Internet surveys** are easily ignored or deleted by a single keystroke, thus contributing to response rates that are dramatically lower than other methods.

An advantage, however, is that the length of Internet questionnaires does not seem to affect response rates, as is the case with mail surveys.[15] If respondents begin the questionnaire, they are likely to finish it. Another advantage is the quick turnaround. Most Internet surveys can be completed in 48 to 72 hours,[16] the time period shortened dramatically by automatic entry of the data into a file readable by statistical software. In addition, there is some evidence that respondents may give more complete answers to open-ended questions.[17]

Although the cost of Internet surveys may potentially be less than mail (or interviewer-conducted surveys), an experiment comparing mail and Internet modes of questionnaire delivery demonstrated that there were substantial costs associated with managing the email sample, including answering technical questions about the electronic questionnaire. Costs of the Internet and mail versions may be close if similar quality and response rates are desired.[18]

Interviewer-Conducted Surveys

The **telephone survey** is quickly supplanting the **face-to-face survey** in the United States, primarily due to the lower costs of the telephone interview and to falling response rates of face-to-face studies in urban areas. Refusal rates often exceed 40 percent of potential respondents.[19] The probability of social class biases due to the availability of telephones has greatly diminished. For studies on subpopulations such as the poor, the elderly, or young adults (who are often transient), the face-to-face interview may provide more reliable data than the telephone survey. Comparisons of data collected by the two methods show few statistically significant differences, although there is some evidence to suggest that respondents find telephone interviews to be less rewarding and more tiresome than face-to-face interviews, even though the telephone interview is generally faster paced.[20]

The telephone survey requires much less administration than the face-to-face survey. For example, a face-to-face study requiring 200 interviewers, each completing 7 or 8 interviews, could be accomplished by 30 to 40 telephone interviewers, each doing 40 to 50 interviews. Supervisory staff is similarly reduced.

If telephone interviews are completed from a central phone facility, the telephone survey can result in higher quality data, because supervisors can monitor interviewers and give immediate feedback to correct interviewer errors. Face-to-face interviewers work under far less supervision.

Telephone surveys have the additional advantage of being far more economical than face-to-face surveys, often averaging half the expense. Telephone surveys also take less time to complete than do an equivalent number of face-to-face interviews.

Comparing Self-Administered to Interviewer-Conducted Surveys

Interviews with a sample of respondents can be conducted through the mail, face-to-face, or over the telephone. Each of these modes of interviewing has its advantages and disadvantages. Which should you use?

Although there is no fixed number of criteria to be weighed, five are suggested here: cost, response rate, kinds of questions, numbers of questions, and social desirability bias. While the mail questionnaire ranks well on cost, it receives relatively low marks on the other four criteria. The face-to-face interview has the advantage in terms of the kinds and number of questions that can be included. For response rate and social desirability bias, there is little difference between face-to-face and telephone interviews. But telephone interviews have a distinct advantage in cost. Increasingly in recent years, the telephone interview has been the mode of choice for most research projects.

Survey Data Collection

Survey questionnaires traditionally have been printed on paper, and the responses have been recorded in pencil or pen directly on the questionnaire. The data were entered in a computer file and proofread before analysis.

Today most large survey research organizations use **CATI (computer-assisted telephone interviewing)** systems for data collection in telephone surveys. In CATI the survey questionnaire is programmed into a computer file. Interviewers sit at computer terminals and directly input the respondents' answers to all questions. In most systems, both numerical (closed-ended) and text (open-ended) responses can be easily handled. Numerical data go directly into a data file, where they can be available for analysis almost immediately after the interview is completed. Text responses require additional coding before statistical analysis can be performed.

The key advantage of CATI systems is the quality of the data collected. In most systems, out-of-range responses will not be accepted by the system (e.g., if the interviewer hits "6" for a scale that ranges only from 1 to 5, the CATI software may ask him or her to "try again," resulting in very clean data). CATI also easily handles complicated questionnaires that require the interviewer to skip around the instrument depending on the respondent's answers to previous questions. Such skips and branches are programmed into the CATI instrument by the study managers and do not rely on the interviewers' judgment and memory. CATI instruments can easily handle questionnaires far too complicated for any human interviewer to correctly follow. Skips and branches occur automatically based on the respondent's previous answers, thereby eliminating many interviewer errors. Previous answers can be inserted into subsequent questions in order to personalize the questions to the respondent.

Some survey directors also feel that CATI increases the speed with which a telephone survey can be completed, but such time savings are debatable for the inexperienced CATI user and for small projects. The major drawbacks to CATI are the complexity of the systems and the time required to master them. The more flexible the system (and hence the more options the study director has), the more complicated the system is to learn and to use. Six months to a year is the usual shakedown period for the more complicated CATI systems. Survey personnel must in general be more highly trained for CATI projects than for pen-and-paper surveys.

Any time savings to be realized with a CATI system fall at the end of the study period. Once the data are collected, they can be made available for statistical analysis almost immediately, since the data entry and proofreading steps are eliminated. More time is required in CATI projects at the beginning of the study, however, since it is far more difficult to program and debug a questionnaire into CATI language than it is to type and photocopy a pen-and-paper instrument.

Computerized data entry is also occasionally used for face-to-face and self-administered surveys. *CAPI (computer-assisted personal interviewing)* software packages are now available for use on laptop computers that can be taken to the respondent's location. *CADE (computer-assisted data entry)* software enables fast and accurate entry of data from mail surveys. Some software packages also have a self-administered feature that permits respondents to sit at terminals and enter their own responses.

Survey Sampling

Sampling is the process of selecting *elements* (in survey research, usually telephone numbers, households, or individuals) from a population (the aggregate of elements about which the researcher wants to make inferences). For example, a nationwide telephone opinion poll might use a sample composed of 1,200 household telephone numbers within the United States. Assuming that interviews are conducted only with adults, the researcher could use the data gathered from people in the sample to make inferences about the opinions of U.S. adults living in households that have working telephones.

Samples can be classified as either probability or nonprobability samples. In **probability samples**, the elements are selected by a random process involving chance; therefore each element has a known chance of being selected into the sample. In *nonprobability* samples, the elements are selected by other means (such as the researcher's convenience or ideas about the types of clients that should be included in the study), and there is no way of estimating each element's chance of selection. Knowing the odds of an element being selected allows us to estimate how representative the sample is of the population. Much error in our data is due to sampling. **Sampling error** can be estimated for probability samples but not for nonprobability samples. Although it is theoretically possible for a nonprobability sample to be representative, we have no way of evaluating its worth.

Types of Probability Samples. The easiest to understand, but the one used least often for surveys, is the *simple random sample*, in which each element of the population has an equal chance of being selected into the sample on each sampling round. Simple random sampling requires a list of all elements in the population, and this is one of the method's

major drawbacks: Some populations are impossible to list, and some are so large that listing all elements would be impractical. For example, a directory of all 40,000 students at a university could be considered a list of the university's student population. Each student in the directory could be sequentially numbered from 1 to 40,000. Numbers (and the students attached to the numbers) would then be selected at random, such as by computer, by using a table of random numbers, or even by selecting from among 40,000 numbered balls. On the first round, every student's chance of being selected is 1/40,000. After the first student has been selected for the sample (without replacement, meaning that the student is ineligible for selection again), each remaining student's chance of being selected is 1/39,999. Although the probability of selection changes between the selection rounds, each student has an equal chance of being selected within each round.

A less cumbersome sampling method, but one that also requires a list of elements, is the *systematic sample*. A sampling fraction is computed by dividing the number of elements in the desired sample by the number of elements in the population. If we want to sample 1,000 students from the university population of 40,000, then our sampling fraction would be 1/40; the sample will include one-fortieth of the population. Using the directory of students, we select a random number between 1 and 40 as the starting point, and then select every fortieth student from that random start. Systematic sampling is an efficient method that can result in a representative sample, unless there is some periodicity in the list, such that the sampling fraction corresponds to a cycle inherent in the list. For example, the interval for a systematic sample of newspaper "days" should never be seven, for such an interval will yield a sample made up entirely of the same day of the week.

Some samples make use of ***stratification***, a process that requires the researcher to identify one or more variables on which the sample elements can be categorized, generally at the time of sampling. Independent sampling procedures are then carried out within each stratum. For example, we might want to compare people from the Midwest with those from the South, West Coast, and East Coast. We would stratify our sample by geographic region and draw separate samples from each region.

If those separate samples yield different sample-to-population proportions (e.g., a Midwesterner's chances of being selected into the sample are 1 in 1,000, whereas a Southerner's chances are 1 in 750), then we have *unequal probabilities of selection*. Generalizations made about the country as a whole will be biased by the fact that Southerners are more likely to be selected than Midwesterners. To make the total sample representative of the nation as a whole, the researcher must *weight* the sample to bring all strata back into proportion to the population.

Stratified samples with unequal probabilities of selection are common in studies where the researcher is stratifying on a variable that is not evenly distributed through the population, such as ethnicity. If 70 percent of the population is Anglo, 20 percent Hispanic, and 10 percent black, then a sample of 1,000 individuals would yield only 200 Hispanics and 100 blacks – samples that are generally considered too small to be representative of those populations (see "sampling error" later in this section). If Hispanics and blacks are *oversampled* in order to have sufficient numbers for separate analysis, then the total data set will require weighting before it is analyzed to yield a picture of the overall population.

Most telephone surveys use *cluster sampling*, which takes advantage of the fact that telephone numbers are assigned in groups. Telephone numbers are assigned first in groups

of area codes (the first three digits of the long-distance ten-digit number), and second in groups of exchanges or prefixes (the first three digits of the seven-digit local number). ***Random digit dialing (RDD)*** is a process of sample generation that uses known lists of area codes and exchanges and combines these with randomly generated suffixes (the last four digits of the seven-digit local telephone number). RDD samples are not dependent on telephone directories, which do not include unlisted telephone numbers and which become out-of-date and inaccurate over time. A nationwide RDD sample also would be a multistage sample. The first stage might include sampling from among the population of area codes. In the second stage the researcher might sample from among all prefixes in the area codes selected in stage one. In stage three randomly generated four-digit suffixes would be added to the prefixes and area codes to complete a list of ten-digit numbers.

The problem with such an RDD sample is that it generally yields only about 25 percent working residential telephone numbers; the remaining numbers are not assigned or not residential. Waksberg[21] suggested an RDD cluster design that generally yields 50 to 60 percent working residential numbers. The researcher first randomly samples from among the population of area code prefix combinations, as described above, then adds a randomly generated four-digit suffix to each area code prefix combination. This number is dialed. Once the number is determined to be a working household number, then the last two digits of the ten-digit number are removed and the eight-digit number remaining is designated as a *primary sampling unit (PSU)*. One hundred two-digit numbers are then generated in a random order for the PSU in order to create 100 potential ten-digit telephone numbers. Interviewers are then instructed to complete a predetermined number of interviews in that cluster.

Respondent Selection. For most telephone and in-person surveys, the sample element (e.g., the telephone number or address) represents a household. We still have the task of selecting the individual to interview within that household. It is not adequate to interview whomever answers the telephone or door, since such individuals turn out to have characteristics in common. For example, women are more likely to answer the telephone than men. The researcher must decide whether his or her population includes all residents of the household (including children), only adults (i.e., at least 18 years old), only heads of households, or some other subset of household residents.

One commonly used method requires that all adults within the household be listed by the interviewer. The selection method is then based on the number of men, the number of women, and the last digits of the telephone number. This results in a probability sample of individuals within households.

Another common method is to select the individual who has most recently had a birthday. This results in a nonprobability sample of respondents, since the probability of a given individual being selected depends on the memory of the individual who answers the telephone and on the distribution of birthdays throughout the year and throughout the population. However, it is easier and quicker to carry out than the household-listing method.

Sampling Error. Table 13.1 includes estimates of sampling error at the 95 percent confidence level—usually called *margin of error*—for simple random samples of 50 to 2,400 elements. The estimates are computed by multiplying the standard error of a *proportion* by 1.96 (the appropriate value for the 95 percent confidence level.)[22]

TABLE 13.1 ***Margin of Error (at the 95 Percent Confidence Interval) for Proportions Ranging from 10/90 to 50/50 and for Simple Random Sample Sizes Ranging from 50 to 2,400***

Sample Sizes	Proportions				
	50/50	40/60	30/70	20/80	10/90
50	13.86	13.58	12.70	11.09	8.32
100	9.80	9.60	8.98	7.84	5.88
200	6.93	6.79	6.35	5.54	4.16
300	5.66	5.54	5.19	4.53	3.39
400	4.90	4.80	4.49	3.92	2.94
600	4.00	3.92	3.67	3.20	2.40
800	3.46	3.39	3.18	2.77	2.08
1000	3.10	3.04	2.84	2.48	1.86
1200	2.83	2.77	2.59	2.26	1.70
1400	2.62	2.57	2.40	2.10	1.57
1600	2.45	2.40	2.25	1.96	1.47
1800	2.31	2.26	2.12	1.85	1.39
2000	2.19	2.15	2.01	1.75	1.31
2200	2.09	2.05	1.91	1.67	1.25
2400	2.00	1.96	1.83	1.60	1.20

The margin of error sets boundaries around an obtained percentage that will contain the true population value in 95 out of 100 samples of the size shown. For example, for random samples with Ns of 400 and an obtained percentage of 50, the true population value of the percentage will be found somewhere between 45.10 and 54.90 in 95 out of 100 samples. The margin of error decreases as sample sizes increase and as percentages move away from 50/50.

Margin of error for a proportion at the 95 percent confidence level = +/− 1.96 * (square root of $p(1 - p)/n$)

where n = the number of elements in the sample and p = one of the percentages in the observed proportion.

If, for example, 60 percent of the 1,000 people interviewed favor candidate A and 40 percent favor candidate B, then the margin of error for this observed 60/40 proportion is +/− 3.04 percent. We are 95 percent confident that candidate A's true support in the population is between 56.96 percent and 63.04 percent.

A similar margin of error for a *mean* can be computed by multiplying the *standard error of the mean* by 1.96.

Margin of error at the 95 percent confidence level for an observed mean = +/− 1.96 * (square root of Var/n)

where Var = the observed variance, the sum of squared deviations from the sample mean and n = the number of elements in the sample.

Table 13.1 shows that the margin of error decreases as the sample size increases and as the observed proportion moves away from 50/50. Not only are data from 1,000 individuals more likely to correctly represent the population than are data from 100, but also it is easier to correctly detect the majority position when the proportion is 10/90 than when it is closer to 50/50.

How big should a sample be? The researcher should select the sample size with the smallest margin of error he or she can afford (within time and monetary constraints). Although it seems counterintuitive, error due to sampling depends almost entirely on the absolute size of the sample, not on the proportion of the sample size to the population size, *when the sample elements constitute less than 10 percent of the population.*[23] Therefore, a sample of 1,200 people can as precisely represent a country of 220,000,000 as it can a city of 100,000.

The formulas given above are appropriate for simple random samples or systematic samples with no stratification. If the sampling plan calls for a stratified sample, then sampling error will be lower "for the variables that are more homogeneous within strata than in the population as a whole."[24] Higher sampling errors are associated with samples involving clustering and unequal rates of selection. It is also important to remember that margin of error estimates only errors due to sampling. A discussion of other types of survey errors is at the end of this chapter.

Nonprobability Sampling on the Internet. Although the Law of Large Numbers requires a randomly selected large sample to reduce bias,[25] the problem with most Internet surveys is that there is no sampling frame of the population being studied. This is especially true in general population surveys, where not only does an unknown percentage of the population have access to the Internet (and the characteristics of these people is generally unknown), but also there is no list of email addresses for those who do have access. It is impossible to select a random sample in such a situation, thus raising questions about validity. For small populations made up solely of individuals who have Internet access with known email addresses, a random sample may be possible. It is important to know that people are not responding more than once, and asking them for their email addresses can help filter out duplicates. However, even where sampling frames may be acquired and duplicates screened out, there is no way to control for the multiplicity of email addresses, some even with false information.[26] In addition, not all email addresses are valid or current.

There have been two basic strategies to general population sampling on the Internet. The first involves selecting a traditional random sample for a telephone survey, and then making calls that screen for Internet users. This method is expensive and fails to make use of the speed of electronic interviewing, requiring telephones, interviewers, data entry, and so on. The second strategy involves sampling directly online through either/or advertising or direct selection.[27] Many such polls result in very small response rates and invalid data sets, but the use of multiple methods of sampling can allow polls to approximate representativeness. Some pollsters have posted announcements about their surveys on newsgroups, placed banners on browsers like Yahoo!, used their own email lists, and advertised in print publications. This approach does not result in a random sample, but by exposing the study to many sources of the online population, it increases the odds of an unbiased sample.[28]

If the goal is to represent a general population, however, we must remember that the online population is not itself representative of the general population. It generally is better educated, with higher income, and younger.[29] A 1998 study conducted by the National Telecommunications & Information Administration of the U.S. Department of Commerce showed that 26 percent of United States households had Internet access and 33 percent of those living in the U.S. had access to the Internet either at home or work.[30]

Nonetheless, Internet polling can yield good data. In the 1998 midterm elections, the Harris Black/Excite poll correctly predicted 21 of the 22 races in the 16 states they studied. Large samples from each state were said to be the key.[31] But for pollsters educated with the traditional routine of random sampling, it is difficult to accept what amounts to a convenience sample as representative. Harris Black International is studying the extent to which carefully weighted Internet polls can be as representative as random polls.[32]

Instrument Design

The term *instrument design* refers to writing and formatting the survey questionnaire. Questionnaires should be easy to read and should have an easy-to-follow format. Pen-and-paper questionnaires should put plenty of space around each question; they should also make clear whether the correct answer is to be circled or checked in a box or circle. Skips and branches need to be clearly marked with text instructions and/or arrows. CATI instruments are usually formatted to have one question per screen, but the programmer should take care to type the text so that ends of lines and punctuation make the questions easy to read. Tabs can be used to make the screen visually attractive and easy to follow. Although CATI skips and branches must be programmed accurately, they are "transparent" to the interviewer and require no special graphic presentation.

The questions in the instrument represent *operational definitions* for the variables under study. It is desirable for the researcher to specify his or her *hypotheses* prior to data collection and to provide explicit *theoretical definitions* for all concepts in the hypotheses. Then the operational definitions relate directly to the theoretical definitions. The researcher should have a clear idea of how every question will be used; respondents should not be abused by asking them questions that may not be used because the researcher did not carefully think through his or her project. It is a good idea to make up dummy tables prior to questionnaire construction that specify the analyses necessary to test the hypotheses of interest. Here are some keys to designing a good questionnaire:

1. Make sure your language is appropriate for the respondent and pick words that have the same meaning for everyone. For example, unless you're interviewing college professors, don't ask, "Do you favor or oppose the tenure system?"
2. Avoid long questions.
3. Limit questions to a single idea. Don't ask, "Do you favor or oppose televised presidential debates and equal coverage for all candidates?" Ask, "Do you favor or oppose televised presidential debates? Do you favor or oppose equal coverage for all candidates?"
4. Don't assume that your respondent has factual information or that he or she can report another's opinions. For example, parents may or may not be able to list

the television shows their children watch, but they are unlikely to accurately report their children's opinions.

5. Establish your frame of reference. For example, when asking about behaviors like television viewing, be sure to set a time frame such as *hours per day.*
6. Be considerate of your respondent's ego. For example, "Do you *happen* to know the mayor's name?" is better than "Do you know the mayor's name?" Also, phrase questions so that they are not objectionable. Don't ask, "Did you graduate from high school?" Instead ask, "What is the highest grade of school that you completed?"
7. Avoid biased or leading questions. Don't ask, "Do you let your children watch violent television shows?" Instead ask, "What shows do your children watch on television? Any other shows?"
8. Decide whether the question should be open- ("What do you think about George W. Bush's performance as president?") or closed-ended ("Some people say that, overall, George W. Bush has been a good president. Do you strongly agree, agree, remain neutral, disagree, or strongly disagree that George W. Bush has been a good president?"). They don't necessarily yield the same results. Be certain that closed-ended response categories are mutually exclusive and exhaustive.
9. Consider whether general or specific questions are needed. For example, is it enough to ask, "How well did you like the television show?" or should you also ask, "Have you recommended the show to anyone?"
10. Decide whether the question should be stated personally or impersonally. Impersonal: "Is the amount of television that your children watch too much, not enough, or just about right?" Personal: "Would you like your children to watch more television, less television, or is the amount that they're watching now okay?"
11. Avoid ambiguous wording. For example, do not ask, "Do you usually vote?" Instead ask, "Did you vote in the last presidential election?"
12. When asking for sensitive information, use introductory statements and transition phrases to ensure that the respondent understands why you need the information. Respondents will answer even the most sensitive questions if they understand the reason for the questions and believe the request to be legitimate.
13. Begin the interview with questions that will be interesting and easy for the respondent to answer.

Before using the questionnaire to collect data, it must be *pretested.* Test interviews are conducted with members of the population being studied; for general population surveys, around fifty pretest interviews are usually adequate. Be sure that respondents who participate in the pretest are not included or interviewed again as part of the actual study. When pretesting, interviewers look for errors in the instrument (e.g., spelling, punctuation, grammar, incorrect skips or branching) and for questions that may be misunderstood by the respondents or to which respondents may take offense. Pretest interviewing is more difficult than the actual data collection interviewing, because interviewers have to think critically about the instrument at the same time they are reading questions and recording answers. Pretest interviewers should always make notes about each pretest interview for review by the study directors.

Hiring and Training Interviewers and Supervisors

Interviewing—whether on the telephone or face to face—is a tough job. The interviewer has to get important factual information from a stranger who probably doesn't want to be bothered and, at the same time, accurately record that information. If the interviewer is working on a CATI or CAPI system, then he or she has the additional challenge of mastering hardware and software that may not always work properly. Being able to distinguish a hardware problem from a software problem from interviewer error is a treasured talent in an interviewer. Unfortunately, such individuals don't remain interviewers very long; they are promoted to supervisors at record speed.

What else makes a good interviewer? A pleasant telephone voice is essential for phone surveys, but not sufficient. The best interviewers seem to be highly empathetic and confident about their ability to carry out the interview. Good interviewers listen to respondents' words and tone of voice and circumvent the respondent's suspicions and uncertainties. Good interviewers convey the legitimacy and importance of the survey project by projecting a professional demeanor. And above all, good interviewers quickly learn procedures and carry them out. If CATI or CAPI systems are used, then typing proficiency is necessary and a familiarity with computers will speed training.

In many parts of the country, it is advisable to conduct interviews in languages other than English. Hiring bilingual interviewers presents some special problems. You must be certain that the individual really is proficient in both languages; being able to speak Spanish when you're face to face with someone may not equal speaking Spanish fluently on the telephone. In addition, you must consider what level of proficiency is necessary (colloquial or formal speech and/or the ability to grammatically write the language), how accurately the individual can translate between the languages, and how similar the person's accent is to that used by the population being studied.

It is advisable to provide bilingual telephone interviewer applicants with a one-page bilingual questionnaire and arrange for a bilingual tester to telephone them at a later date for a test interview. The tester should evaluate the applicant's ability to switch easily from one language to the other, translate accurately, write in both languages coherently, answer questions from the interviewer that are not on the script, and speak in an acceptable manner *before* training the individual in survey procedures.

Some organizations pay interviewers an hourly rate for training and some pay a flat fee or nothing. Paying the same hourly rate for training as for actual interviewing conveys to the interviewer that his or her time is valued and that the training is important. This may translate into more learning and better efficiency.

Most training programs include the following: administrative procedures, such as filling out time cards and other record-keeping; using CATI or CAPI technology or recording answers on pen-and-paper instruments; tone of voice, pacing, and reading questions; probing and clarifying answers; how to elicit respondent cooperation; answering respondent questions and comments; dealing with hostile respondents and converting refusals; setting up callback appointments (a scheduled time for completing the interview with the respondent) and making callbacks; how to use the sample; and lots of practice interviews.

The best supervisors are promoted from among the interviewers; they already know the system and can anticipate interviewers' problems. Supervisors need additional training

on sample control and maintenance, personnel supervision, hardware and software use, monitoring interviewers, and verifying interviews.

Additional training for interviewers and supervisors is generally needed before beginning a new study. Interviewers should be given a chance to make several practice interviews in order to familiarize themselves with the questionnaire, and supervisors need to practice sample control. Both interviewers and supervisors must be briefed about the study's purpose and goals. Frequently, *question-by-question objectives* are provided in writing by the study director for each item in the questionnaire. Several hours of training may be necessary to explain each question's objectives to interviewers and supervisors.

Conducting the Interview

The interviewer has at least seven basic tasks to perform in the interview. His or her first job is to *elicit the cooperation* of the person who answers the phone or doorbell. Professional dress can be important for face-to-face interviews, and a professional tone of voice is critical for telephone interviewers. The interviewer may have to answer the respondent's questions or comments before proceeding with the interview. Respondents often want to know who is sponsoring the study, why someone else in the household can't be interviewed instead, what the information collected will be used for, and how long the interview will take. The project director should provide the interviewer with answers to the most common questions and comments; giving the interviewer some stock phrases to rely on can help her or him persuade the respondent to cooperate.

Second, interviewers must *read the questions exactly as written* in exactly the order given. Even slight deviations can cause the respondent to provide a completely different answer. Standardization in interviewer performance is important. Some words in the questions may be underlined or written in capital letters; these should be emphasized by the interviewer. Occasionally a respondent will provide an answer before the full question text has been read; if there is *any* doubt that the respondent would answer differently after hearing the rest of the question, the full question text should be reread.

Third, the interviewer must *record* the respondent's answers exactly. This usually means writing down (or typing on the CATI/CAPI keyboard) the appropriate number for a response, but it may also include recoding verbatim open-ended responses. If the respondent talks quickly or at length, the interviewer may need to interject a phrase such as "I'm writing this down (typing what you've said), please wait a moment." It is very important that interviewers write down exactly what the respondent has said; many times the exact words used by the respondent are crucial to interpreting the respondent's meaning.

Fourth, the interviewer must *clarify* questions when the respondent is uncertain what a question is asking for. The question-by-question objectives provided by the study director should include definitions of key words and phrases when allowed—sometimes the study director wants the respondent to provide his or her own interpretation: "Whatever _____ means to you." The question-by-question objectives should also tell the interviewers what they are allowed to say about the goals of each question.

Fifth, interviewers should be prepared to *probe* for more complete answers or when they think that the respondent has given an inappropriate answer. Sometimes probes are scripted on the instrument, with explicit instructions for their use. For example, in an open-ended question like, "Why do you watch television?" interviewers would probably be in-

structed to probe "Any other reason?" after each response. Most probes, however, cannot be scripted, because they must be used to improve a particular respondent's performance. For example, if the interviewer asks, "How many days per week do you read a daily newspaper?" and the respondent says, "Oh, three or four," then the interviewer should probe for a specific answer: "Would you say it is closer to three or to four?" The key to probing is that the probe used is neutral; that is, it does not lead the respondent toward a particular answer. Often the best probe is silence. Other good probes include repeating the question, "anything else?", "in what ways?", or "how is that?"

Sixth, interviewers should provide neutral *feedback* occasionally to let respondents know that they are performing adequately. Feedback is not used to congratulate respondents on the type of response they are giving (i.e., a certain attitude), but rather to reward them for responding at all. Short feedback is used most often: "Thank you," "Uh huh," and "That's helpful." Longer phrases should be mixed in every few items: "Thank you, that is helpful to our research," "That is useful information," or "I appreciate your cooperation." Feedback should not include any phrases that indicate approval of the type of response; interviewers should avoid such terms as *okay*, *right*, and *good*.

Finally, the interviewer may have to set up a *callback* appointment if the respondent is unable to begin or complete the interview at the current time. The interviewer should try to make the appointment for a specific date and time. If that is not possible, then he or she should try to set up a general time, such as "tomorrow afternoon."

If the survey questionnaire has been prepared in more than one language, the interviewer must also evaluate which language the respondent wishes to speak. Often this can be deduced from the way the telephone or door is answered, but sometimes the interviewer may need to ask. Once the choice of language is made, however, the interviewer must also be sensitive to the needs of the respondent, switching back and forth between languages as is necessary to communicate effectively.

In addition, for Internet surveys, it may be necessary to have someone answer telephone calls about the study, often of a technical nature.

Quality Control

Collecting data of the highest quality should be the primary goal of the survey researcher. There are several ways in which the quality of the data can be maintained. First, the project director must provide sufficient *training* for both interviewers and supervisors. Training is important not only for basic survey procedures, but training specific to each new study is also crucial.

Second, the interviewers should be monitored for at least one hour's worth of interviewing on each study. For telephone surveys, this generally means having a supervisor listen to interviews over an extension telephone that allows the supervisor to hear the interviewer, but not the reverse. The supervisor should follow the interview with a copy of the questionnaire being used. For CATI studies, a "slave" terminal can be tied into the interviewer's terminal: This allows the supervisor to hear what the interviewer says, what the respondent says, *and* see exactly what the interviewer records when she records it. After an interview is monitored, the supervisor should meet with the interviewer to give feedback on the performance. Frequently, an evaluation sheet is used, with a copy being put in the interviewer's personnel file.

Third, a small percentage of interviews conducted in a study (such as 5 or 10 percent) should be verified. To verify an interview, a supervisor calls a randomly selected respondent to confirm that the interview actually was conducted. Verification is more important in face-to-face surveys and in those telephone surveys that are not conducted in centralized phone facilities because falsification of data is easier when supervisors are not present when the interviews are conducted.

Fourth, the researcher should use sample control methods that result in the highest possible percentage of completed interviews and the lowest possible percentage of refusals. There are many formulas for response rates, completion rates, and refusal rates, and the selection of one is more a matter of choice than science. Frey[33] gives a simple formula for each:

- The ***response rate*** reflects the percentage of potential eligible respondents who cooperated in providing information. It may be calculated as the number of completed interviews divided by the number of eligible respondents. "Eligible respondents" include those who completed interviews, who partially completed interviews, who refused, who could not speak the language, who were ill or otherwise indisposed, or who were away for the duration of the study. "Ineligible respondents" might include those who do not meet the respondent profile (e.g., not 18 years of age or older), but "ineligible elements" in the sample could also include (in a general population household survey) businesses, nonworking numbers, and numbers that are *always* busy or that ring but are never answered.
- The *completion rate* reflects the proportion of cooperating respondents as a percentage of all sample elements contacted, eligible or not. It is calculated by dividing the number of completed interviews by the total number of sample elements (e.g., telephone numbers or households) used.
- The *refusal rate* is the opposite of the response rate. It reflects the percentage of eligible respondents who refuse to be interviewed. It is calculated by dividing the number of refusals by the number of eligible respondents.

Evaluating Survey Research

Measurements derived from survey research are subject to two main sources of errors—*sampling error* and *nonsampling error*.[34] Sampling error, as we discussed earlier, can be empirically estimated for data collected from a probability sample. Nonsampling errors are equally important, but they are extremely difficult to estimate empirically with any degree of confidence.

Nonsampling errors are of two types. *Random errors* reduce the reliability of measurements, but they generally cancel each other out over numerous measurements (e.g., one person overestimates the number of hours per day that he watches television and another underestimates her viewing). *Nonrandom error* or ***bias*** is more troublesome, because these errors are systematic and do not cancel out over repeated measurements. One type of nonrandom error is *nonresponse bias*, which results from differences between those who complete