CHAPTER SIX

NONRESPONSE IN SAMPLE SURVEYS

6.1 INTRODUCTION

Chapters 1 and 2 discussed how survey researchers use the term "nonresponse" to describe the failure to obtain measurements on sampled units. Sometimes, the failure is complete—the person chosen for the sample refuses to cooperate with the survey request entirely (e.g., Sample Person: "I never participate in surveys; please don't call me again"). Sometimes, the failure affects only one item in the survey measurement (e.g., Interviewer: "What was your total family income last year?" Respondent: "I don't know that; my wife keeps those records"). The total failure is termed "unit nonresponse"; the partial failure is called "item nonresponse."

unit nonresponse

item nonresponse

Nonresponse can affect the quality of survey estimates. If the nonrespondents have different values on variables that are components of the estimate, its value based on the respondents may differ from that based on the total sample. When the departure is a systematic one, endemic to all implementations of the survey design, it is called "nonresponse bias." On some simple estimates (like the sample mean), nonresponse bias is a function of the correlation between the survey variable and the likelihood of participating in the survey (see Section 2.3.6).

Nonresponse rates (the percentage of eligible sample cases that are nonrespondent) in most household surveys have increased over time in the United States and Western Europe. There are three principal sources of nonresponse that appear to have different causes: failure to deliver the survey request to the sample person, failure to gain the cooperation of a contacted sample person, and inability of the sample person to provide the requested information. They appear to affect different types of estimates produced by surveys.

This chapter provides the reader with the essential concepts and practices involving nonresponse. It first presents information on response rates and their trend over time; then it discusses the link between nonresponse rates and bias. Next, it dissects the phenomenon of nonresponse into distinct types. It ends with a discussion of how survey design features can affect nonresponse error in different estimates.

6.2 RESPONSE RATES

In simple terms, a survey's response rate is merely the percentage of eligible sample cases that were measured. The nonresponse rate is its complement. Before we examine some typical values of response rates, we need to note that this simple definition belies the complexity of computation of response and nonresponse rates.

6.2.1 Computing Response Rates

Since nonresponse rates have traditionally been viewed as direct measures of survey quality, there is a checkered history of unethically presenting deflated estimates of nonresponse to give the appearance of high-quality statistics. This problem has been the focus of several committees of professional associations (Frankel, 1983; American Association for Public Opinion Research, 2000). The guidelines of the American Association for Public Opinion Research (AAPOR) can be found at www.aapor.org. That website provides several different response rate computations, depending on the survey design used. In general, there are three complications in calculating response rates:

1) Some sampling frames contain units that are not target population members, requiring a screening step to determine eligibility (e.g., a telephone surveys of households when the frame contains business numbers). With such designs, there is uncertainty about the eligibility of nonrespondent cases and, hence, what the denominator of the response rate should be.
2) Some sample frames consist of clusters of sample elements in which the number of elements is unknown at the time of sampling (e.g., a survey of school children chosen from a sample of schools). When a full cluster is nonrespondent, it is unclear how many sample elements are nonrespondent.
3) Unequal probabilities of selection are assigned to different elements in the sampling frame (e.g., oversamples of minority ethnic groups). In this case, it is unclear whether selection weights should be used in the computation of response rates (see Groves, 1989).

One way to approach the first two problems is to estimate the value of the denominator, using either external information or information from other cases. Thus, the response rate might be

$$\frac{I}{I+R+NC+O+e(UH+UO)}$$

where

I = Complete interview
R = Refusal and breakoff
NC = Noncontact
O = Other eligible
UH = Unknown if household/occupied household unit
UO = Unknown eligibility, other
e = Estimated proportion of cases of unknown eligibility that are eligible

An estimate of e can be obtained from the current survey, e.g., $(I + R + NC + O)/(I + R + NC + O +$ ineligibles chosen into sample). Alternatively, some special methodological studies might be mounted to learn e, by studying a sample of cases with initially unknown eligibility to learn more about them. Finally, if no

estimate of e can be obtained, it is recommended that two response rates are presented: one including $(UH + UO)$ in the denominator and the other excluding the term. This produces a range of response rates within which the true response rate must lie. More sophisticated modeling approaches to estimating e in an RDD telephone survey are presented in Brick, Montaquila, and Scheuren (2002).

The third problem in estimating response rates arises when (as described in Chapter 4) unequal probabilities of selection are used. For example, if a community survey about city services oversampled areas where poor persons lived (stratum 1), using twice the sampling fraction as other areas (stratum 2), how should the response rate be computed? There are often two response rate issues in such a design. One is a comparison of the response rates in the poor areas (stratum 1) with the response rates in the nonpoor areas (stratum 2). These would be relevant to the analysis that compares means of the two strata. In this case, the two response rates would use the same approach as above. If, however, the interest was in the overall sample mean, which, as noted in Chapter 4, Section 4.5, must use selection weights, w_i, then the response rate for the overall sample might use the same weights, adjusting each sample element for the selection probability it received.

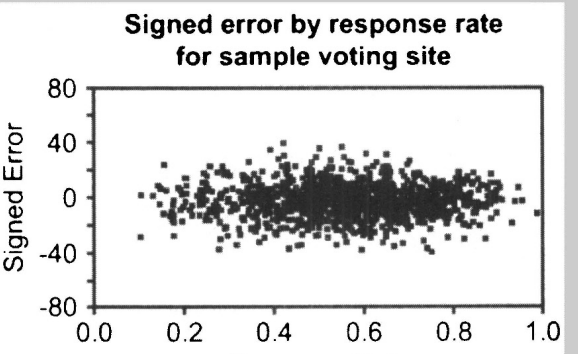
A variety of other rates are used for different purposes. For example, refusal rates [e.g., $R/(I + R)$] and refusal conversion rates (initial refusals that were subsequently interviewed) are used to evaluate interviewer performance. For establishment surveys, coverage rates (the proportion of the total being measured that is accounted for by the respondent units) are used to reflect that, when estimating things like output or number of employees, missing Walmart is not the same as missing the corner convenience store. For surveys such as NAEP, in which selection occurs over several levels (e.g., schools and pupils), compound rates are calculated to reflect nonresponse at each level.

Merkle and Edelman (2002) on How Nonresponse Rates Affect Nonresponse Error

Merkle and Edelman (2002) present an observational study that finds no relationship between nonresponse rates and nonresponse error.

Study design: Exit polls use a probability sample of voting places, with interviewers selecting a systematic random sample of voters exiting the voter station. For each sample voting place, the difference between the Democratic and Republican vote percentages was compared between respondents and the public vote totals.

Findings: The response rates at the sample places varied from 10 to 90%, mostly falling in the 45–75% range. The plot below has a point for each voting place; the x axis is the response rate for the sample place; the y axis is the error of the difference between Democratic and Republican vote percentages. There is no apparent relationship between response rate and total error of the estimated difference. A good predictor of cooperation was how far away from the exit door the interviewer was asked to stand.



(Source: Merkle and Edelman, 2002.)

Limitations of the study: Because of lack of randomization of interviewers to sample place, there could be a confounding of interviewer effects and true differences in cooperation likelihoods. Further, the exit poll is a unique setting, unlike most other surveys.

Impact of the study: The study offers an example of the absence of nonresponse error when nonresponse causes are unrelated to the statistic.

6.2.2 Trends in Response Rates Over Time

Some notion of variation in response rates can be attained by examining the trends in response rate over time for ongoing surveys. For example, Figure 6.1 shows the response rate for the NCVS. The NCVS attempts interviews with all household members 12 years old and older. The NCVS reports response rates at the household level and the person level. The household nonresponse rate measures the percentage of households in which no household member was successfully interviewed. This is the top (dashed) line in Figure 6.1. It shows relatively stable values, ranging between 3 and 4 percentage points (a household response rate of 96–97%). The lowest line (shadow line) is the household refusal rate; it similarly shows a relatively consistent level over the years. The solid black line is the person-level refusal rate, which is the ratio of all sample household members enumerated who refused the interview request to all those enumerated. In short, it is an estimate of the percentage of persons providing an interview, once access to the household has been achieved. This nonresponse component is rising over the years, nearly doubling in the years covered by the chart.

Figure 6.2 shows the response rate trend for the Current Population Survey conducted for the Bureau of Labor Statistics by the US Census Bureau, which produces the monthly unemployment rates for the United States. A very different trend is displayed—for many years an overall consistent nonresponse rate of about 4 to 5%, but a continually increasing refusal rate. This is an example of a survey that kept the overall nonresponse rate low by reducing the noncontact rate over the years, in the face of higher refusals. What happened in 1994? There was a large change in the interviewing protocol in that year, moving to a computer-assisted personal interviewing design and a new questionnaire. One possible explanation for the impact is that the short questionnaire (8–12 minutes) was often taken on the

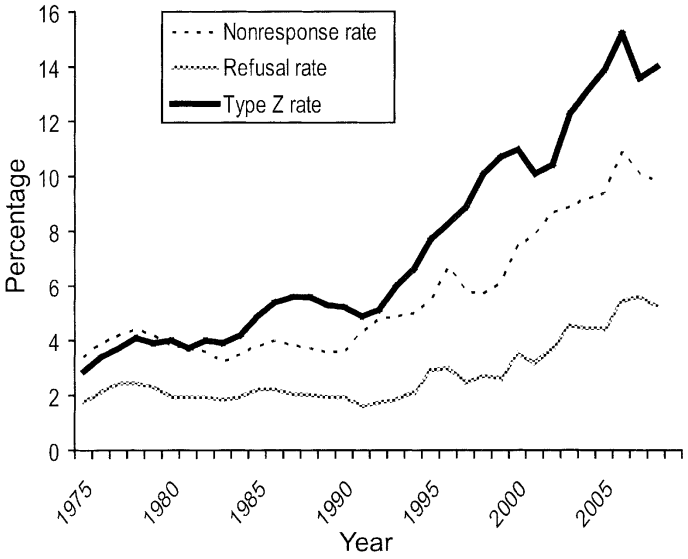


Figure 6.1 Household nonresponse rate, household refusal rate, and person refusal rate for the National Crime Victimization Survey by year. (Source: U.S. Census Bureau, 2007.)

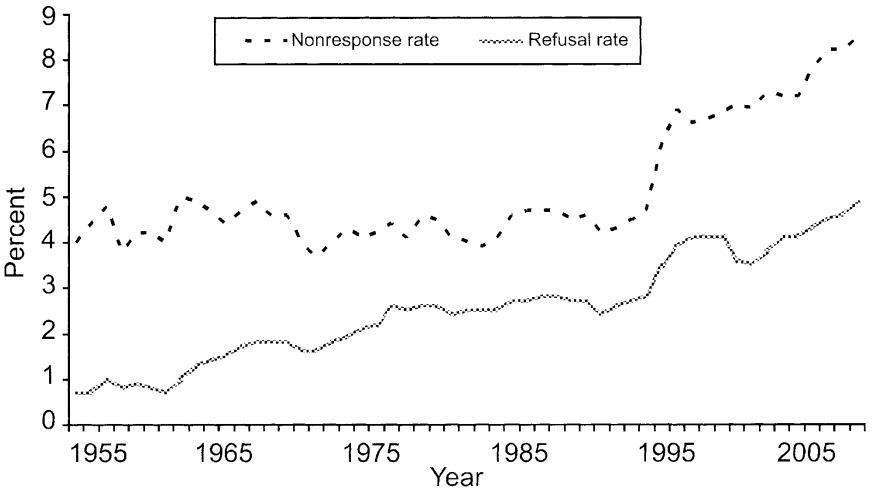


Figure 6.2 Nonresponse and refusal rates for the Current Population Survey by year. (Source: U.S. Census Bureau.)

doorstep with a paper questionnaire, but could not be done that way with a laptop computer and, hence, refusal rates increased (see Couper, 1996).

Both of the nonresponse rates above are quite low but the increasing trends are common. Both studies are well-funded face-to-face surveys; both are conducted by the Federal government. In general, response rates for other surveys are lower, with academic surveys somewhat lower and private sector surveys considerably lower.

Figure 6.3 shows overall nonresponse and refusal rates for the Survey of Consumers. The survey has much higher nonresponse rates than NCVS or CPS.

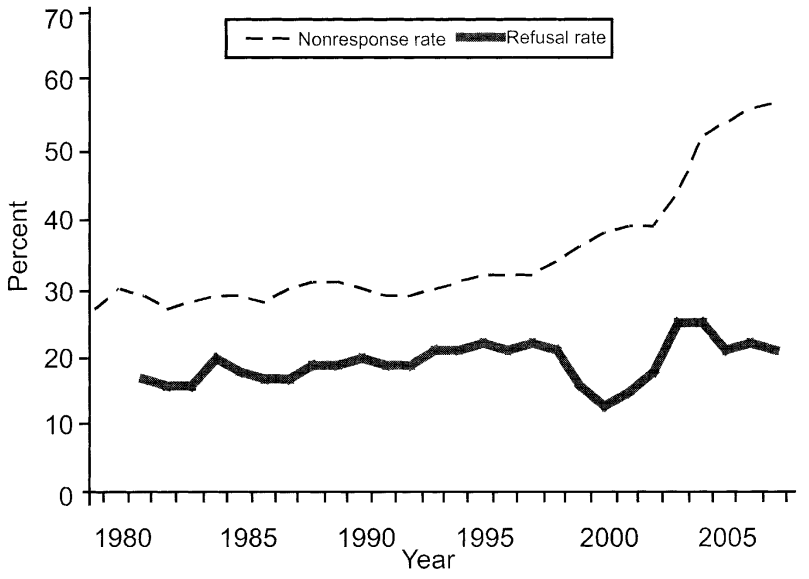


Figure 6.3 Nonresponse rate and refusal rate for the Survey of Consumers by year. (Source: Survey of Consumers.)

The overall nonresponse rate shows a steadily increasing trend, from percentages in the 30s in the 1980s to percentages near 60 more recently. The refusal rate shows a much smaller rise. At the same time, the percentage of the interviews that were taken only after an initial refusal climbed from about 7% to about 15% of the interviews. Again, the lesson is clear: response rates have declined over time.

Figure 6.4 presents nonresponse rates for the Behavioral Risk Factor Surveillance System (BRFSS). Because there are individual nonresponse rates for each state conducting the BRFSS, the graph shows the median nonresponse rate among all reporting states. The figure shows nonresponse rates that are more like the SOC than the NCVS. This probably results both from the telephone mode (versus the face-to-face mode of the NCVS) and the fact that the U.S. government is only an indirect sponsor of the surveys. The overall nonresponse rate increases from percentages in the 30s to percentages near 50% in later years.

Increasing nonresponse rates are not only a phenomenon of the United States. A study of nonresponse trends in 16 European countries over a 20-year period ending in the 1990s found that the noncontact rate increased on average 0.2% per year and the refusal rate 0.3% per year (de Leeuw and de Heer, 2002).

These are results from household surveys. Surveys of business establishments are most often conducted using mail questionnaires or the telephone (or some combination of both). There is less evidence of increasing nonresponse in business surveys (Atrostic and Burt, 1999).

6.3 Impact of Nonresponse on the Quality of Survey Estimates

Perspectives on the role of nonresponse rates in survey quality are undergoing rapid change. For some decades, the dominant goal of survey researchers was to

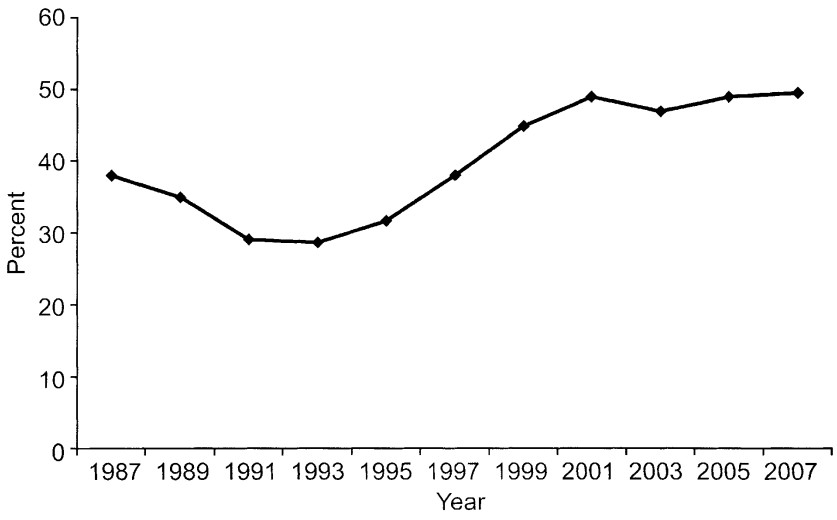


Figure 6.4 Median nonresponse rate across states, Behavioral Risk Factor Surveillance System, 1987–2007. (Source: BRFSS.)

minimize nonresponse rates. For example, Alreck and Settle (1995, p. 184) say, "It's obviously important to do as much as possible to reduce nonresponse and encourage an adequate response rate." Babbie (2004) is bold enough to say "I believe that a response rate of at least 50 percent is adequate for analysis and reporting. A response rate of 60 percent is good; a response rate of 70 percent is very good" (p. 261). Finally, Singleton and Straits (2005) note, "Therefore, it is very important to pay attention to response rates. For interview surveys, a response rate of 85 percent is minimally adequate; below 70 percent there is a serious chance of bias" (p.145). All of these quotations come from books used to teach students about survey methods.

The goal of minimizing nonresponse rates as the sole method of reducing nonresponse error resulted from an interpretation of the expression

$$\bar{Y}_r = \bar{Y}_n + \left(\frac{M}{N}\right)(\bar{Y}_r - \bar{Y}_m)$$

where $\bar{Y}_r$ is the unadjusted respondent mean, $\bar{Y}_n$ is the full sample mean, M/N is the proportion of the population that is nonrespondent, and $\bar{Y}_m$ is the nonrespondent mean (a value unknown in most surveys). Many researchers examined the two-factor term that is the bias of the respondent mean, inferred that $(\bar{Y}_r - \bar{Y}_m)$ was fixed in the population and concluded the only way to reduce the nonresponse bias was to reduce the nonresponse rate.

nonresponse bias

The formula above is sometimes labeled the "deterministic" view of survey nonresponse. It is compatible with the notion that there is a fixed set of respondents and nonrespondents in the population. For all sample persons, posing the survey request reveals to the researcher who are respondents and who are the nonrespondents. Instead of this viewpoint, a more modern view of nonresponse is that each person is potentially a respondent and potentially a nonrespondent—the decision to participate is the realization of a stochastic process (i.e., a process subject to random variability). In this perspective the expression above changes to

$$\bar{Y}_r = \bar{Y}_m + \frac{\sigma_{yp}}{\bar{p}}$$

where σ_{yp} is the covariance between, y, the variable of interest in the survey, and p, the propensity to respond, among units of the population; and $\bar{p}$, the mean propensity in the population. The covariance measures the joint variation of two variables,

$$\sum_{i=1}^{N}(y_i - \bar{y})(p_i - \bar{p})/(N-1)$$

so the expression above implies that when the likelihood of responding is strongly related to the variable of interest in the survey, then nonresponse bias for the respondent mean will be large. In short, nonresponse bias should vary over different estimates in a survey (as a function of the covariance of y and p). Instead of focusing on the response rate solely, the researcher has to focus on whether response propensity and the survey variable are correlated.

response propensity

Is there empirical support for large nonresponse error variation within surveys across different estimates? Yes. A "meta-analysis" synthesizes the findings of many different scientific studies of the same phenomenon. There are a set of surveys specially designed to measure the nonresponse bias of estimates from the survey. Sometimes these compare respondents and nonrespondents on variables contained on the sampling frame, sometimes on supplementary data matched to the sample cases, sometimes on data from a prior screening interview, and sometimes on data from a followup effort to measure nonrespondents. We note that across all the estimates, the nonresponse rates of the studies range from 14% to 72%, with a mean nonresponse rate of 36%. Most of the estimates come from studies using nonsurvey records (24% from the sampling frame, 32% from a supplementary dataset); 28% come from studies using follow-up of nonrespondents with some extraordinary effort. Most of the remaining studies use screener interview data (14%). A very small percentage (2%) use reports of intentions about responding to a future survey request.

Figure 6.5 presents a scatterplot of 959 estimates of the absolute value of the relative nonresponse bias:

$$\left| \frac{100 \times (\bar{y}_r - \bar{y}_n)}{\bar{y}_n} \right|$$

where the numerator contains the difference between respondent and full sample means, and the denominator is the full sample mean. This figure contains a point for each of the means reported in the 59 studies, with complementary percentages for binary variables. For each binary variable, two percentages can be computed. The smaller of the two tends to generate higher relative nonresponse bias. Hence, the figure presents the nonresponse bias of both complementary percentages. The plot displays vertical sequences of points, representing different estimates com-

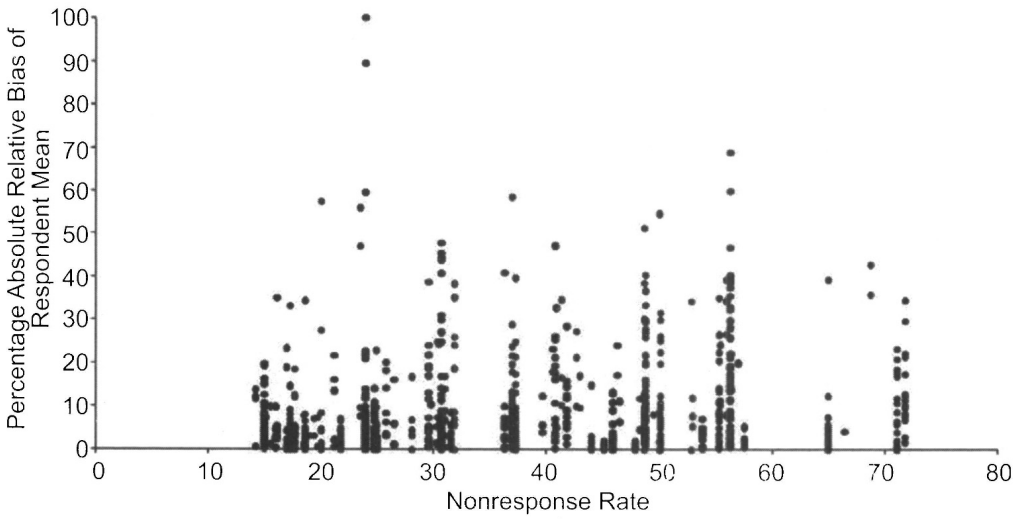


Figure 6.5 Estimates of the absolute value of the relative nonresponse bias for 959 estimates by nonresponse rate of survey. (Source: Groves and Peytcheva, 2008.)

puted from the same survey. The figure clearly shows: (a) large relative nonresponse biases exist in the studies, (b) most of the variation in nonresponse lies across estimates within the same survey, and, as implied by that observation, (c) the nonresponse rate of a survey, by itself, is a poor predictor of the absolute relative nonresponse bias. If a naïve OLS regression line were fit to the scatterplot, the R^2 would be 0.04. In short, insight into the linkage between nonresponse rates and nonresponse bias requires more information about the circumstances of each survey measurement.

There are two practical implications for survey researchers from the figure above. First, often there are very small σ_{yp} terms; that is, the variable of interest is not correlated to response propensity, so there is little nonresponse bias (regardless of the response rate of the survey). Second, finding out how response propensity is related to the important survey variables is a new task for researchers attempting to reduce nonresponse error. This requires thinking about nonresponse error in a causal way. There is an important note of caution about the figure. Since the scatterplot comes from different surveys, it does not provide an answer to the question, "For my specific survey, will increasing the response rate reduce my nonresponse error?" Sometimes, the answer will be "yes"; sometimes, "no." There are even examples in surveys in which increasing the response rate *increased* the nonresponse error (see Merkle, Edelman, Dykeman, and Brogan, 1998). The answer depends on whether the types of nonrespondents brought into the respondent pool as the response rate is increased act to increase σ_{yp} or decrease it.

It is important to note that nonresponse can affect both descriptive statistics and analytic statistics (like a regression coefficient). The expressions for the effect of nonresponse error on such statistics are more complex than those above, often functions of covariances of the variables involved. A large set of modeling techniques have been developed to address nonresponse error in such estimates, most drawing on developments in econometrics (Heckman, 1979; Berk, 1983).

In the next section, we argue that, to become more sophisticated in our understanding of when nonresponse rates produce nonresponse error in an estimate, we need to think causally about nonresponse mechanisms.

6.4 Thinking Causally About Survey Nonresponse Error

Three alternative causal models underlie the relationship between response propensity and nonresponse bias (Groves, 2006). As graphically shown in Figure 6.6, the "separate cause" model asserts that the vector of causes of the Y variable is independent of the causes of response propensity, P. In this case, expected values of Y among respondents would be unbiased estimates of those among all sample persons and corresponds to the "missing completely at random" case (Rubin, 1987). The "common cause" model asserts that there are shared causes (Z) of response propensity and the Y variable; this model corresponds to the "missing at random" case. The "survey variable cause" model asserts that Y itself is a cause of response propensity; this is the "nonignorable" condition of nonresponse.

nonignorable nonresponse

In each model nonresponse bias in the simple respondent mean can be portrayed as $\sigma_{yp}/\bar{p}$, where σ_{yp} is the covariance between a given survey variable, y,

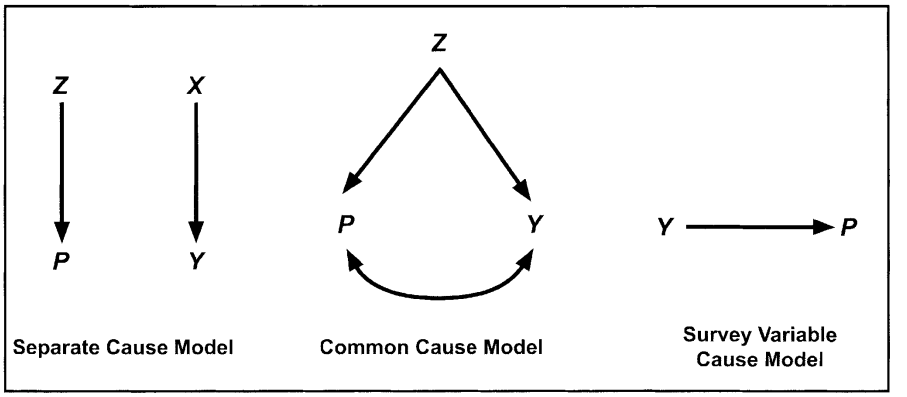


Figure 6.6 Alternative models for relationship between response propensity (*P*) and survey variable (*Y*), involving auxilliary variables (*S*, *Z*). (Source: Groves, 2006.)

and the response propensity, p; and $\bar{p}$ is the expected propensity over the sample members to be measured (Bethlehem, 2002). The separate cause model would produce a zero covariance, the common cause model would produce a nonzero covariance (but a zero covariance controlling for Z), and the survey variable cause model would have a nonzero covariance.

The expression above reminds us that nonresponse bias varies over different estimates within a survey, as a function of whether the likelihood of survey participation is related to the variable underlying the estimate. The scientific question associated with this expression is "what causes a correlation between Y and P" or "what causes a survey variable to be correlated to the likelihood to respond?"

6.5 Dissecting The Nonresponse Phenomenon

Methodological research has found that three types of unit nonresponse have distinctive causes and, for many surveys, distinctive effects on the quality of survey statistics. They are:

1) The failure to deliver the survey request (e.g., "noncontacts," failure to locate the sample unit, or postmaster returns mail surveys).
2) The refusal to participate (e.g., a contacted person declines the request).
3) The inability to participate [e.g., a contacted person cannot understand the language(s) of the questionnaire].

6.5.1 Unit Nonresponse Due to Failure to Deliver the Survey Request

Nonresponse due to noncontact or failure to deliver the survey request misses sample persons whose activities make them unavailable in the specific mode of data collection. The key concept is the "contactability" of sample units—whether

they are accessible to the survey researcher. Figure 6.7 shows a basic diagram of the influences acting on the contactability of sample units.

In household surveys, if we knew when people were at home and accessible to us, we could make a successful contact in one attempt. However, the accessible times of sample units are generally unknown; hence, interviewers are asked to make multiple calls on a unit until a first contact is made. Some sample units have "access impediments" that prevent strangers from contacting them (e.g., locked apartment buildings or telephone answering machines). People who throw away mail from unfamiliar sources often are missed in mail questionnaire surveys. People who are rarely at home often remain uncontacted even after repeated call attempts by interviewers. People who have call blocking services on their telephones often are not aware of the attempts of telephone interviewers to reach them.

For example, about 2% of sample households in the NSDUH are never contacted at the screening stage. The same statistic for the SOC is much higher. In the NSDUH, these noncontacted units are likely to be disproportionately in multiunit structures and other structures with access impediments. Because the SOC is a telephone survey, such structures do not necessarily pose contact problems; instead, houses with caller ID and call blocking devices tend to be disproportionately noncontacted.

In practice, the percentage of successful calls declines with each successive call. For example, Figure 6.8 shows the percentage of sample U.S. households contacted by call number, among those never yet contacted over five different household surveys, some telephone, some face-to-face. About half of the contacted households are reached in the first call. With each succeeding call, a smaller and smaller percentage is reached. Variation across the surveys in the figure probably reflect effects of sample design variation and calling rules after the first call.

What predicts how many calls are required to gain first contact in household surveys? There are two principal answers:

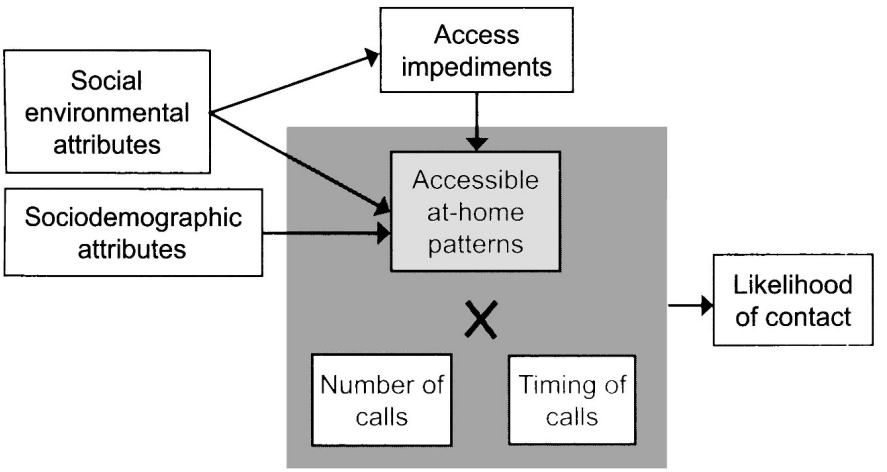


Figure 6.7 Causal influences on contact with sample household.

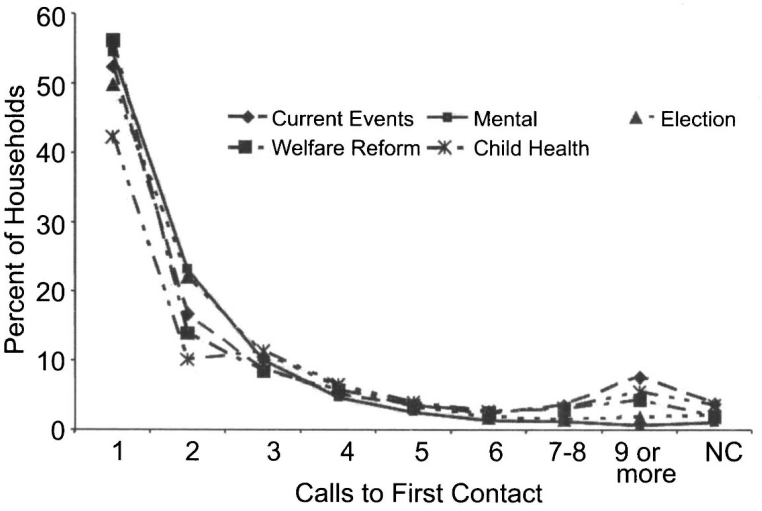


Figure 6.8 Percentage of eligible sample households by calls to first contact for five surveys. (Source: Groves, Wissoker, Greene, McNeeley, and Montemarano, 2001.)

1) Calls in the evening and on weekends are more productive than calls at other times.
2) Some populations have different accessibility likelihoods than others.

Sample persons tend to be more accessible to interviewers when they are at home. When are people at home? Most households have very predictable schedules. For those who are employed out of the home, most are away from home at set times, often the same periods each week. Most employed persons in the United States are away from home in the period of 8:00 AM to 6:00 PM Monday through Friday. If interviewers call at those times, proportionately fewer persons are reached. As Figure 6.9 shows, the best times appears to be Sunday through Thursday evenings from 6–9 PM local time, no matter what call attempt number is considered. Those evenings share the feature that the next day is a work day. Friday and Saturday evenings are different, with lower rates of contact in general. Times during the day on the weekends are better than times during the day during the work week. As it turns out, there are very few households in the United States where no one is ever at home in the evening.

There are systematic differences in noncontact rates across subpopulations in the United States. The easiest households to contact tend to be those in which someone is almost always at home. These include households with persons who are not employed outside the house, either because they are retired, they care for young children not yet in school, or some other reason.

One measure of difficulty of accessing different types of persons is the mean number of calls required to achieve the first contact with someone in the household. Persons in households that have some access impediments require more calls to first contact. These include apartment buildings with locked central

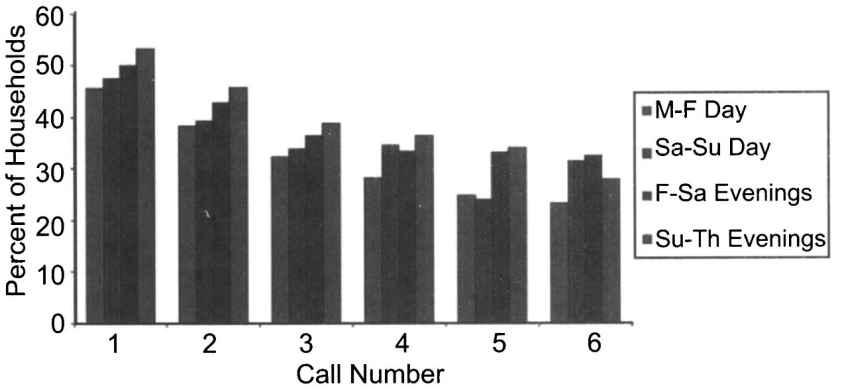


Figure 6.9 Percentage household contacted among those previously uncontacted by call number by time of day. (National Survey of Family Growth, Cycle 6.)

entrances, gated communities, or rural residences with locked gates. In telephone surveys, numbers connected to caller ID or other screening devices require more calls to first contact. Persons who live in urban areas tend to require more calls to first contact (partly because they tend to be single-person households in units with access impediments).

Mail, e-mail, and Web surveys (those modes that do not use interviewers to make contact) make the survey request accessible to the sample person continuously after it is sent. That is, in a mail survey, once the mail questionnaire is received by the household, it remains there until some household member does something with it. This can be at any time of the day or day of the week. Thus, the number of attempts required to gain contact with the sample unit has a different pattern. Chapter 5 notes that one attribute of these self-administered modes is that the researcher cannot easily differentiate a questionnaire that has not been seen by the sample household from one that has been seen and rejected.

In short, in interviewer-assisted surveys some of the noncontact arises because of at-home patterns of different groups and the number of different persons in the household. Other noncontacts arise from access impediments. The former affects all modes of data collection. The latter is specific to individual modes. For example, locked apartment buildings affect the productivity of face-to-face interviewers but not necessarily telephone interviewers. Thus, the composition of noncontacts can vary by mode. A corollary of this is that surveys that combine multiple modes can reduce the noncontact rate.

There are unanswered research issues about noncontact nonresponse that need to be tackled in the future. Most noncontact nonresponse is likely to be independent of the purpose of the survey. That is, the sample unit is not difficult to contact because of the topic of the survey but rather because of a set of influences that would be present for any survey request. This implies that nonresponse error would arise only for statistics related to those influences. Research into why sample units are difficult to contact may yield practical guidance to researchers on when nonresponse error might arise from noncontact and when it would not.

Estimates Biased by Noncontact Nonresponse. Bias flows from nonresponse when the causes of the nonresponse are linked to the survey statistics measured. For example, if one is mounting a survey whose key statistic is the percentage of persons who are employed outside the home and live by themselves, as Figure 6.10 shows, the value of the statistic is heavily affected by noncontact nonresponse. The figure shows estimated nonresponse bias for four different surveys. Imagine that we were interested in estimating what percentage of households contained only one person. For statistics like that, there are external sources of estimates from the Census Bureau, which have very high accuracy. If one would make only one call attempt on the households, the survey would disproportionately miss households with one person because they tend to be at home less frequently. The value of the statistic is underestimated by 27 to 35% if the design calls for a maximum of one call. A two-call survey overestimates the statistic by 20–32%. This statistic is sensitive to noncontact error because noncontact is partially caused by the amount of hours the sample person spends at home, accessible to the interviewer. In contrast, it is quite likely that another statistic, say the percentage of persons who report an interest in politics, is not as heavily affected by the noncontact rate.

6.5.2 Unit Nonresponse Due to Refusals

It could well be argued that the essential societal ingredients for surveys to gain cooperation of sample persons are rare in human history (Groves and Kahn, 1979, p. 227). Success requires the willingness of persons to respond to a complete stranger who calls them on the telephone, mails them a request, or visits their home. The persons must have little fear of physical or financial harm from the

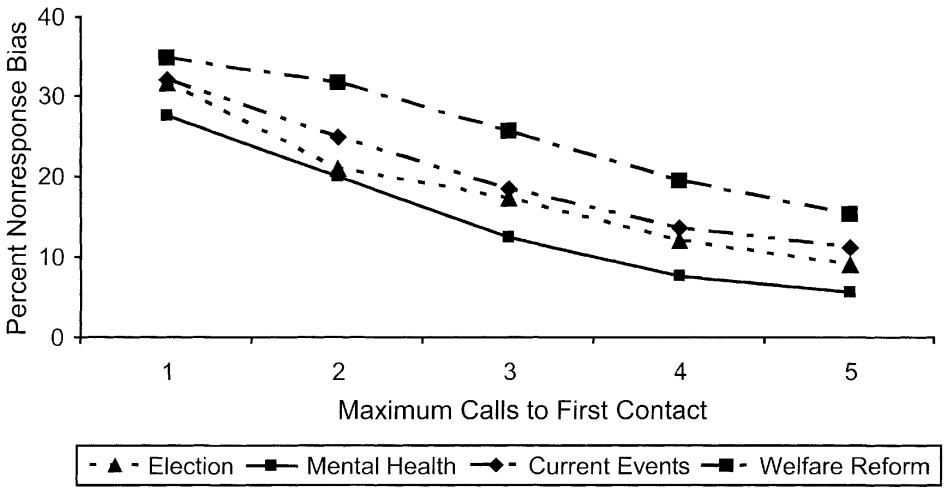


Figure 6.10 Percentage nonresponse bias for estimated proportion of single person households, by number of calls required to reach the house hold, for four surveys. (Source: Groves, Wissoker, Greene, McNeeley, and Montemarano, 2001.)

stranger, of reputational damage from the interaction, or of psychological distress caused by the interview. The sample persons must believe the pledge of confidentiality that the interviewer proffers; they must believe that they can speak their minds and report intimate details without recrimination or harm. Think, for a moment, about whether there are societies right now in the world that miss one or more of those ingredients.

Despite the ubiquity of surveys, in most countries they are not daily (or even frequent) experiences of individual persons. But the frequencies of surveys have increased over time. For some years, a survey was conducted that asked persons whether they had participated in any other survey in the past few years. In 1980, the percentage reporting participating in the last year was about 20%; in 2001, about 60%. (These are likely to be overestimates because they do not represent the answers of nonrespondents to the survey that asked about past participation) (Sheppard, 2001). In the United States, two types of survey efforts are most common. First, many service-providing organizations conduct customer satisfaction surveys, which collect ratings of the quality of the service. Second, every two years legislative elections produce large numbers of surveys, especially in congressional districts with close races. Outside of these circumstances, the frequency of survey requests to most people is low.

Groves and Couper (1998) put forward the hypothesis that when persons are approached with a survey request, they attempt quickly to discern the purpose of the encounter with an interviewer. Since the encounter was not initiated by the householders, they value returning to their original pursuits. The fact that there are common reasons for requests, purposes of requests, and institutions that are making requests, may lead to standardized reactions to requests. These might be default reactions that are shaped by experiences over the years with such requests. Survey requests, because they are rare relative to the other requests, might easily be confused by householders and misclassified as a sales call, for example. When this occurs, the householder may react for reasons other than those pertinent to the survey request. The fact that surveys often use repeated callbacks is probably an effective tool to distinguish them from sales calls. When surveys are conducted by well-known institutions that have no sales mission, interviewers can emphasize the sponsorship as a means to distinguishing themselves from a salesperson. There are several findings from survey methodological research that support the hypothesis that some sample persons may misidentify the intent of the survey interviewer:

1) Decisions to decline a request are made quickly (on the telephone most refusals take place in less than 30 seconds).
2) Persons who refuse at one point often accept when recontacted (so-called refusal conversion rates often lie between 25–40%).
3) Some survey interviewers are trained to attempt to avoid the misattribution (e.g., they say "I'm not selling anything" early in the call).

The speed of decision making suggests that persons focus on a small number of features of the request that are salient to them. Interviewer introductions on telephone and face-to-face surveys are often quite brief (see box). Some persons might choose the sponsorship of the survey as most salient; others, the vocal attributes of the interviewer; still others, the topic of the survey. In some self-

administered surveys (e.g., mailed paper questionnaires), seeing the questions makes more salient the purpose of the measurement.

Why Unit Nonresponse Occurs. The causes of nonresponse are of growing interest to survey methodologists. Many theoretical frameworks have been applied to the survey participation problem, but they all involve influences arising from four different levels:

1) The social environment [e.g., large urban areas tend to generate more refusals in household surveys; households with more than one member generate fewer refusals than single-person households (Groves and Couper, 1998)].
2) The person level [e.g., males tend to generate more refusals than females (Smith, 1983)].
3) The interviewer level [e.g., more-experienced interviewers obtain higher cooperation rates than less-experienced interviewers (Groves and Couper, 1998)].
4) The survey design level (e.g., incentives offered to sample persons tend to increase cooperation).

The first two influences are out of the control of the researcher. For example, events that have nothing to do with a survey request affect how people react to the request (e.g., the Tuskegee experiment on syphilis among African American men is often cited as an influence on lowered response rates in that population, see Section 11.5.2). The last two influences, the interviewer level and the survey design level, are features that the researcher can manipulate to increase response rates (this is discussed in Section 6.6).

The theoretical perspectives that are most commonly applied to survey participation include "opportunity cost" hypotheses, based on the notion that busy persons disproportionately refuse to be interviewed because the cost of spending time away from other pursuits is more burdensome than for others. They include notions of "social isolation," which influence persons at the high and low ends of the socioeconomic spectrum to refuse survey requests from the major institutions of the society; notions of "topic interest," which fuel hypotheses about how the "interested" disproportionately responding may induce nonresponse error in key statistics; and notions of "oversurveying" that suggest fatigue from survey requests. There are many variants of these concepts that arise in different disciplines. Unfortunately, there seems to be spotty support for any one theory explaining the phenomenon. Most of the theories offer explanations that fit into the "person level" or the "social environment level" mentioned above.

Most theories do not inform how the diverse influences on participation to cooperate manifest themselves

The "I'm Not Selling Anything" Phenomenon

Evidence that sample persons misjudge the intent of interviewers has led to what some interviewers are sometimes trained to say in their introduction: "Hello, I'm Mary Smith from the RCDF Research Services. This is not a sales call. I'm conducting a survey on your recent experiences with telephone service." This is clearly an attempt to correct a feared misinterpretation of why the interviewer is calling.

Does it work? Experiments in using the technique show mixed results (van Leeuwen and de Leeuw, 1999). The effects are probably dependent on whether other information provided in the introduction underscores the credibility of the interviewer when she says, "I'm not selling anything."

at the moment the decision is made. One theoretical perspective that attempts to describe the underpinnings of these behaviors is called "leverage-salience theory" (Groves, Singer, and Corning, 2000). Under the theory, different persons place different importance on features of the survey request (e.g., the topic of the survey, how long the interview might take, the sponsor of the survey, or what the data will be used for). Some persons may positively value some attribute; others, negatively. Of course, these differences are generally unknown to the survey researcher. When the survey approach is made to the sample person, one or more of these attributes are made salient in the interaction with the interviewer or the survey materials provided to the sample person. Depending on what is made salient and how much the person negatively or positively values the attribute, the result could be a refusal or an acceptance.

In Figure 6.11, two different sample persons are represented by a scale, which if tilted to the right implies an acceptance, and to the left, a refusal. Prior to the exposure to the survey request, Person 1 most positively values the topic of the survey (the position of the "topic" hook is to the right of the fulcrum); Person 2 is quite uninterested in the topic. Person 1 has limited free time and is very sensitive to the time demands of a request; Person 2 has no sensitivity toward the burden of the interview. Person 1 is only mildly sensitive to an incentive offer; Person 2 is quite positively disposed to receiving a cash incentive. When the interviewer makes contact, she emphasizes the sponsor of the survey and the incentive (this emphasis is represented in the figure by the size of the ball weights placed on the scale hooks). The result of the request is that Person 1 is more likely to accept the request than Person 2. Using the metaphor of a scale, the value the sample person places on a specific attribute of the request is called the "leverage" of the request. How important the attribute becomes in the description of the request is called its "salience." There are several implications of this theory:

What Interviewers Say

Interviewers are trained to provide various pieces of information quickly in their introduction:

On the telephone, "Hello, my name is Mary Smith. I am calling from the University of Michigan in Ann Arbor. Here at the University we are currently working on a nationwide research project. First, I would like to make sure I dialed the right number. Is this 301-555-2222?"

OR

In face-to-face encounters, "Hello, my name is Mary Smith, and I work for the University of Michigan's Survey Research Center. Let me show you my identification. The University of Michigan is conducting a nationwide survey, and we are interested in talking to people about their feelings on a variety of topics, including their feelings about the economy, the upcoming presidential elections, and some important issues facing the country these days. You should have received a letter from the University of Michigan telling you about this survey. "

1) People have many different reasons for acceding to, or declining, a survey request, and these are not known at the outset by the requestor.
2) No one introduction is suitable to address the concerns of diverse sample persons.
3) Interviewers must have ways of quickly discerning the concerns to make salient those attributes given positive leverage by the sample person.

We discuss this in more detail in Section 6.6.

Surveys that make their requests without using interviewers make more or less salient various attributes of the request, generally through words or symbols

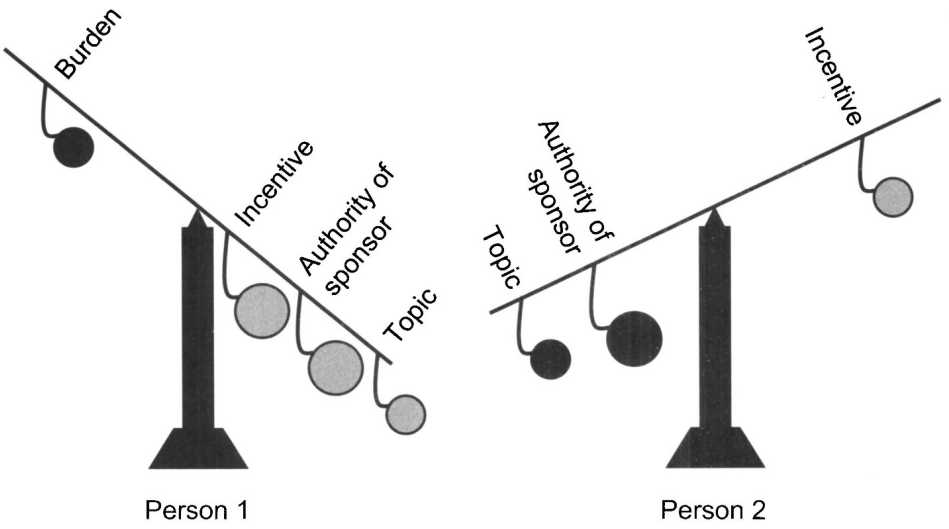


Figure 6.11 Two sample persons with different leverages for attributes of a survey request.

in the survey materials. These might include prominent display of the sponsor of the survey through envelopes and stationery, or muted display of the sponsorship. They might include a cash incentive affixed to the cover letter in a prominent position. It might involve placing sensitive questions toward the end of the questionnaire, attempting to mute their saliency prior to the decision of the sample person.

Estimates Biased from Nonresponse Due to Refusals. Similar logic underlies how refusal nonresponse error is generated. If the cause of the refusal is related to key statistics of the survey, the refusal rates portend nonresponse error in those statistics. Evidence for this is not plentiful but one example involves a comparison of a survey design that used an incentive for respondents and one that did not. The incentive acted to increase the cooperation of sample persons who were not interested in the survey about physician-assisted suicide, a topic then being debated in the community. One statistic was whether the respondent was involved in the community (e.g., a member of a civic organization or an attendee at political events). In the survey without the incentive, about 70% of the respondents reported such activity. In the survey design without the incentive, about 80% reported such activity (Groves, Singer, and Corning, 2000). Levels of community involvement are overestimated in the design without the incentive.

There are some ubiquitous correlates of the tendency to refuse a survey request. That is, many persons place "hooks" on the left of the scale above because they find these relevant to a survey request. For example, sponsorship effects are common, with central government surveys generating higher response rates than academic than commercial surveys (Groves and Couper, 1998). Burden of the survey as measured by pages in a self-administered questionnaire produces lower response rates (Goyder, 1985; Heberlein and Baumgartner, 1978); burden as measured by length of telephone and face to face interviews shows less clear

effects (Bogen, 1996). Males tend to refuse more than females (Smith, 1983). Urbanicity is a powerful indicator of response rates in all modes (de Leeuw and de Heer, 2002). Adults who live alone tend to be refusers (Groves and Couper, 1998); households with young children show higher response rates than others (Lievesley, 1988). When the key variables of the survey are related to these attributes, we can anticipate nonresponse biases in the respondent-based estimates.

6.5.3 Unit Nonresponse Due to the Inability to Provide the Requested Data

Sometimes, sample persons are successfully contacted and would be willing to be respondents, but cannot. Their inability stems from a myriad of sources. Sometimes, they cannot understand any of the languages that are used in the survey. Sometimes, they are mentally incapable of understanding the questions or retrieving from memory the information requested. Sometimes, they have physical health problems that prevent them from participating. Sometimes, because of literacy limitations, they are unable to read or understand the materials in a mail survey. Sometimes, in business surveys, establishments do not have the necessary information available in the format or time frame required by the survey.

Since the reasons for their inability to comply with the survey request are diverse, it is also the case that what statistics are affected by the nonresponse are diverse. For example, in a survey measuring the health characteristics of a population, the inability to respond for health reasons portends nonresponse biases. Estimates of the well-being of the population would be overestimated because of the systematic failure to measure the unhealthy. On the other hand, estimates of the political attitudes of the population in the same survey may not be as severely affected by the same causes of nonresponse.

There has been relatively little methodological research into the causes and effects of unit nonresponse of this type. In many household surveys, the relative amount of such nonresponse is small. However, in studies of the elderly or immigrant populations, the rates can rise to substantial levels. For such designs, research discoveries have yet to be made regarding the role played by the interviewer, the method of data collection, features of the languages used, and characteristics of the sample persons.

6.6 Design Features to Reduce Unit Nonresponse

At this point in the chapter, we have noted that nonresponse error can harm the quality of survey statistics, that the extent of harm is a function of how the influences toward nonparticipation relate to the statistic in question, and that the larger the nonresponse rate, the larger the risk of nonresponse error (other things being equal). In general, the only visible part of the quality impacts of nonresponse is the nonresponse rate. Hence, much effort of survey researchers has been focused on reducing nonresponse rates.

Figure 6.12 decomposes survey participation into three steps: contact, the initial decision regarding participation, and the final decision regarding participation. The last step arises because many survey designs use repeated efforts to address the concerns of reluctant respondents.

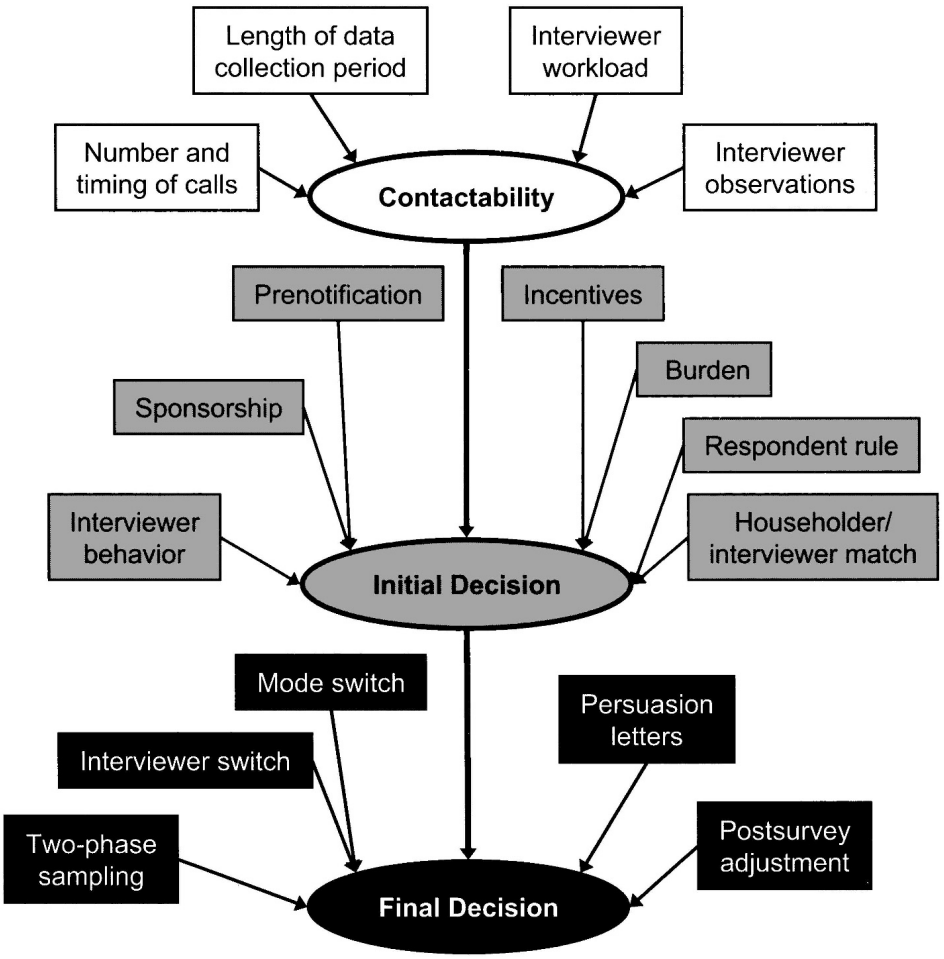


Figure 6.12 Tools for reducing unit nonresponse rates.

We have learned in Chapter 5 that the different modes of data collection tend to have different average response rates. The typical finding is that face-to-face surveys have higher response rates than telephone surveys. Telephone surveys have higher response rates than self-administered paper and Web surveys, other things being equal. It is a common finding that the use of interviewers in face-to-face and telephone surveys increases response rates, both because of higher success at delivering the survey request and because of their effectiveness in addressing any concerns about participation that sample persons may have.

There are several features in Figure 6.12 that address interviewer actions. First, leverage-salience theory of survey participation offers several deductions about interviewer behavior. Recall that different sample persons are likely to vary in how they evaluate the survey request (assigning different "leverages" to different attributes). Since these are unknown to the interviewer, the interviewer must somehow discern them in order to gain their cooperation.

One further deduction from leverage-salience theory is that training interviewers to recite the same introductory description to each sample person will not

be effective. This appears to have empirical support. When Morton-Williams (1993) compared response rates for interviewers instructed to recite a standard script with those trained to address the sample persons in a freer manner, the unscripted interviewers had higher response rates (see box on the Morton-Williams study on p. 206). Groves and Couper (1998) propose two principles of interviewer behavior that may underlie the Morton-Williams experimental findings: maintaining interaction and tailoring. Expert interviewers appear to engage the sample persons in extended conversations (whether or not they are pertinent to the survey request). The interviewers "maintaining interaction" in this way attempt to gain information about the concerns of the sample person. Effective interviewers then "tailor" their remarks to the perceived concerns of the sample person. This tailoring appears to explain some of the tendency for experienced interviewers to achieve higher cooperation rates than novice interviewers. They carefully observe the verbal and nonverbal behavior of the sample persons in order to discern their concerns. When they form hypotheses about those concerns, the interviewers "tailor" their behavior to the concerns. They customize their description of the survey to those concerns.

maintaining interaction

tailoring

Figure 6.12 also indicates that if the initial decision of the sample person does not yield an interview, surveys seeking high response rates often make further efforts to bring the person into the respondent pool. These can involve switching interviewers, changing to a different mode, or sending persuasion letters. Finally, when the refusal decision appears final (or when all efforts to contact the sample unit have failed), then a variety of postsurvey features can be used. First, two-phase designs can be introduced whereby a sample of nonrespondents are followed up using a different recruitment protocol. Second, statistical adjustments can be made at the analysis stage (these are discussed in Chapter 10).

The remainder of this section briefly reviews the literature on methods to increase response rates, whenever possible making comments about when the method might affect the nonresponse bias of different kinds of estimates. The literature is a large one, consisting of several books (Goyder, 1987; Brehm, 1993; Groves and Couper, 1998; and Groves, Dillman, Eltinge, and Little, 2002), as well as hundreds of articles in scientific journals. The research that produced these methods often used randomized experimental designs to compare two ways of designing a survey protocol. One sample of persons received one protocol; another, the other. That protocol that produced the highest response rate was judged to be preferable. Sometimes, the researcher had available external indicators of the characteristics of sample person and, thus, could estimate whether increasing the response rate increased the quality of the survey statistics. Usually, however, the purpose of the studies is to demonstrate how to achieve higher response rates, whether or not they improved the survey statistics. We use Figure 6.12 to organize the discussion.

What Interviewers Say about Approaching Sample Households

In focus groups, interviewers describe how they prepare for a visit to a sample household:

"I use different techniques depending on the age of the respondent, my initial impression of him or her, the neighborhood, etc."

"I try to use a 'gimmick' in my attire when visiting HUs. Bright colors, interesting pins, jewelry, nothing somber or overly 'professional' or cold looking—fun items of attire like ceramic jewelry, scarves tied in odd ways. If my [initial drive through of the neighborhood] spots cats or dogs in windows or doors, I make a note and wear something like a cat pin on my coat, etc."

Number and timing of attempts to access the sample person. In both self-administered questionnaires and interviewer-assisted surveys it has been repeatedly found that the larger the number of sequential attempts to deliver the survey request to the sample person, the higher the likelihood of successfully contacting them. Goyder (1985) and Heberlein and Baumgartner (1978) show that repeated efforts to gain access to the sample units tend to reduce nonresponse. For U.S. telephone and face-to-face surveys, contact with household samples is easier on Sunday–Thursday evenings (the nights preceding work days) and during the day on weekends. Only a tiny minority of households can be reached only during the weekday day hours. As Figure 6.10 showed, these efforts radically reduce the nonresponse bias of estimates of the prevalence of single-person households. Estimates based on other strong correlates of "at-homeness" would show similar effects.

Data collection period. The longer the data collection period, the higher the likelihood that all persons are made aware of the survey request. How short is too short? It is noteworthy that the Current Population Survey is conducted in about 10 days and achieves near 100% contact rates. Thus, with proper staffing of interviewers such surveys can make initial contact with the vast majority of persons in a relatively short time. Mail surveys need longer periods because of the time requirements of the postal service. Choosing the data collection period probably affects the bias of estimates in similar ways to choosing the number of calls to make on sample cases.

Interviewer workload. Each sample case assigned to an interviewer requires time to contact. In household surveys, only about 50% of the first contact attempts are successful using the usual calling rules. If too many cases are assigned to an interviewer, insufficient effort will be given to some cases. Botman and Thornberry (1992) note when there is insufficient time or too many cases to work, nonresponse can increase, both noncontacts (because of insufficient effort to access the sample units) and refusals (because of inadequate refusal conversion effort). Since this feature also limits calls, its biasing effects should be focused on similar estimates to those of callback rules (e.g., single person households of employed persons tend to be nonrespondent).

Interviewer observations. The face-to-face mode has a distinct advantage in making contact with sample households—the ability of an interviewer to observe characteristics of the sample unit. Sometimes, these are visual observations (e.g., toys in the yard imply that children are resident); sometimes they are the result of neighbors providing information about the sample unit. Interviewers documenting what informants say in intermediate contacts can be useful to survey managers. When informants ask questions about the survey, it portends higher odds of their eventually consenting (Groves and Couper, 1998). To the extent that interviewers can observed attributes of the sample that are correlates of the key variables, the researcher can attempt to reduce bias in estimates based on those variables.

Sponsorship. In most countries of the world, the central government survey organizations attain higher cooperation rates than academic or private sector survey organizations. When the sponsor of the survey has some connection to the target population (e.g., a membership organization) the strength of the connection

is related to the response propensities. For example, random half samples in a face-to-face survey were assigned to U.S. Census Bureau and University of Michigan Survey Research Center interviewers. The cases with survey requests from the Census Bureau refused at a 6% rate; those with survey requests from the University of Michigan refused at a 13% rate (National Research Council, 1979). Many speculate that the higher response rates of governments arise from (a) the widespread understanding that democratic governments legitimately need information from the residents to benefit them, or (b) a belief that participation is mandatory.

Prenotification. An advance letter mailed to a sample household can generate higher rates of cooperation than no such prenotification (e.g., see Traugott, Groves, and Lepkowski, 1987; for a contrary finding, see Singer, Van Hoewyk, and Maher, 2000). If the letter explicitly notes the sensitive content of the interview, however, it can depress response rates [as in ACSF (1992) cited in deLeeuw, Hox, Korendijk, Lensvelt-Mulders, and Callegaro, 2007]. Further, its effect may be dependent on the agency of the letter author, perhaps as evidenced by the stationery. For example, in a randomized experiment the same letter written on a market research letterhead received a lower response rate than one on university letterhead. In fact, the market research company letter generated a lower response rate than no letter at all (Brunner and Carroll, 1969). Interviewers appear to value such letters, and for that reason alone they are standard practice in most scientific surveys.

What kinds of estimates might have their nonresponse bias affected by advance letters? Unfortunately, there are few studies showing differential effects of letters across subgroups. One could speculate that nonresponse biases sensitive to advance letters would be located among correlates of likelihoods of receipt or readership of the letters. For example, since advance letters are a device requiring literacy for some of their

Berlin, Mohadjer, Waksberg, Kolstad, Kirsch, Rock, and Yamamoto (1992) on Incentives and Interviewer Productivity

Berlin et al. (1992) discovered that offering incentive can *reduce* total survey costs.

Study design: Three randomized treatment groups promised no incentive, a $20 incentive, or a $35 incentive upon the completion of literacy assessments in a pretest of the National Survey of Adult Literacy. The interviewer began with background questions and then gave the respondent the assessments for self-completion. The study used an area probability sample of over 300 segments of census blocks with about 2800 housing units, with each segment assigned to one of the treatments.

Findings: The incentives brought into the respondent pool less educated persons who in the past tended to be nonrespondent. As shown below, the $35 incentive achieved the highest response rates; the no-incentive condition, the lowest.

	Incentive Level		
	None	$20	$35
Response Rate	64%	71%	74%
Interviewer costs	$130	$ 99	$ 94
Incentive costs	0	20	35
Total costs	$130	$ 119	$129

With fewer interviewer hours per case, the $20 incentive more than paid for itself, relative to no incentive. The $35 incentive group was not as cost-effective, even though its response rate was higher.

Limitations of the study: Applying the results to other studies is limited by the fact that the survey task (i.e., self-completion of literacy assessment packages) is unusual, the survey was sponsored by a government agency, and the number of primary areas was small (16). This produced standard errors that themselves were rather unstable.

Impact of the study: The finding that giving incentives could save money was counterintuitive and important.

Morton-Williams (1993) on Tailoring Behavior by Interviewers

Morton-Williams (1993) presented the results of a randomized experiment in Britain that supports the value of interviewers' tailoring their introductory remarks to the concerns of the sample person.

Study design: Fourteen of 30 interviewers were asked to use a scripted introduction to a face-to-face survey; the complement were asked to use their own judgment about the content of their remarks. All interviewers were asked to give their own names, identify their organization, show an identity card, mention the indepedence of their organization, describe the survey, and note how the address was selected. The scripted introduction specified the exact words the interviewers were to use.

Findings: The scripted interviewers obtained a 59% response rate; the unscripted, a 76% response rate.

Limitations of the study: The description does not specify how interviewers or housing units were assigned to the treatment groups. Hence, there may be some confounding with true differences among units that cooperate and the treatment groups. The statistical analysis does not appear to account for the interviewer component of variability; it is not clear that the response rate differences are beyond those expected by chance, given the small numbers of interviewers.

Impact of the study: The experiment was the first to support the growing evidence that scripting interviewers in the recruitment phase of a survey may harm cooperation rates.

effects, one would expect that letters would have lower effects in semiliterate populations.

Incentives. Offering an extrinsic benefit to participation increases cooperation rates (Singer, 2002). Cash incentives tend to be more powerful than in-kind incentives of similar value. Incentives paid prior to the survey request are more powerful than those offered contingent on the survey being completed. Higher incentives are more powerful, but some studies report diminishing returns as the amount rises. If the incentive effect is powerful enough, the total cost of the survey can actually decline because of lower interviewer or follow-up costs (see the box on p. 205).

There is some evidence that when an incentive is used, it disproportionately affects the participation of persons less interested in the topic (Groves, Couper, Presser, Singer, Tourangeau, Piani Acosta, and Nelson, 2002). Without an incentive, the respondents consist of those interested in the topic (and often reporting different attributes on the key variables). With an incentive, the respondent pool better reflects the full population.

Burden. The evidence for how survey length or cognitive burden affects cooperation is mixed (Singer and Presser, 2007). However, there is evidence that perceived length of time or complexity of self-administered instruments reduces cooperation. For example, in examining a large number of surveys, both Heberlein and Baumgartner (1978) and Goyder (1985) found that each additional page of a self-administered questionnaire reduced the response rate by 0.4 percentage points.

Respondent rules. Surveys that permit several alternative persons in a household to provide the survey information tend to have higher cooperation rates than those selecting a random adult respondent. Similarly, surveys that permit proxy reporting have higher response rates than those requiring self-report. There is some indication of bias effects of such rules. For example, in health surveys the prevalence of impairments and illnesses is often higher in surveys accepting proxy reports (since illness is both a cause of nonresponse and a survey variable).

Interviewer introductory behavior. Especially in telephone surveys, the first few seconds of interviewer interaction affect cooperation rates. There is little

empirical research on this, but what exists suggests that variation in inflection and speed of communication are related to higher cooperation rates (Oksenberg, Coleman, and Cannell, 1986). Morton-Williams (1993) shows that rigid scripting of interviewers' introductory remarks can increase refusals (see the box on p. 206). Given the causal approach to nonresponse bias reviewed above, one would worry about biasing effects of these behaviors if interviewers tended to use the survey content as a persuasive tool.

Interviewer/household matching. There is no research but much speculation that matching interviewers with sample persons in a way that improves the likelihood of trust and acceptance of the interviewer improves cooperation. For example, Nealon (1983) shows that female interviewers obtained higher cooperation rates than male interviewers for a sample of spouses of farm operators. Similar practices of using "indigenous" interviewers are common in studies in anthropology. When thinking about nonresponse bias and this practice, the researcher has to consider whether the same interviewer attribute that improves participation may affect the answers of the respondents. (We note in Chapter 9 that interviewer gender and race have been found to affect answers for survey topics related to gender and race.)

Interviewer switches. It is common for data collection supervisors to replace the interviewer who receives an initial refusal with another interviewer who has characteristics that might be more acceptable to the sample person (e.g., often attempting to match attributes as the prior paragraph noted). The new interviewer reviews the history of the case, including any documentation on the initial refusal, then approaches the sample unit again. This can be viewed as a naïve form of tailoring, by manipulating some characteristics of the interviewer to be more amenable to the concerns of the sample person. For example, a male face-to-face interviewer who appears to frighten an elderly woman living alone can be replaced with an older female interviewer.

Mode switches. There are many designs that begin with a cheaper mode of data collection (e.g., mail questionnaires and Web surveys) and then use more expensive modes (e.g., telephone and face-to-face) for nonrespondents from the first mode. In single-mode surveys, the clear pattern is that face-to-face surveys can achieve higher response rates than telephone surveys or mailed self-administered surveys. Mixing the modes allows one to optimize resources to improve cooperation. Estimates related to literacy might have nonresponse bias properties sensitive to such switches.

Persuasion letters. It is common to send persons who initially refuse to be interviewed a letter that reinforces the serious purpose of the survey and notes that an interviewer will again call on the household to address any concerns. Common practice attempts to tailor the letters to expressed concerns of the sample unit (e.g., sending letters emphasizing confidentiality to those persons concerned about their privacy).

This brief review suggests that there are a number of methods employed to reduce nonresponse in surveys. Some of these are based on firm research evidence; others are based on practical experience. Not all techniques work in all circumstances. Examples of failed replications of experiments can be found. The

challenge for the practitioner is to determine the strategies likely to be most successful given a particular design and target population.

two-phase sampling

postsurvey adjustment

The final two boxes in Figure 6.12 are tools that require different statistical analysis techniques. Two-phase sampling draws a probability sample of nonrespondents, for whom new methods of contact and requests for participation are used. Those sampled who are thereby measured are used to estimate the characteristics of all the nonrespondents existent at time of the sampling. Postsurvey adjustment (see Chapter 9) is a technique using the existing respondents in ways that compensate for those nonrespondents who were missed (e.g., giving more weight in the analysis to urban respondents than rural ones when the response rates in urban areas are lower).

Despite the well-developed methodological literature on ways to increase response rates, there are many unanswered research questions. These include:

1) When efforts to interview reluctant respondents succeed, do they provide responses more tainted by measurement errors?
2) When do efforts to increase response rates affect nonresponse errors and when do they not?
3) How should efforts to reduce noncontact versus refusal rates be balanced against one another?
4) Considering both sampling error and nonresponse error, given a fixed budget, when can the researcher justify less than full effort to reduce nonresponse rates?

The research in this arena is likely to be very important for the future of the field, because survey designs are now devoting large portions of their research budgets to nonresponse reduction, with little scientific basis for this practice.

6.7 Item Nonresponse

The discussion above centers on unit nonresponse (the failure to obtain any measures on a sample unit). Item nonresponse is a severe problem in some kinds of surveys. Item nonresponse occurs when a response to a single question is missing. For example, a respondent to the Survey of Consumers may willingly participate and start answering questions, but when the interviewer asks about family income in the last year, they may refuse to give an answer to that question.

The impacts of item nonresponse on a statistic are exactly the same as that for unit nonresponse, but the damage is limited to statistics produced using data from the affected items. Thus, the expression on page 189 for a sample mean applies to the combined effects of unit and item nonresponse.

It often appears that the causes of item nonresponse are different from those of unit nonresponse. Whereas unit nonresponse arises from a decision based on a brief description of the survey, item nonresponse occurs after the measurement has been fully revealed. The causes of item nonresponse that have been studied in methodological research include: (a) inadequate comprehension of the intent of the question, (b) judged failure to retrieve adequate information, and (c) lack of willingness or motivation to disclose the information (see Beatty and Herrmann, 2002; Krosnick, 2002). However, research in this area is in its

infancy. Most of the methodological research on question wording (see Chapter 7) has focused on properties of the questions that change the substantive answers. A research program in the causes of item-missing data would be of great utility to practitioners.

There is evidence for item-missing data arising from respondents' judging that they do not have adequately precise answers. Some experiments show persons failing to give a specific income value but willingly providing an estimate within an income range (e.g., between $50,000 and $75,000; Juster and Smith, 1997). Evidence for the effect of motivation for the respondent role is that open questions (requiring the respondents to write in answers they invent) tend to have larger missing data than closed questions (requiring the respondents to choose from a list of answers).

Figure 6.13 is a model of the response process posited by Beatty and Herrmann, which distinguishes four levels of cognitive states regarding the information sought by the survey question: available, accessible, generatable, and inestimable. The four states are ordered by level of retrieved knowledge suitable for a question response. They posit both errors of commission (reporting an answer without sufficient knowledge) and errors of omission (failing to report an answer when the knowledge exists). Social influence to give an answer may produce data with measurement errors. Item-missing data can arise legitimately (for those in an "inestimable" cognitive state) or as a response error (for those with the knowledge available). The latter situation might arise when social desirability

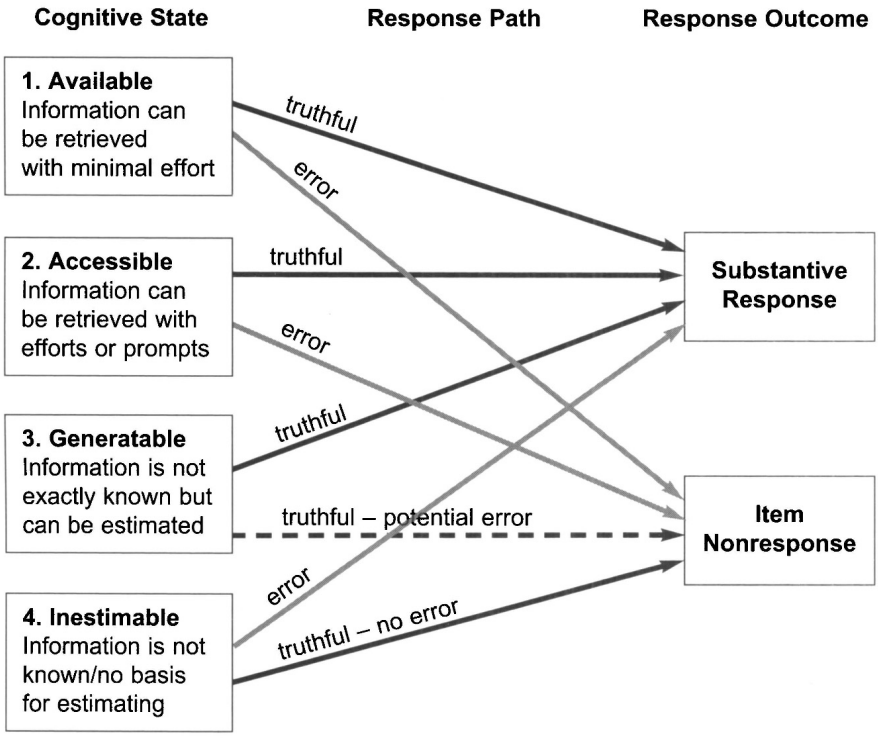


Figure 6.13 Beatty-Herrmann model of response process for item-missing data.

influences a respondent to refuse to answer a question (or answer, "don't know") instead of revealing a socially unacceptable attribute.

The tools used to reduce item nonresponse are reduction of the burden of any single question, the reduction of psychological threat or increase in privacy (e.g., self-administration), and interviewer actions to clarify or probe responses. The strategies used to compensate for item nonresponse are often quite different from those for unit nonresponse, as in the former case the analyst usually has a rich vector of other responses with which to adjust. Thus, imputation is most often used for item-missing data, whereas weighting class adjustments are more common for unit nonresponse (see Chapter 10).

6.8 Are Nonresponse Propensities Related To Other Error Sources?

Before we close out this chapter, the reader should be alerted to some survey methodological research that shows that nonresponse may be linked to other errors, especially coverage and measurement errors. From the literatures on censuses, we see that young, single adult males, who often have tenuous ties to several households (and thus tend to be missed in address based sampling frames), also tend to be nonrespondent. Similarly, there is evidence that those most reluctant to respond sometimes provide answers to questions that are less thoughtful and subject to greater measurement error (see Olson, 2006). Although this research is just now emerging, it is another important reminder that merely increasing response rates is an overly simple reaction to fears of nonresponse bias in survey estimates.

6.9 Summary

Surveys produce numbers that attempt to describe large populations by measuring and estimating only a sample of those populations. When the designated sample cannot be completely measured and estimates are based only on responding cases, the quality of survey statistics can be threatened. There are two types of nonresponse: unit nonresponse and item nonresponse.

In their simplest form, nonresponse rates are the ratios of eligible sample units that were not measured to the total number of eligible units in the sample. In practice, nonresponse rates are sometimes difficult to compute because the eligibility of nonrespondents remains unknown and the sample design assigns varying probabilities of selection to different frame elements.

Not all nonresponse hurts the quality of survey estimates. Nonresponse produced by causes that are related to key survey statistics is the most harmful kind (e.g., failure to contact persons who are rarely at home in a survey of how they use their time). Such nonresponse is termed "nonignorable" nonresponse. Nonresponse can harm the quality both of descriptive and analytic statistics. Different statistics in the same survey can vary in their magnitudes of nonresponse error.

There are three classes of unit nonresponse that have different causes and, hence, affect the quality of survey statistics in different ways: inability to gain

access to the sample unit, failure to gain the cooperation upon delivery of the survey request, and the sample unit's inability to provide the data sought.

Nonresponse rates by themselves do not predict the nonresponse error of individual estimates in a survey. Indeed, there is evidence of large variation of nonresponse error among estimates within the same survey. Despite continuing ignorance about when nonresponse matters and when it does not, there is a strong set of professional norms to increase the response rates of surveys. Guidelines of most survey professional organizations describe efforts to obtain high response rates.

There are many tools that survey researchers have to increase the response rates in surveys. These include repeated callbacks, long data collection periods, small interviewer workloads, using interviewer observations to guide their behaviors, advance letters, using trusted sponsors, short questionnaires, use of proxy respondents, tailoring of interviewer behavior to the concerns of the sample person, matching interviewer and sample person characteristics, persuasion letters for initial refusals, mode and interviewer switches for reluctant respondents, and two-phase samples for nonresponse. Almost all of the methods require spending more time or effort contacting or interacting with the sample units. This generally increases the costs of surveys.

An important remaining challenge to survey researchers regarding nonresponse is determining when it hurts the quality of survey statistics and when it does not. More research is needed on this issue. Without it, there is no guarantee that efforts to increase response rates are wise. Without it, there is no way to justify being satisfied with low response rates.

KEYWORDS

access impediments
contactability
ignorable nonresponse
inability to participate
item nonresponse
leverage-salience theory
maintaining interaction
noncontact
nonignorable nonresponse
nonresponse bias
opportunity cost
oversurveying
postsurvey adjustment
refusal
response propensity
social isolation
tailoring
topic interest
two-phase sampling
unit nonresponse

FOR MORE IN-DEPTH READING

Groves, R., and Couper, M. (1998), *Nonresponse in Household Interview Surveys*, New York: Wiley.

Groves, R., Dillman, D., Eltinge, J., and Little, R. (eds.) (2002), *Survey Nonresponse*, New York: Wiley.

Särndal, C., and Lundström, S. (2005), *Estimation in Surveys with Nonresponse*, New York: Wiley.

EXERCISES

1) Shown below is the distribution of final case results for an RDD telephone survey conducted on a sample of 2127 randomly generated telephone numbers covering the contiguous 48 United States (excluding Alaska and Hawaii) plus the District of Columbia. The target population consists of households in the 48 states and the District of Columbia. The topic of the survey is the household's recycling activity: availability of recycling pickups in the household's city/town, and the household's use thereof. All phone numbers were dialed a maximum of 20 times, covering weekdays and weekends, daytime and evenings, until one of these final results was achieved.

Completed interviews	614
Refusals	224
Answering machine with residential message on every call	180
Never answered on every call	302
Eligible household contacted, but no interview for other than refusal reasons	127
Businesses/nonresidences	194
Nonworking numbers	486
Total	

Compute the response rate in three different ways, as follows (You might want to consult the "Resources for Researchers" section of www.aapor.org).

a) Assume all unanswered numbers are eligible.
b) Assume all unanswered numbers are ineligible.
c) Estimate eligibility among the unanswered numbers
d) Estimate the response rate among the unanswered numbers.

2) You conducted a survey of 1500 randomly selected county school systems nationwide. Your study aims to examine factors related to whether or not school systems offers sex education programs/classes, comparing school systems located in areas where the majority of the local population belongs to a conservative religious group (CRG) to those in areas where the members of conservative religious groups represent a minority of the local population. The sample yielded the following results, and from external sources you already know which school systems offer sex education programs/classes.

Population Within School System	Sample Size	Response Rate	Responding School Systems	Percent Offering Sex Education	
				Respondents	Non-respondents
Majority CRG	500	50%	250	5%	0%
Minority CRG	1000	60%	600	50%	35%

Assume that the sample sizes represent the proportions in the population (i.e., an equal probability sample). The estimate of the overall percent of school systems offering sex education, as computed based on responding school system reports, is thus 36.8%, computed as follows:

Majority CRG: 5% × 250 = 12.5
Minority CRG: 50% × 600 = 300
12.5 + 300 = 312.5 312.5/850 = 36.8%

Estimate the nonresponse bias in the estimate of "percent of school systems offering sex education" of the full sample.

3) You have studied the effects of incentives on cooperation in surveys.

   a) Describe the reasoning behind the common effect that prepaid incentives have larger net effects on cooperation rates than promised incentives.
   b) Describe why incentives can sometimes reduce the cost of a survey.

4) You are completing a telephone survey of members of a professional organization, regarding their level of activity in support of the organization. You face the decision about whether to use some funds to increase the response rate over that currently achieved, which is 80%. A key indicator to be estimated from the survey is the percentage of members who attend every monthly meeting of the local chapter of the organization. With the 80% response rate, the current estimate is that 42% of the membership attends monthly meetings. (For purposes of your answer, ignore sampling error differences in the estimates.)

   a) Is it possible that, in truth, over half of the membership attends monthly meetings?
   b) What arguments would you make regarding the likely characteristics of the nonrespondents on the attendance measure?

5) Briefly explain in words the relationship between the nonresponse rate and the bias of nonresponse for an estimate of the population mean.

6) You are planning a survey about health care practices in a population of participants in the past year's U.S. masters' swimming competitions. The survey interview consists of questions about what diet and exercise regimens the respondent pursues. The same questionnaire was used in a general population survey recently. You discover that a randomized experiment regarding a $10 incentive produced a 20 percentage point increase in response rates and a decline in the estimated proportion of respondents reporting they exercised. Identify one reason to expect the incentive to have similar effects and one reason why the incentive may have different effects for the survey you are planning.

7) Given what you have learned about the effect of advance letters on survey cooperation rates, make a choice of which organization's stationery to use for the letter in the following three surveys.

a) Assume that your goal is to maximize cooperation rates. Choose whether to use the sponsoring organization's letterhead (the organization paying for the survey) or the data collection organization's letterhead (the organization collecting the data, which you can assume to be relatively unknown among sample persons); then explain the reason for your decision.

Sponsor	Data Collection Organization	Target Population	Which organization's stationery for advance letter?	Reason for answer
Commercial credit company	Academic survey center	U.S. households		
Federal government	Commercial market research firm	Low-income households		
Highway Construction lobbying organization	Nonprofit research organization	Presidents of highway construction companies		

b) Now assume that you are interested in minimizing the nonresponse bias of the following estimates for each survey.

Sponsor	Data Collection Organization	Target Population	Which organization's stationery for advance letter?	Reason for answer
Commercial credit company	Academic survey center	U.S. households	Mean amount of credit card debt	
Federal government	Commercial market research firm	Low-income households	Percent with favorable attitudes about food stamp program	
Highway Construction lobbying organization	Nonprofit research organization	Presidents of highway construction companies	Percentage judging lobbying effort effective	

8) What subpopulations within the household population in your country tend to be noncontact nonrespondents in telephone surveys with restricted time periods for the data collection? Why?

9) Using leverage-salience theory, describe one hypothesis of why estimates of support for a given political candidate may be affected by the sponsorhip of the survey.

10) In studying the effect of incentives, an experiment was conducted among new college graduates. The sample was randomized into two groups, one group receiving questions related to prescription drug coverage by govern-

ment health plans and the other group receiving questions related to environmental problems. Each of these two groups was further divided into two: one receiving no incentives and another receiving $5 incentive, thus yielding a 2 × 2 experimental design.

Briefly describe one possible finding from this experiment in relation to incentives and survey topic.

11) Give three reasons why face-to-face surveys typically achieve higher response rates than telephone surveys.

12) This chapter described alternative ways to increase response rates in surveys.

a) Identify three methods used to increase response rates.
b) Thinking of the adult household population, identify a subgroup for whom one of the methods you listed in (a) does not appear to be as effective as for other subgroups, and briefly explain why.

13) The chapter shows that reducing nonresponse rates can sometimes increase nonresponse error on simple statistics like the respondent mean. Identify a survey design feature than might create such an outcome in a survey of members of a social organization (a local dance group) when attempting to measure the mean number of years that a member has been active in the organization.