PART I

Introduction

1

Defining hard-to-survey populations

ROGER TOURANGEAU

1.1 Introduction

This book is about populations that are hard to survey in different ways. It focuses on populations of people rather than establishments or institutions. In an era of falling response rates for surveys (Brick & Williams, 2013; Curtin, Presser, & Singer, 2005; de Leeuw & de Heer, 2002), it may seem that *all* household populations are hard to survey, but some populations present special challenges of various sorts that make them harder to survey than the general population. Some of these hard-to-survey populations are rare; others are hidden; some are difficult to find or contact; still others are unlikely to cooperate with survey requests. This chapter tries to distinguish the major challenges that make populations hard to survey and reviews attempts to quantify how hard to survey different populations are.

One way to classify the various sources of difficulty is by what survey operation they affect. In this chapter, we distinguish populations that are *hard to sample*, those whose members who are *hard to identify*, those that are *hard to find or contact*, those whose members are *hard to persuade* to take part, and those whose members are willing to take part but nonetheless *hard to interview*. These distinctions reflect the main steps in many surveys. First, a sample is selected. Often, the next operation is identifying members of the target population, for example, through screening interviews. Then, the sample members must be found and contacted. Once contact is made, sample members have to be persuaded to do the survey. And, finally, the willing respondents have to have whatever abilities are needed to provide the requested data or special steps have to be taken to accommodate them. As we shall see, with any given population, problems can arise with each of these operations, making the population hard to survey. And, as will become clear, some hard-to-survey populations present combinations of several kinds of trouble.

1.2 Hard-to-sample populations

In the ideal case, there is a complete and up-to-date list of the target population and the sample can be drawn from this list. Unfortunately, this ideal is rarely realized in practice; for most populations of interest in surveys, there is no list frame and sampling begins with some general purpose sampling frame, such as an area, address, or random digit dial

(RDD) frame. Problems arise when the target population represents a small fraction of the frame population. Kalton (2009; see also Chapter 19 of this volume) distinguishes major subgroups or domains (constituting more than 10 percent of the total population), from minor subgroups (1 to 10 percent) and from mini-subgroups (less than 1 percent of the total population). To pick out the members of the target population from the other members of the general population, surveys often begin by administering a short battery of screening questions. In the absence of a special frame or frames, then, one reason that a population can be hard to sample is that its members are rare, representing a small fraction of the larger frame population, often the general population. (Another source of difficulty, to which we return later, is that it may be hard to identify the members of the rare population in a short screener.)

Discussions of the issues involved in sampling rare populations (e.g., Chapter 19; Kalton & Anderson, 1986; Sudman, Sirken, & Cowan, 1988) often point to two other population characteristics, apart from overall prevalence within the general population, that affect the level of difficulty in finding members of the population in a screening survey. The first is the level of variation across areas or sampling strata in the prevalence of the rare subgroup. It is sometimes possible to increase sampling efficiency by oversampling strata where the prevalence of the rare subgroup is relatively high and undersampling areas where the prevalence is relatively low. It is easier to find members of the rare population when a substantial proportion of them is concentrated in a small number of areas or strata that can be identified prior to sampling. For example, a recent experiment in the National Household Education Survey attempted to boost the number of Hispanics in the sample by targeting census tracts in which at least 13 percent of the population was Hispanic (Brick, Montaquila, Han, & Williams, 2012).

The other variable affecting the difficulty of locating members of a rare population is the cost of a screening interview relative to the cost of the main interview. If screening interviews are relatively cheap (for example, only a few questions are needed to identify members of the target population), then having to carry out a lot of them will not affect the final data collection costs so much as when screening is relatively expensive. Consider a situation in which members of the rare population constitute 5 percent of the total population. If we ignore the effects of nonresponse to the main interview, this implies that twenty screeners will have to be done for each main interview. However, if the screening interviews cost only one twentieth of the main interview, then the total costs per case are only doubled by the screening costs (that is, twenty screeners plus one main interview cost twice as much as a main interview alone). But if the screening interviews are expensive – say, half the cost of the main interview – then the need to complete twenty screenings per main interview will drive up the total cost per case by a factor of 11. Screening costs can be high if medical tests or a long series of questions are needed to identify members of the target population or if it is difficult to get people to complete the screener. Some surveys use a two-phase screening process, where the first-phase screener casts a broad net and the second phase screener applies more stringent criteria. Clearly, sampling efficiency matters more when the screening process is expensive.

Kalton (2009) provides a measure (R) of the gains in sampling efficiency that can be achieved with a disproportionate allocation of the initial sample across strata that vary in the prevalence of the rare population:

$$R = \frac{\left[\Sigma W_h \sqrt{P(c-1) + P/P_h}\right]^2}{P(c-1)+1}, \tag{1.1}$$

in which W_h is the proportion of the rare group in stratum h, P is the overall prevalence of the rare group, P_h is its prevalence within stratum h, and c is the ratio of data collection costs for a member of the rare population to the costs for the nonmembers (that is, for cases who screen out).

One way to measure the difficulty of sampling members of the rare population is by the added cost per case due to the need to conduct screening interviews. With a proportionate allocation of the screening sample, the added cost (Δ_c) per case, expressed as a proportion of the total cost per case, depends on the prevalence (P) of the rare group and the cost ratio parameter (c) described earlier in Equation 1.1:

$$\Delta_c = 1 + c/P.$$

Under an optimal allocation across strata, the added cost would be $R \times \Delta_C$, where R is the efficiency gain factor defined in Equation 1.1. For example, if the efficiency factor was .8 and screening increased the data collection cost per case by a factor of 1.5, the net effect would be an increase of 20 percent (that is, .8 × 1.5 = 1.2). Δ_C and $R \times \Delta_C$ provide measures of the sampling difficulty associated with a rare population. In summary, then, a population is harder to sample as its overall prevalence becomes lower, as its prevalence varies less across the sampling strata, and as the screening costs increase relative to the cost of a main interview. In the best case, most of the rare target population falls within a few strata or a single high prevalence stratum and the screeners are relatively inexpensive.

A related situation involves selecting the sample from two frames – a general purpose frame with low prevalence but high coverage of the rare population, and a special frame with higher prevalence but less complete coverage of the rare population. The latter might be a list of known members of the rare population. The dual frame sample yields the highest gains compared to the general purpose frame alone when the special frame has a much higher prevalence than the general purpose frame and when it includes a large fraction of the rare population (e.g., Lohr & Rao, 2000).

Another type of population that presents particular difficulties for sample designers are mobile or "elusive" populations. These are populations, such as the homeless and similar groups (e.g., migrant workers), that are not easily linked to any one place. Here, the best sampling strategy often involves sampling places where the members of the elusive population are likely to be found rather than sampling the members of the population directly. Kalton (2009; see also Chapter 19 in this volume) describes this approach as "location sampling." Examples include sampling homeless shelters and soup kitchens as a strategy for capturing the homeless (e.g., Ardilly & Le Blanc, 2001) or sampling oases or waterholes to capture nomadic herdsman. Sampling is likely to continue for some period of time and, precisely because such populations are mobile, members may have multiple chances for selection. Moreover, the frame of locations is likely to be incomplete; thus, elusive populations may well be undercovered even when a location sample is selected. If the main goal of the survey is to estimate the size of the population, capture-recapture methods can be used. These methods, initially developed for estimating the size of nonhuman populations, are now used in estimating the coverage of censuses of human populations (see Mulry, Chapter 3 in this volume). Two samples are taken; in the best case, the samples are completely independent. (With the census, one of the samples is the post-enumeration survey sample; the other is a sample from the census enumerations.) The estimate of the size of the population reflects the proportion of cases found in both samples. A potential problem with this method is "correlation bias" – that is, the violation of the assumption that the capture and recapture probabilities are independent. When members of the rare population systematically vary in their elusiveness (or when they vary in their elusiveness within sampling strata), this variation will produce correlation bias. Imperfections in the sampling frame can also lead to correlation bias. For example, if the frame for a survey of the homeless omits certain sites, then the homeless linked only to those sites are likely be missed in both the initial and recapture survey.

Mobility presents challenges not only for sampling a population, but also for locating the members of the group. We have more to say about these problems in Section 1.4 below.

1.3 Hard-to-identify populations

A screening survey is predicated on the assumption that the respondents are both willing and able to answer the screening questions accurately. Screening data are often provided by household informants, who provide information about themselves and about the other members of the household. In some cases, a neighbor may be used as a last resort when screening is based on age, race, or some other visible characteristic. And, in network samples, screener respondents may be asked not only about their own households but also about the members of linked households (e.g., the households of their siblings; see Sudman *et al*., 1988, and Chapters 23 and 24 in this volume, for discussions). Regardless of the exact method of screening, the accuracy of the screening data will depend on the screening respondents knowing the relevant characteristics of each person they are asked about and their willingness to report that information. Unfortunately, these conditions may not always be met, creating a second type of hard-to-survey population.

1.3.1 Stigma, sensitivity, and motivated misreporting

Consider the difficulties in identifying the members of some cultural or religious minority, such as immigrants (see Massey, Chapter 13 in this volume), men who have sex with men, or Muslims (Keeter, Smith, Kennedy, Turakhia, Schulman, & Brick, 2008). Members of a highly stigmatized population, such as illicit drug users, may keep this characteristic secret even from other household members. And household informants may be reluctant to identify persons with the relevant characteristics to outsiders.

Even when the characteristic of interest is not sensitive (for example, when the population of interest is a specific age group), screening interviews often miss members (Horrigan, Moore, Pedlow, & Wolter, 1999; Judkins, DiGaetano, Chu, & Shapiro, 1995). Although almost all surveys are prone to *some* undercoverage (see, for example, Shapiro, Diffendal, & Cantor, 1993, on the coverage of the Current Population Survey, or CPS), the undercoverage in screening surveys seems to be worst for the very groups targeted by the survey. One of the best documented instances of such underreporting involves the National Longitudinal Survey of Youth, 1997 Cohort (NLSY-97). The eligible population for this survey was young people, aged 12–23. Horrigan and his colleagues (Horrigan *et al.*, 1999) compared the numbers of persons found in the NLSY-97 screening effort with the expected numbers based on CPS figures for the different age groups. The NSLY screening data show roughly the same numbers as expected for the age groups above 23 and slight undercoverage for those below 12 (roughly 90 percent coverage relative to the CPS). For the age range targeted in the screening effort (12–23 years old), however, the coverage dropped to about 70 percent. Similar problems have been found with several other national surveys (see Judkins *et al.*, 1999); in each case, undercoverage was considerably worse for the survey's target population than for other groups. To avoid the biases produced by this sort of underreporting, surveys sometimes retain some of the households that screen out for further data collection. Of course, this increases data collection costs.

Tourangeau, Kreuter, and Eckman (2012) argue that the underreporting of eligible household members in screeners is an example of *motivated misreporting*, in which respondents, interviewers, or both, shade the answers to minimize the work they have to do (see also Kreuter, McCulloch, Presser, & Tourangeau, 2011). When eligible households screen out, they do not have to complete the main interview, reducing the burden for both the potential respondent and on the interviewer. My co-authors and I carried out an experiment in which we varied how much the screening questions in a telephone survey disguised the target population (Tourangeau *et al.*, 2012). Some households got questions that asked directly about the eligible population ("Is anyone who lives there between the ages of 35 and 55?"); a second group of households got questions about younger and older age groups ("Is everyone who lives there younger than 35? Is everyone who lives there older than 55?"); a final group got a series of questions for each member of the household, including their sex, race, and age. The last method is known as the full roster approach. The full roster clearly beat both the direct questions and the complement questions for finding members of the target population. With the full roster version of the screening questions, 45 percent of the households screened in versus 32 percent with the direct questions and 35 percent with the complement questions. We knew from the frame data that some of the sample households included an eligible household member; the full roster led to the least underreporting within these households.

The downside was that the full roster also produced the lowest overall response rates (24 percent versus 32 percent for the direct question group and 29 percent for the complement question group); these response rates reflect nonresponse to both the screener and the main interview. Both interviewers and nonrespondents seem to contribute to the shortfall in eligible household members. There was a highly significant negative correlation (-.58) across interviewers between their screener response rates and their screener eligibility rates. The interviewers with the highest response rates to the screener also found the lowest proportions of eligible households.

So, there is clear evidence that members of even nonstigmatized groups can be hard to identify in screening interviews. It seems quite likely that the undercoverage of members of stigmatized groups will be even worse. At least one line of evidence provides support for this conjecture. Tourangeau, Kearney, Shapiro, and Ernst (1996) carried out an experiment that varied the procedures used to roster the members of sample households. We found that an anonymous rostering procedure led to better coverage of young Black males, a group often underrepresented in surveys and censuses. This study was done mainly in poor neighborhoods, where coverage is often low. The respondents in our screening sample may have deliberately omitted some household members (especially Black male members) because they were worried about losing welfare benefits or incurring some other penalty if they included them. Such concerns may lead to concealment on the part of respondents; my colleagues and I argued that the anonymous rostering procedure helped allay such concerns and reduced omissions from the rosters. These results suggest that omissions may occur more often the more respondents that are worried about the potential costs of reporting a member of the target population.

1.3.2 Metrics for the hard to identify

There are several ways to quantify the level of difficulty in identifying members of a given population. My discussion of the prior work in this area has already mentioned some of these potential metrics.

The most commonly used measure of the difficulty of identifying members of a specific population is its *coverage rate*. The coverage rate is the estimate of the size of the population from the survey to the estimated size based on some benchmark survey or the census:

$$CR = \frac{\hat{N}_i}{N_{Bi}}, \tag{1.2}$$

in which $\hat{N}_i$ is the estimated size of population group i from the survey (typically, the sum of the weights for the respondents in that group after any nonresponse adjustments) and N_{Bi} is the benchmark for that group (such as the estimate of the subgroup's size from the American Community Survey).

The coverage rate reflects the joint effects of all sources of error (including frame problems, screener nonresponse, and so on), not just misreports in the screening interviews;

in addition, it captures the net impact of all these forms of error. That is, overreports and underreports can cancel out so that a coverage rate near 1.0 may mask a high level of offsetting errors (see Mulry, Chapter 3 in this volume, who describes additional measures used to assess coverage in a census). The screening classifications can sometimes be compared to more accurate measures of the relevant characteristics. This allows the proportion of those who should have screened in but were incorrectly classified as ineligible to be computed (this is the *false negative rate*); similarly, it also allows the proportion of those who screened in but should have been classified as ineligible (the *false positive rate*) to be computed. False negatives are generally more problematic than false positives, since the latter can be removed once they are identified in the main interview.

1.3.3 Other methods for hard-to-identify populations

Snowball sampling, and its more recent outgrowth respondent-driven sampling (RDS; see Chapters 23 and 24 for discussions), are methods intended to reduce the problems of identifying members of rare or stigmatized populations. As Goodman recently pointed out (Goodman, 2011), snowball sampling was originally introduced by Coleman (1958–59) as a method for selecting a sample of the members of a social network, such as groups of friends at a school. Coleman started with a random sample of network members and used this initial sample to identify other members of the network. As Goodman noted, his method yielded a probability sample. Over time, however, snowball sampling has come to mean recruiting a convenience sample of members of some population, typically members of a “hidden” population (such as illicit drug users or illegal immigrants); these initial “seeds” then recruit additional members of the population, who then recruit additional members, and so on. In a series of papers, Heckathorn (1997, 2007, 2011) has explored the statistical properties of RDS and introduced several estimators that can be used with such samples. Under certain assumptions, Heckathorn argues, the estimators are unbiased. For our purposes here, three of the assumptions underlying RDS are crucial (these quotations are all taken from Heckathorn, 2011, p. 363):

(1) “Respondents know one another as members of the target population, as is typical of groups such as drug users or musicians”;
(2) “The network of the target population forms a single component”; and
(3) “Respondents can accurately report their personal network size, i.e., the number of those they know who fit the requirements of the study such as drug injectors or jazz musicians.”

If these assumptions are met, the members of the hidden population are not hidden to each other, but only to members outside the population. Of course, even if members of the hidden population know each other, this does not mean they are willing to reveal each other to the researchers. (Consider using RDS to recruit a sample of illegal immigrants.) It remains to be seen how often these and the other assumptions on which RDS rests are met in practice and how robust the method and associated estimators are when its assumptions are violated (see Chapter 24).

1.4 Hard-to-reach populations

So, some populations are rare or elusive and, as a result, hard to sample. With other populations, the challenge is picking out the members of the target group from some larger population (such as the general population), particularly when the members of the target group do not want to be identified. But there is still another source of difficulty that can make a population hard to survey – the members may be hard to locate or hard to contact. For example, Kelleher and Quirke (Chapter 10) describe a survey of Irish Travellers, a group that is hard to survey for several reasons, not the least of which is their mobility.

1.4.1 The hard to locate

There are at least four types of mobile populations that may be hard to locate:

- Members of traditionally nomadic cultures (such as the Bedouins of Southwest Asia and the Tuareg of North Africa);
- Itinerant minorities (such as the Romani in Europe or the Travellers in Ireland);
- Persons who are temporarily mobile or displaced (recent immigrants, homeless persons, refugees); and
- Persons at a mobile stage in their life cycle (college students).

Some of these populations are quite large. Passel (2006) estimates that there are 11.1 million "unauthorized migrants" in the United States (although these are probably mostly in households and thus not especially mobile) and estimates of the size of the Romani population in the US range up to a million. Mobility can make the members of some populations hard to locate. As we noted earlier, one strategy for capturing the members of mobile populations is to sample places where they are likely to be found. For example, in the United States, the 2010 Census sent enumerators to migrant worker camps, soup kitchens, and homeless shelters in an effort to count these mobile populations.

Mobility can also be a problem for longitudinal, or panel, surveys. There are a few papers on movers in such surveys (e.g., Couper & Ofstedal, 2009; Lepkowski & Couper, 2002). Couper and Ofstedal examined sample members who moved between rounds of the Panel Study of Income Dynamics (PSID) and the Health and Retirement Survey (HRS). They note that some 13.7 percent of the US population moved in 2004; the corresponding rates in Western Europe were somewhat lower. Both of the surveys that Couper and Ofstedal looked at were quite successful at finding sample members who had moved. The PSID located 96.7 percent of the 1,441 cases that needed to be tracked for the 2003 round and the HRS located 98.7 percent of its 1,294 movers for the 2004 round of that survey. Still, although these tracking efforts were very successful, they also required considerable resources. On average, it took 10.2 tracking calls to find the movers in the PSID and 7.4 tracking calls to find the movers in the HRS. Still, as these results suggest, the vast majority of movers are eventually found.

The correlates of being found, according to Couper and Ofstedal (see also Lepkowski & Couper, 2002) are, not surprisingly, related to the person's level of attachment to a specific place. People who are married, employed, older, and engaged in community activities are more likely to stay put and are easier to find if they do move. Despite a tendency to change their surnames, women seem to be easier to track than men are. In general, populations that are only loosely attached to a specific home or place are difficult to find. Thus, the homeless are notoriously difficult to count and to interview and they are missed by virtually all general population surveys (although see Chapter 9 in this volume). A less extreme case involves persons with weak attachments to several households. They are at risk of being omitted from household rosters and thus missed by surveys; Martin (1999) estimated that some 4 million persons in the United States might have such tenuous connections to a household. And people displaced by storms, other natural disasters, and wars can require extraordinary efforts to find and interview (see Chapters 6, 7, and 8 in this volume).

1.4.2 Barriers to access

Even when sample members can be found, it may still be difficult to contact them. One long-term trend that has probably contributed to the decline in response rates throughout the developed world over the last two decades is the widespread adoption of lifestyles and devices that shield people from unwanted solicitations. More and more Americans live in gated communities, locked apartment buildings, or other residential settings in which they are protected by gatekeepers, and the trends are similar in Western Europe. By the mid-1990s, nearly 40 percent of new residential developments in the US were gated (Blakely & Snyder, 1997). Even before cell telephones became popular, Americans used caller-ID and answering machines to screen out their telephone calls; now, as the population shifts to cell telephones, almost everyone is able to filter his or her calls.

It is not clear whether this shift to cell telephones has made it harder or easier to reach potential respondents. According to Blumberg & Luke (2012), about 25 percent of the adult population in the US was cell-only by mid-2010. Hispanics, young adults (18–34 years old), people living with roommates, poor people, and renters were more likely to be cell-only than the rest of the population. The figure for Hispanics was nearly 35 percent; for 25–29 year olds, it was more than 51 percent; and for adults living with unrelated adults, it was 69 percent. Although cell phones do encourage the screening of incoming calls, they are mobile devices and many cell users have their telephones with them all the time. In general, though, it seems that many of the same groups that are hard to survey for other reasons (such as young adults) are also getting harder to contact; these groups seem to be overrepresented in the cell-only population. At the other end of the spectrum, Groves and Couper (1998) suggest that two groups are relatively easy to contact – the elderly and parents with young children. Members of both of these groups are more likely to be at home than members of other subgroups of the general population. On the other hand, access to elderly in assisted-living settings may be limited by gatekeepers.