8

Sampling in Special Contexts

P revious chapters have covered general issues in sampling. In addition to these general points, any given research context is likely to present its own specific sampling issues. Here, we discuss various contexts that are of interest to many social researchers. We consider the following:

- Sampling for online research
- Sampling visitors to a place
- Sampling rare populations
- Sampling organizations
- Sampling groups such as influence groups or elites
- Sampling for panel research
- Sampling in international contexts
- "Big data" and survey sampling
- Incorporating new technologies

Each of these contexts presents its own special sampling problems and its own characteristic procedures. In this chapter, you will learn about those problems and procedures.

8.1 SAMPLING FOR ONLINE RESEARCH

As noted earlier in this book, online data collection is increasingly common and may present significant sampling problems.

The first sampling problem associated with online surveys of the general population is the potential for coverage bias because many people are not online. Only 75% of the U.S. population currently uses the Internet (U.S. Census Bureau, 2012), and the figure is lower in most other countries. Coverage is disproportionately low for elderly people, African Americans, Hispanics, and people with lower levels of education and income, which implies corresponding coverage biases in general population surveys (Couper, 2000).

A second problem in general population surveys is the potential for coverage bias because available sampling frames cover only a fraction of the online population. As noted in Chapter 2, one can buy demographically balanced samples from opt-in online panels: however, these panels are typically assembled through volunteerism rather than random selection, and even if they contain millions of names, this is still a relatively small fraction of the general population, so the theoretical potential for coverage bias is high. Alternatives such as using Amazon MTurk or assembling a frame through social media may present additional coverage and selection problems, as discussed in Chapter 2.

It must be emphasized that these coverage problems relate to surveys of the general population. If the population of interest is found online, such as visitors to a website or members of a social media site, or if it is a special population for which a list of e-mail addresses is available, such as students in a college directory, then coverage is typically not an issue.

The third sampling problem in online data collection is the potential for nonresponse bias stemming from low response rates. In our experience, response rates in online surveys are often less than 5%. Response rates are better for special populations surveyed on topics of interest, such as teachers surveyed about educational issues or college students surveyed about campus issues: here, response rates for well-executed surveys may be 40% or better. Response rates also are better for samples drawn from opt-in panels, typically 25% or better, but one must consider that people are in the panel because they have already agreed to participate in research.

Overall, we can say the following:

• The problems of sampling online populations are lowest when (a) the target population has high levels of online access, which minimizes potential coverage bias from this source; (b) one has (or can assemble) a relatively complete list of e-mail addresses for the population, which minimizes potential coverage bias from this source; and (c) the research topic is of interest to the target population and the study is well executed, which minimizes potential

nonresponse bias. These conditions may be met for special populations such as professional groups or college students at a particular campus.

• Online studies of the general population are more problematic; however, in this context, panel samples are generally the best option because of their ability to provide balanced samples for which historical norms may be available. (Online panel samples also may be useful in the context of dual-frame designs when rare groups are of interest, as discussed later in this chapter.)

• Because of frame and nonresponse problems, online surveys of the general population typically must be viewed as having some form of nonprobability sampling, so estimates drawn from these studies must rely on some form of model-based estimation as discussed in Chapter 7.

Despite the problems associated with online samples, researchers increasingly rely on them. In defense of online samples, as noted in Chapter 2, the samples obtained in online research may be as good as or better than the alternatives being considered; for example, a study conducted with a sample of MTurk panelists may be at least as defensible as a study conducted with a sample of college students. Also, online surveys are wholly appropriate for special populations that are found online or for which a list of e-mail addresses is available. Finally, online samples may be good enough for the purposes at hand. We discuss the question "How good must the sample be?" in Chapter 9.

Some online studies are done by posting an invitation to participate on a listserv, discussion board, or Facebook page. Samples obtained in this manner are, of course, nonprobability samples—samples of volunteers with no controls on sample quality. They may be acceptable in contexts where other volunteer samples are accepted; for example, in academic experimental research where participants are randomly assigned to experimental conditions and wide latitude is given regarding sample quality, or perhaps for exploratory purposes. However, these samples are not appropriate for situations in which one wishes to generalize numerical estimates from the observed sample to a specifically defined population.

8.2 SAMPLING VISITORS TO A PLACE

Intercept samples (also called *location samples* or *site samples*) are samples of visitors to a place. Intercept samples might be taken of visitors to a shopping

mall, a store, a business district, a museum, a park, a stadium, a street fair, a polling place, and so on.

The most common examples of intercept research are shopping mall studies, also known as mall surveys or mall intercept surveys. These are used as an inexpensive form of market research, especially for studies that require the presentation of physical stimuli such as products, advertisements, or trademarks. In this type of research, the intercept sample is not meant to profile visitors to the specific place: Rather, it is meant as a convenience sample of a broader population, and nonprobability methods are the norm. If multiple sites are used, the selection of locations is based on judgment; for example, in a study concerning the likelihood of confusion between two trademarks, data might be gathered in four shopping malls in different regions of the country, with the selection of specific malls being based on the availability of interviewing services. Within sites, interviewers are usually left to their own devices, subject to any quota requirements.

There may be a desire for intercept samples with higher quality, especially when the research is meant to profile visitors to the particular location. A common example is political exit polling, where the goal is to characterize voting in selected precincts for purposes of analyzing the election and predicting the result. Other examples that we have seen include (1) profiling the trade area from which a store, shopping mall, or business district draws patrons; (2) profiling the characteristics and/or interests of visitors to a street festival, amusement park, athletic event, museum, play, symphony concert, zoo, and so on; (3) profiling how visitors to a place move through it and use the facilities; and (4) in public health surveys, profiling people who visit places with certain health risks, such as gay men who frequent establishments where sex is permitted. In these situations, since the intercept sample is intended to be specifically representative of visitors to the particular place, better sample quality may be desired.

It is possible to use standard sampling procedures to select careful intercept samples, both in the selection of data collection sites and the selection of visitors within sites, as follows.

8.2.1 Selecting Places for Intercept Research

In many applications of intercept research, a single place is of interest. For example, if the purpose of the research is to map the trade area of a particular retail store or to learn the characteristics of visitors to a local arts festival, then the research location is set by definition and there is no need to draw a sample

of places (although one may wish to sample places within the place, as described below).

In other applications, the places where data will be collected are intended to represent a broader population of sites. For example, in exit polling, the voting precincts in which data are gathered may be intended to represent the entire state. Here, the sites can be viewed as clusters and either (a) chosen with equal probabilities if one plans to apply a fixed sampling rate within sites, such as every 20th voter, or (b) chosen with probabilities proportionate to size if one plans to gather a fixed number of observations at each site. For example, in an exit polling context where all votes in a state are of interest, the votes can be viewed as being clustered by precinct, and precincts (clusters) can be selected for purposes of polling. Such samples also can be stratified if desired; for example, in election research, "swing" precincts that historically have shown more variation in party preference and/or turnout levels might be assigned to a separate stratum and sampled at a higher rate than "safe" or stable precincts.

A problem that may arise is inability to obtain access to selected places; for example, in drawing a careful sample of shopping malls, one will encounter the problem that malls are private business establishments and many will not permit interviewing on their premises. If this problem becomes too large to ignore, one's general options are to weight the data for place characteristics or substitute places in the same general location with similar characteristics.

8.2.2 Sampling Visitors Within Places

Within sites, the simplest way to obtain a probability sample is to sample people systematically as they enter or leave the site. For example, if 100 interviews are desired at a particular site, and 3,600 people are expected to visit during the interviewing period, one might select every 36th visitor after a random start.[1] Sampling issues that arise in this context include the following:

• *How long should the intercept period last (and when should it be)?* In some cases, there may be a natural period in which data should be collected; for example, in exit polling, data should be collected on election day while the polls are open. In other cases, the intercept period may not be obvious; for example, in studying visitors to a shopping center, there is no natural start and end to the

---

1. This is easiest to do if the task of sampling is separated from data collection—in other words, if the person who counts and selects visitors is not also responsible for soliciting participation or gathering data.

research period. Should data be collected in a single morning? A single day? Three days? A week? More? Obviously, this decision will affect the sampling interval, as the total number of visitors will increase with the length of the intercept period. A rule of thumb is that the intercept period should capture any hourly or daily variations that one might expect in the nature of visitors: for example, one might expect differences in employment status, age, and sex between voters who visit the polls during normal office hours and those who come early or late, and Blair (1983) shows differences between weekday, weeknight, and weekend visitors to shopping malls. It also may be desirable to capture broader time differences, such as differences in museum visitors during the school year versus vacation periods or the tendency of lower income workers to shop around payday.

• *What if you don't have an estimate of population size?* To establish a sampling interval or sampling rate, you have to know how many visitors to expect. This information may or may not be available. In sampling voters, there should be historical data on voter turnout at the precinct level. In sampling visitors to a zoo, or amusement park, or museum, there should be historical ticket and/or turnstile counts. In sampling visitors to an athletic event or symphony performance, there should be data on advance sales and historical walk-up sales. In sampling customers at a store, there should be historical transaction counts. However, in sampling visitors to a shopping mall, business district, or open-air street festival, there may not be any close estimate of population size. In such situations, a preliminary traffic count is needed. If such a count is not possible—for example, if one cannot count the visitors to a 3-day street festival until it is time to do the research—then one must rely on rough estimates or use another sampling method as described later in this section.

• *How should the sample size be adjusted for nonresponse?* Some people will refuse to participate in the research, and the sampling plan must be adjusted accordingly. For example, if you desire a sample size of 100, and you expect a 50% response rate, you need to intercept 200 people to get the desired 100, and the sampling interval should be set accordingly. Response rates for intercept studies can vary widely depending on the nature of the place, the quality of the interviewers, the time of day or year, and the nature of the task, and the best way to estimate response for any given study will be through a pilot test.

• *What if your estimates of population size or response rate are wrong?* Since any estimate of population size will be based on historical data, it is likely to

contain some level of error, and you are likely to encounter more or fewer visitors than expected. Likewise, your estimate of response rate may be off. As a result, you may reach the desired sample size before the end of the intercept period or reach the end of the period without enough observations.

If you reach the desired sample size before the end of the intercept period and the discrepancy is not large, carry on with the sampling plan to avoid bias against later time periods. If the discrepancy is large and you cannot afford the larger number of observations, hopefully such a large discrepancy will be apparent early in the data collection process and the sampling plan can be adjusted; otherwise, if you stop data collection early, you risk bias. Similarly, if you reach the end of the intercept period without enough observations and the discrepancy is not large, accept the discrepancy rather than oversampling later time periods, and if the discrepancy is too large to accept, hopefully this will be apparent early in the process and the sampling plan can be adjusted.[2]

- *What if the number of entrances and exits is large?* In trying to draw a probability sample of visitors to a place, our preference is to intercept them as they enter or leave. There may be various advantages in doing so—for example, people may be less rushed and more likely to participate when they first arrive, and people who are leaving can report completely about their activities at the site—but our primary reason from a sampling perspective is that everyone is likely to enter and leave once, so we can account for the entire population of visits with equal probabilities at the points of entrance or exit.

Sampling visitors as they enter or leave is easy if the place has a single entrance or exit, and not too difficult if the place has a small number of entrances or exits, but becomes increasingly difficult as the site becomes more complex. Shopping malls, for example, may have dozens of entrances (including those directly into stores). For these situations, where it may not be practical to sample continuously at every entrance, Sudman (1980) proposed a scheme in which visitors are grouped into clusters based on entrance and time period (e.g., Entrance 1 from 10:00–11:00 a.m. on Saturday), and clusters are sampled with probabilities proportional to size. A major drawback to this scheme is that it requires separate estimates of population size (number of visitors) for each

---

2. In theory, some protection from large errors may be allowed by sampling over several days or even weeks, which allows time to adjust the sampling plan. However, this doesn't always work. For example, if visitor counts are influenced by the weather, it may be difficult to predict counts from one day to the next.

entrance by each time period. Such estimates are unlikely to be available, and generating them may be costly; as a result, this procedure receives little use. We discuss other methods for sampling at complex sites, including sites that do not have formal entrance and exit points, later in this section.

• *How should children and groups be treated?* Visitors to most places will include a mixture of adults and children, as well as individuals and groups. Their treatment depends on the nature of the research and the corresponding definition of the population. In some cases such as exit polling, the population of interest consists of adult individuals. Here, children are ineligible and should be ignored, both for purposes of counting and selection. Likewise, groups are irrelevant, and people who enter in groups should be counted and selected as individuals. In other situations, children and group composition may be of interest: For example, in profiling visitors to the zoo, there is a good chance that we will be interested in the number of children who attended, their ages, and other information. In these situations, we might define a population of adults, count only individual adults, and essentially treat the number and nature of children as a characteristic of the adult; alternately, we might define a population of groups, count and select groups, and use a single adult respondent to provide information about the group.

• *Are visits the desired population unit?* In sampling visitors to a place, the implicit sampling unit is the visit, not the person or the money they might spend (Blair, 1983). Consider, for example, Person A who visits a store 50 times a year and Person B who visits once a year. Holding aside systematic patterns in the timing of these visits, if we intercept visitors to the store on any given day, the chance of catching A is 50 times larger than the chance of catching B. This is entirely appropriate if we wish to describe the population of *visits:* From this perspective, each of A's visits is a separate population unit that deserves its own chance of selection. However, if we wish to describe the population of *people* who visit the place, then A and B should have equal chance of selection; from this perspective, A's multiple visits can be viewed as duplicate elements in the sampling frame, which produce selection bias in favor of the more frequent visitor. This bias can be corrected by measuring visit frequency—for example, by asking respondents how many times they have come to this place within the past year—and weighting for the inverse of frequency. Similarly, if one is interested in a population of dollars spent on a per annum basis, it will be appropriate to measure visit frequency and level of spending, weight down for visit frequency to get to the level of people, and then weight back up for each person's expenditure (see Blair, 1983).

• *Should data be weighted for length of visit?* Nowell and Stanley (1991) suggest that visitors' probabilities of being intercepted at a place are related to the length of time spent on site, so intercept data should be weighted for (the inverse of) visit length. In general, though, we do not recommend "length weighting." If intercepts are conducted at entrance or exit points, most people will pass these points once as they enter and/or once as they leave, regardless of how long they stay, and the probability of being intercepted will not relate to visit length: Therefore, length weighting will not reduce bias. If intercepts are conducted elsewhere, such as the central court of a shopping mall, then the probabilities of people passing an intercept point may indeed correlate with visit length: However, (a) since the visit is not yet complete, people may not be able to give good estimates of its length, and (b) for length weighting to reduce bias, visit length must also correlate with the variables being adjusted, which may be problematic. Background variables such as age, income, and home ZIP code typically do not correlate well enough with visit length to justify weighting. Variables that clearly relate to time on site, such as the probability of visiting any particular store within a mall or the total number of stores visited, correlate better with visit length, but we have found that the reduction in bias resulting from weighting is not consistent or large enough to justify the associated increase in sampling error.

If there is no reliable estimate of the number of people who will visit a place, a procedure that might be used is to sample systematically by time. For example, if 100 interviews are desired, and the interviewing period is scheduled for 10 hours or 600 minutes, one might select a visitor every 6 minutes after a random start. The advantage of this approach is that it does not require any estimate of the population size (number of visitors). The disadvantage is that it implicitly samples visitors from low-traffic time periods at a higher rate than visitors from high-traffic periods, which might result in selection bias. This bias can be managed by counting the number of visitors while the research is being done and weighting data from different time periods accordingly; for example, if it turns out that weekend visitors were sampled at half the rate of weekday visitors, then weekend observations should be weighted up by a factor of 2.

At complex sites, time sampling can be combined with the entrance-by-time-period clustering suggested by Sudman (1980). Clusters are defined in a similar manner (e.g., Entrance 1 from 10:00–11:00 a.m. on Saturday), selected with equal probabilities, and a fixed time-sampling scheme is applied within

each cluster. The advantage of this procedure is that it allows one to draw a probability sample even at a complex site without estimates of population size. The disadvantage is that it implicitly applies a higher sampling rate to visitors from low-traffic time periods and low-traffic entrances compared with high-traffic times and high-traffic entrances. Again, this bias can be managed by counting visitors and weighting.

In some cases, it may not be possible to intercept visitors at points of entrance and exit. This might occur at places that do not have formal points of entry and exit, such as a business district or a street fair or a park, or at places where intercepts are allowed only in constrained areas, such as the central court of a shopping mall. In these situations, one can at least try to establish intercept points that cover different geographic areas and different time periods. For example, in a shopping mall with an east-west orientation, one might establish western and eastern intercept points and collect data at each point during weekday, weeknight, and weekend periods, to capture differences among visitors that are likely to occur across locations and time (Blair, 1983). Traffic counts can be taken at each intercept point for each time period and the data weighted accordingly. This is essentially a quota sampling procedure with quotas defined on location and time, whether or not an effort is made to sample individual visitors at each intercept point through some random procedure.

8.3 SAMPLING RARE POPULATIONS

Rare groups are often of interest to researchers. By rare groups, we mean groups that account for no more than 20% of the general population and usually much less. Examples include men who have sex with men, who are of interest for (HIV) health risk studies; low-income households, who are of interest for welfare policy research and other purposes; and purchasers of specific goods or services.

In sampling a rare population, the first thing to determine is whether a good list is available. If so, sampling and locating respondents is straightforward. However, this is the exception rather than the rule. In most cases, screening of the general population is necessary, and the costs of screening can equal or far exceed the costs of interviewing. For example, if the target group comprises 2% of the general population, 50 screeners are needed to locate each group member unless sampling efficiency is improved in some way.

Exhibit 8.1 Snapshot of Methods for Locating Rare Populations

Method	Conditions in Which It Might Be Useful
Telephone cluster sampling	There are a substantial number of telephone exchanges with no members of the target population
Disproportionate stratified sampling	Prevalence of the target population varies substantially across geographic areas such as telephone exchanges
Network sampling (multiplicity sampling, snowball sampling, respondent-driven sampling)	Members of the target population can be accurately identified and reached through others in a well-defined network such as members of their immediate family
Dual-frame sampling	A special frame is available that has high prevalence of the target population
Location sampling	Members of the target population tend to congregate at identifiable locations
Use of online panels	You are willing to treat the online panel members as representative of a broader population such as all online users

As a result, survey researchers have long been interested in cost-effective methods for sampling rare groups within the general population. Methods that have been used in this regard include telephone cluster sampling, disproportionate stratified sampling, network sampling, dual-frame sampling, location sampling, and, most recently, online data collection (cf. Kalton, 2001). All of these methods are useful under certain conditions, as indicated in Exhibit 8.1 and discussed below.

8.3.1 Telephone Cluster Sampling

Telephone cluster sampling (TCS) for rare groups, a variation of Mitofsky-Waksberg sampling (see Chapter 2), was described by Blair and Czaja (1982) based on a suggestion from Sudman, and further described in Sudman (1985). It works as follows. A random number is dialed within a bank of telephone numbers: This number can be selected via list-assisted random-digit dialing or

any other probability sampling procedure.[3] If the number is found to be a working household number, the household (or person) is screened for membership in the target group. If the household is not a member of the target group or if the number is not a working household number, then no further sampling is done within the bank. However, if a group member is found, further sampling is done within the bank until a prespecified number of group members is identified.[4] This procedure has the effect of rapidly dropping telephone banks with no target group members and "fishing where the fish are."

The potential usefulness of TCS for rare groups depends two factors: (a) the extent to which members of the target population are clustered within telephone exchanges and (b) how rare the group is. The harder it is to find the group in the general population, and the greater the extent to which group members are clustered, the greater the benefits of searching near one group member for another.

If TCS is used, one faces the question of appropriate cluster size. Sudman (1985) suggests that the optimal number of group members to be taken within each telephone bank is typically in the range of 8 to 10. However, it is difficult to reach cluster sizes of 8 to 10 rare group members in banks of 100 telephone numbers. For example, Blair and Czaja (1982) used TCS to sample African Americans, who comprised about 9% of working household numbers at the time of their study, and found that 18% of the retained telephone banks could not produce a cluster size of 10. If some banks fall short of the desired cluster size, the total sample will likewise be smaller than desired. This is easily addressed through additional sampling or by increasing the initial sample size. More important, TCS requires equal cluster sizes for the sample to be EPSEM and hence unbiased. If some clusters fall short, weights must be used to compensate, and sampling variance increases as a result. In the case of Blair and Czaja (1982), the need to weight respondents from "short" banks produced an increase in sampling variance that threatened to completely offset the increase in screening efficiency (Waksberg, 1983). Sudman (1985) suggested dealing with this problem

---

3. The original purpose for Mitofsky-Waksberg (M-W) sampling was to find working household telephone numbers and eliminate banks of nonworking numbers, as discussed in Chapter 2. For this purpose, list-assisted RDD (which also eliminates nonworking banks) and M-W sampling are competing alternatives. However, when the goal is to find members of a rare group, the two procedures are complementary. List assistance almost completely eliminates banks of nonworking numbers, and telephone cluster sampling can further eliminate working banks in which the target group does not occur.

4. The cluster size is defined as identified, not cooperating, eligible households or individuals. If the cluster size is *k*, then calling in the bank stops after *k* group members are identified by screening, whether or not they consent to the main interview.

by increasing the size of the telephone banks; for example, defining banks of several hundred rather than 100 telephone numbers. Alternately—or in addition—one can cut the cluster size, and our experience is that cluster sizes of 2 to 3 are more appropriate than the 8 to 10 suggested by Sudman (1985).

While TCS for rare groups is theoretically appealing, it has not worked well in practice and is rarely used. Furthermore, as people switch from landlines to cellphones, geographic clustering within telephone exchanges declines. However, Blair and Blair (2006) suggest some conditions under which TCS may be useful in dual-frame designs that also use online panels to locate rare groups.

8.3.2 Disproportionate Stratified Sampling

Disproportionate stratified sampling for rare groups may be effective if the prevalence of the target group is higher in some areas than in others. The high-prevalence areas can be assigned to a separate stratum and sampled at a disproportionately high rate to raise the overall efficiency of the design. For example, in sampling men who have sex with men, which is of interest in public health studies, certain areas such as the Castro district in San Francisco have higher than average prevalence of the target group, and the relevant telephone exchanges can be given disproportionate assignments (cf. Blair, 1999). Likewise, groups such as affluent households, households living in poverty, or buyers of specialized products may be disproportionately prevalent in identifiable areas.

The usefulness of disproportionate stratified sampling for rare groups depends on the extent to which the prevalence of the target group varies across geographic areas. One also would like to have prevalence data for the target group (or some reasonable proxy) by telephone exchanges (or some reasonable proxy) to define strata a priori and make efficient strata allocations; however, if such data are unavailable or of poor quality, it is possible to use a two-phase adaptive sampling in which prevalence information acquired in Phase 1 is used to determine near-optimal strata definitions and allocations in Phase 2, where most of the data collection will be done (cf. Blair, 1999). For example, the number of calls needed to find a member of the target group within an area (or the number of calls between group members) provides information about the group's prevalence. The adaptive approach may be generally useful in addressing the common problem in sampling rare special populations—that there are often insufficient or poor-quality secondary data to use in developing the sample design.

Note that the conditions under which disproportionate stratified sampling (DSS) is effective for locating rare groups are parallel to the conditions under which telephone cluster sampling (TCS) is effective: TCS is effective when the target group is geographically clustered, which is tantamount to saying that the group's prevalence varies across geographic areas. This might be taken to imply that TCS and disproportionate stratified sampling are competing procedures and that one or the other should be used where they are applicable. However, this is not necessarily so, because the two procedures work through different mechanisms. In fact, the methods may be complementary.

TCS works by eliminating banks of telephone numbers in which the target group does not appear. There is a common misconception that it also works by undersampling banks of numbers with relatively few group members, but this is not so. The procedure is EPSEM, and it does not oversample banks with high levels of group prevalence or undersample banks with low prevalence. Once empty banks are eliminated, the effectiveness of TCS is unaffected by the target group's distribution across the populated banks. In contrast, disproportionate stratified sampling *does* oversample and undersample populated banks with higher and lower prevalence, and the procedure's effectiveness does not require that any banks be completely empty. In effect, stratified sampling works by over-sampling telephone banks where the target group's prevalence is high, while TCS works by eliminating banks where the group does not occur.

The two procedures are redundant if the available stratifying information allows one to identify telephone banks in which the target group does not occur: In this situation, one will simply allocate zero sample to the empty stratum, and there will be no further gains from TCS. The procedures also are redundant if there is little variation in prevalence among the nonzero banks: In this situation, once TCS eliminates the empty banks, there will be little to gain from dispropor-tionate stratified sampling.

Usually, if there are sufficient variations in prevalence to justify the use of stratification, we can separate higher versus lower prevalence areas (either a priori or through adaptive procedures), but we cannot state with certainty that any given telephone bank is empty of the target group. For example, we know that the Castro district has a higher prevalence than other areas for men who have sex with men, and we know that Beverly Hills has a higher prevalence for extremely affluent households, but we cannot a priori identify banks of 100 telephone numbers in which those groups are guaranteed not to occur. In such situations, it may be use-ful to combine stratification with clustering and oversample the high-prevalence areas while using TCS to eliminate empty banks in the low-prevalence areas.

8.3.3 Network Sampling

Another way to enhance the efficiency of screening for rare groups is *network sampling,* also called *multiplicity sampling.* Much of the early work on network sampling was done by Sirken and his colleagues (Sirken, 1970, 1972; Sirken & Levy, 1974; Levy, 1977). A brief summary on this topic (and other methods for sampling rare groups) is found in Sudman, Sirken, and Cowan (1988) and Blair (1990).

Network sampling for rare groups works as follows. Members of a random sample drawn from the general population are screened for the defining characteristic(s) of the target group and also are asked whether the members of some prespecified social network such as their brothers and sisters have the characteristic(s). If any member of the network falls into the target group, the respondent is asked for contact information, and the researcher attempts to interview those network members. The effect of measuring these social networks, as well as the respondents themselves, is to identify more members of the rare group in the initial screening interviews.

For network sampling to be useful, the following conditions must be met.

• First, the informant must be able to report accurately whether each network member is or is not in the target population. This requires that the defining characteristics of the group are visible (or known) to other members of the network. Reporting errors (producing both false-negative and false-positive target population member identification) are often a serious source of bias in network samples and a drain on the design's efficiency.

• Second, if network members are to be interviewed, network sampling requires that the initial respondents be willing to provide referrals to other network members and, perhaps more important, that they be able to provide adequate contact information such as telephone numbers or e-mail addresses.

• Third, it must be possible to obtain an accurate estimate of network size for weighting purposes. The probability of any given member of the target group being identified is proportional to the number of people who might identify him: For example, if the network is defined as siblings, a rare group member with one sibling has two ways to be identified (he might be drawn in the initial sample or his sibling might be drawn and identify him), while a group member with three siblings has four ways to be identified and an only child has one. In effect, each member of the network functions as a duplicate listing for every other member,

which creates a selection bias in favor of people with larger networks. To correct this bias, it is necessary to measure network size and weight by its inverse.

Note that the key issue in this regard is the not number of people that a first-stage respondent might identify (the initial respondent's network size) but rather the number of people that might identify each target group member. Therefore, the initial respondent usually cannot provide useful information about network size unless the network is reflexive; for example, if the network is defined as siblings, and the initial respondent says that he has three siblings, each network member also has three siblings.

Empirically, it has been shown that these conditions are best fulfilled when the network consists of close relatives who are likely to know the most about the member of the rare population and who can report accurately regarding the network size. While it is tempting to define networks more broadly so as to cast a wider net, one pays a price in rapidly increasing reporting error when more distant relatives or other types of networks are used.

This type of network sampling is particularly useful when the purpose of the research is to estimate the prevalence of a rare group rather than contact its members. If all one needs is prevalence data, then network sampling has the potential to expand the effective sample size without imposing the difficulty of obtaining referrals and finding networked respondents. This, in fact, is the purpose for which the method was developed and used in its early days (cf. Sirken, 1970).

Network sampling for rare populations also can be used in a snowballing fashion with multiple networking stages: This type of sampling is called *snowball sampling, chain referral sampling,* or *respondent-driven sampling.* Here, the initial respondent is screened for membership in the rare population of interest: If and only if he is a member of that population, he is asked to identify additional group members, who in turn are asked to identify other group members, and so on until no new members are identified. The logic of this procedure is that if members of the target population know each other, and these social networks are exhaustively pursued when encountered, then, for any given network, it should be possible to give each person in the network a 100% probability of being selected (given that the network is selected) as long as the person knows at least one other member of the network who can identify him. The probability that the network is selected will be a function of its size, measured through the number of identified members, and data should be weighted by the inverse of network size to correct the resulting selection bias in favor of larger networks.

If snowball referrals are incomplete, and the probability that any given network member will be identified through referral falls below 100%, there will be a selection bias in favor of members who know more people in the network and hence are more likely to be identified. Referrals may be incomplete for various reasons: for example, respondents don't take the time to list everyone they know, or they overlook someone, or the researcher stops early.

To be effective, snowballing requires not that members of the rare population are known to people in the general population, as in the first type of network sampling that we described, but that they are known to each other. This condition is most likely to be met for small, closed populations, such as social or political elites or specialized professional groups. Snowballing also has been used to identify groups of gay men (Heckathorn, 1997), although, in such an application, one might question whether the procedure is limited by respondents' willingness to identify other members of their social networks, especially if they know that those members are not "out" and prefer not to be identified.

A drawback of snowballing is that the size of the research task cannot be well specified in advance, because it depends on the number and size of networks encountered, and once the project begins, you must exhaustively map each identified network to maintain control of the sampling probabilities.

8.3.4 Dual-Frame Sampling

Dual-frame sampling for rare populations is a special case of stratified sampling that is useful when there is an efficient but incomplete frame of the target group, such as a membership list. It combines a sample from the efficient but incomplete frame with a sample from the general population to augment coverage. Consider the following example that was introduced in Chapter 2:

CASE STUDY 8.1

A researcher wishes to conduct a telephone survey of 200 people who visited an open-air "art festival" last weekend. These people can be found by screening the general population, but only 5% of the local population is estimated to have attended the festival. Of those who did attend, an estimated 20% registered for a prize drawing, and the registrations are available to the researcher.

(Continued)

(Continued)

Assume, for purposes of example, that festival attendees who registered for the drawing can be contacted and interviewed at a cost of $12 per completed interview (including the costs of noncontacts and nonresponse). Assume that screening for attendees in an RDD sample of the general population will cost $5 per screening interview and $10 per main interview (again including the costs of noncontacts and nonresponse, which in this case are primarily borne at the screening stage).

If the sample is drawn entirely from the list of registrants, 200 interviews can be obtained for $2,400 (200 @ $12 per interview). However, there is substantial room for coverage bias because the frame only covers 20% of the population.

If a sample of 200 attendees is screened from the general population, the coverage problem is eliminated, but the cost is much higher. If 5% of the population attended the festival, it will take 20 screeners to get an attendee. Therefore, the cost per completed main interview will be $110 (20 screeners @ $5 each to get the respondent, plus $10 for the main interview), and the cost of 200 main interviews will be $22,000 (200 @ $110 per).

A dual-frame design will allow full coverage of the population of attendees at a lower cost than general screening. Under this procedure, the attendees are stratified into those who registered and those who did not. Nonregistrants, who account for 80% of attendees, will be screened from the general population at a screening rate of 4% (80% of the 5% of the population who attended). It will take 25 screeners to get a nonregistrant, and the cost per completed main interview will be $135 (25 screeners @ $5 each, plus $10 for the main interview). Registrants, who account for 20% of attendees, will be contacted directly from the registration list, and the cost per interview will be $12, as before. Using the formula shown in Chapter 5 for optimal stratum allocations when costs differ by strata, it may be seen that the optimal procedure in this example is to allocate approximately 45% of the sample to attendees who registered for the drawing and 55% to attendees who did not register. So, for a sample of 200, the researcher should take 90 registrants from this list and screen 110 nonregistrants from the general population, for a total interviewing cost of $15,930 (90 @ $12 per, or $1,080, plus 110 @ $135 per, or $14,850).

Note that in the process of screening the general population for 110 people who attended the festival but did not register for the drawing, we will encounter some people who attended the festival and *did* register (specifically, we should hit 110 * 20/80, or 27 such people). Here, we assume that these people are treated as ineligibles when encountered in the general population frame. It also would be possible to interview these people, and reduce the sample from the registration list accordingly, but in this example, there would be relatively little savings from doing so (27 * $12 = $324, or about 2% of the current expected cost of $15,930), and there is some administrative appeal in having each stratum drawn from one and only one frame.

In general, since dual-frame sampling is a form of stratified sampling, some way of distinguishing the strata will be necessary. This is not a problem if the efficient frame and general frame can be physically cross-checked to identify members

of the general frame that are also in the efficient frame (so they can be treated as ineligible in the general sample or reassigned to the efficient sample). If the identifying information in the two frames is not consistent or accurate enough to allow physical cross-checking, then we will need for respondents from the general population to be able to self-report whether they are in the efficient frame. If they cannot do so reliably—for example, if general population respondents in our example can tell us whether they went to the festival but cannot reliably tell whether they registered for the drawing—then errors in these reports will create errors in stratum assignments and a corresponding potential for errors in estimates.

8.3.5 Location Sampling

Rare populations that tend to congregate at particular places can be sampled at those places. For example, recreational groups such as deep-sea fishermen might be contacted at harbors from which they embark. Likewise, researchers interested in HIV risk behaviors have gone to gay bars to obtain samples of gay men. Location sampling is particularly useful in contexts where face-to-face contact and "street credibility" are important in getting members of the rare population to identify themselves, as when gay men are asked about HIV risk behaviors.

In many applications, location sampling by itself may not give satisfactory population coverage: For example, samples obtained at gay bars will omit gays who do not patronize these places, and nonpatrons may differ systematically from patrons. If coverage is a concern, location sampling can be combined with a general probability sample that screens for target group membership and availability at the interview location. This is a form of dual-frame sampling.

Location samples also can be combined with snowballing (as described earlier in this chapter) to reach outside the location frame. However, this is not likely to be a complete solution to coverage problems, because the population members who are not accessible via location sampling may not socialize with the population members who are: For example, gay men who do not patronize gay bars may not socialize with those who do. Also, note that in a location-driven snowball sample, probabilities of selection for members of any given network will depend not on total network size, as in regular snowball sampling, but rather on the number of network members who patronize the selected places.

Ideally, in location sampling for rare groups, some form of probability sampling will be applied at the location, rather than haphazard intercepts. Our earlier discussion of sampling visitors to a place is relevant in this regard.

8.3.6 Online Data Collection for Rare Groups

Earlier in this chapter, we noted problems with online samples of the general population: Response rates are low, and there is substantial exposure to coverage bias because (a) online access varies widely across income, education, and ethnic groups, and (b) there is no complete frame of the online population. However, given the cost challenges of studying rare groups, if data can be gathered at lower cost on the web, then one might consider dual-frame designs in which the web is used for target group members who are accessible online, and telephone (or some other method with broad coverage) is used for those who are not online (Blair & Blair, 2006). The logic of using online data collection for rare groups is that the web should have lower screening costs than other methods because no incremental labor or postage is needed to contact potential respondents.

A key issue in this regard is the potential for coverage bias stemming from the fact that available sampling frames cover only a fraction of the online population. A theoretically clean approach to this issue is to stratify the target population on the basis of "in frame versus not in frame" rather than "online versus not online." However, this approach is of little practical use, because if the "in frame" stratum contains only a small percentage of the total population (such as an online panel with 2 million members out of a U.S. population of more than 300 million), then the optimal allocation to this stratum will be too small to justify the incremental costs of web-based data collection. There also is the practical difficulty of getting telephone respondents to give an accurate indication of whether they appear in any given online frame, so they can be sorted out. Overall, the most practical approach is to stratify on the basis of online versus offline, but this requires a leap of faith that the online frame represents the broader online population.

In this regard, an important feature of dual-frame web-phone designs is that they will allow a comparison of web and phone results. If results from web and phone do not differ, one generally might assume that the effects of coverage bias in the web data are negligible. If the results do differ substantially by mode, that difference might be caused by coverage bias or might reflect legitimate differences between the two strata (e.g., if the survey has a subject matter, such as frequency of online purchases, that relates to presence in the web frame). Either way, if mode differences are observed, some form of weighting adjustments may be used to correct for potential bias.

8.4 SAMPLING ORGANIZATIONAL POPULATIONS

The most significant issue that distinguishes organizational samples from those of individuals or households is the enormous variability in the size of organizations. For example, a researcher trying to get an estimate of the potential demand for a new type of industrial equipment will quickly realize that demand is greatly affected by a few very large firms. As a result, it usually is optimal to stratify organizations by size and oversample the larger organizations, as discussed in Chapter 5. A further simplification is to sample *all* the very large organizations that account for most of the sales and/or variance in a category (see Hansen, Hurwitz, & Madow, 1953).

A second and related issue is how to measure the size of organizations for efficient stratification. Common standards include the value of annual revenues, the value of company assets, and the number of employees. In general, these standards are correlated, but there are differences: For example, a manufacturing company is likely to have a high value of assets relative to the number of employees, while a service company is likely to have a high number of employees relative to assets. The choice of a measure usually will depend on two factors. First, what measures are available? For example, if the sampling frame lists companies along with their annual revenues, it will be easy and inexpensive to stratify using revenues as a measure of size. Second, what measure is likely to have the best correlation with the phenomenon of interest? For example, if the research concerns employment issues, then stratification based on number of employees may be more efficient than stratification based on revenues.

A third sampling issue is determining the appropriate unit to study within organizations (i.e., the appropriate population unit). For businesses, is it a plant, a regional office, a division, or the entire firm? For educational institutions, is it a classroom, an academic department, a school, or a school district? The choice depends primarily on the topic of the study and whether decision authority and/or policies vary within organizational subunits. For example, in a study of educational policies, the defining issue will be the level at which relevant policies are set.

A fourth issue is determining who within the organization will be an appropriate informant and whether multiple informants are needed to provide accurate results. In highly formal organizations, such as many school systems, it may

be easy to identify decision authority through job titles and choose respondents in this manner. In many other organizations, things are not so obvious, and it may be necessary to contact selected organizations to identify the right informant(s). This is usually done by telephone, starting with the switchboard or possibly a relevant job title, and being transferred from phone to phone until the right person is reached. Even if a list is available, it is necessary to confirm that you have the right person, because organizational lists can go out of date rapidly, especially for business managers.

8.5 SAMPLING GROUPS SUCH AS INFLUENCE GROUPS OR ELITES

Earlier in this chapter, we discussed how *snowball sampling* might be used to find members of a rare population that are linked in social networks. The same procedure can be used to identify other types of social groups.

Consider, for example, sociometric studies where the topic of interest is communication or influence patterns in small groups. Here, the population unit of interest is the group, but there is usually no list of groups. Rather, a sample is drawn of individuals, and these individuals are asked to identify other people who stand in some relationship to them, who may in turn be asked to identify other members of the group, and so on. Similarly, consider research where the topic of interest is the nature and activities of social or political elites. Here, initial respondents may be selected by their formal roles (e.g., city councilperson), but informal members of the elite are found through snowballing.

If the initial sample is random (as opposed to a list of role titles), then the probability of identifying any given group through snowball sampling depends on size of the group (because each member provides a chance of identifying the group), so there is a selection bias in favor of larger groups. This bias can be corrected by measuring group size and weighting for its inverse. Also, there may be a selection bias at the individual level in that the person who is known to more people has a higher probability of being mentioned than does the isolate. This bias is least important if the snowball procedure is continued until no new names are mentioned, because an exhaustive listing of the group should ensure that every member is identified as long as there is at least one other group member who can identify him or her. Also, if the group is exhaustively listed, its size can be measured by simply counting the listed members.

8.6 PANEL SAMPLING[5]

A panel is a group of individuals, households, or organizations that provide information for more than one period in time. The simplest panels involve respondents who report before and after an event, such as studies that track voters through an election. More complex panels involve respondents who report weekly on a continuous basis on household expenditures or other behavior. Panels are used for the following purposes:

• *To measure change.* A fundamental advantage of panels is that the sampling variances of measures of change are much smaller for panels than for a series of independent samples, and it is possible to measure changes in individual as opposed to group behavior over time. For example, "scanner panels" monitor participants' grocery purchases over time, and the results can be used to measure general brand switching and changes in behavior resulting from promotions.

• *To measure behaviors that accrue over time.* By measuring behaviors as they occur, panels may be able to provide more accurate data than retrospective measurements, even if change per se is not of interest. For example, panels are used to measure television viewing or radio listening, not to document program switching but simply to gather more accurate data about this ephemeral phenomenon.

• *To provide balanced sampling frames.* As noted in Chapter 2, some companies maintain online panels from which samples can be drawn for one-time surveys. The use of these panels is motivated primarily by convenience, cost, and speed: For example, it is possible to draw a geographically and demographically balanced sample from an online panel and obtain a reasonable number of responses within days.

A researcher who wishes to establish and maintain a panel, especially a long-term panel, faces three principal sampling issues: (1) possible nonresponse bias associated with the initial request to participate, (2) possible nonresponse bias associated with differential panel mortality over time, and (3) possible bias associated with panel aging.

---

5. This section is based on Sudman (1976).

8.6.1 Initial Nonresponse in Panels

Regarding initial nonresponse, no research project achieves full cooperation from selected respondents. However, the problem is larger for panels because of the greater burden placed on respondents. For example, which would you be more likely to accept: a request to participate in a one-time survey about your purchasing habits or a request to maintain a purchasing diary for the next year? Initial response rates for panel studies are usually at least 1/3 lower than for surveys, with the difference depending on the nature of the panel request.

Initial response rates for panel studies also vary across different types of respondents, creating potential bias. For example, Jordan (2004) describes problems that the A. C. Nielsen company has had in recruiting Hispanics to its U.S. consumer panels, which have, among other things, led to criticism that Nielsen's well-known television ratings are culturally biased because they undercount Hispanic audiences. Likewise, Sudman and Wansink (2002) show that people who cooperate with consumer panels are less likely to be in one- or two-person households, more likely to be in households with young children, more likely to have a nonemployed woman, and more likely to report being price conscious: In other words, households that participate in consumer panels are more likely to have stay-at-home mothers with some discretionary time during the day and an interest in the topic of home economy. More generally, willingness to participate in a panel is positively related to free time and interest in the task.

The level of cooperation achieved with a panel request is, of course, not independent of the recruiting methods used and the tasks required of panel members. Interestingly, though, continuing cooperation tends to be similar regardless of the initial response rate. When greater efforts are made to get initial cooperation from respondents, there seems to follow a higher dropout rate on a continuing basis.

8.6.2 Differential Mortality Over Time

As just noted, panels suffer from mortality (dropouts) over time. For periodic, repeated interviews, an additional 5% or 10% loss should be expected from the remaining sample on each subsequent interview. Mortality may result from changes in a respondent's life, such as getting married or having a baby, or simply from fatigue or loss of interest in the task.

Keeping mortality to a minimum is not an automatic process but one that requires considerable effort and experience in the techniques of maintaining

panel cooperation. Even though most panels find it important to compensate participants with money or prizes, a continuing program of communication with panelists is equally essential to establish and maintain the high level of morale that reduces panel turnover.

Despite one's best efforts, some level of mortality will occur in a panel. If this mortality is evenly spread across different types of panel participants, it will not threaten the representativeness of the panel. Unfortunately, it usually is not evenly spread across participants. Just as people who are busier or less interested are less likely to accept the initial request to participate, they are more likely to drop out over time. As a result, differential mortality over time tends to exacerbate any nonresponse biases found in the initial panel sample.

8.6.3 Panel Aging

The third problem encountered in panel sampling is loss in representativeness due to panel aging. Imagine, for example, a consumer panel formed 50 years ago. Even if the sample for this panel was randomly drawn, and the initial response rate was 100%, and nobody dropped out, the panel would not be representative of today's general population. We would have panel mortality in the truest sense of the world, as many of the initial panel members would have died over the years, and the survivors would all be old. There would be no panel members in their 20s, 30s, 40s, or even 50s. Their shopping baskets would be low on children's cereals, baby diapers, chips, and soft drinks and perhaps heavy on high fiber cereals and denture adhesive.

8.6.4 Implications for Panel Sampling

The basic point to remember in maintaining a long-term continuing sample is that panels are dynamic and should not be treated as a sample drawn for a one-time survey. The population is changing continuously because of new household formations (or new business startups), dissolution of old households (or businesses), and household (or business) moves. A panel, if it is to remain representative of this changing population, must reflect these changes.

In this regard, panelists should be followed when they move. This may be conceived as a dynamic system that is frozen for just an instant to allow a sample to be drawn from it and then released, and the subsequent motion of the

sample represents the motion of the population. If one made the mistake of sticking to a fixed sample frame (e.g., by dropping households that move and replacing them with households that move into the same dwellings), it would be difficult to locate and include dwelling units that were built after the sample frame was designed, and it would be difficult to allow for shifts in population from place to place. Also, dropping households that move will lead to mortality bias because certain types of families (e.g., young, small households) are more likely to move than others.[6]

Accounting for moves is not enough. Some method must be designed for continually rejuvenating a long-term panel by bringing into it the proper number of new households and dropping dissolved households. Dissolved households are easy to handle. The only necessity is to drop them when they are observed. An example is when one member of a couple dies and the other moves in with relatives or to an assisted living facility: If this person has been a panel member, she is dropped at this time, and there is no need to replace her in the panel with a new household. Regarding new household formations, panel members may be periodically asked whether there has been any change in the number of adults or children living in the home and if anyone has moved away to set up a new household. Family members who move away to set up new households are recruited with probabilities inversely proportional to the number of persons who will constitute the new household (this is done so that all new households have the same probability of being added, regardless of the size of the new household). Thus, in the case of marriages, half the split-offs are recruited. Empirical evidence has indicated that this recruiting method brings new young households into a panel at the proper rate.

Even after allowing such evolution in a panel, the panel is likely to lose representativeness over time as a function of aging and differential mortality. To maintain representativeness, replacement panel members must be recruited disproportionately from geographic and demographic categories that become underrepresented. So, for example, if 18- to 24-year-old unmarried men have disproportionately high dropout rates, replacements should be disproportionately recruited from this group. If Hispanics are becoming a larger portion of the general population and are underrepresented in the panel, they should be disproportionately recruited to bring the panel in line with the broader population.

---

6. For some studies, a panel may be limited to residents of a specified geographic area, such as a city, county, or state. In this case, if a household moves from the area, it becomes ineligible and is dropped from the panel.

Given that continuing response rates in a panel are likely to be well under 50% after allowing for initial nonresponse and subsequent mortality, and given that disproportionate, stratum-driven recruitment will be needed to maintain panel representativeness over time, one might ask whether it is worthwhile to use probability sampling to recruit panel members. Why spend top dollar on probability methods if they are undermined by low response: Instead, why not use looser methods such as quotas to control panel composition and reduce cost? The answer is that probability samples are preferable because, even if they are not perfect, they are the best samples available. They are theoretically measurable and they minimize potential biases associated with convenience sampling or volunteerism.

Having said that, we should acknowledge that quota-based panels are the norm in the field of market research. For example, all of the large online panels except for the GfK KnowledgePanel are recruited through nonrandom methods and controlled through demographic balancing. The argument in favor of these panels is that their repetitive use makes it possible to benchmark their performance over time, so that one can come to know and accommodate any biases. This is a reasonable argument.

8.6.5 Other Issues in Panel Sampling

In the case of panels that are used primarily as sampling frames, a sampling question that arises is how to manage disproportionate response across geographic and demographic groups (and consequent nonresponse bias) in individual studies. Two methods may be used: *disproportionate outgo* and *back-end weighting*. In disproportionate outgo, groups with a history of lower response rates are oversampled as needed to balance the resulting data. In back-end weighting, a proportionate sample is drawn, and any resulting discrepancies in sample composition are controlled through weighting. Disproportionate outgo is somewhat more complicated at the sampling stage but allows one to use the data without weighting, which can be an advantage in a panel context where one might wish to compare results from parallel studies over time and such comparisons are most easily done with natural (unweighted) data.

Also, in situations where one wants to track changes in a population over time but does not wish to face the issues associated with a fixed panel, an alternative procedure that might be used is a rotating panel, or rotating surveys with partially overlapping samples. For example, to study the process by which people buy home appliances, a well-known company has used a rotating design in

which respondents are surveyed once, then a second time 6 months later, then a third and last time 6 months after that, and in any given month, 1/3 of the total sample is participating for the first time, 1/3 for the second time, and 1/3 for the third time. This design produces response rates similar to surveys while still allowing the company to study how purchase intentions flow into purchases of home appliances over a 1-year period. A discussion of rotating designs may be found in Kish (1965, 1987).

8.7 SAMPLING IN INTERNATIONAL CONTEXTS

In general, this book has discussed survey sampling in a U.S. context. Here we consider some points of difference across countries. Of course, the basics of sampling do not change—that is, inaccuracy can stem from sampling error or sample bias, sampling error is controlled through sample size, stratification and/or clustering may improve efficiency, sample bias is controlled through sampling process, and so on.

The principal differences seen across nations from a sampling perspective relate to frame availability, sample design issues related to frames and data collection modes, and response rates.

Sampling frames that are available for the general population can be quite different from one country to another. For example, European countries typically have publicly available registers of residents that can be used for sampling. This can improve sampling coverage for the general population, although it may not be of value in sampling subgroups within the population. On the other hand, information for sample design that can be found in censuses in some countries may be completely unavailable or unreliable in developing countries with infrequent or poorly executed censuses.

Within the United States, high female participation in the workforce and relatively high labor costs for interviewers have made it very expensive to do in-home interviewing, while high telephone ownership means that telephone surveys can be done with little coverage bias (although nonresponse is a significant problem). This has led to extensive use of telephone surveys and thus extensive use of sampling techniques that address the challenges of telephone surveys, such as random-digit dialing. In developing nations, telephone ownership may be lower or skewed toward cellphones, making telephone surveys less attractive, while the economics of in-home interviewing may be much more

favorable. This leads to greater use of in-home surveys and thus greater use of techniques such as cluster sampling with geographically defined areas.

Similarly, the coverage problems faced by online surveys are more severe in developing nations. Exhibit 8.2 shows that while Internet access has grown in developing as well as developed nations, it lags far behind in the developing world.

Also, shopping mall surveys are common in the United States for purposes of market research. Mall-based shopping is less common in other, more densely populated countries, where people are more likely to shop on commercial streets. It is still possible to sample and interview people in public locations, but the working environment for interviewers is different, and the ways in which respondents are selected and screened may change accordingly.

A final sampling difference across nations relates to response rates. Response rates have been dropping in the United States and can present major concerns for potential nonresponse bias. In many developing nations, this is less of an issue. There can be major challenges in accessing secure properties, but if interviewers are able to reach households, they are less likely to encounter nonresponse.

Exhibit 8.2   Coverage Problems for Online Surveys Are More Severe in Developing Nations
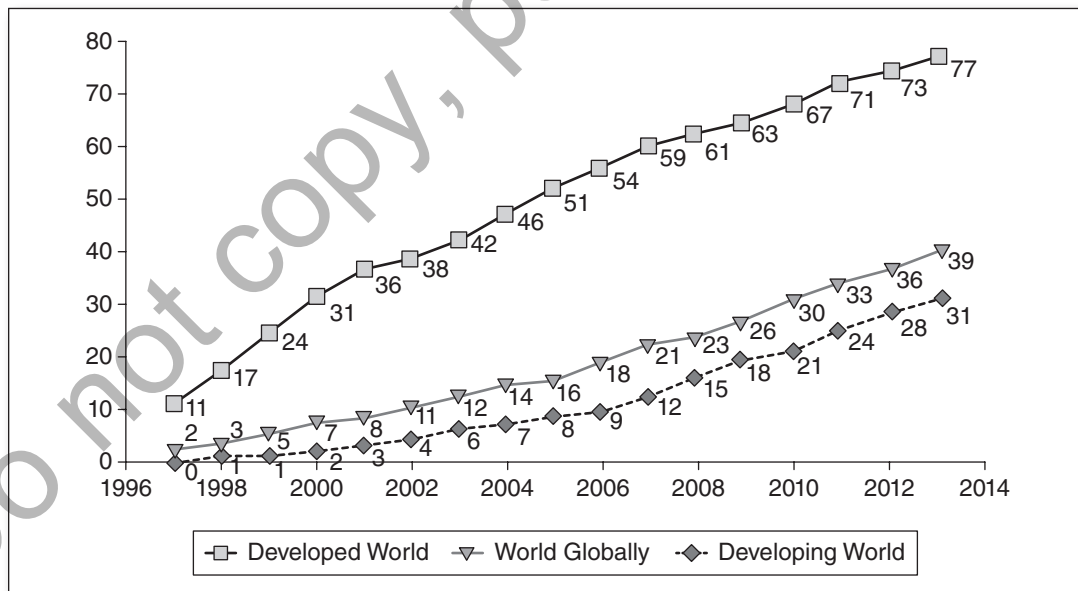
Source: Courtesy of Jeff Ogden (W163) with CC-BY license on Wikimedia Commons.

8.8 "BIG DATA" AND SURVEY SAMPLING

In recent years, increased computing capabilities have led to increased ability to store, analyze, and cross-reference data and a corresponding use of *big data* or *data science*. While there is no standard definition of *big data,* in general it is data that reside in a database or combination of databases that may contain millions of records and require specialized software to manage and analyze. Some of these databases have existed for years, such as government databases of federal program participants or, in the private sector, medical records databases. More recently, researchers have assembled large databases from sources such as Google search data, social media websites, and websites with product reviews. Many of these newer databases are generated automatically by user activities.

In some phases of standard surveys, big data may supplement the usual survey sampling and operations. A big data set may

- provide additional information about a target population to aid survey design or analyses,
- help compensate for item nonresponse,
- provide information about unit nonrespondents for analysis and weighting,
- be merged with a survey data set to enrich analysis possibilities.

This type of use for big data is simply an extension of the procedures that surveys commonly use. A more radical use of big data sets is as a fundamental, sometimes defining, resource for data collection. We will discuss each type of use in turn.

8.8.1 Big Data as a Survey Complement

We will illustrate the possible use of big data as a survey complement in the context of government surveys. Many large and complex surveys are conducted by the U.S. federal government and its contractors. As large as many of these samples are, and as lengthy as the interviews can be, the surveys could make use of additional information, especially at the individual case level, if it was easily and cheaply available. These data could be used in the ways just listed to improve the quality of the surveys.

In some cases, data that a survey is intended to collect may already reside in government databases. For example, it is typical for government surveys to ask questions about individual or household income. Such questions are sensitive to many respondents and consistently have item nonresponse that exceeds most other variables. Also, the answers to such questions can have significant response error. However, much of the desired information is already available in Internal Revenue Service (IRS) databases. The IRS data are not perfect but are likely to be at least as good as answers to survey questions. If it were possible to access and merge individual financial information from IRS records—it won't be, for legal reasons, but if it were—then questions could be eliminated from the survey, thus shortening the questionnaire or making space for other items while simultaneously reducing potential bias from item nonresponse or response errors.

There are substantial barriers to using government records in this way, including technical difficulties in merging administrative records with survey data, and regulatory or ethical obstacles that must be addressed before such records can be used at the individual level. There are also issues on the survey side in collecting identifiers that are specific enough to merge such data sets. The general idea, though, is that administrative records might contain a variety of information that could be appended to survey responses to enrich the data or simplify data collection, in much the same way that an online panel can automatically add previously collected demographic information to data obtained from any panel member.

Administrative records also might help in sample design. For example, if an addressed-based sample is selected from the U.S. Postal Service Master Address File (MAF), IRS data corresponding to the same people at the same addresses might be used to stratify the sample on financial variables.

Similarly, to the extent this sample has nonrespondents (as any survey will), IRS or other administrative records might provide some information about the nonrespondents that is useful in assessing the types of people who responded and hence in estimating nonresponse bias and constructing weights to adjust for nonresponse.

8.8.2 Big Data as a Survey Replacement

The second category of big data applications give big data a central role in data collection and possibly replace surveys (or their sampling stage)

altogether. For example, consider a company that wishes to measure consumer attitudes toward its products. A traditional approach is to draw a sample of consumers and conduct a survey in which such questions are asked. A big data approach is to scrape websites with product reviews, blogs, and other websites with user-generated content to capture and code comments about the company's products. Some sampling of websites might be done to narrow the data collection task, and the collected data might be sampled to reduce the coding task, or the full data might be used. The obvious appeal, at least on a first look, is the possibility of using the entire data set and foregoing sampling altogether.

Under some conditions, this approach could be very cost-effective. There may be, however, some problems that have to be resolved. For example, if the researcher is interested in a sample of individual persons, then people who make more comments will have proportionally higher inclusion probabilities. This is the common frame problem of duplication. We can try to clean the data of duplicates, but this becomes cumbersome as the data set becomes enormous, and duplicates will be difficult to identify unless the same person has made the exact same comments under the exact same name.

More important, although the comments may contain a key variable of interest, there is likely to be information of interest that is not included. For example, the company might wish to measure attitudes among users and non-users of its products, heavy users versus light users, and younger people versus older people, and the dataset may not contain the relevant information.

Most important, the people who post comments are essentially a sample of volunteers. Who are these people? For example, are fans of the products over-represented? Are people with extreme attitudes (positive or negative) overrepresented, in the same way that online political comments may overrepresent partisans relative to the "silent majority"?

These issues reveal a basic difference between the "data first" approach sometimes associated with big data and a traditional survey design approach as reflected throughout this book. The survey design approach begins with the survey purpose, followed by defining the relevant population and making decisions about the method of data collection and sampling. This approach does *not* begin with the features of data that happen to be available. It is tempting to capture comments about the company's products made by people on the web, because those data are essentially free, but that advantage comes at the cost of changing the population definition from "consumers" to "people who make comments about these products on the web."

Essentially, big data can be mistaken for the entire population when it is, in fact, a sample that has gone unrecognized. The question is whether the process that has given rise to the data makes it the *right population* or at least a proxy without obvious biases.

Once we think of big data in this way, it becomes clear that just as big data may be useful as a survey complement, so surveys may be useful as a complement to big data. In the above example, online comments might be taken as the primary data source, and a survey might be conducted to assess the nature of any biases in those data relative to the consumer population of interest. The survey might measure attitudes of interest, whether respondents had posted comments online, and the number and location of such posts. Such information would allow us to weight the online data for "nonresponse" (i.e., nonposting) and duplications; more generally, it would allow us to improve model-based estimates drawn from the online data. The analogy is using data from a sample of nonrespondents to evaluate and improve survey estimates. Like any such enterprise, the size and/or frequency of the survey would be less than if the survey was used as a primary data source.

Another issue that sometimes arises with big data is the fact that every analysis of data patterns finds statistical significance because the number of observations is so large. For example, we might observe that people who post comments about a company's products on Tuesday are more likely to be negative than people who post on Saturday—or people who post between 9:00 and 10:00 a.m. are more likely to be negative than people who post between 10:00 and 11:00 a.m.—or people who post between 9:00 and 9:05 a.m. are more likely to be negative than people who post between 9:05 and 9:10 a.m. Such results may reveal a meaningful phenomenon if you can explain them. The concern, though, is that an abundance of significant findings with extremely large samples can lead the researcher to see patterns where none really exist.

8.9 INCORPORATING SMARTPHONES, SOCIAL MEDIA, AND TECHNOLOGICAL CHANGES

Survey sampling and data collection, like all aspects of society, are influenced by computing and communications technologies. We are currently witnessing an unprecedented wave of new devices (smartphones, tablet computers), software (mobile apps), services (web-based social media), and technical capabilities

(fast processors, cloud computing). These developments have rapidly transformed how people connect with friends, peers, and strangers (e.g., Facebook, Google +, LinkedIn), communicate generally (blogs, Twitter), share information (YouTube), and interact (Second Life). They also are affecting survey practice in ways that range from minor tweaks to radical changes.

As with big data, these resources can be used in the context of standard survey methods—for example, to supplement a standard data collection mode to improve coverage or reduce costs. It is also feasible that a technology or service may replace one or more standard survey methods by providing new frames, new data capture methods, or access to special populations that traditionally have been exceedingly difficult to identify and survey, as well as creating special populations of interest in themselves (e.g., Facebook members).

8.9.1 Smartphones and Surveys

Cellphones in general and smartphones in particular have affected survey research because of their prevalence and their rapidly expanding features. For example, surveys no longer have to depend on contacting sample members in their residences but may potentially access them anywhere. This has a variety of implications, including a shift in optimal calling times to maximize response rates (Brick et al., 2007) and a need to verify that the respondent is not engaged in some activity such as driving a car that would make it inappropriate to proceed with a survey at that time.

A significant implication of cellphone use is that the telephone sampling frame now contains a mixture of landline numbers corresponding to households and cellphone numbers corresponding to individuals. If the desired population unit is individuals, it is necessary to deal with clustering in the landline frame and sample within households. If the desired population unit is households, it is necessary to determine whether the cellphone holder is the appropriate informant for their household, deal with potential duplication in self-reported informants, and deal with the possibility that the cellphone holder is an ineligible minor.

Increasingly, cellphones are smartphones that permit survey participation requests to be sent by text message or e-mail. As part of this sort of use, it may be possible to screen a sample to identify target subpopulation members. The device also gives respondents the option to download an app so that they can answer the survey questions on the phone's screen. Follow-up contacts or refusal

conversion can be administered in the same way as the initial contact. In addition, smartphones have the capacity to add visual features to surveys or collect location information using GPS. These capabilities can clearly benefit some survey objectives. On the other hand, the researcher can lose some control over the quality of the realized sample (e.g., ensuring that minors are not inadvertently included in a survey).

To learn more about the implications of smartphones for surveys, Peytchev and Hill (2011) provided smartphones to a small probability sample of adults after an initial in-person interview and subsequently conducted a number of methodological experiments over the course of a 19-week longitudinal study. Their results are limited in a number of ways, including the sample size ($n = 92$), but raise important questions about how factors such as the screen display of questions, visuals, and response options, as well as using the keyboard to enter answers, can all potentially affect response behaviors. From a sampling perspective, the use of these devices may differ across population members, producing unintended changes in the sample composition.

At this time, smartphones are owned by only about 35% of adults in the United States (Smith, 2011), although coverage is growing rapidly. This means that a general population survey must use another mode(s) along with the smartphone to achieve reasonable coverage of the general population.

8.9.2 Social Media and Surveys

Social media currently relate to survey sampling in four ways. First, they may be used as a vehicle for establishing contact with respondents who are missing other contact information or who decline to participate. For example, longitudinal surveys almost always have some amount of attrition due to losing contact with panel members, and Facebook and other social media can be useful in reestablishing contact with panel dropouts.

Second, social media may be used as a vehicle for recruiting respondents. For example, the personal data that are part of social media may be very helpful in lowering the cost of recruiting a target population with particular characteristics such as educational background, type of employment, or leisure activities. Social media also might be used for respondent-based sampling, where one respondent leads to another. Of course, using social media to recruit respondents has the effect of using social media as the sampling frame for a broader population, and one must consider the possible coverage bias that may result.

Third, the personal data found on social media may be useful for stratifying the sample or simply enriching the data set.

Fourth, in some cases, the population of interest may be defined by social media usage. Tessitore and Macri (2011) considered methods to use Facebook for sampling when the target population is partly defined by Facebook membership. They describe severe difficulties in applying standard probability sampling and concluded that quota sampling or some type of convenience sample was necessary.

As one might expect, the effectiveness of using social media for locating or recruiting respondents varies substantially across population subgroups and across nations. For example, within the United States, social networking site use drops significantly across age groups, as shown in Exhibit 8.3.

Two potential ethical issues should be recognized in sampling from social media databases. One is the possible inclusion of minors. When minors are

Exhibit 8.3 Percent of Internet Users in Each Age Group Who Use Social Networking Sites
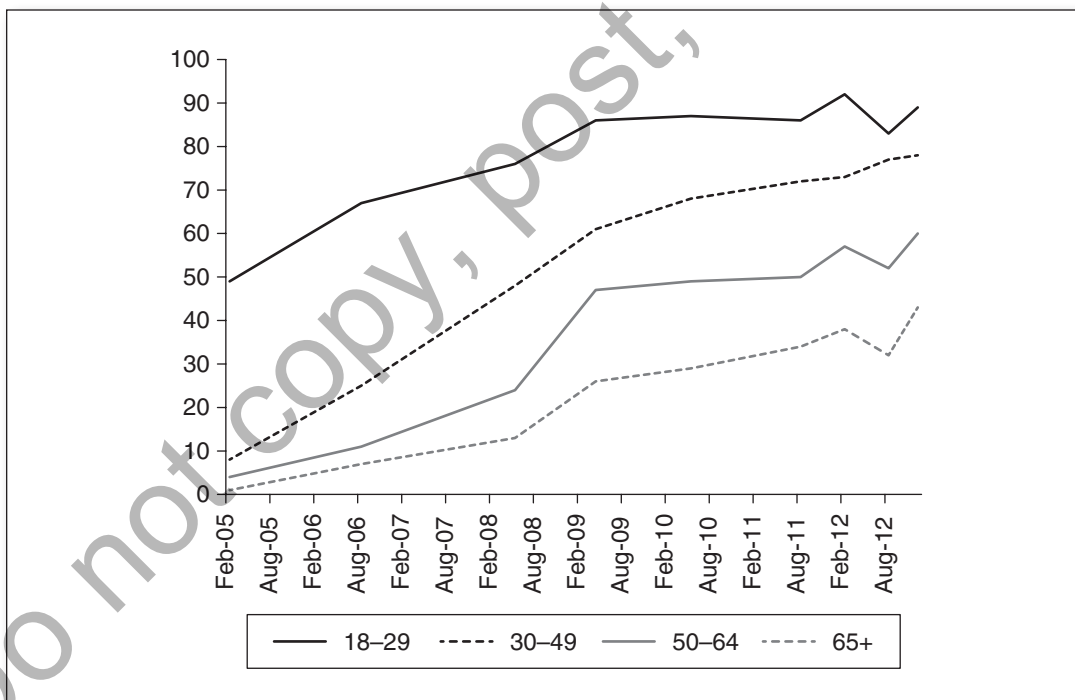
Source: Data from Pew Research Center "Social Networking Use." Pew Research Center, Washington, D.C. (May 1, 2013) http://www.pewresearch.org/data-trend/media-and-technology/social-networking-use/, accessed on 9/23/14.

surveyed, special procedures and safeguards are required by university institutional review boards (IRBs), including parental approval for the minor to participate in the research. In most surveys, people under the age of consent are excluded from the survey population, but if samples are compiled by opt-in or other voluntary procedures or recruited from social media sites, it is possible for minors to be inadvertently included. The fact that the researcher is unaware of minors being selected does not obviate the researcher's responsibility not to survey persons under the age of consent without parental approval.

Also, an unintended consequence of sampling from social media databases is the possible inclusion of information about respondents that is not relevant to the research. Once personally identifiable information is collected, the researcher is responsible for safeguarding it. That obligation can become difficult or impossible if the sample and data are stored on a server that the researcher does not control. When using free or inexpensive commercial resources, such as cloud storage and processing, the researcher can, without having specifically considered the shift, move her research data, sample, and respondent identifiers into an unsuitable research environment.

8.9.3 A General Framework for Incorporating New Technologies

Technological developments beyond smartphones and social media are likely to have implications for survey sampling. Here we present a general framework for responding to such developments.

In the opening chapter of *Envisioning the Survey Interview of the Future,* Conrad and Schober (2008) provide an excellent summary of many issues, potential benefits, and cautions generally applicable to using new technologies for survey interviewing. Conrad and Schober recognize that the same error sources—from coverage, selection, nonresponse, and measurement—will be of concern, but how each source is affected will require research and experience with the new methods. Some of the potential decisions, benefits, and risks they list include the following:

- How prevalent is the technology?
- What are the "costs" of *not* adopting a new technology (e.g., on respondent perceptions and response)?
- What can one assume about how well people can use a new technical tool (e.g., answering questions on a smartphone screen)?

- Will technologies have positive or negative effects on population groups with varying cultures and languages?
- What are the pros and cons of using the full capabilities of new technologies—for example, to link (potentially in real time) the survey interview to other information that is available about the respondent? This includes ethical issues such as the need for informed consent to employ some types of available information or technical capabilities.

Conrad and Schober's (2008) thorough presentation (as well as the remainder of their book) is informative, important, and engaging. Their framework leads to the following suggestions.

Start by reviewing available literature on how the technology at issue has performed for uses similar to yours. In general, if the literature on the technology is sparse, weak, or nonexistent, the risk of extreme or unanticipated effects is greater.

In considering the likely positive and negative effects on various sources of error, keep in mind that those effects are likely to differ depending on whether the proposed technology will be used to supplement more traditional methods versus relying solely on the new tool. Also ask whether there are design changes that may alleviate negative effects. If, for example, you know that a new method will reduce cost but also yield a lower response rate, perhaps you can "move" the negative effect. It may be feasible to shift the balance of sample size between modes or compensate with additional follow-ups—such as reworking smartphone nonresponses with more callbacks or using a different mode—while retaining most of the cost advantages.

What if lower response and poorer coverage both disproportionately affect a particular demographic subgroup? If that group is important overall or for separate analysis, lower cost may be a false savings.

The bottom line is to consider the main threats to your survey's reliability and validity, as well as how the technology will affect them. If there is insufficient information to make this assessment, then some type of pilot study, even a very small trial, may be wise. Often, a small amount of real data is far more informative than a large amount of speculation or conjecture.

8.10 CHAPTER SUMMARY

This chapter discussed issues in sampling special populations. We considered (1) online populations, (2) visitors to a place, (3) rare populations, (4) organizational

populations, (5) groups such as social influence groups and elites, and (6) panel sampling. We also considered sampling aspects of "big data" and smartphones, social media, and other technological changes.

Regarding online populations, we noted three principal sampling problems: potential coverage bias stemming from the fact that many people are not online, potential coverage bias stemming from the incomplete coverage of available frames, and potential nonresponse bias stemming from low response rates. Because of these issues, online data collection is best suited to situations in which the population of interest has a high level of online access, a reasonably complete list of the population is available, and the topic of the research is of interest to the population. Online data collection also may be used in dual-frame sampling of rare populations, as discussed later in the chapter.

Regarding visitors to a place, we noted that nonprobability sampling is common in this context, but careful samples can be drawn. Cluster sampling may be used to obtain a probability sample of places, and probability samples of visitors within places may be sampled via systematic sampling of people or time (with clustering by entrances and time periods if appropriate). Issues that arise in this type of sampling include when to do the intercepts, what to do if you don't have an estimate of population size, how to adjust the sample size for nonresponse, what to do if your estimates of population size or response rate are incorrect, what to do if the number of entrances and exits is large, how to treat children and groups, what to do if visits are not the desired population unit, and whether to weight the data for time on site.

Regarding rare populations, we described various methods that might be used to improve screening efficiency and reduce the cost of studying such groups. These methods are (1) telephone cluster sampling, which may be useful when the group has heavy geographic clustering and is very rare; (2) disproportionate stratified sampling, which is useful when the group's prevalence varies substantially from one place to another; (3) network sampling, which is useful when the defining characteristics of the rare population would be known to others; (4) dual-frame sampling, which is useful when you have an efficient but incomplete list of the rare population; (5) location sampling, which is useful when the rare population tends to congregate in certain places; and (6) online sampling, which may be useful for very rare populations or if one has access to a productive sampling frame.

Regarding organizational populations, we noted four principal sampling issues. First, organizations usually vary enormously in size, and therefore it is

appropriate to stratify samples on the basis of size. Second, one must choose an appropriate measure of size. Third, decisions must be made regarding the organizational unit that is appropriate for research purposes. Fourth, one must determine who speaks for the organization—that is, who is the appropriate informant and whether multiple informants are needed.

Regarding groups such as social influence groups and elites, we noted the applicability of snowball sampling in identifying and mapping these groups.

Regarding panels, we noted three sampling issues: potential nonresponse bias associated with the initial request to participate, potential nonresponse bias associated with differential mortality over time, and potential loss in representativeness associated with panel aging. In general, panel response is biased in favor of people with more time and/or more interest, and these biases tend to become greater over time. To maintain representativeness over time, a well-designed panel will track household moves, household dissolutions, and new household formations: In addition, it usually will be necessary to stratify the panel by geographic and demographic groups and disproportionately recruit groups as needed to maintain panel balance.

Regarding sampling and "big data," we noted that cross-referencing databases may hold great promise for reducing interview length and cost, enriching data, stratifying samples, and assessing possible nonresponse bias. However, there is danger in assuming that a database will provide accurate results just because it is large. A "data first" mentality runs the risk of changing the implicit population that is studied.

Finally, regarding smartphones, social media, and other technological developments, we noted the specific benefits and problems associated with smartphones and social media, and we presented a general framework for evaluating whether and how to incorporate new technologies into your research. Key points are to (1) check the literature and learn from others, (2) consider whether you will be using the new technology to complement proven methods or using it to replace methods, (3) think about how you might mitigate negative features of the technology, (4) think about how the positives and negatives of the technology will affect not just your broad population but also subgroups of special interest, and (5) if you don't have enough information to know how the technology will work, consider a pilot study.

Exercise 8.1

Imagine that you are on a team of students that wishes to conduct a survey of student opinions regarding various issues on your campus. The team's plan is to conduct intercept surveys with at least 50 students. How would you obtain this sample? What location(s) would you use? When would you do the research? How would you select respondents?

Exercise 8.2

A public health researcher wishes to conduct a U.S. national telephone survey of households that are (a) headed by a woman living without an adult partner, (b) with at least one child present younger than 14 years, and (c) a household income under $25,000 per annum. How would you design the sample for this survey? Would you expect any of the methods described in Section 8.3 to be useful?

Exercise 8.3

A health scientist at a university wishes to conduct a panel study of dietary practices, exercise practices, and weight changes among students. The plan is to conduct an initial survey with entering freshman, with online follow-up questionnaires administered monthly to all participants for the following 3 years (including summers). What sampling plan would you propose for this study? How would you draw the initial sample? How would you maintain the panel? Would you propose any changes to the intended data collection procedures?