PART III

Respondent Selection Issues
Nonresponse error at the Unit Level
THE RESPONSE RATE CHALLENGE

Nonresponse error can be at the unit level as well as at the item level. Unit nonresponse occurs when some people in the designated sample are not interviewed. The most intriguing controversy on nonresponse error is the extent to which efforts should be made to maximize response rates. A high response rate has traditionally been considered essential to guarantee that the sample is representative enough to be credible, but studies are beginning to find that lower response rates do not necessarily lead to greater response bias.

Statistically, as with nonresponse error at the item level, nonresponse error at the unit level can bias results if the response rate is low and/or if respondents and nonrespondents are dissimilar. As stated in chapter 7 for item nonresponse, the true mean of a variable equals its sample mean plus a bias equal to the missing data rate multiplied by the difference between the means for respondents and nonrespondents. Thus, the amount of bias depends on the nonresponse rate and on how different respondents and nonrespondents are. Nonresponse error causes bias if nonresponse and survey variables have a common cause (Groves and Couper 1998, chap. 1). Survey results would be biased, for example, if fear of crime caused people to refuse to answer a survey measuring fear of crime, or if Democrats were less likely to participate in a voting survey. Similarly, postelection survey estimates of voter turnout are invariably biased upward because the people who are most likely to vote are also most likely to participate in surveys.
This chapter begins with a discussion of the sources of survey nonresponse. The second section discusses theories of survey participation, and the third section summarizes empirical studies of cooperation along with methods for increasing cooperation. The response rate controversy and statistical solutions to unit nonresponse will be reviewed in the fourth section.

**Survey Nonresponse**

There are three types of unit nonresponse. *Noncontact* refers to situations in which the designated respondent cannot be located, such as when no one answers the phone or when a mail questionnaire cannot be delivered. Noncontact is basically a logistics problem: it is “ignorable” nonresponse (see chapter 7) if being at home is essentially random with respect to the variables being measured, which is usually the case.

*Incapacity* is when the designated respondent is incapable of being interviewed, due to such reasons as physical or mental health, speaking a different language from the interviewer’s, illiteracy in self-administered surveys, or technological inability to deal with an Internet survey. Incapacity is largely beyond the researcher’s control; it rarely causes nonresponse bias, except when it is related to the subject of a survey, as when health surveys cannot interview people too ill to participate.

*Noncooperation* occurs when the designated respondent refuses to participate in the study. Noncooperation is likely to be nonignorable if people are not willing to participate in a survey because of its topic, such as if people with conservative social values were less willing to answer a survey on sexual behavior. The cooperation issue involves important psychological considerations that will be detailed in this chapter.

Survey reports often include separate contact and cooperation rates. The contact rate is the proportion of cases in which some member of the household unit was reached, whereas the cooperation rate is the proportion of all eligible units ever contacted that were interviewed. These rates can be combined into an overall response rate, which is the proportion of eligible reporting units that led to interviews. Some surveys also report a refusal rate—the proportion of all potentially eligible cases in which the respondent refuses to be interviewed. These statistics are often thrown around loosely, without clarity as to how they were calculated, so the American Association for Public Opinion Research (AAPOR 2004) has prepared a standard set of definitions, which are detailed in the appendix to this chapter.

**Contactability**

Although the main focus of this chapter is on cooperation, it is appropriate to turn briefly to the other main aspect of unit nonresponse: the problem of contacting the respondent. People are often not at home when interviewers attempt to contact them. Groves and Couper (1998, chap. 4) construct a model of contactability in which the relevant factors are physical impediments (such as locked gates or caller ID machines that impede access), at-home patterns, and when and how often the interviewer tries to contact the household. These factors are influenced by social-environmental factors and social-demographic attributes.

It is sometimes possible to obtain information about contact patterns. When contact is made with the appropriate household but not with the designated respondent, the interviewer may be able to obtain supplemental demographic information about the intended respondent. For example, using information ascertained from neighbors or other family members, Brehm (1993, 44–47) reports that the elderly, people with children, women, and those with higher incomes are more likely to be accessible in face-to-face surveys.

There are also time-of-day and day-of-week patterns to survey participation. Telephone interviewing is more successful during the evening than during the day, except on Saturdays and Sundays. Also, different types of people are home at different times of day; with older people more likely to be at home during the daytime. Lavrakas (1993) gives information on the best times for telephone interviewing. Fortunately, a study of telephone interviewing does not find significant differences between months of the year (Losch et al. 2002).

Tracking polls that report daily statistics are affected when particular types of people are less likely to be home some days than others; for example, younger people are less likely to be at home on weekend evenings. One statistical solution to this problem is to report only three- or four-day moving averages. That solution averages out the effects of sampling any particular day so as to avoid a Friday-night or Saturday-night effect.

**Panel Studies**

Panel surveys inevitably lose some respondents at subsequent waves (known as mortality, attrition, or wave nonresponse). Panel mortality is partly a recontact problem if the person has moved. The cumulative effect can be considerable: for example, not being able to follow 10% of the participants in the previous wave of a panel study would cumulate to being
able to recontact less than half of the original sample by wave eight. Panel mortality can be a serious problem, biasing the sample toward residentially stable households (Winkels and Withers 2000). Some panel studies try to follow movers, but others do not. Young people are especially mobile, so panel studies become unrepresentative in terms of age if they do not try to track young people.

Following respondents who move is itself a complicated operation (see Burgess 1989). One approach, called forward tracing, involves asking people at the initial interview for the name and address of someone who will know their address if they move. An alternative is “tracking” respondents, such as by sending them periodic reports or greeting cards between waves, so they can be tracked by the change-of-address notifications that the post office provides.

Panel studies can examine what types of people drop out by examining the data for their first interview. Attrition is not random: O’Muircheartaigh (1989) points out that the attrition in SIPP was around 20% in five waves for people who rent their room for cash, respondents under age twenty-four, nonwhites, and married people whose spouse was absent. Lepkowski and Couper (2002, 265–68) found that ability to locate the respondent again was related to sociodemographic and geographic variables (with African Americans less likely to participate again), while on an election study panel survey (Lepkowski and Couper 2002, 268–71) the ability to locate the person again was related to accessibility (people who provided phone numbers being easier to locate again) and survey experience (people with less missing data being easier to locate again).

Callbacks

A simple response to survey nonparticipation would be substitution for the people who were not interviewed. When someone cannot be contacted, the interviewer would just substitute a neighbor or a respondent who seems similar. Statisticians, however, strongly argue against this approach. Not only is there no way of telling how well the replacements substitute for those who were not interviewed, but there is every reason to believe they would be different. At the extreme, consider a one-evening telephone poll on a breaking news story. Substituting new respondents for people who could not be contacted would further increase the poll’s dependence on the type of people who are at home and easy to contact in the evening, making the sample even less representative of the population. An example of an exception when substitution makes sense is when needed information can be obtained by talking to a knowledgeable informant, as when one person monitors the health of his or her entire household. Substitution is most common in marketing research, where the conventional wisdom is that purchase patterns would not be affected by the variables that cause survey nonparticipation.

The alternative to substitutions is callbacks—trying again to contact the desired respondent. This approach is common in most survey modes, though cost often constrains the extent of callbacks. Repeated callbacks can be expensive, which creates a trade-off between boosting the response rate and keeping survey expenses within budget. Survey organizations differ considerably in how many callbacks they do, with five being a common number for phone interviews. Market opinion surveys often do not bother with callbacks at all, whereas some academic survey organizations will call phone numbers back more than a dozen times. Telephone surveys routinely find that the percentage of people successfully contacted goes down with each callback, suggesting a rule of diminishing returns (Groves and Couper 1998, chap. 4). In any case, the number of callbacks is often set to vary depending on the disposition on early calls, for example with more callbacks when a respondent is selected than when no one was at home.

Callbacks are made at different times of the day and days of the week to increase the chances of contacting the household. In face-to-face interviewing, the interviewer will seek to return to the area later and try again to locate the respondent. In computer-assisted telephone interviewing, call-scheduling software can keep track of codes for call outcomes, schedule calls, set priorities on calls, assign calls to interviewers, keep track of callbacks, keep track of refusal conversions, close out cases, refer calls to supervisors, and generate reports for monitoring interviews. This leads to a more systematic history of the calls than is likely when the records are kept by hand.

To give an idea of the success of callbacks in telephone surveys, Groves et al. (2004, 171) show a summary of five separate surveys in which only 40%–60% of those eventually interviewed were interviewed on the first call. In other words, callbacks can double the contact rate. An additional 10%–20% were interviewed on the second call, another 10% on the third call, and about 5% on the fourth, with fewer on each of the following calls. It is difficult to generalize from these results without more knowledge of the specifics of when calls were being made in these individual surveys, but this study has important implications for surveys with short polling periods. Overnight phone surveys will clearly have much lower contact rates than surveys that have time for callbacks. The news media often
conduct overnight polls to gauge public reaction to an important news event. The inability to attempt callbacks in such polls is a problem because the type of people who happen to be at home one evening are not likely to be fully representative of the population at large, particularly underrepresenting young men. As a result, overnight polls should not be viewed as careful surveys. Even three-day surveys that the media often conduct are likely to have low response rates, because they are not able to conduct many callbacks and refusal conversions.

Research has also been done on callbacks. Groves and Couper examined how households that were contacted in the first attempt differ from those that were not contacted until later attempts. For example, they found higher contact rates at the first attempt for households with more adults, young children, and elderly people in the household. As to social-environmental factors, Groves and Couper report lower contact rates in urban areas, though that is explained in some of their models by multiunit structures and the crime rate. Several telephone surveys have compared the demographics of participants according to the number of callbacks required in order to contact them. Typical findings are that women, unemployed, married, older and less educated people are more likely to be reached with only a few callbacks (Merkle, Bauman, and Lavrakas 1991), and more callbacks help mitigate these biases. That finding implies that the extra effort to reduce noncontacts would not introduce bias.

An important caution is that nonparticipants in a survey can differ from people who were more difficult to contact. Studies on different survey topics support this conclusion. Analysis of six British surveys (Lynn et al. 2002) found that the households added by efforts to reduce the noncontact rate were different from those that were not contacted originally, especially as regards age and education. Similarly, when Lin and Schaeffer (1995) used official records on child support payments, they determined that people who could not be contacted differed from those who were difficult to contact. Those who could not be located owed or paid the least amount of support payments.

Unfortunately, working to increase the response rate can also increase the refusal rate. The case in point is a 1986 phone survey by the Census Bureau that found that doubling its usual survey period for the National Crime Survey from two weeks to four increased both the response rate and the refusal rate by 3% (Sebold 1988). Thus, an additional 6% of phone numbers were eventually answered after repeated calls when the survey period was extended, but this increased the cost of the study with only a marginal increase in response rate and a corresponding increase in the refusal rate. Researchers must decide whether the increased effort is cost effective under these circumstances.

Theories of Survey Participation

While contactability is important in surveys, cooperation is a more serious problem. The theories of survey participation focus on what leads people to cooperate by participating in surveys and therefore how to increase the chances that people will cooperate. There are economic, sociological, and psychological theories of survey participation.

The simplest approach is a cost-benefits economic perspective, viewing participation as a matter of costs versus benefits. People are more likely to participate if the costs are low (such as a short survey) and if the benefits are high (such as being paid to participate). The cost-benefits logic is certainly important, but by itself it does not provide a sufficient handle on survey cooperation.

Social exchange theory (Blau 1964) is the basis for one sociological approach to survey participation. This theory assumes that people’s actions are motivated by the return that people expect to obtain. The exchange concept emphasizes that the survey must have value to the respondents in order to interest them. Social exchange theory is based on establishing trust that the rewards for participation will outweigh the costs. From this perspective, giving the respondent a monetary incentive to participate in the survey can be seen as a kindness that evokes a norm of reciprocity. The designated respondent feels that he or she must participate to reciprocate the kindness of the incentive. This interpretation implies that the nature of the incentive is less important than the act of giving an incentive, so that a token incentive should work as well as a large one.

Several psychological approaches to survey participation have been advanced. The most basic is the notion of altruism, viewing survey participation as helping behavior for a stranger. Interviewers often include in their request for participation "It would really help us out” as a way to get people to think altruistically. Another psychological approach is based on opinion change, trying to convince the person that the survey topic is salient, relevant, and of interest to the respondent. This approach also implies that simple messages requesting participation are likely to be more successful than ones that are hard to comprehend. However, in line with Petty and Cacioppo’s (1986) distinction between central and peripheral
routes for persuasion, a direct approach to obtain participation may not work as well as a peripheral approach would.

The request for a person to participate in a survey can be seen as a request for compliance. Looking at the factors that affect compliance with requests more generally, Cialdini (1988) has differentiated six influences on compliance, each of which suggests a peripheral persuasion approach to get a designed survey respondent to participate in the survey. The first is reciprocation: incentives for respondents are a way to invoke this mechanism by making them feel they owe something to the interviewer. The second is consistency: interviewers use this when they first get their foot in the door with an initial question and use that as a lever to get people to continue with the interview. Third is social validation. Interviewers sometimes attempt to invoke this mechanism by mentioning to respondents that many of their neighbors have participated or saying that most people who have participated have enjoyed doing the interview. Fourth is liking, which is invoked if the respondents like the interviewer. Fifth is authority, getting people to comply with people in power, as when a survey emphasizes that it is sponsored by the federal government. Sixth is scarcity, making people value a rare opportunity, as when emphasizing that survey participation is important because the researcher needs to hear the experiences or views of people like the participant. Thus, there are several different possible ways of getting people to comply with the request of being a survey participant. Groves, Cialdini, and Couper (1992) review these compliance approaches in more detail.

Brehm (1993, chap. 4) includes several psychological factors relating to the individual in his model of survey compliance. Brehm’s model consists of four predictive factors: the respondent’s relationship to strangers (including suspicion and fears regarding confidentiality), relationship to the interviewer (rapport and empathy), relationship to the interview (time and interest in the topic), and relationship to self-image (if the respondent views himself or herself as a helper, is healthy, and is free from family stress).

Subsequent models of survey participation have combined economic, sociological, and psychological considerations. Groves and Couper (1998, chap. 2) develop a conceptual framework that includes four factors that affect survey participation. The social environment is relevant in terms of economic conditions and neighborhood conditions for face-to-face interviews. The survey design, particularly the topic of the survey and how it is administered, affects participation. The household being approached is a third factor, including its sociodemographic characteristics. Fourth, the interviewer is an important aspect of obtaining cooperation, with the interviewer’s characteristics, experience, and expectations all affecting the contact between the interviewer and the household. Groves, Cialdini, and Couper (1992) call attention to a fifth factor as well: the interaction between the respondent and the interviewer. The initial conversation between the interviewer and the respondent plays a crucial role in the decision to participate. The decision of a person to participate in a survey thus depends on the interaction between the social environment, the survey design, the household, and the interviewer.

This model is extended in Groves, Singer, and Cornings (2000) in their “leverage-saliency theory.” This theory recognizes that different individuals will be affected by different matters in considering whether to participate (leverage). Some people give considerable weight to incentives or to the topic of the survey or the authority of the sponsor, while others give more weight to the how big a burden participating will be. Additionally, the request to participate makes particular points salient to people when asking them to participate. The authors model the decision to participate as a function of the products of those factors. For example, in a survey about community matters, the survey has high leverage for a person with high community involvement, so that can positively affect the chances of participation if the community aspect is made salient in the request to participate, whereas making that salient would not increase the likelihood of participation of a person with low community involvement. Note that this logic works to keep response rates low on mail questionnaires: respondents see the topic at a glance, so a mail survey may lose people not interested in the topic unless other factors such as incentives or authority prevail in their minds. By contrast, respondents on interviewer-administered surveys cannot gauge the topic of the survey as easily, so the topic has less negative potential. This leverage-saliency model is likely to provide a comprehensive framework for future work on this topic.

The leverage-saliency theory has an important implication for getting people to cooperate: the request for participation should ideally be tailored to the individual respondent, using an appeal that fits the person’s concern about the survey topic, the identity of the sponsor, and so on. The tailoring strategy will be discussed at length later in this chapter.

The Noncooperation Problem

The noncooperation problem has become very serious. Refusals were not a serious problem in the 1950s and 1960s but have increased since then.
For example, a 2001 survey by the Council for Marketing and Opinion Research (Sheppard 2001) found that just under half of the sample said that they had refused to participate in a survey in the previous year, compared to 36% in 1990 and 19% in 1980. The National Election Studies (NES) surveys had refusal rates in the low teens through the early 1970s but has had 20% and above since then (Luevano 1994). The 1990 NES survey had 20% refusal, compared to only 2.5% noncontact and 6% “other noninterview” (Couper 1997). The telephone interview response rate for the Survey of Consumer Attitudes stayed about 70% from 1979 to 1996, but maintaining that rate by 1996 required twice as many calls to complete an interview and required refusal conversion on twice as many cases (Curtin, Presser, and Singer 2000). A Pew Research Center study (2004) found that the response rate on a standard telephone survey that was in the field five days fell from 36% in 1997 to 27% in 2003, with the difference being due to 20% less cooperation of households that were reached. Thus the refusal rate has increased as surveys have become more common. Yet high response rates are demanded in some situations, most notably the 75%-80% rate expected by the Office of Management and Budget (OMB), which must approve most surveys by the U.S. federal government.

Internet surveys were originally seen as a possible panacea, with the ability to contact large numbers of potential respondents at minimal cost. However, as they have become ubiquitous, unit nonresponse has become a problem with Internet surveys too (Vehovar et al. 2002). Krosnick and Chang (2001) provide rare detail on response rates on two Internet pre-election surveys taken in June–July 2000. Only 18% of the 12,573 Harris Interactive panel members who were contacted participated in the survey. A Knowledge Networks survey had a response rate of 28%, calculated as the proportion of people who were originally phoned and asked to join an ongoing panel who agreed to join the panel, had the WebTV receiver installed, and participated in this particular survey. Calculated as a proportion of 7,054 ongoing Knowledge Networks panel respondents who were asked to participate in this survey, the rate of participation was 70%. For comparison, a parallel telephone survey by the OSU Center for Survey Research obtained interviews from 43% of the 3,500 eligible phone numbers.¹

These figures show that the phone survey had a much higher response and the Harris Interactive survey had the lowest. However, the lower participation rates for the Internet surveys are partially balanced by the larger size of their ongoing panels, usually around 10,000 for Harris Interactive,

so that there were more participants in the Internet surveys than in the telephone survey at a lower cost. Survey researchers have to decide how they view this trade-off between response rate and sample size.

Studies of Cooperation

Several studies have sought to assess the importance of some of the factors that have been hypothesized to affect survey cooperation. Some studies obtain a record of the participation process by having interviewers document their initial interactions with people, both those who agree to participate and those who do not. This permits the researcher to be able to estimate a response propensity function and develop models about nonparticipants. Although it is usually not possible to find out information about people who do not participate in a study, some estimates can sometimes be obtained. For example, people who would not cooperate may be like the actual respondents who were more reluctant to be interviewed.

Some research looks at single factors that affect cooperation. For example, studies show interviewer effects on unit nonresponse. Groves and Couper (1998, 211–13) examined face-to-face surveys and found that interviewers with more experience and more confidence had more success in obtaining respondent cooperation. They also found (chap. 8) more negative comments about survey participation when people were contacted on the phone versus more questions about the survey in face-to-face contacts, with fewer of both types of behavior being obtained by experienced interviewers, probably because of their skills, including their ability to anticipate concerns.

In a rare study of survey administration effects on cooperation, Marton (2004) analyzed a series of 53 monthly Buckeye State Poll surveys taken by the Ohio State University’s Center for Survey Research from late 1997 through early 2002. After detrending the time series, she found that cooperation rates were lower in months when the survey was longer and in months when there were more new interviewers.

Brehm (1993, chap. 4) was able to do a multivariate test of parts of his survey compliance model through secondary analysis of past surveys, though few surveys have all the needed measures of the respondent’s relationship to strangers, to the interviewer, to the interview, and to self-image. As an example of his findings, the relationship to the interview was much more important than relationship to strangers for the 1989 Detroit Area Study. Relationship to the interview was again the most important factor for the 1988 NES, relationships to the interviewer and to self-image were of moderate importance, and relationship to strangers was the least
important factor. Although generalizing from such few studies is hazardous, Brehm's analysis suggests that interest in the topic, time for the interview, and having had a bad past experience with interviews are among the most important factors in affecting survey compliance.

An experiment by Groves, Presser, and Dipko (2004) finds some support for the leverage-saliency theory. The researchers sampled from lists of four groups of people (e.g., teachers), developed tailored introductory survey requests that were appropriate for each group, and then randomly assigned the tailored requests to respondents so that the effects of tailored requests could be measured experimentally. They found that the odds of cooperation were about 40% higher on the topics of interest and that monetary incentives did not have significant added effects. There is a potential biasing effect if tailoring leads people who are more interested in a topic to be more likely to agree to be surveyed, but the authors estimate that effect as being only in the 1%–2% range. Thus this study implies that tailoring the request to participate would not lead to serious response bias.

Some types of surveys encounter specific problems. Nonresponse in the election exit polls conducted by the Voter News Service (VNS) was primarily due to refusals. In 1996, for example, the average refusal rate was 33%, plus 10% "misses" when the interviewer was too busy to intercept the voter or when the voter did not pass by where the interviewer was stationed (Merkle and Edelman 2000). Laws in some states regulate how close exit interviewers can be to polling places; in Texas, for example, barring interviewers from being within 100 feet means that people who park close to the polling place never have a chance to be interviewed. By 2000, the VNS response rate fell to 51% (Konner 2003). Merkle and Edelman (2002) find that the response rate was affected by such factors as the number of questions on the front of the questionnaire and how far from the polling place the interviewer must stand. Fortunately, response rate by precinct had little power in explaining deviations of the precinct's exit poll result from its actual vote. Still, nonresponse is considered to be one of the main factors that contributed to exit polls' overestimating the Kerry vote by about 5.5% nationally in 2004.

Human interviewers increase response rates. A recent development in phone surveys is interactive voice response (IVR), the system in which people are phoned and expected to answer recorded questions. That approach is not likely to be very effective, since respondents often hang up as soon as they realize they are listening to an impersonal recording. Beginning with a human contact to request their cooperation for a survey is likely to be more effective, even if the person is switched to the IVR system for the actual survey, though people may still not stay with a very long IVR survey. Tourangeau, Steiger, and Wilson (2002) evaluated four IVR studies by the Gallup Organization that began with a human interviewer. They found that IVR breakoffs were more common than in parallel CATI interviews; breakoffs varied with the length of the survey, with just over 1% on a 6-item survey and 31% on a survey averaging 24 items.

The factors that affect cooperation in surveys of businesses are somewhat different. Cooperation is likely to be adversely affected by the company's staff resources, which can get strained when the economy is poor and as government filling requirements increase. Response rates in surveys of businesses have not fallen as much as they have for household surveys, but Willimack, Nichols, and Sudman (2002) stress the importance of burden reduction in order to obtain cooperation in business surveys, such as through providing alternative response modes and relaxation of time and data needs.

Reasons for Noncooperation

Some surveys record the concerns that people raise when they refuse to be interviewed. Brehm (1993, 53–56) finds that blacks are more likely to be rated as suspicious of the interviewer whereas Hispanics are more likely to raise confidentiality concerns. Higher-income people are more likely to say they are "too busy" to be interviewed, while older people are less likely to give that explanation (Brehm 1993, 59–61). People who are less interested and informed about politics are more likely to refuse at first to participate in a political survey (Brehm 1993, 61–64).

The comments that people make when being asked to participate can also provide useful information about refusals. Cooper (1997) analyzed the comments that respondents made in the introductory conversations in the 1990 NES. People who had said they were "not interested" were more likely to have missing data, reply "don't know," and say little on open-ended questions, suggesting a lack of cognitive involvement in the interview. By contrast, people who said they were "too busy" were less likely to engage in satisficing behavior, with less missing data. Cooper agrees with Brehm that people who are less interested in politics are more likely to refuse to participate in a political survey, so that the final sample overstates the level of political interest among the public.

Cooper and Groves (2002) analyzed comments made in the introductory interactions in 101 telephone interviews. Time-delay statements (such as "I'm busy now") were common (63% of the calls), as were
questions (35%), and negative statements (27%). As to be expected, people making negative statements were less likely to cooperate, with a cooperation rate of only 56%, 26% lower than the rate for people who did not make such comments. Yet that still means that most people who made negative statements still cooperated in the end. Interestingly, the cooperation rate was not lower for people who made time-delay statements or who asked questions.

Demographic Correlates of Cooperation

Much of the literature on survey cooperation looks at the demographics of cooperation, since it is easier to assemble data on demographics than on psychological correlates. Some studies use information in the sampling frame about the nonrespondents or ask others (proxies) about the nonrespondents. The demographics of participants are often compared with Census Bureau data to ascertain characteristics of nonparticipants, though it is important to realize that such comparisons combine contact and cooperation issues. Additionally, it is possible to see what types of people are most likely to drop out of multiwave panel studies in order to gain information about patterns of cooperation.

Several examples of common findings show what we know about the correlates of survey participation in the United States. Age is one common correlate, though the pattern is complicated because the age groups with higher cooperation rates have lower contact rates. Young people are more willing to participate if they are successfully contacted, but they are less likely to be at home, which results in underrepresentation of young people in some surveys. For example, a greater cooperation rate was found in government face-to-face surveys among households in which everyone was under age thirty (Groves and Couper 1998, 133–37) and in families with young children (Groves and Couper 1998, 140–41). Still, the greater difficulty in finding younger people at home led to an underrepresentation by as much as 6%–9% of people under thirty in the face-to-face surveys taken by the General Social Survey (GSS) and National Election Studies (NES) from 1978 to 1988 (Brehm 1993, 26–29). The patterns reverse for older people: they are less likely to cooperate but are more likely to be at home, which typically balances out in overrepresentation of older people in surveys. Kaldenberg, Koenig, and Becker (1994) obtained a lower response rate from a mail survey of retired people with increasing age. Older voters are also less willing to participate in election exit polls, though that difference diminishes when older interviewers are used (Merkle and Edelman 2000). Though older people were more likely to refuse to participate (Groves 1989, chap. 5), their greater likelihood of being at home led an overrepresentation by 2%–9% of people over sixty-five in the face-to-face 1978–88 GSS and NES surveys (Brehm 1993, 26–29).

Surveys routinely underrepresent men. Brehm (1993, 30–31) finds that NES and GSS face-to-face surveys undercount men by 4%–8%. He concludes that men are less likely to be at home and more likely to refuse. Also, phone surveys tend to get a higher proportion of women than men because there are many more female-only households than male-only households and because women are more likely to be at home and to answer the phone (see also the discussion of gender and within-household selection of respondents in chapter 10). Women are also more willing to participate in election exit polls (Merkle and Edelman 2000).

The differences are less consistent on race. NES face-to-face surveys overestimate the proportion of blacks by as much as 5%, whereas its phone surveys underestimate the proportion by 2%–6% (Brehm 1993, 29–30). Brehm suggests this pattern is due to blacks' being less likely to refuse to be interviewed, along with a lower proportion of blacks having telephones. Groves and Couper’s (1998, 137) multivariate analysis finds a greater cooperation rate among Hispanics in government face-to-face surveys, whereas racial differences were not significant when the age composition of the household was controlled.

Educational differences are also complicated. NES and GSS face-to-face surveys have higher proportions of people who did not graduate from high school than Census Bureau statistics show, but NES telephone surveys have lower proportions of non–high school graduates (Brehm 1993, 31–32). People with higher education are more likely to return mail questionnaires (Groves 1989, chap. 5).

In Marton’s (2004) analysis of 53 monthly Buckeye State Poll surveys described above, months with higher response and cooperation rates generally had distributions of race, gender, and income that were closer to census values. The exceptions involve age and education, partially echoing the Keeter et al. (2000) result that a study with a lower response rate had an education level more similar to census values. There was not much variance between months in some of these monthly series, but this study still suggests that surveys with higher response rates are likely to be more representative demographically.

Some plausible factors have not been found to affect cooperation rates. In particular, Groves and Couper’s (1998, 123–31) analysis of several federal government surveys find that cooperation rates were not affected by the number of adults in the household or amounts of discretionary time.
However, there was greater cooperation among people with lower housing costs, which suggests that higher cooperation was obtained with lower-income people.

As to Internet surveys, a 2001 telephone and Internet study by the Council for Marketing and Opinion Research (Sheppard 2001) found that males were more likely to refuse the Internet study. In both modes, males reported having refused during the previous year more often than did females. Also, those with higher educational levels and higher income were more likely to report having refused to participate in surveys in the past year, whereas older people were less likely. No differences were found among race or ethnicity categories.

Note that some of the separate correlates of cooperation discussed above overlap with one another. For example, face-to-face surveys report less success in areas with high crime rates and high population density, in large metropolitan areas, and in housing units that are in bad repair (Groves and Couper 1998, chap. 4), situations that overlap to a considerable degree. Also, Groves and Couper (1998, chap. 4) find greater success in households with several adults, with children under age five, and with adults over age seventy, all three of which may simply be instances of greater success with households in which there are more likely to be adults at home.

Multivariate analysis is required to determine which factors have separate effects on cooperation. The correlates that have been discussed so far are important, but some may be spurious artifacts of other correlates. Two studies show that many demographic correlates survive such multivariate controls. Brehm’s (1993, 132) multivariate analysis of demographic factors in the 1988 NES study found significant effects associated with income, gender, Hispanic ethnicity, race, and whether the respondent and interviewer were of different races (which depresses participation), with the participation status of 84% of the cases predicted correctly from these factors. Groves and Couper’s (1998, 146–50) analysis found low housing costs, young and elderly households, and households with young children had greater cooperation in government surveys, whereas single-person households had lower cooperation, with several other factors controlled.

While there have been many studies of the demographics of nonparticipation, demographic correlates do not necessarily translate into response bias. For response bias to result from demographic imbalance, the demographic in question would have to relate to the behavior being measured. For example, an election survey that overrepresented women would be off the mark if there is a gender gap in voting, but an election survey that overrepresented left-handed people would be unlikely to produce biased results. As discussed later in this chapter, it is possible to weight the sample to correct for some demographic differences, but weighting makes assumptions about nonresponse that may be wrong. As a result, understanding the factors that lead to nonparticipation is more important than focusing on the demographics of participation.

It is worth emphasizing too that willingness to participate in surveys may differ between countries, so the extent of nonresponse error is likely to differ across countries in multicity surveys. The correlates of cooperation and effects on response bias are also likely to vary across countries.

Panel Surveys

Panel surveys face special problems maintaining cooperation. People may refuse to be reinterviewed, depending in part on recollections of their feelings about the previous interview. The greatest threat to reinterview refusal is usually at the first attempt. For example, the 1984 Survey of Income Program Participation (SIPP) experienced a 5% nonresponse rate in the first wave, 3% more in the second wave (not counting people who moved), 2% more in each of the next two waves, and so on, until the total attrition rate not due to moving was 16% by wave 9 (Bailar 1989, 6–7). In practice, people who refuse to be reinterviewed in one wave are sometimes willing to be reinterviewed in a subsequent wave. SIPP does try to recapture nonrespondents from previous waves; for example, 5% of fifth-wave respondents had missed at least one previous interview (Kalton, Kasprzyk, and McMillen 1989). Kalton and Brick (2000) detail the issues involved in weighting panel surveys to handle wave nonresponse.

When looking at what types of people are most likely to continue cooperating on panel studies, Lepkowski and Couper (2002, 265–68) found that their cooperation was related to community attachment (with renters less likely to participate). The pattern was somewhat different on an election study panel survey (Lepkowski and Couper 2002, 268–71), with cooperation being affected by survey experience (people who were reluctant in the first wave being less cooperative in the second) and accessibility (people who moved in the past three years participating less).

Panel studies have both additional advantages and potential difficulties as regards unit nonresponse issues. Panels with overlap and/or rotation allow comparison of the remaining part of the panel to the new part of the sample. However, there are potential definitional problems with household panels, since the composition of many households changes over time as people move in or out.
Techniques for Increasing Response Rates

The most common strategies for increasing the response and cooperation rates include advance letters, incentives, tailored requests to participate, refusal conversions, and easing the technological burden. These strategies will be discussed in the remainder of this section, with discussion of how they play through for interviewer-administered surveys, self-administered surveys, and Internet surveys. Note that substitutions to compensate for noncooperation are considered undesirable for the same reasons they are undesirable as a means to compensate for noncontact.

Advance Letters

Advance letters (also called prenotice letters) are sometimes sent for surveys so that the respondent recognizes that the interview attempt is legitimate rather than a disguised sales attempt. Trangott (1987) finds that advanced contact increases response rate on face-to-face surveys by more than 10%. There do not seem to be backfire effects against advance letters, except for a study showing one if the letter is sent a month in advance of contact (Groves and Couper 1998, chap. 10). Although studies differ on the impact of advance letters for telephone surveys, a study in Arizona (Goldstein and Jennings 2002) found a 16% gain in cooperation rates among contacted households with letters, with a more accurate representation of older people and men. Often telephone numbers are selected randomly for phone surveys, without knowledge of the person’s name or address, but advance letters can be sent by using services like Telematch, which do reverse searches to obtain names and addresses for a list of sampled phone numbers (see http://www.gannetofset.com/html/telematch.html). Advanced notification is also useful in e-mail and Internet surveys (Cook, Heath, and Thompson 2000; Kaplowitz, Hadlock, and Levine 2004), to the extent to which it is possible to do so. Advanced contact is especially necessary for interviewing elites.

As to the content of advance letters, Dillman (2000, 156) suggests that they be short, personalized, and worded in a positive manner. The letter should give only a general description of the survey topic but should stress how important the survey is. If an incentive is being sent with the questionnaire, the letter can indicate that a “token of appreciation” will be provided.

One approach to avoid in Internet polls is bombarding newsgroups and other e-mail lists with announcements of them. This approach is considered equivalent to spamming, and it has resulted in flaming of even well intentioned researchers.

Incentives

An increasingly popular approach for dealing with noncooperation is giving respondents incentives to participate. The social exchange interpretation of an incentive is as a kindness that evokes reciprocity as well as demonstrating the thoughtfulness of the giver, so long as it is not seen as an indication that the survey will be unusually burdensome. An alternative interpretation is in terms of the cost-benefit trade-off, viewing the incentive as a payment for services, so that it may be more effective with low-income respondents. Incentives can also be targeted, given only to respondents who are deemed the least likely to cooperate. For example, Groves, Singer, and Corning (2000) show that incentives have a larger effect for people with low involvement in the topic of the survey but have less effect for people with high involvement, who have a high cooperation rate even without incentives (also see Groves, Presser, and Dipko 2004).

There is now considerable research on the types of incentives that are most productive. The usual advice is to use prepaid incentives rather than the promise of an incentive (Church 1993). Singer, van Hoewyk, and Maher (2000), for example, found that prepaid incentives increased response rates even for telephone surveys by 10%, whereas promised incentives did not increase responses. The effect was not an interviewer expectancy effect, since it occurred even when the interviewer did not know that the person received the incentive. Five-dollar incentives were cost effective in that study, reducing the average number of phone calls per case from 10.22 to 8.75. Fortunately, prepaying incentives did not decrease the chance of the person’s participating in a later survey, even if that later survey did not include an incentive (Singer, van Hoewyk, and Maher 1998). There is some fear that people who participate only with incentives will not bother answering all items, but item nonresponse was not increased in this study; indeed incentives seemed to decrease item nonresponse among older people and nonwhites.

Incentives generally are $2 or $5, with the amount obviously increasing over time with inflation. An early study by James and Bolstein (1990), for example, reports a response rate in the 70% range in a mail survey with a $1–$2 incentive versus 54% when no incentive was given. Larger incentives lead to higher response rates, but there seems to be a diminishing effect with larger incentives. With small incentives, it is usually easier to include cash in the envelope than to write checks, which would have to be kept on the survey organization’s books for several months if they are not cashed. Larger incentives are increasingly being used, as when the
2000 NES survey gave a $20 incentive, which was increased to $40 in the week before the election.

A large-scale study by Trussell and Lavrakas (2004) provides further detail on the usefulness of monetary incentives. Respondents for a mail survey were randomly assigned prepaid incentives from zero to $10, with respondents contacted in advance by telephone. Respondents who were given some incentive were more likely to cooperate, with higher incentives generally leading to more cooperation. Larger incentives had the greatest effects among people who could not be contacted on the phone and those who refused to participate when contacted on the phone; this suggests that incentives should be targeted to people who have not shown a willingness to cooperate.

Nonmonetary incentives can be effective for particular types of respondents. For example, movie tickets are reasonable incentives for students, whereas bottles of wine are better incentives for elites than money would be. Material incentives such as pens are sometimes used and have been found to increase the response rate (Willimack et al. 1995), but funds used for the added effort and for packaging may increase participation more if distributed as financial incentives. Promising a contribution to a charity in the respondent's name costs less than prepaid incentives, since only people who participate are rewarded, but the evidence seems to be that charity incentives are not effective (Warriner et al. 1996; Hubbard and Little 1988; Furse and Stewart 1982; cf. Robertson and Bellenger 1978).

Some surveys offer lotteries as incentives, giving participants a chance to win a large reward. Lotteries have the advantage of rewarding only actual participation, whereas some people who are sent monetary incentives for mail surveys still do not participate. Also with lotteries it is necessary to send money to only one person rather than go to the trouble of sending out money to a large number of people. If it is a large survey in which the chances of winning the lottery prize are small, however, people may not participate if they understand the small chance of winning—though this does not seem to deter people from participating in state lotteries. Still, it is not clear that lotteries are as effective as prepaid incentives for survey participation.

Incentives can be given very efficiently in Internet surveys. Participants in one-time surveys are often awarded a gift certificate number that they can use at an Internet merchant's site. People who have joined ongoing survey panels are sometimes given points that can be accumulated for large prizes. These techniques reward only people who finish the surveys, while turning over the distribution of incentives to firms that specialize in that.

Tailoring the Survey Request

In understanding cooperation with interviewer-administered surveys, it must be kept in mind that this is a very quick decision by respondents. Typically, people have no idea who is on the phone or at the door when they answer. They must quickly figure out how to categorize the survey attempt: if it is real rather than a disguised sales attempt, if the source is one they respect (such as their home state university), if the interviewer seems friendly and competent, if they have some time at the moment, and so on. This makes the introductory survey request particularly important.

The importance of how the interviewer begins the conversation with the respondent is shown by Oksenberg and Cannell's (1988) finding that 40% of refusals occurred during the first few sentences, 50% later in the introduction, and only 10% after the interview has started. Thus the first 15–30 seconds of an interviewer-administered survey should be seen as crucial. Unfortunately, the respondent selection techniques that yield the most representative samples (see chapter 10) take longest, are more intrusive, are more likely to cause suspicion, and therefore are most likely to lead to nonresponse.

It was originally felt that there should be standard introductions to surveys, using the best possible justification of participation. Few interviewer-administered surveys still adopt that approach. Even survey researchers who believe in standardization now recognize that scripting is not effective. A scripted introduction can actually lead to problems, raising concerns that would not otherwise come to the person's mind. Morton-Williams (1993) was one of the first studies to show a much higher response rate with unscripted introductions, 17% higher in that instance. Going back to the discussion of conversational interviewing (chapter 4), a Dutch study found greater success in obtaining interviews when interviewers began with a conversational approach than when they followed a standardized script in asking for the interview (Houtkoop-Steenstra and van den Bergh 2002), possibly because the interviewers sounded more natural and/or were more motivated.

The modern approach to survey participation is to "tailor" the pitch to participate to the potential respondent. Different respondents have different values, so no single strategy for achieving cooperation will be optimal for all potential respondents. The interviewer is supposed to learn to read cues about the best way to motivate particular types of people. Tailoring is particularly effective in face-to-face interviews in which the interviewer can size up the respondent and his or her household. It is also used in
telephone interviews, with interviewers being told to use their natural skills to respond as the conversation develops. The tailoring approach can be taught as part of interviewer training. Groves and McGonagle (2001) report a study in which interviewers were trained to learn to identify and classify the cues of reluctance from householders. Interviewers with that training had better response rates than the control group. In their analysis of the introductory interactions in 101 phone interviews, Couper and Groves (2002) found that almost all interactions included at least one statement by the respondent that was informative; only 8% of the calls they analyzed had no comments at the introductory stage that could be used in tailoring the request to participate in the survey. The cooperation rate was higher when interviewers followed up with a tailored approach, though the number of cases in key cells is too small to prove this effect beyond question.

The most common approach is a brief introduction that gives essential information but is tailored to the respondent’s concerns. Guidelines from the American Association for Public Opinion Research (AAPOR) suggest that telephone interviewers begin by identifying themselves, stating their affiliation, indicating the sponsor of the survey, summarizing its purpose, explaining how the person was selected (e.g., randomly), encouraging cooperation, and verifying the phone number. Lavrakas (1993) favors a brief introduction, establishing credibility, and quickly getting the respondent to answer the first question so that she or he has become committed to answering. As an example, the box on the opposite page gives the introduction used by Ohio State University’s Center for Survey Research for one of its polls.

The social exchange theory has been blended with Cialdini’s compliance principles in Dillman’s (2000, chap. 1) tailored design approach for self-administered questionnaires. Dillman lists several possible ways of providing rewards to survey participants, including showing respect to them, thanking them for participating, asking for their advice, supporting their values as represented by their group attachments, giving them tangible incentives, making the survey interesting, providing them social validation by emphasizing that people similar to them have participated, and telling them that the opportunity to respond is scarce. The costs of participating can be reduced by not making the respondent seem dependent on the researcher, not embarrassing the respondent (as by making sure the questionnaire is not too hard to fill out), making participation convenient, keeping the questionnaire short with easy formats, minimizing requests for sensitive information that people may not want to disclose, and making

---

Introduction-Selection Sequence

GENERAL INTRODUCTION:

(PLEASE TAILOR WORDING TO REDUCING REFUSALS!!!)

· Hello, my name is _____ and I’m calling from Ohio State University (in Columbus).

· Your household has been randomly sampled to participate in the ____ Poll. This is a very important scientific survey of Ohio residents that we are conducting this month for professors at Ohio State and . . .

· (The interview takes about 10–15 minutes to complete. It’s entirely confidential, your household’s participation is voluntary, and I really need and will appreciate your help.)

· (This month’s questionnaire asks a series of questions related to . . . In addition, there is a series of background questions to help the professors who will be analyzing the data.)

WITHIN HOUSEHOLD RESPONDENT SELECTION:

· For this survey, I’d like to interview the person in your household who is at least 18 years old and who had the last (i.e. most recent) birthday. (Is that you or someone else?)

NUMBER VERIFICATION:

· Before I begin the interview, may I please verify that I have reached you on ____

PROCEED TO QUESTIONNAIRE WHEN SELECTED RESPONDENT IS AVAILABLE. DETERMINE WHEN BEST TO CALL BACK IF SELECTED RESPONDENT IS NOT AVAILABLE AND RECORD INFORMATION ON CALL-SHEET.

Source: Ohio State University Center for Survey Research

the requests to participate similar to those in other surveys that the person has participated in. Trust can be established by emphasizing an authoritative sponsor, making the survey seem important (such as with professional-looking questionnaires), and giving incentives in advance.

Several approaches are used to boost the cooperation rate, in line with the different compliance approaches reviewed early in this chapter. Some studies identify their sponsorship as a means of justifying participation. Surveys from state universities usually emphasize their school’s name when interviews are conducted in their state, and surveys from the nation’s most prestigious universities use their school’s name even in interviews outside their state. Some research shows the highest response to surveys sponsored by the federal government, next to those with university sponsorship, and lowest to those with commercial sponsorship. A related approach is using endorsements of the survey by people or organizations that respondents would recognize and value positively.

A study by Dijkstra and Smit (2002) found that personal appeals (“I would like your cooperation”) were more effective in inducing respondents to participate in the survey than referring to authority (their sponsoring university) or social validation (that most people enjoy the interview).
Their results on the effectiveness of tailoring were not consistent. Repe-
ting the person's objection diminished participation.

The evidence accumulated in a meta-analysis by Singer, Von Thurn, and Miller (1988) points to higher response rates with assurances of con-
fidentiality, but the effect is small and holds only when sensitive questions
are involved. With nonsensitive topics, they found that confidentiality
assurances decrease response rates, but that effect is very small. However,
mail return of the census, according to a Survey of Census Participation
after the 1990 census, was significantly affected by confidentiality and pri-
vacy assurances, with an estimated 3–6 percentage point increase in
response rates (Singer, Mathiowetz, and Couper 1993). The importance of
privacy concerns is clear in a finding that mail return of the 2000 census
was significantly less among people concerned that it might be used for
law enforcement purposes, such as by the FBI to trace troublemakers and
for locating illegal aliens (Singer, van Hoewyk, and Neugebauer 2003).

Telephone interviewers are usually given standard "fallback" argu-
ments to use when respondents ask questions or give specific reasons for
not participating. Fallback statements used by the Ohio State University
Center for Survey Research are shown in the box on the opposite page. For
example, the person who says he or she is too old to participate is told that
it is important to represent the views of older people in the survey, whereas
the person who asks about sponsorship will be assured that the sponsor of
the survey is legitimate. Obviously an important part of the interviewer's
job is to figure out the most appropriate way to tailor the request to this
particular respondent. For that reason, interviewers can be seen as
"request professionals" (Groves and Couper 1998, 18), and the best inter-
viewers have learned how to request cooperation so as to succeed.

Of course even tailoring an appeal for cooperation requires main-
taining the interaction of the interviewer and the respondent. In face-to-face inter-
views, this is known as the problem of keeping the interviewer's foot in the
door. So long as the door is not completely slammed in the interviewer's
face, there is a chance of convincing the person to give the interview. This
problem is even more serious for telephone surveys, since it is very easy for
people to hang up when interviewers introduce themselves. If the person
stays on the phone, the interviewer can attempt to tailor the pitch to the
respondent, but that requires the person to stay on the phone. Survey
organizations that try to convert refusals instruct their interviewers to be
very polite in initial encounters, so there is a possibility of a later conver-
sion by another interviewer if the respondent initially refuses. Surveys with
nonprobability sampling that permit substitutions for refusals instead seek
to maximize their number of interviews per hour, which means the interviewer is better off contacting the next person on the list than trying to maintain the interaction with a reluctant respondent.

Tailoring the survey request is more difficult in mail surveys. Using different approaches for different subpopulations is not feasible, since little will be known about differences between respondents before they are sent the survey instrument. At best, tailoring to subpopulations may be possible when they appear on separate lists or if side information can be obtained before mailing them the survey.

Tailoring is similarly hard to achieve in Internet surveys, but minimally it is important to appeal directly to the respondent to participate, such as by sending e-mail surveys in a personalized manner rather than as part of a mass mailing that shows e-mail addresses of large numbers of people. Additionally, Internet surveys require an inviting welcome screen so as to minimize unit nonresponse. Similarly, the first question should be one that will interest people and be easy to answer, as a way of drawing people into the survey.

The Council of American Survey Research Organizations (CASRO) suggests that online e-mail surveys have been most effective when the subject line indicates the topic (e.g. "Survey Invitation for Pet Owners"), there is an explanation of how the e-mail address was obtained, the sponsor and company conducting the survey are identified, the purpose of the research is explained, the approximate completion time is stated, the incentive for participation is described, and valid contact information is given.

Refusal Conversion

Many survey organizations recontact people who initially refuse and try harder to get them to participate. Interviewers who excel at conversion of refusals often make this second attempt. More than a quarter of the people who originally refuse to participate will often cooperate on a second attempt, which shows that the quick choice not to participate is not necessarily a permanent decision. This partially suggests that the initial interviewers were not fully effective, but it also points out that it can be useful to try different approaches on the same person. Studies show that the verbal reasons people give for refusals are not stable under alternative cues given to the person, so it is often worthwhile to try again to get the interview.

The initial interviewers can be given "refusal forms" to fill out when someone is not willing to be interviewed. The interviewers describe the type of appeal they used and the person’s response, so that the conversion attempt can try a different appeal or a different way to mollify the person’s concerns. The box on page 186 shows the Refusal Report Form used by Ohio State University’s Center for Survey Research CATI unit. Conversion attempts are not usually made when the person was rude, mad, or obscene, or directly said not to call back ("hard refusals"), but many refusals are not that clear-cut, as when the person said they did not have time at the moment or when the person answering the phone hung up before the interviewer could ascertain whether he or she was the intended respondent ("soft refusals"). Some surveys switch to a different interviewing mode to try to convert refusals, trying a phone interview when a face-to-face interview failed, or vice versa. Sometimes a refusal to a questionnaire is followed up with a personal interview.

Survey organizations must decide how much effort to put into refusal conversions. Refusal conversions may be costly without being very successful, and, even worse, some studies find them not to be representative of all refusals (Lynn et al. 2002). It may be possible to use the initial comments during refusals to decide who is worth recontacting. Groves and Couper’s (1998, 261-65) multivariate analysis shows that people who ask questions during the first attempt to survey them are more likely to cooperate in later contacts, whereas those who make negative comments or say they’re too busy are less likely to cooperate eventually.

To give an idea of the stakes involved, a report on a CATI survey at the time of an election in 1996 required 7.56 calls on average to complete interviews with refusal conversions versus 4.01 for nonrefusals, with a cost 1.6 times the cost for nonrefusals (Mason, Lesser, and Traugott 2002). The refusal conversions increased the response rate from 46% to 61%, but with more item-missing data—19% on the crucial vote intention item versus the 12% level for nonrefusals and with about 25% having missing data on explanatory variables compared to 11% for nonrefusals. This example nicely illustrates the considerations in deciding whether refusal conversions are cost effective. They are costly and many do not provide full data anyhow, but they can boost the response rate considerably, and imputation can help handle the item-missing data that can result.

Instead of trying to convert refusals, another approach is to seek to conduct a very short interview (possibly with someone else in the household) to collect some demographic information that can be used for later analysis. The discussion below of postsurvey statistical adjustments shows how this information can sometimes be employed. A variant of this is to sample a subset of the refusals to get more information about them.
Refusal Report Form

Interviewer # __________  Case id __________

1. Did the person who refused have the last [most recent] birthday?  
   Yes  
   No  
   Uncertain

2. Demographics of the person refusing  
   GENDER  
   Female  
   Male  
   Uncertain  
   AGE  
   Child  
   Adult < 30  
   30–59 Yrs  
   60 or older  
   Uncertain  
   RACE  
   Asian  
   Black  
   Hispanic  
   White  
   Uncertain

3. How much interaction between you and the person refusing has gone through?  
   No interaction at all  
   (immediate hang-up)  
   1 2 3 4 5 6 7  
   Much interaction (conversations beyond the standard introduction, questions and answers, explanations, etc.)

4. Reason for refusal (check all that apply):  
   Immediate hang-up/uncertain of the reason  
   Time-related concerns/reasons/excuses  
   Objects to surveys/topics/sponsors  
   Concerned about confidentiality/legitimacy of survey  
   Other reasons (specify) __________

5. Strength of refusal:  
   VERY WEAK 1 2 3 4 5 6 7  
   VERY STRONG

   Respondent attitude:  
   VERY POLITE 1 2 3 4 5 6 7  
   VERY RUDE

   NOT AT ALL ANGRY 1 2 3 4 5 6 7  
   VERY ANGRY

6. Did you tell the person:  
   A. How s/he was sampled?  
   Yes / No

   B. The nature/purpose of survey beyond the standard intro?  
   Yes / No

   C. Confidentiality?  
   Yes / No

   D. How the data would be used?  
   Yes / No

   E. Verification with supervisor/sponsor?  
   Yes / No

7. What can you recommend, if anything, for gaining respondent/household cooperation if a conversion attempt is made? __________

Source: Ohio State University Center for Survey Research

particularly when fully studying the refusals would be expensive. Statistically, data obtained from the refusals can then be treated as a stratum (see chapter 10), using stratified sampling formulas (Kalton 1983). A related approach would be to do extensive refusal conversion in the first phase of a survey, to estimate what the optimal level of persuasions would be for the rest of the survey.

While refusal conversion is generally associated with interviewer-administered surveys, attempts should also be made to elicit responses from people who do not respond to Internet surveys (Best and Krueger 2004, 77) and those who do not send back mail questionnaires on the first mail-
ing. Dillman’s Tailored Design Method will next be explained as a method that combines tailoring with persistence in seeking survey cooperation.

The Tailored Design Method

Mail surveys are notorious for low response rates, often as low as 20%. As a result, there has been considerable focus on methods to increase cooperation with mail surveys. In particular, Dillman (2000, 27–28) claims that mail surveys can average a 75% response rate when his Tailored Design Method is followed. This method combines insights on how to obtain compliance from respondents and very practical advice gained from large numbers of studies. Five features are emphasized (Dillman 2000, chap. 4): a questionnaire with a “respondent friendly” design (see chapter 6), stamped return envelopes, personalized correspondence, token financial incentives given in advance, and five mailings as described below.

Dillman’s five mailings begin with an advance letter, alerting the respondent to a survey that will arrive in a few days and requesting cooperation. The questionnaire mailing includes the incentive and a cover letter explaining the importance of responding. A thank-you postcard is sent a few days later, asking for the questionnaire to be returned soon if it has not yet been. This is followed in two to four weeks by a replacement questionnaire for people who have not responded by that date. Those first four mailings are all by first-class mail, but the fifth is by a different mode, such as special delivery or Federal Express (or even by telephone if the person’s phone number can be obtained).

Dillman (chap. 4) also gives detailed advice on each mailing. The cover letter should have a short introductory paragraph indicating what is being requested and a second paragraph explaining why the survey is useful and important. The letter must also stress that answers will be confidential and that participation is voluntary. Dillman recommends a casual mention of the incentive, such as “We have enclosed a small token of appreciation as a way of saying thanks for your help” (162). The cover letter should provide information on how to contact someone if the respondent has questions. Dillman urges that the letter have a real signature and a postscript that again expresses thanks, as well as explaining what to do if the person is not eligible for the study.

When the questionnaires are mailed, they are usually given identification numbers so that follow-up mailings can be employed to increase the response rate. It is usually best not to conceal identification numbers, because of the problems that could ensue if respondents realize that there is a concealed number. For example, an identification number under the
return postage stamp in a survey of Michigan state legislators in the early 1970s was detected, and the story circulated immediately around the legislature resulting in bad feelings toward the sponsoring university. If sensitive data, such as on drug use, are being collected, the identification number can be put on a separate postcard that respondents are asked to send back when they return the completed questionnaire.

The message on the postcard at the third mailing should be much simpler than in the original cover letter. It is intended to remind people of the survey. The card should begin by saying that a questionnaire was sent and why, which helps in situations in which the person never noticed the original mailing. Then comes thanking people who have sent the questionnaire back and asking others to do so “today.” The message should end with an explanation of how to get a replacement questionnaire if it is needed.

The replacement questionnaire in the fourth mailing should also be accompanied with a cover letter, but the letter’s message at this stage should be phrased more strongly. It should emphasize that other people have responded and again highlight the importance of the survey. The fifth contact should be viewed as the final attempt to obtain cooperation, using more expensive contact methods if increasing the response rate is urgent.

Dillman (2000) gives the specific evidence on which the Tailored Design Method is based, but it is useful to mention a few findings regarding how much effect particular parts of the process have. A meta-analysis of journal articles based on mail surveys from 1961 through 1985 (Fox, Crask, and Kim 1988) finds the largest effects were university sponsorship (9% effect), advance letters (8%), and stamped return postage (6%). Smaller significant effects were found for postcard follow-up (3.5%), using first-class outgoing postage (2%), and a green questionnaire (2%). Notification of a cutoff date and the postscript requesting cooperation did not have significant effects. Incentives had effects but with diminishing returns; the added response rate increment was largest for using a small incentive and was more minimal with large incentives.

Further evidence on the size of effects comes from a series of experiments testing approaches to increase the response rate to the U.S. decennial census (Dillman 2000, chap. 9). Government surveys are not immune to the response rate problems of mail surveys, even when compliance is mandatory by law, as shown by the 75% response rate to the census in 1980 and the 65% rate in 1990. The experiments showed that response rates could be improved by using a respondent-friendly design (a 2%–4% effect), an advance letter (5%–7%), a postcard thank-

you reminder (5%–8%), a replacement questionnaire (6%–11%), and a prominent reminder that response is mandatory (9%–11%). These procedures, other than the replacement questionnaire, were implemented for the 2000 census and helped bring the response rate up to 67%.

**Easing the Technological Burden**

A special concern for Web surveys is the extent to which unit nonresponse is due to technical problems. The respondent’s computer hardware or software may not be able to handle the questionnaire. Browser problems and slow connect times may discourage survey participation. These situations may improve over the years as browsers are improved and standardized and as fast connections become more common, but researchers have to remember that many respondents do not have Internet capabilities that permit the same multimedia access that the researchers themselves have.

The respondent may also lack the technical skills that the survey assumes. Dillman and Bowker (2001) point to several such factors that lead to nonresponse in Internet surveys. First, some people do not know how to provide answers, as in how to handle a drop-down menu. People may also not understand how to change answers, since some formats require only choosing a different answer whereas others require clicking a second time on the original answer. It is awkward when not all the answer options show up on the screen at the same time so that viewing the options requires scrolling up or down the page. Also, sometimes it is necessary to take several actions to answer a question (such clicking on an answer and then having to click elsewhere to go on to the next question).

An early problem with Internet surveys was that respondents could not tell how long the survey would be. If people got the impression that the survey was endless, many would break off early. Therefore it has become common to show respondents how far along they are in the survey through progress indicators, such as a bar in a corner in the screen which darkens as the person gets further through the questionnaire. Allowing people to skip questions rather than requiring each question to be answered before the respondent can go on to the next one can also decrease breakoffs.

An unusual danger of Internet surveys is that respondents seem to be especially sensitive to the survey administration. A 2001 study by the Council for Marketing and Opinion Research (CMOR) found that Internet respondents tended to be more critical of the mechanics of the survey than were telephone respondents (Business Market Research
Experience in being members of Internet panels can further exacerbate this tendency. One result is that people may self-select out of Internet surveys that do not satisfy their view of how such surveys should be presented.

Some of these factors involve item nonresponse as well as unit nonresponse, but respondents who get discouraged by factors such as these are probably less likely to complete the questionnaire and to participate in future Internet surveys. Skilled programmers can prevent some of these potential problems, and standards are developing for others, but differences in respondents’ familiarity levels with computers will inevitably persist. Therefore it remains essential to seek to ease the technological burden for respondents in Internet surveys.

**Response Bias**

Nonresponse has pervasive statistical effects (Brehm 1993, 93–117). It leads to underestimates of the variance of a variable, which can lead to false statistical inferences. The bias for a regression coefficient can either be positive or negative, depending on whether the variables are less or more related to each other among nonrespondents than respondents (Groves and Couper 1998, 10–11). Nonresponse can also bias estimates of multivariate relationships, particularly when people who are extreme on one of the variables do not respond (Brehm 1993, 93–117). For example, losing people who are low on an independent variable such as political interest (so the sample is truncated) can bias survey estimates of variables affected by political interest if the relationship between interest and the dependent variable is different for people with low political interest. Sampling to exclude extreme low cases on the dependent variable (so that variable is “censored”) would decrease estimates of the effects of independent variables, an attenuation effect. Studies that use variables related to survey participation as either their independent or dependent variables are particularly susceptible to these effects.

A very minimal strategy is to examine the survey results to see how big a difference there has to be between respondents and nonrespondents to reverse sample findings. If 55% of the sample is satisfied with a service provider, for example, but that is based on only a 60% response rate, the researcher calculates how much nonparticipants would have to differ from participants for the true proportion to be less than a majority. Or if positive change were found for 55% of those who change on a question between two waves of a survey, but that is based on reinterviewing only 80% of the original sample, one can calculate how much those who were not reinterviewed would have to differ from those who were reinterviewed for the true proportion changing in a positive direction to be less than a majority.

**Unit Nonresponse and Response Bias**

The underlying assumption of most of the research on unit nonresponse is that it is vitally to make heroic efforts to maximize the response rate. A low response rate indicates that the sample size is smaller than desired (which can affect statistical significance testing), that there may be nonresponse bias, and that people will be suspicious of the study. Readers of a report on a survey with low response rates may doubt the representativeness of its sample. Sponsors may feel that a low response rate indicates lack of effort by the survey organization.

There is, however, a revisionist view regarding the importance of high response rates. As surveys began to obtain lower response rates, researchers began to test whether that really matters. A set of studies now suggest that higher response rates do not necessarily lead to more accurate results, because lower response rates do not always translate directly into bias.

In an important experiment, Keeter et al. (2000) examine the effects of vigorous efforts to increase the response rate in a telephone survey. They compared a “standard” five-day survey in which people at home were interviewed (the youngest adult male at home, otherwise the oldest female at home) with at least five callbacks and one follow-up for a household refusal to a “rigorous” eight-week survey with random selection among all household members and a prepaid incentive. The response rate was only 36% for the standard procedure versus 61% for the rigorous procedure; the contact rate was 68% for the standard versus 92% for the rigorous. The key concern is whether these differences lead to sizable differences in results, and the important finding is that they do not. Regardless of the large differences in response and contact rates, the samples obtained in the standard and rigorous approaches were very similar on most variables, with an average difference of only 2%, with no difference being greater than 9%. There were only slight differences on demographics, with the respondents in the standard version being more nonwhite (4%), less educated (5% less college), lower income (4% less high income), and more renters (6%). Also, the standard procedures obtained more political independents, which fits with Brehm’s (1993) study and others showing that politically uninvolved people are harder to contact. Significant differences were not obtained on
attention to media, engagement in politics, social trust, connectedness, social and political attitudes, or attitudes toward surveys.

Although the differences in results between the standard and rigorous approaches in the Keeter study were minimal, there were, of course, substantial cost differences between them. The standard approach yielded about one interview per hour compared to only 0.8 interviews per hour for the rigorous approach. The implication of this study is that we may not need to conduct many callbacks or use rigorous procedures, since the differences between the two approaches were few. Still, this study does not examine the limits to that logic—such as whether comparable results would have been obtained with only three days of interviewing instead of five days. Also, the report of this study focuses only on single-variable results, without examining any implications for multivariate analysis. Yet the result is important, because it implies that expensive procedures for increasing response rates may not be necessary. Furthermore, it is not the only study with this claim.

A similar implication derives from a study by Curtin, Presser, and Singer (2000) on the surveys on which the Index of Consumer Sentiment is based. They find that excluding respondents who required more effort to contact and to interview would not have affected monthly analysis, since nonresponse bias was constant and did not affect comparisons over time. At most, some differences appear when responses to multiple consumer sentiment surveys are aggregated, since the impact of nonresponse bias is more evident when larger sample sizes reduce the sampling error.

As mentioned earlier, Merkle and Edelman’s (2002) analysis of 2000 VNS election exit polls found that there was no relationship between the response rate in a precinct and the deviation of the precinct’s exit poll result from its actual vote. This result again suggests that lower response rates do not necessarily lead to higher response bias.

Perhaps the most provocative of these studies is one involving the accuracy of predictions of pre-election surveys. Visser, Krosnick, Marquette, and Curtin (1996) compared the forecasting accuracy of mail and telephone surveys in Ohio. Over a fifteen-year period, they found that the mail surveys predicted more accurately than phone surveys (average error of 1.6% versus 5.2%) even though the mail response rate was only about 25%, compared to a response rate around 60% for the phone surveys. The explanation the authors give is that people who were interested enough in the election to return the mail questionnaire were showing the same type of initiative that is required to go to the polls to vote and therefore are more predictive of actual voting.

This result suggests that low response rates are not necessarily bad. On the basis of this study, Krosnick (1999) directly argues that phone surveys would forecast election outcomes more accurately if they did not aggressively seek high response rates. Further, he generalizes this to argue that high response rates are not necessary for survey representativeness, citing other evidence that correcting for demographic biases in the sample has little impact on correlational analysis (Brehm 1993) and that improvement in response rate does not necessarily alter substantive conclusions of a study (Traugott, Groves, and Lepkowski 1987).

However, there are reasonable limits to this argument. One could interpret the results of the Visser et al. study as unique to election forecasting and could view the evidence from the nonelection studies cited in this section as slim. Additionally, simply accepting low response rates would permit survey organizations to be lazy in contacting respondents, which would let response rates fall even lower than they have already fallen. Yet it is interesting to see that several studies converge in arguing that low response rates do not matter, a conclusion that completely contradicts the conventional wisdom in the field.

**Statistical Solutions to Unit Nonresponse**

Attempts to eliminate survey nonresponse can never be totally successful, but there are statistical means for dealing with the problem through weighting and through modeling the nonresponse process.

**Weighting**

Survey respondents are sometimes weighted in a differential manner so that the obtained sample better resembles the population of interest. Three weighting systems will be mentioned here: probability of response weights, weighting class adjustment, and poststratification adjustment.

The simplest weighting method is using probability of response weights. Observations are weighted inversely to their ease of acquisition, with people who were easiest to interview being weighted less than the hardest. For example, the Politz-Simmons method uses weights based on the number of days during the previous five days that the respondent would have been available to be interviewed at the same time of day. People who would have been home at that time every day would receive the lowest weight, whereas people who were home only on that one day would receive the highest weighting, being treated as representative of people who are hard to contact. This assumes that people who are at home only on Friday night are similar to those at home only on Monday night, for example, which
can be a false assumption, but it recognizes that people who are rarely at home differ from those who are always at home.

A generalization of this procedure is to weight cases by the difficulty of obtaining them: so, for example, an interview obtained on the first mailing, visit, or phone call would receive less weight than one obtained in a much later attempt. The logic of this extrapolation for difficulty procedure is that people who are impossible to contact may be most like those who are difficult to contact—but though some survey analysts would argue that the unreachables are not like actual respondents. In any case, the high proportion of refusals among nonrespondents makes this weighting less useful, since those who refuse are not necessarily similar to those who are difficult to survey (Smith 1983).

A weighting class adjustment method uses variables that were involved in the sampling design. For example, if one was sampling from lists of different types of doctors who responded at different rates, one could then weight them back appropriately. Say that one were interviewing off two lists of doctors, with 55 specialists on one list and 45 family doctors on the other list, and that interviews were taken with 33 specialists and 40 family doctors. As shown in table 8.1, the response rate for specialists would then be 60% and that for family doctors would be 89%. Weights are then calculated as the reciprocal of those response rates: 1/0.60 = 1.667 for specialists and 1/0.89 = 1.125 for family doctors. When the 33 interviews with specialists are weighted at 1.667, the weighted number of cases (n) for them becomes 33 x 1.667 = 55, and when the 36 interviews with family doctors are weighted at 1.125, the weighted n for them becomes 40 x 1.125 = 45. Thus the weighted n’s for the two types of doctors together are back up to their numbers in the original lists that were to be interviewed.

Weighting treats the interviews with the harder-to-interview group as more valuable than those with the easier-to-interview group. The basic assumption is that the data are missing at random (MAR) and that people in the category who were interviewed are like the people in the same category who were not interviewed, so they can be used to represent those who were not interviewed. That means, however, that weighting gives false impressions to the extent that nonrespondents in a category differ from respondents in that category.

The most common weighting method for unit nonresponse is called poststratification adjustment. It differs from weighted class adjustment in that information about nonrespondents is not available, so these weights are instead population based. The weighting is based on external information about the number of people in the category, usually population values from the census. A typical instance involves taking a random-digit-dialing telephone sample and then having to adjust the results on the basis of sex, race, and age. Poststratification weighting assumes that the data are missing completely at random (MCAR), and it fails to the extent that the people who were not interviewed differ from those with the same demographics who were interviewed.

It is not possible to weight on variables for which the true population value is not known, and it is not advisable to weight on variables whose values are highly changeable. For example, weighting preelection polls on the basis of past party identification distribution can bias poll forecasts, since that distribution can change during the election campaign (Altsop and Weisberg 1988). Thus, in 2004 there was much concern in preelection polling about whether to weight up Democrats, since their numbers seemed to be fewer than in previous years, but in the end some of the most accurate polls were those that did not weight, because there had indeed been a drop in the percentage of Democrats among actual voters. The best rule is to weight on the basis of factors involved in selecting the sample while being cautious about weighting on other variables. In any case, it is important that weighting be disclosed in reports on surveys, so readers can judge for themselves whether the results are vulnerable to faulty weighting decisions.

Poststratification weighting works best when the poststratification variables are strongly correlated with the dependent variable of interest. However, multiple variables in a study could be used as dependent variables, which means that different variables might be useful for poststratification adjustment for different data analyses. It is common to adjust on several variables at once, using computer programs that adjust for each variable in turn iteratively until the weights stabilize (a procedure known as raking). When there are multiple design variables, the differences of the dependent variable within combinations of these variables should be made small while the differences between different combinations of these

<table>
<thead>
<tr>
<th>Table 8.1 An Example of Weighting Class Adjustments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specialist</td>
</tr>
<tr>
<td>Population</td>
</tr>
<tr>
<td>Responded</td>
</tr>
<tr>
<td>Response Rate</td>
</tr>
<tr>
<td>Weight</td>
</tr>
<tr>
<td>Weighted N</td>
</tr>
</tbody>
</table>
variables should be made large. This procedure adjusts for biases in the sample, but the variance increases, so there is a trade-off between bias and variance.

The estimates of the vote given by the Voter News Service (VNS) on election night were based on poststratification weights, using correlations between exit poll results and past election results for the sample precincts, with actual vote totals for the sample precincts substituting for exit polls when they become available (Mitofsky and Edelman 1995). The problems that VNS had in dealing with the 2000 presidential election led to reconsideration of that procedure (though the Florida debacle—in which Florida was first called early for Gore and then called in the middle of the night for Bush before the original call was retracted, sending the Florida election into recount territory—was more likely due to the differences between how people thought they voted and whether and how their votes were counted). Similarly, the overstatement of the Kerry vote in some key states in the 2004 exit polls seems to have been partially due to the use of problematic mathematical models.

Weighting is a common solution, but a caution should be emphasized: some statisticians believe that weights should be used only for univariate analysis, not for multivariate modeling such as regression analysis or factor analysis. If education affects results, this argument would be, then education should be used as a predictor in the regression equation rather than as a weighting factor. Even weighting by probability of response can attenuate effects, so that weighting leads to missing significant effects (see, for example, Brehm 1993, 119–21). In fact, weighting often makes little difference in actual practice. The results are often so similar with and without weighting that weighting is sometimes not considered worth the trouble and the added difficulty of explaining the weighting to clients. However, as Voss, Gelman, and King (1995) show, weighting sometimes does have substantial effects, with support for George H. W. Bush appearing 2%–4% higher in preélection polls in 1992 when weighting was not used.

**Modeling Nonresponse**

Another set of statistical procedures for dealing with unit nonresponse is model based, modeling the nonresponse process. Nonresponse is treated as stochastic in this approach. This approach is especially appropriate when the likelihood of responding is related to variables of interest in the study. There are actually several statistical procedures of this type, including Little’s maximum likelihood estimation and Bayesian methods that incor-

corporate prior information about the underlying distribution of the variable (see Lessler and Kalsbeek 1992, 199–207; Little 2003). These methods are controversial in that they make assumptions about the nonresponse, and they are difficult to employ because they must be estimated separately for each variable of interest rather than simply deriving a single weight as above. They require having some predictors of participation in the survey, including at least one predictor that affects participation but not the dependent variable of interest.

The model-based procedure that will be described here is Heckman’s method for correcting for selection bias. It involves developing a pair of equations, a probit equation for the chances of entering the survey and a regression equation for the variable of interest in the survey, with the two equations being solved simultaneously through maximum likelihood procedures.

A useful example of the Heckman method is John Brehm’s (1993) analysis in his book *The Phantom Respondent.* Brehm develops three separate models of survey participation, making use of different sources of information about cooperation. One is an administrative model, using predictors for the survey participation equation based on aspects of the field administration, such as the number of calls to the household in order to get the interview, whether or not a persuasion letter was sent, and whether or not refusal conversion was necessary. Brehm admits that the administrative model can lead to inconsistent estimates of residuals because these predictors are endogenous to the survey participation model, but he finds that these effects are minimal in practice. His second approach is behavioral, using predictors such as the respondents’ relationship to strangers, to the interviewer, to the interview, and to self-image, based on respondents’ reasons for refusal or compliance. In the studies he examined, the most important of these was relationship to the interview, with relationship to the interviewer also having a large effect.

The third possibility would be using demographic predictors in the selection equation, but demographic information about nonrespondents can usually be obtained only while the study is being administered, at which point it is easier to use the field administration model (particularly if demographics are to be part of the second substantive equation). As a result, Brehm relies mainly on the administrative model.

Some examples of Brehm’s results usefully illustrate the potential of this approach. The Heckman correction for estimating voter turnout as a function of demographics leads to a larger estimate of the effect of education on turnout, due to the censoring effect of survey participants’ being
more likely to participate in voting (136–38). The largest change in regard to the turnout equation is a more negative constant term after correction, meaning that turnout is overestimated in surveys when a correction for nonresponse is not employed, with an effect estimated at 18% in the 1986 National Election Study. On the NES abortion question, the correction shows that the survey understates opposition to abortion by blacks, after controlling for other demographics, and a change in the constant term implies that NES underestimates the proportion of people who are adamantly against abortion (154–57). On the other hand, estimation of income as a function of demographics does not change when a Heckman model is employed. All in all, Brehm finds that although not all variables are affected, our understanding of some variables is affected by unit nonresponse, so assuming that nonrespondents are the same as respondents can lead to serious error.8

A final statistical approach for dealing with unit nonresponse is imputation, in a manner analogous to how imputation is used for item nonresponse (see chapter 7). Rubin and Zanutto (2002) propose a “matching, modeling, and multiple imputation” method in which matched substitutes are used to impute data for nonrespondents (modeled using predictions from background covariate variables that are available for the nonrespondents), though this assumes that appropriate covariate variables can be obtained for the nonrespondents.

Unit Nonresponse and Survey Constraints

Unit nonresponse can be minimized by aggressive attempts to get the cooperation of designated respondents. There are several methods for this. In the case of noncontact, the solution is to do repeated callbacks in the hope that contact will eventually be made. In the case of noncooperation, many survey organizations have some interviewers who specialize in refusal conversions—convincing people to participate in the survey. Another way to deal with the potential problem of noncooperation is to offer incentives to participate: money, other material items (from pens to DVDs), or a chance to participate in a lottery. Additionally, tailoring the request to participate is becoming the most common approach to securing cooperation in interviewer-administered surveys.

It is also useful to collect as much information as possible on nonresponse, especially if corrections for nonresponse are contemplated. Obviously there are limits to this possibility, but it is often possible to keep track of simple demographics (at least gender on phone surveys) and

reasons given for nonresponse as well as number of callbacks and whether refusal conversion was attempted.

There is also an important choice to be made between weighting the data for nonresponse and trying to model the nonresponse, especially when multivariate analysis is planned. Survey researchers are more familiar with the use of weights, but a strong case can be made for modeling instead. Obtaining a single set of weights for adjusting for nonresponse takes less time and effort, whereas deciding on an intricate set of modeling choices for each dependent variable is more costly. Menard (2002, 41) summarizes that “weighting may be preferable when nonresponse rates are low, while modeling nonresponse may be preferable when nonresponse rates are high (e.g., over 30%), but in this latter case, neither is likely to be entirely satisfactory.”

Minimizing unit nonresponse in a survey is not an inexpensive task. The trade-offs are clear here, since spending more money on avoiding unit nonresponse directly takes away money that could otherwise be spent on increasing the sample size. Academic survey organizations tend to be more likely to do large numbers of callbacks and to have conversion specialists on their interviewing staff, while inexpensive commercial phone banks are least likely to do callbacks or bother trying to get conversions. Survey clients will sometimes require a particular level of response rate, which survey organizations will factor into their bids for the project.

Academic researchers usually desire to maximize their response rates, expecting that this will yield more valid data and that it will minimize the chance of journal reviewers’ rejecting their work on the basis of having too small a response rate. Market research operations that require timely results are least likely to emphasize response rates, particularly if they need daily data on topics such as satisfaction with services provided. Survey researchers who employ the total survey error approach tend to take the middle route, doing some callbacks and conversion attempts but emphasizing to their clients that doing either extensively would take away too much from the total sample size that could otherwise be afforded. As long as this does not become a justification for low response rates, it is certainly advisable for researchers to be realistic about response rates.

The most challenging controversy is whether response rates really matter. Low response rates raise credibility problems, since poll consumers will assume that samples with low response rates are not representative of the target population. Studies are now finding that higher response rates do not necessarily lead to less response bias. Yet the credibility problem remains: will a survey be believed if its response rate is low? The public is
already somewhat skeptical of surveys, and reporting surveys with low response rates could further exacerbate that skepticism. Clients and journal reviewers similarly will often not believe survey results unless the response rate is reasonably high. Accepting lower response rates would lower survey costs, but caution should be exercised so that survey credibility is not sacrificed.

Appendix: Response Rate Calculation

Response rate statistics for surveys are often stated, without explaining how they were calculated. As a result, the American Association for Public Opinion Research (AAPOR) has prepared a standard set of definitions for response rates, cooperation rates, contact rates, and refusal rates (AAPOR 2004). This appendix will summarize those definitions, which are available on the AAPOR Web site.

Disposition Classification

Since surveys can involve thousands or even tens of thousands of contact attempts, it is important to give a clear code as to the disposition of every interview attempt. The AAPOR standard definitions distinguish four results of interview attempts: (1) interview, (2) eligible cases that are not interviewed (unit nonresponse), (3) unknown eligibility, noninterview, (4) not eligible. Each is divided into various types, as follows (see fig. 8.1, p. 202).

1. Interviews can either be complete interviews (I) or partial (P). Some survey organizations include partials as a successful interview in their calculation of response and cooperation rates, whereas other organizations are more conservative and do not consider them successful interviews. Both categories are counted when one is looking at refusal and contact rates. Breakoffs, which occur when the respondent terminates the interview so early that less than half of the questions were answered, are usually instead treated as eligible, no interview.

2. Unit nonresponse ("eligible, no interview") consists of three situations. The first is refusals (R), including household refusals before identifying the proper respondent and breakoffs. The second is noncontacts (NC). For phone interviews, this occurs when the interviewer never talked to the selected respondent. For in-person household interviews, this occurs when the interviewer is unable to gain access to the building, no one was reached at the housing unit, or the respondent was away or otherwise unavailable. For mail surveys of specifically named persons, noncontacts occur when there is notification that the respondent could not fill out the questionnaire during the field period and when the questionnaire was mailed back too late for inclusion. The other (O) category consists of situations where the designated respondent had died or was physically or mentally unable to do the interview, or there was a language problem, plus literacy problems for mail surveys. All eligible, no interviews count as failures when the response rate is computed.

3. The unknown eligibility includes two subcategories. One is when it is unknown if the household is eligible (UI). In telephone surveys this occurs when the number is sampled but not dialed, is always busy, never answers, has an answering machine that does not make clear if the location is residential, has a call-screening or call-blocking device, or in the case of technical phone problems such as circuit overloads. For face-to-face household surveys, the unknown household eligibility situation occurs when contact was not attempted, when the interviewer was unable to reach the household (as because of weather) or an unsafe area, and when it was not possible to locate the address. For mail surveys, this occurs when the questionnaire was never mailed and when nothing was ever returned so there is no information about the outcome.

The second subcategory of unknown eligibility is a broad "unknown, other" (O) category. It includes instances when it is not known whether an eligible respondent resides at the household. For phone and in-person interviews, this is the case when it is not found out whether there is an eligible respondent and when there was failure to complete the needed screener. There are several additional situations included under this category for mail surveys, partly because the U.S. Postal Service (USPS) puts a variety of different explanations on returned mail, such as "illegible address," "insufficient address," "no mail receptacle," and "no such address." Mail that is returned with forwarding information is also treated in this category.

Different survey organizations treat the unknown eligibility, noninterview view category differently in computing the response, refusal, and cooperation rates. The most conservative approach is to count each of these as a failure, whereas the survey effort appears more successful when they are totally omitted from the calculation.

4. The last disposition category, not eligible, does not count in computing response and related rates but is included so that every interview attempt is coded in some manner. This category occurs when the person lives outside of the designated geopolitical boundary. For phone samples, it also occurs when the phone number is a fax/data line, is a nonworking (not yet assigned) or disconnected number, or is a beeper or cellular/mobile
phone (if not the household’s only phone). For in-person interviews, it also occurs when the housing unit is vacant as well as when there are other security measures such as a doorman excluding entry. Not eligible also includes contacts with nonresidences (such as business or government offices,) institutions (such as prisons and sanitariums), and group quarters (such as military barracks and phones in the hallway of a college dormitory, fraternity, or sorority). Not eligible also includes situations when there is no eligible respondent at the household, as when all residents are under age eighteen or when no one passes the screening questions for sampling subgroups. Some samples have quotas for particular subgroups (such as a given number of older men), so a household is categorized as not eligible if the quota for that subgroup is filled by the time that household was contacted. Finally, not eligible includes duplicates in large mailings of mail surveys. The frequency of noneligibles is sometimes reported to clients, as when the client wants to know the proportion of people screened out in a survey of grandparents because they have no grandchildren.

### Response Rate Formulas

AAPOR recognizes a variety of formulas for response rates (RR), cooperation rates (COOP), contact rates (CON), and refusal rates (REF), depending on how particular subcategories are treated. These formulas vary in how conservative they are, depending on whether partial interviews are treated as interviews and how unknown eligibility is considered. Academic and government surveys sometimes require using the most conservative formulas, whereas commercial organizations likely prefer reporting to clients higher response, cooperation, and contact rates and lower refusal rates. The formulas will be summarized here, using the category number above as part of the notation. That is,

- \( I = \text{Complete Interview} \)
- \( P = \text{Partial Interview} \)
- \( R = \text{Refusal} \)
- \( NC = \text{Noncontact} \)
- \( O = \text{Other eligible nonresponse} \)
- \( 2 = \text{Unit nonresponse} = R + NC + O \)
- \( UH = \text{Unknown if household eligibility} \)
- \( UO = \text{Unknown other eligibility} \)
- \( 3 = \text{Unknown eligibles, noninterview} = UH + UO \)

The response rate indicates the proportion of eligible reporting units that led to interviews. The most conservative version is \( RR_1 = 1/(1 + 2 + 3) \), whereas the highest response rate is obtained with the formula \( RR_6 = 1/(1 + 2) \). CASRO uses \( RR_3 = 1/(1 + 2 + 3\epsilon) \), where \( \epsilon \) is an estimate of the proportion of category 3 that is eligible.\(^{11}\)

The contact rate is the proportion of cases in which some member of the household unit was reached. The different versions of the contact rate formula differ only in whether to count category 3 as relevant to the calculation. The most conservative contact rate is \( CON_1 = (1 + R + O)/(1 + 2 + 3) \), whereas the highest contact rate is \( CON_3 = (1 + R + O)/(1 + 2) \).\(^{12}\)

The cooperation rate is the proportion of all eligible units ever contacted that were interviewed. Two choices are whether to count partial interviews as cooperation successes and whether the other subcategory of eligible, noninterview counts as a failure. The most conservative household-level cooperation rate is \( COOP_1 = 1/(1 + R + O) \), whereas the highest cooperation rate is obtained with the formula \( COOP_4 = 1/(1 + R) \).\(^{13}\)

Finally, the refusal rate is the proportion of all potentially eligible cases in which the respondent refuses to be interviewed, including breakoffs. As
with contact rates, the different versions of the refusal rate formula differ only in whether they count unknown eligibility, noninterviews as relevant to the calculation. The most conservative refusal rate is REF3 = R / (1 + 2), whereas the lowest is REF1 = R / (1 + 2 + 3).  

Table 8.2 illustrates the calculation of the conservative versions of these formulas for the 1998 The Ohio Political Study (TOPS) conducted by the OSU CSR.


<table>
<thead>
<tr>
<th>Table 8.2 Response Rate Calculations for 1998 TOPS Survey</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID</td>
</tr>
<tr>
<td>IP</td>
</tr>
<tr>
<td>2R</td>
</tr>
<tr>
<td>3R</td>
</tr>
<tr>
<td>2NC</td>
</tr>
<tr>
<td>2NC</td>
</tr>
<tr>
<td>2O</td>
</tr>
<tr>
<td>2O</td>
</tr>
<tr>
<td>3UH</td>
</tr>
<tr>
<td>3UH</td>
</tr>
<tr>
<td>3UO</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

**SUBTOTALS**

<table>
<thead>
<tr>
<th>1total</th>
<th>Interview</th>
<th>534</th>
</tr>
</thead>
<tbody>
<tr>
<td>2R</td>
<td>Refusal</td>
<td>488</td>
</tr>
<tr>
<td>2NC</td>
<td>Noncontact</td>
<td>122</td>
</tr>
<tr>
<td>2O</td>
<td>Other eligible nonresponse</td>
<td>58</td>
</tr>
<tr>
<td>2total</td>
<td>Eligible, no interview</td>
<td>668</td>
</tr>
<tr>
<td>3total</td>
<td>Unknown eligibility</td>
<td>106</td>
</tr>
<tr>
<td>4total</td>
<td>Not eligible</td>
<td>812</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>2120</td>
</tr>
</tbody>
</table>

Response Rate #1 = 1 / (1 + 2 + 3) = 510 / (534 + 668 + 106) = 39.0%

Contact Rate #1 = (1 + R + O) / (1 + 2 + 3) = (534 + 488 + 58) / (534 + 668 + 106) = 82.6%

Cooperation Rate #1 = 1 / (1 + R + O) = 510 / (534 + 488 + 58) = 48.3%

Refusal Rate #1 = R / (1 + 2) = 488 / (534 + 668) = 40.6%

Source: Ohio State University Center for Survey Research

---

This chapter will cover several relatively obscure types of survey error that are associated with the sampling frame, the actual set of units from which the sample will be taken. Frame error is the error that can arise when the elements in the sampling frame do not correspond correctly to the target population to which the researcher wants to make inferences. The most important frame error is coverage error, the mathematical difference between a statistic calculated for the population studied and the same statistic calculated for the target population. A simple example of a coverage problem involves sampling from phone books, because they exclude people with unlisted numbers who are certainly part of the target population. Sampling frames and coverage error will be explained in detail in the first section of this chapter.

The second section of this chapter deals with a frame error problem of ineligibles that arises when some elements in the sampling frame are not part of the target population. For example, random-digit-dialing (RDD) methods for telephone surveys call some business phone numbers even when the target population consists only of households. There is often a trade-off between coverage error and eligibility, since increasing the coverage may also bring in cases that are ineligible. This often necessitates finding an efficient sampling frame with a level of coverage that is high but not so high as to require contacting many cases that prove to be ineligible.

The third section describes two additional types of frame errors. Clustering happens when there is one element in the sampling frame for