California Center for Population Research
University of California - Los Angeles

New Approaches to Demographic Data Collection*

Donald J. Treiman
Yao Lu
Yaqiang Qi

CCPR-2009-022

September 2009
Last Revised: September 2009

California Center for Population Research
On-Line Working Paper Series

New Approaches to Demographic Data Collection*

Donald J. Treiman
University of California at Los Angeles

Yao Lu
Columbia University

Yaqiang Qi
People's University of China

(last revised 9/21/09)

(Word count, excluding title page, References, and Appendix: 15,189)

Paper prepared for *Frontiers of Demography*, edited by Zai Liang
to be published by People's University Press, Beijing, in Chinese

*Donald J. Treiman, California Center for Population Research, University of California at Los Angeles (UCLA), 4284 Public Affairs Bldg., Box 957236, Los Angeles, CA, 90095, USA, email: treiman@ucla.edu; Yao Lu, Department of Sociology, Columbia University, 501 Knox Hall, 606 West 122nd St., New York, NY 10027, USA, email: email: yl2479@columbia.edu; Yaqiang Qi, Department of Sociology, Renmin (People's) University of China, Zhongguancun Dajie 59, Haidian District, Beijing, China, 100872, email: qiyaqiang@ruc.edu.cn. Preparation of this paper was supported by a grant from the National Science Foundation to UCLA: SES0551279 ("Internal Migration in China: Data Collection and Analysis"). We are grateful to Harry B. G. Ganzeboom, Perry Peifeng Hu, William M. Mason, Shige Song, and the volume editor, Zai Liang, for very helpful comments, and to Ted Mouw for pointing us to some useful references.

ABSTRACT

As population scientists have expanded the range of topics they study, increasingly considering the interrelationship between population phenomena and social, economic, and health conditions, they have expanded the kinds of data collected and have brought to bear new data collection techniques and procedures, often borrowed from other fields. These new approaches to demographic data collection are the concern of this chapter. We consider three main topics: new developments in sampling procedures; new developments in fieldwork procedures; and new developments in the kind of information collected in demographic surveys. We conclude with some comments on data sharing in the demographic research community. Where possible we illustrate our points with Chinese examples.

INTRODUCTION

In recent years the field of demography has been expanding rapidly and broadly, so that current demographic research now extends well beyond studies of fertility, mortality, and migration, the traditional concerns of the field. These developments have been of two main kinds.

First, much recent work in many social science and biomedical disciplines (sociology, economics, geography, public health, and medicine) has adopted a "population-based" approach (Duncan 2008:763-772). Sometimes this is as simple as basing analysis on data collected from samples of general populations, for example, studies of the incidence of disease based on probability samples of all residents of geographic areas rather than, say, members of a *danwei* or visitors to a particular clinic. But sometimes demographic concepts are merged with substantive ideas drawn from other disciplines, such as studies of social mobility that take account of fertility differentials by social status (Mare and Maralani 2006).

Second, there has been a great deal of "cross-fertilization" within these fields, resulting in the creation of new sub-specialities. Two good examples are the recent emergence of "bio-demography," the incorporation of biological data and research questions into demographic studies; and "spatial demography," the incorporation of spatial data (geo-coded information and satellite-generated measures and images) into population studies.

As is clear from these two examples, an important locus of these developments has been in the expansion of the kind of data collected and in data collection techniques. These developments—new approaches to demographic data collection—are the concern of this chapter. We consider three main topics: new developments in sampling procedures; new developments in fieldwork procedures; and new developments in the kind of information collected in demographic surveys. We conclude with some comments on data sharing in the demographic research community. Although these developments are worldwide, where possible we will refer to Chinese examples.

SAMPLE DESIGNS AND SAMPLING PROCEDURES

This section describes several developments in sample designs and sampling procedures, including the move toward true probability samples of general populations; sample design issues; and new sampling procedures, particularly those designed to sample rare or difficult-to-find populations. In order to limit the scope of our review we omit discussion of the literature on telephone interviews and mail questionnaires, since these are relatively uncommon mechanisms for surveying the general population in China.

Probability sampling

The conventional design for sampling general populations is multistage area probability sampling, in which a hierarchy of geographical units is identified (say, counties, townships, villages/neighborhoods, and households); units at each level are selected at random but (except at the household level) with the probability of selection proportional to the population size of the unit (PPS); and then individuals are selected to be interviewed. A national sample of 8,000 people might be created by selecting 100 counties, then two townships within each selected county, then two villages or neighborhoods within each selected township, then 20 households within each village or neighborhood. Within each selected household, one individual (or more) would be selected at random to be interviewed. Multistage samples often are more complex than just described because they are "stratified" by features of the geographical units, for example, whether the county is in a coastal, middle, or western province; whether it is predominantly urban or rural; the average level of education of the population; etc.

Probability sampling is not a new method. However, it has increasingly come to be recognized in China that neither convenience samples (for example, visitors to a clinic) nor samples of "typical" places provide an adequate basis for inference to broader populations; see the useful discussion by Duncan (2008:764-767) showing how such samples can result in misleading inferences. While casual samples still are widely used in health research, because of the difficulty of collecting health information from general populations, there is increasing emphasis even among medical and public health researchers in securing population-based samples because of known biases in hospital- or clinic-based samples. Two studies of the effect of prenatal malnutrition caused by the Great Leap Forward famine on the prevalence of young-adult schizophrenia provide a useful contrast. St. Clair et al. (2005) first suggested a link between the famine and schizophrenia, combining hospital records and historical population data for a single county. But Song, Wang, and Hu (2009), in motivating their study based on data from the 1987 *Chinese National Disability Sample Survey*, a very large national probability sample survey, noted that hospital-based studies in China are vulnerable to the selectivity of hospital admissions, which favors urban high status people. Because of the selectivity of hospital admissions, generalizations to the general population based on hospital records will inevitably be biased. Indeed, Song and his colleagues found sharp urban-rural differences in the linkage between exposure to the famine and schizophrenia that were entirely missed by St. Clair et al.

Even when Chinese surveys purport to be based on general population samples, all too often the samples are chosen on a casual basis. Even major studies do this. For example, the well-known *China Health and Nutrition Study*, a joint project of the Carolina Population Center and the Chinese Center for Disease Control and Prevention, which has followed the same people since 1989, is based on semi-probability samples of nine provinces concentrated in Eastern and Central China (with some primary sampling units (PSUs) chosen at random but others chosen purposively, and small areas within the PSUs chosen randomly).[1] Thus, although results from this study are often presented as pertaining to China as a whole, they hardly do so, since, with the exception of Guizhou Province, the entire Western Region of China, where health and nutrition problems are most severe,[2] is omitted from the sample.

Fortunately, many studies conducted by the Ministry of Health and the Family Planning Commission do sample the entire population of China; but many of these studies are not readily available for use by the research community, a point we will return to below. True national samples are less common in university-based studies, but they are becoming more common. One of the first such studies was Treiman, Walder, and Li's 1996 survey of *Life Histories and Social Change in Contemporary China (LHSCCC)*, carried out at People's University. In 2003, the *Chinese General Social Survey (CGSS)*, a Chinese version of the U.S. *General Social Survey (GSS)*, was initiated by Bian Yanjie, at the Hong Kong University of Science and Technology, in

[1] Chinese surveys available to the research community, usually without charge, are described in the Appendix, together with access information. Other studies referred to in the text are listed in a special subsection at the end of the **References**. Each survey is listed by both its full name and its abbreviation the first time it is mentioned in the text and subsequently by its abbreviation.

[2] As one indicator, 22 per cent of those living in China's Eastern and Central Regions report themselves to be in excellent health, compared to only six per cent of those living in China's Western Region (computations from the survey of *Internal Migration and Health in China (IMHC).*

partnership with Li Lulu at People's University. So far, this survey has been repeated five times, in 2003, 2004, 2005, 2006, and 2008, with another wave scheduled for 2010.

Design issues

The increasing opportunity for Chinese population scientists to carry out sample surveys lends added importance to the need for good sample design. Unfortunately, many Chinese surveys exhibit design flaws that make them less valuable than they could be. Thus, here we offer some points for consideration when designing new sample surveys.

Comparisons are the essence of social research

Unfortunately, researchers interested in a particular population often make the mistake of collecting data only from members of that population. The problem is that you cannot study a constant. If you want to know how the urban population compares to the rural population, you need a sample of both urban and rural people. Similarly, if you want to study, say, the determinants and consequences of internal migration, you need a sample of both migrants and non-migrants. This is an extremely simple point, but one that often is ignored in data collection efforts. For example, in 2002-2003 the National Bureau of Statistics carried out the *National Survey of Temporary Migrant Children in China*, a 9-city survey of 6,343 migrant households with children age 0-18. The restriction of the sample to the children of temporary migrants makes it impossible to assess the effect of migration on children's lives. Of course, a description of migrant children is possible, but without a comparison to non-migrant children it is difficult to interpret the meaning of the descriptive statistics. From an analytic perspective, all that can be done is to make internal comparisons, e.g., comparing the conditions and experiences of migrant children residing in different cities (Lu 2007). But, presumably, the question that motivated the data collection was how migration affects children. Given the design, this question cannot be answered.

Partial samples are subject to sample selection bias

There is an additional problem with partial samples, such as samples of migrants or samples of the urban population only—even descriptive statistics may be biased due to sample selection bias. Consider the claim that in urban China intergenerational social mobility is particularly pronounced relative to what is observed in other nations (Whyte and Parish 1984). The difficulty with such a claim is that in China rural-to-urban migration is often a *consequence* of upward mobility, the outcome of a process in which the brightest and most ambitious of those from peasant origins do well and go far in school and thus qualify for urban jobs. Thus, *de jure* urban-only samples tend to be comprised both of those from urban origins, who experience typical patterns of intergenerational mobility, and formal rural-to-urban migrants, who experience extreme upward mobility. Computations of mobility rates that fail to distinguish between these two groups will then overstate the true rate of upward mobility in the Chinese population as a whole. See Wu and Treiman (2007:416-417,440-443) for a more detailed discussion of this issue.

NBS rural and urban surveys do not jointly cover all of China

Apart from the sample selection bias inherent in analyzing samples selected on the basis of the dependent variable, there is an additional problem with conducting separate surveys of the urban and rural sectors, which is the conventional practice of the National Bureau of Statistics (NBS). The NBS has two separate data collection units, one for rural China and the other for urban China. In both cases, samples are drawn from the lowest administrative unit, "village committees" and "neighborhood committees" respectively. However, not all of China is covered by village committees and neighborhood committees. In particular, commercial buildings, factories, hospitals, universities, other institutions, and construction sites, as well as roadways and parks, are excluded, yet people live on these sites. Thus, sampling only within villages or neighborhood committees will exclude a portion of the population. This means that even when it is possible to combine NBS rural and urban surveys (which is seldom, because in general common questionnaires are not used), the combination will not yield complete coverage of the population of China. Moreover, a 2003-2004 pilot study for a national survey of migration and health encountered situations in urban fringe areas in which long-time residents were regarded as under the purview of the village administration but newly-arrived migrants were not (Treiman et al. 2006). In these areas, a part of the *de facto* resident population would be excluded from the sampling frame by definition.

So what to do? There are two possibilities. First, villages and neighborhoods could be sampled in the usual way, but in addition portions of a township not covered by either villages or neighborhood committees could be sampled. This would require careful estimation of the fraction of the population of a township living in such places to ensure that they are sampled in proportion to that fraction. However, in urban China there is no longer a consistently defined unit below the *jiedao*: "neighborhood committees" have in some instances been replaced by larger "community committees," high rise buildings often have their own "property owners' committees," and so on. Thus, given that townships are too large to be enumerated (in 2000, they averaged about 25,000 in population size), a way of sub-sampling small areas ("enumeration districts" (EDs)) within townships would need to be devised. One way to do this is to sample administrative units down to the township level and then, as an additional stage, to sample smaller areal units within each township, using a version of the **spatial sampling** procedures described below. For an example applying such a procedure, see the documentation for the 2008 *IMHC* survey.

Many surveys include only the easy-to-reach population

Until recently, the conventional way of carrying out general population surveys in China has been to sample from the local population register (*hukou*) and, sometimes, from lists of temporary residents. This is no longer an adequate sampling strategy, at least in urban areas, because there is so much residential mobility in contemporary China and many people fail to change their registration when they move. Computations from the *IMHC* survey show that 39 per cent of those living in cities changed their residence in the past five years and that of these 61 per cent moved within the same city the last time they moved. Moreover, 82 per cent of the within-city movers lack local *hukou*, presumably because their *hukou* is still held by their old neighborhood committee. Informal rural-to-urban migrants also commonly lack local registration. Fully 75 per cent of peasants (those with rural *hukou*) who reside in cities lack local *hukou*. These results are consistent with the findings of Landry and Shen (2005), who found, in a 2002 survey in Beijing, that 45 per cent of their respondents would have been missed had the survey team relied on registration lists. Because of the growing inadequacies of *hukou* lists, adequate sampling of the Chinese population now requires household enumeration procedures. We will discuss these in the section on **Fieldwork procedures**.

Apart from reliance on registration lists, another shortcoming of many Chinese surveys is their propensity to conflate families and households. Adequate coverage of the population requires that all members of a *dwelling unit* or *household* be enumerated.[3] The exclusion of roomers, servants, and employees from those eligible to be interviewed means that the sample does not adequately cover the population. Sometimes this is done on the ground that such people are "temporary" members of their households and will be counted in their permanent households. But sample surveys almost never interview members who are living elsewhere, even when they are recorded in household rosters. Because of this, the migrant population is often under-counted. Migrants also tend to be undercounted in panel surveys because few surveys attempt to interview those who have moved away between waves. But this, too, results in inadequate coverage of the population, especially in contemporary China where more than 40 per cent of the population lives other than where they are registered and 21 per cent of the population has moved within the last three years, the typical length of time between waves of panel studies (both computations are from the *IMHC*).

A final way in which surveys fail to cover the entire population is a propensity to overlook people living in non-standard dwellings. We will discuss this point in the section on

Fieldwork procedures

Clustering and stratification

Good sample designs strike a balance between coverage and cost. As already has been noted, general population surveys often are designed as multistage probability samples, with cases clustered within particular small geographic areas. The primary reason for clustering cases is to reduce costs—it is much less expensive to conduct interviews in several households within

---

[3]It is convenient to define a household or dwelling unit as "a living space with a doorway opening into a common or public area. Thus, for example, a room rented out to a migrant that opened into a family's living space would be regarded as part of the family household. A room opening into a courtyard or passageway would be regarded as a separate household." (Treiman et al. 2006:89-91).

a limited geographical area than to send interviews to widely scattered households. However, the typical consequence of clustering is to inflate standard errors.[4]

Sometimes surveys are excessively clustered, so that the standard errors become very large; or, to state the same point differently, the standard errors become similar to those expected from a much smaller simple random sample. A case in point is a recent cardiovascular study said to be based on a “nationally representative sample of 15,540 Chinese adults aged 35-74 years in 2000-01” (Gu et al. 2005:1398-1399). Upon close inspection of the sampling description, it turns out to be based on 20 sampling points, a rural township and an urban “street committee” selected in each of 10 provinces via multistage sampling procedures that did not take account of differential population size. Even ignoring the fact that the provinces “were selected to be representative of the geographic and economic characteristics in their regions,” that is, were selected purposively, this sample was so highly clustered that it almost certainly had the precision of a much smaller sample.[5]

Also as previously noted, multistage samples often are stratified, by region, type of place, the socioeconomic level of the community, and so on. Stratification is used both to ensure that there is adequate coverage of all segments of the population, sometimes sampling segments at different rates (e.g., over-sampling migrants in order to have enough migrants to be able to compare them with non-migrants), and to reduce the deleterious effect of clustering, since

[4]A second reason for clustering cases is to permit multilevel analysis (DiPrete and Forristal 1994; Mason 2001), studying whether some micro-process, say the relationship between mother’s education and fertility, varies across communities (Entwisle and Mason 1985). A powerful analog to multilevel analysis for experimental designs is cluster randomized trials (ALLHAT 2002; Miguel and Kremer 2004; Hayes and Moulton 2009).

[5]The authors report (p. 1400) that they assume a design effect “(the ratio of the variance of a statistic from a complex sample to the variance of the same statistic from a simple random sample of the same size)” of 1.5; but this almost certainly is far smaller than if they had correctly calculated actual design effects.

stratification generally reduces standard errors. Indeed, it sometimes is possible to dramatically reduce the size of design effects by stratifying on variables strongly correlated with the main analytic variables (for a Chinese example, see Treiman et al. 1996; see also discussion of the same study in Treiman 2009:207-212).

Household vs. person samples

Many social processes of interest to population scientists pertain to the behaviors and decisions of families and households rather than or addition to those of individuals. For example, the decision to migrate is often seen as the outcome of family risk-diversification strategies—some grown children are kept home to work on the family farm while others go out for wage work, thus providing two alternative ways to protect against downside risk (Stark 1992; Massey et al. 1993, 1998; Roberts 1997). Thus, surveys that take individuals as the sampling units may not be able to capture family structures and dynamics that affect outcomes. One common solution to this problem is to collect extensive information about other family and household members; but usually it is necessary to restrict the information collected to socio-demographic items since other family members cannot be expected to be reliable reporters regarding attitudes, values, preferences, and the reasons for making particular decisions, nor regarding retrospective information such as work, residential, medical, and other histories.

For this reason, surveys increasingly are designed to interview all members of each sampled household. The *China Health and Nutrition Survey (CHNS)* uses such a design, as do the *Indonesian Family Life Survey (IFLS)* and the *Mexican Family Life Survey (MxFLS)*. Although relatively expensive since the interviewer must spend many hours with various members of a household (but not nearly as expensive as samples of the same size where one adult per household is interviewed), such surveys yield rich information on the characteristics and beliefs of each member of the household and also permit direct biometric, anthropometric, and cognitive assessments (see the discussion below of developments regarding the kind of information collected in sample surveys). An additional advantage is that such designs permit the estimation of household fixed-effect models, which provide a useful way of purging analysis of the effect of all unmeasured variables that are constant across members of a household (see Treiman (2009:Ch. 15) for a brief introduction to such models and Allison (2005) and Wooldridge (2006:Chs. 13 and 14) for more extended explications). Finally, household samples do not suffer from a bias inherent in the most common household interview design, in which households are chosen at random within small areas and one eligible person (say, one adult) per household is chosen at random to be interviewed. The problem with such designs is that in large households each individual has a smaller chance of being selected than do individuals in small households. The usual way this bias is corrected is to weight the data for each person interviewed by the number of eligible respondents in the household. While this is a reasonable approach, it is not optimal since weighted data have larger standard errors than non-weighted data.[6]

The use of household samples becomes problematic when family members live apart, as is generally true of families in China from which some members have "gone out" for work, or when individuals live in large collective households with non-family members, as is also true of many migrants in China. In such cases, it may be preferable to sample individuals and then also

[6] An alternative to weighting is to include all variables used to construct weights in one's estimation equations and then to carry out the estimation without weights. The optimal way to carry out estimation for complex sample designs is an unsettled and rapidly evolving topic in statistical research. For a useful recent discussion, see Gelman (2007), plus five comments and Gelman's rejoinder published with Gelman's paper.

to sample selected relatives, e.g., siblings, spouses living apart, parents, etc. Doing this will, of course, increase the expense of carrying out surveys, since it sometimes will be necessary to send interviewers to other locales to interview the selected relatives. But it is a useful way to gain the advantages of household samples in situations in which many families have members living in more than one dwelling unit.

New sampling methods

In this section we consider some relatively new sampling strategies designed for situations in which conventional probability samples of households do not work well. These include spatial sampling, contingent quota sampling, respondent driven sampling, and other methods for sampling hard-to-reach populations (targeted sampling, time-and-place sampling, screening, capture-recapture methods, and adaptive sampling).

Spatial sampling

When an adequate sampling frame is not available, a useful alternative is to sample spatial units identified by "global positioning system" (GPS) coordinates, a method first used in studies of the environment—ocean temperatures, vegetation types, air pollution, etc. In such applications, the area to be studied is subdivided into a set of regular polygons, which are then selected at random. In studying human populations, however, investigators often want to take account of such factors as population density, in order to represent populations of people rather than populations of space, and also other aspects of population heterogeneity. These considerations add complexity to the sampling process, requiring measurements of density and heterogeneity and their incorporation into sampling designs. Two interesting applications are a respiratory health and demographic survey of households carried out in the Delhi, India, metropolitan area in 2004 (Kumar 2007) and a survey of a U.S. county (King County, Washington, which includes Seattle), where the interest was in achieving a sample of different types of built environments in order to study factors conducive to walking and bicycling as means of transportation (Lee, Moudon, and Courbois 2006).

In a different sort of application, Landry and Shen (2005) used spatial sampling as a way of overcoming the limitations of *hukou* lists in contemporary Beijing, especially in the coverage of migrants. Their strategy was to overlay a map of Beijing with a grid showing degrees of latitude and longitude and to randomly sample, but with probability proportional to a separately derived estimate of population size, 50 “square minutes” (trapazoidal areas of approximately 1.4 by 1.8 kilometers at Beijing’s latitude) as their PSUs and then, within each PSU, to randomly sample four “square seconds” (areas of approximately 54 meters by 54 meters) as “secondary sampling units” (SSUs). In each of the selected SSUs doors were counted to create a list of addresses from which to select households for interviewing. This strategy has clear advantages with respect to the completeness of coverage. Landry and Shen compared their results to data obtained from the Beijing Area Survey, which (by design) had sampled the same districts in Beijing, and showed that, as noted earlier, 45 per cent of the respondents to their spatial survey would have been missed by the Beijing Area Survey. See Gibson and McKenzie (2007:222-225) for additional examples of the use of spatial data as a sampling tool.

But the method also has some disadvantages, of which two are substantial. First, the PSUs and SSUs do not conform to any administrative boundaries, making it difficult to utilize administrative data as a supplement to the data collected from respondents. Second, the SSUs selected often result in buildings being divided, partly in and partly out of an SSU, which requires awkward adjustments. A third limitation, reliance on a simple count of doors, can be overcome by changing enumeration procedures, a point we will return to below. Still, it is a method worth considering, as an alternative or supplement to more conventional ways of implementing multistage sample designs.

Contingent quota-sampling

Quota sampling is an approach that assists in obtaining sufficient numbers of respondents within sample strata to achieve a specified level of statistical power. It is similar to stratified probability sampling in that a specific number of subjects from different subgroups are recruited. The difference is that subjects are recruited by convenience rather than randomly. The quotas are pre-determined (i.e., according to sex and age groups) and interviewers are instructed to fill their quotas as best they can for the randomly selected household. Once the quota for a subgroup is met, the interviewers no longer recruit subjects for that subgroup. More detailed discussions of the method can be found in Moser and Kalton (1971) and Kalton (1983).

While the non-random element is a major weakness of this approach and quota samples are essentially convenience samples, several earlier studies suggest that quota sampling often represents a fair balance between technical rigor and the production of useful data (Moser and Kalton 1971; Cumming 1990; Owen, McNeil, and Callum 1998). Although it is not as rigorous as probability sampling, it is often chosen as the most cost effective means of sampling hard-to-reach or hidden populations. One experimental study explicitly evaluates the performance of quota sampling versus probability sampling and finds that in many situations quota sampling can be considered an acceptable alternative to probability sampling when the latter is impossible or difficult to accomplish (Cumming 1990). Quota sampling is especially useful when a sampling frame is not available. Another advantage of quota sampling is that researchers can be specific about the type of subjects desired for the study and can be assured that specific subgroups are represented adequately in the final study sample.

With respect to the disadvantages of the approach, in addition to non-randomness quota samples are vulnerable to bias resulting from giving interviewers excessive latitude in the way they meet the defined quotas. For this reason, a modification, “contingent quota sampling,” is often employed, in which a small geographic area is specified (for example, the “square seconds” used by Landry and Shen 2005) and interviewers are instructed to secure interviews with specified numbers of people with particular characteristics, e.g., men under age 40, men age 40 and over, etc. in each small area. The implicit assumption of this procedure is that, within small areas, there is no correlation between inclusion in the quota and other attributes of respondents. This assumption certainly is more reasonable within small geographic areas than over wider, and presumably more heterogeneous, sampling points, but may nonetheless be problematic. This procedure was utilized for part of the *IMHC* and in several other recent surveys for which the data collection was done by the Beijing-based survey research firm, Millward Brown ACSR. Specifically, if an interviewer was unable to gain entrance to a high rise apartment building (an increasingly common problem in Chinese cities), s/he was instructed to stand outside the building and approach residents as they entered the building, alternating males and females. The assumption of this approach is that those who agreed to interviews would be similar to members of the randomly selected households in the same buildings that interviewers were supposed to visit.

Respondent-driven sampling (a modern approach to "snowball" sampling)

For certain kinds of population research it is necessary to collect data on difficult-to-find populations, such as prostitutes, HIV-positive individuals, and so on. Such populations cannot be sampled in conventional ways since usually there is no complete list of the population to be sampled nor are targeted individuals sufficiently prevalent in the general population, even within small geographic areas, to make area probability sampling feasible. And when special sites are targeted, they tend to produce biased coverage because the kinds of individuals found at such sites tend to be a selected subset of all eligible individuals. (As a simple illustration, it is evident that, in the U.S., the kinds of Blacks who live in Black neighborhoods differ in substantial ways from the kinds of Blacks who live in non-Black neighborhoods.) Moreover, when the criterion for inclusion in the sample involves stigmatized behavior, such as prostitution or drug use, individuals may be reluctant to be interviewed.

Until recently, the only available sampling method was "snowball" (also called chain-referral or network-based) sampling. In this method, a few individuals in the group to be sampled, say, prostitutes, are identified and interviewed. In the course of the interview, they are asked to identify other prostitutes, who are in turn interviewed and also are asked to identify still other prostitutes. In this way the sample is expanded, like a snowball rolling downhill, until the target size is reached.

The obvious limitation of this method is that the "sample" is not a probability sample of the targeted population. "Sociometric stars," that is, those who are central to networks, are much more likely to be named, and they may well differ in their characteristics from more isolated individuals. Moreover, because of the tendency for people to be more likely to interact with others similar to themselves, the choice of initial contacts is likely to affect the distribution of characteristics in the realized sample. Thus, valid generalization from the sample to the population of interest is usually impossible (Erickson 1979).

However, in 1997 Douglas Heckathorn, a sociologist at Cornell University, proposed an important improvement, a method he called "respondent-driven sampling" (RDS) Heckathorn combines snowball sampling with a set of procedures and a statistical model that generates weights for individuals in the derived sample that correct for the fact that some individuals are more likely to be included than others. Initial contacts (usually chosen purposively) are interviewed and are then asked to nominate others, customarily by distributing coupons which are returned to the investigators by those willing to be interviewed. A crucial aspect of the procedure is to collect information from nominators on the number of eligible individuals in their networks, to be able to estimate population coverage. In a series of papers (Heckathorn 1997, 2002, 2007; Salganak and Heckathorn 2004; Volz and Heckathorn 2008), Heckathorn has shown that, when certain (strong) assumptions are met, his method yields asymptotically unbiased population estimates when conventional statistical inference procedures are applied; see also Heckathorn's web site: http://www.respondentdrivensampling.org/. For discussion of a closely related method, "network sampling" or "multiplicity sampling," see Sudman, Sirken, and Cowan (1988:993-994); see also the set of papers by Thompson and his colleagues cited in the references, which include "link-tracing" designs as one general category of "adaptive sampling" strategies.

There now has been considerable work evaluating RDS procedures, and in particular the assumptions required for correct statistical inference; see, in particular, Heimer (2005), Gile and

Handcock (2008), and Merli et al. (2008a). Much of this work has focused on the impact of non-random selection of initial respondents, and the effect of two properties of the resulting networks in overcoming bias introduced by initial non-random selection: the degree of subgroup segmentation within the target population (that is, the extent to which subsequent nominations approximate random selection), and the number of links in the referral chain (the more the better).

Because it meets a strongly felt need—to be able to survey hard-to-locate populations—RDS has been widely adopted; Malekinejad et al. (2008) review more than 100 RDS-based studies in more than 30 nations. Merli's research on prostitutes in Shanghai is a nice example of the application of the method. She and her colleagues interviewed 522 female sex workers (FSW) in Shanghai in 2007 using RDS procedures (Merli et al. 2008b), which enabled them to reach FSWs who were not employed in institutional settings (karaoke bars, hotels, and saunas). However, as she shows (Merli et al. 2008a), her data did not meet the assumptions specified by Heckathorn for valid statistical inference. This is a common outcome and suggests that RDS must be used cautiously. For additional discussion of the implementation of respondent-driven sampling designs, see Johnston et al. (2008).

Other methods for sampling hard-to-research populations

Several other methods for sampling hard-to-reach populations have been utilized by researchers. These include:

Targeted samples. In this method, populations of interest, or locales where members of such populations are known to congregate, are identified on the basis of ethnographic fieldwork, interviews with informants, e.g., service providers, and then members of the targeted populations are chosen for interviews, either through complete censuses of the selected populations or areas, probability sampling of generated lists of individuals, or purposively. While this method provides coverage of populations likely to be missed by other methods, it suffers from the absence of a sampling frame, which precludes statistical inference. See Robinson et al. (2006) and Peterson et al. (2008) for further discussion of targeted samples.

Time-and-place samples (also known as **time-location** samples). In this variant of targeted samples, selected locales are visited at specific times of the day or days of the week. Such samples often are collected at venues where members of the target population are expected to be concentrated, e.g., gay bars to locate HIV-infected individuals, “crack houses” to locate intravenous drug users, etc. But time-and-place sampling need not be restricted to such places. For example, it might be reasonable to sample migrants by approaching people on the street at particular times and places. But, of course, this approach suffers from the same limitations as other kinds of targeted samples. See Watters and Biernacki (1989), Muhib et al. (2001), and Ramirez-Valles et al. (2005) for further discussion of targeted time-and-place samples.

Screening. Under certain circumstances it may be possible to screen for members of rare populations in the course of carrying out surveys for other purposes. If for example, a very large general population survey is mounted, as is done fairly frequently by the Ministry of Health and the Family Planning Commission, it would be possible to insert questions asking whether people have certain rare conditions or diseases. If the answer is affirmative, contact information would be solicited so that people with the disease or condition can be followed up with a specialized questionnaire. For example, the U.S. *National Survey of Family Growth (NSFG)* was based on a sample of reproductive age women previously interviewed in the U.S. *National Health Interview*

Survey (NHIS) (National Center for Health Statistics 1987). Also, Harris et al. (1984) located victims of serious accidents by adding a screening question to a large number of omnibus surveys until a sufficient number was obtained. Of course the adequacy of this strategy turns on the completeness of the original sample. If the initial sample is truly representative of the population, then individuals with the rare condition will also be representative of the population of such individuals. It also turns on the ability to elicit honest answers from potential respondents. Questions about whether individuals are HIV positive or have had homosexual sex or extramarital affairs simply inserted into questionnaires concerned mainly with other topics are not likely to elicit candid responses.

Capture-recapture methods. This method, originally developed to count populations of animals, is particularly helpful in evaluating the completeness of coverage of censuses (Fienberg 1992), surveys, or other enumerations, e.g., disease prevalence rates (McCarty et al. 1993), but also can be used to estimate the size of a rare or difficult-to-find or mobile populations, such as the homeless (Fisher et al. 1994). The basic idea is to make two independent observations of the same population (e.g., two independent probability samples, or a sample plus administrative records, or a resample of the population near in time to the first sample), determining the number of individuals who are included in the first sample only (N_1), the number who are included in the second sample only (N_2), and the number who are included in both samples (M). An estimate of the number of individuals in the population (N_T) is then given by $N_T = (N_1*N_2)/M$. However, the validity of estimates derived from this procedure depends on strong assumptions that are not always easy to meet in human populations: (1) that every individual in the population has an equal probability of being counted (the completeness of coverage of surveys and censuses often varies by ethnicity, age, and sex); (2) that the counts are independent, so that the probability of being counted in each of the samples is uncorrelated (in human populations, this is usually implausible—those who tend to be missed in one data collection effort, e.g., the kinds of people who have no settled address, are likely to be missed in another). (3) where the population is re-sampled, that the size and composition of the population do not change between the initial data collection and the resampling exercise (an assumption that would be violated if, say, the number of migrants increased between initial and followup data collection). See Sudman, Sirken, and Cowan (1988:994-995) for additional detail on this method and Darroch et al. (1993) for a way to obtain valid estimates while relaxing the first assumption.

Adaptive sampling refers to designs in which data from early observations in a survey are used to improve the efficiency of subsequent data collection. That is, in adaptive sampling, the sampling procedure is modified on the basis of observations made during the survey. Whenever targeted subjects are observed, additional neighboring or linked units may be added to the sample. Adaptive sampling procedures for which statistical estimation procedures have been developed can be divided into two general categories: link-tracing designs, which operate primarily based on social network information; and other designs, which operate primarily based on geographic information (Thompson and Collins 2002). Since link-tracing designs already have been discussed in the section on **Respondent driven sampling**, we here focus on other designs (Thompson 1997:299-302). "Sequential stopping designs" are those in which sampling continues until a given criterion, a specified number of cases or a specified sampling variance on a variable of interest, is obtained. "Adaptive allocation designs" start from an initial stratified probability sample. But then, based on the initial observations, an additional stratified sample is drawn based on the observed values for the initially selected units. For example, in a migration study, a set of initial strata could be specified to yield a fraction of the total desired sample. But then a second stratified sample could be drawn that gives larger sample sizes to strata in which many migrants were observed in the first sample. A variant of this approach is "adaptive cluster sampling." In this design an initial sample is selected based on a conventional probability sampling design. Then if any individuals in the sample meet the criterion (say, are migrants), additional cases are added to the target size for the ED. If any of the added cases meet the criterion, still other cases are added, and so on.

Adaptive sampling has been effectively used in surveys where the target population is rare and not easily identifiable but is clustered geographically (Thompson, Ramsey, and Seber 1992). When properly implemented, this procedure can achieve a probability sample at reasonable cost; it also can improve efficiency (particularly for populations that are rare and clustered) and can provide unbiased estimates of population characteristics using available algorithms; and it can substantially increase the yield of targeted units in the sample (Thompson 1997). Overall, adaptive sampling strategies have been found to produce remarkable increases in precision or efficiency compared to conventional sampling designs of equivalent sample size. But the pitfall is that the approach can be expensive and technically difficult. In-depth discussions of the statistical properties of these designs and of the procedures required to achieve efficient unbiased estimates can be found in papers by Thompson and his colleagues (Thompson, Ramsey, and Seber 1992; Thompson and Seber 1996; Thompson 1997; Thompson and Frank 2000; and Thompson and Collins 2002).

Although it often is used for animal or plant studies, the procedure has been modified for use in human participants research and has been used successfully to sample intravenous drug users, persons at risk for HIV/AIDS, people involved in underground economic activities, and adolescent smokers (Thompson 1997; Thompson and Collins 2002; Chaudhuri, Bose, and Ghosh 2004).

FIELDWORK PROCEDURES

In this section we discuss new approaches to implementing sample surveys, including improved enumeration procedures; the use of incentives to increase response rates; methods for converting refusals; computer-assisted personal interviewing (CAPI) methods for recording responses; procedures for handling sensitive questions in the absence of privacy; and methods for keeping track of respondents between waves in panel studies.

Improved enumeration procedures

In many nations, and increasingly in China, the conventional way to select households or individuals to be interviewed in sample surveys is first to select small areas, usually in a multistage process and usually with probability proportional to size (see the discussion in the section above on **Probability sampling**). Then, within each small area, all households or all individuals are enumerated and a list is created from which households or individuals are selected at random to be interviewed. There are basically three ways to carry out the enumeration task (see also Treiman et al. 2006:84-95).

The simplest and least expensive option is to simply move through the area, listing each address. This method works well in residential neighborhoods consisting entirely of single family houses, such as are common in the U. S., although even in such neighborhoods it is likely that "mother-in-law" apartments, occupied guest houses, etc., will be missed. It also works well in high rise apartment areas, provided the enumerator is able either to gain entry to the building or to obtain information on the number of units in each building from the management company or from building guards. But, of course, what is generated is a list of households, which then requires a screening step with an informant in the selected households to determine which individual household members are to be interviewed.

A better approach is to develop a list of addresses together with information on the composition of each household—minimally, the number of eligible residents, e.g., adults. This can be done by engaging as an informant a neighborhood or village committee official who accompanies the interviewer, helping to locate addresses and approaching householders when necessary to obtain information on household composition. The advantage of a local informant is that non-standard dwelling places (sleeping spaces in shops, lofts over workshops, etc.) and "doors-behind-doors" (extra rooms in courtyards, etc.) are less likely to be missed. A second advantage is that a random sample of people can be drawn from the enumeration list, even though the exact individuals to be interviewed are not known and must be determined at the dwelling place by listing all eligible household members and randomly selecting the required number using a Kish table or similar procedure. A disadvantage of this method is that often it is not possible to accurately determine the size or composition of households simply by using informants.

The "gold standard" is to conduct a mini-census, visiting each household, determining who lives there, and recording for each resident the age, sex, and, where necessary, other identifying characteristics. For example, if a household consists of the employees of a small workshop, there may be several individuals of the same age and sex, in which case they need to be distinguished, e.g., the "tall girl from Sichuan," etc. There are several advantages to the mini-census approach: (a) it makes it possible to draw a genuinely random sample of residents of the area, without the clustering that is inherent in sampling first households and then individuals within households; (b) the sample can be drawn by the supervisor, making it less possible for interviewers to cheat by substituting immediately available household members for those requiring a return visit; (c) and it improves the likelihood of complete coverage because household members can act as informants not only about their own households but about the presence of sub-households ("doors behind doors") that the interviewer may have missed. But this method is seldom used because it is very expensive.

No matter which option is used, a map of the area needs to be created with addresses listed on the enumeration sheet also identified on the map. Without such a map, it often is extremely difficult for interviewers to find the households selected for the survey. This is particularly true in crowded neighborhoods in which apartment units are subdivided and extra rooms are built in courtyards, on rooftops, and in storage areas to rent out to migrants; such separate dwelling units are difficult to identify as such and often lack separate addresses. A useful supplement to a map, which is now technically feasible, but which could arouse suspicion in the Chinese context, is to identify each address by geo-coordinates and a photograph (although the accuracy of GPS receivers needs to be kept in mind).

The use of incentives to improve response rates

In principle, the validity of inferences drawn from sample surveys depends on having high response rates (see three special issues of journals devoted to the problem of nonresponse to sample surveys: in 1999, vol. 15, no. 2, of the *Journal of Official Statistics,* edited by Edith de Leeuw*;* in 2006, vol. 169, no. 3, of the *Journal of the Royal Statistical Society*: *Series A (Statistics and Society)*, edited by Peter Lynn; and in 2006, vol. 70, no. 5 of the *Public Opinion*

Quarterly, edited by Eleanor Singer; see also Groves and Couper (1998) and Groves et al. (2002)). When some individuals in a probability sample cannot be interviewed (because they cannot be contacted or refuse to participate), the danger arises that statistics computed on the portion of the sample for which data are available are biased. This occurs if those not interviewed differ systematically from those who are interviewed with respect to the variables under consideration. Although substantial nonresponse does not always result in substantial bias, it often does (Groves 2006). Thus, the best protection is to secure interviews from as high a fraction of targeted individuals as possible. Unfortunately, response rates have been declining in recent years, both in China and elsewhere, mainly because of an increase in refusals (de Leeuw and de Heer 2002).[7] In China, a combination of increased personal freedom, possibly rising levels of mistrust, rapid urbanization (around the world response rates are lower in urban areas (de Leeuw and de Heer 2002)), and changes in the housing environment (an increasingly large fraction of the urban population lives in restricted-access buildings), has resulted in a sharp reduction in response rates. Indeed, the only Chinese surveys able to secure high response rates these days are "official" surveys that people feel uncomfortable refusing; in the U.S., as well, government-sponsored surveys tend to have higher response rates than academic surveys, which tend to have higher response rates than commercial surveys (Groves and Couper 1998).

Because response rates have been declining throughout the world, there has been a fair amount of research on the effect of incentives on response rates. Singer and her colleagues

[7] Valid comparisons of response rates across surveys require a standardized way of measuring non-response. This has led to the development of a set of standards by the American Association for Public Opinion Research that has been widely adopted internationally (AAPOR 2006). Unfortunately, as two recent reviews of response rates make clear (Feng 2007; Hao 2007), the AAPOR standard does not appear to be followed by Chinese investigators.

conducted a meta-analysis of 39 U.S. studies using face-to-face interviewing in which the effect of incentives was experimentally assessed (Singer et al. 1999; see also Singer 2002). Their main conclusions were that incentives improve response rates, especially in studies with low response rates (where presumably other motivations for participation are absent); the larger the incentive, the larger the increase in response rates; both monetary and gift incentives help, but monetary incentives are more effective even when the value of the different types of incentives is equated; incentives matter more when the survey is burdensome because the questionnaire is long; and prepaid incentives appear to be more effective than promised rewards for completed interviews, although the limited number of comparisons makes this result less than definitive.

The use of lotteries as incentives is favored by some investigators. However, the evidence on the efficacy of lotteries, all based on mail or self-administered surveys, is mixed (Singer 2002:6). No evidence exists regarding interview-mediated surveys. Moreover, there is some question as to whether lotteries would be effective in contemporary China because of the low level of trust.

Methods for converting non-responses.

It long has been understood that repeated "call-backs" are effective in reducing non-contact non-responses (Goyder 1985). Of course, remaining in an area for several days or even weeks in an attempt to eventually find someone home can add substantially to the cost of surveys, which is why survey houses, especially commercial survey houses, often are reluctant to do repeated call-backs.

Refusals are more difficult to "convert" than are non-contacts because the latter is mainly a matter of returning enough times to finally find someone at home. Three strategies are used to convert those who initially refuse: offering incentives, or increasing the value of the incentive if incentives were initially offered; elaborating on the value of the study, often by sending a different interviewer; and employing especially skilled interviewers. Sometimes, because of the costs involved, non-respondent conversion is attempted only for a subsample of initial refusers, and the results are combined using the principles of "double sampling" (Hansen, Hurwitz, and Maddow 1993; Levy and Lemeshow 2008). Starting with its 2004 survey, the U.S. *GSS* has been using double-sampling methods to convert non-respondents, typically sampling half of the "temporary non-respondents" as of some cutoff date, usually 10 weeks after the start of interviewing (Davis, Smith, and Marsden 2009:2200). For another example of the use of double-sampling methods to improve coverage, see Jenkens et al. (2004).

An interesting example of the use of skilled interviewers to convert non-responses is the second (2005) wave of the *MxFLS*, conducted three years after the first wave of data collection. By 2005 about 850 of the approximately 35,000 initial respondents had emigrated to the U.S. When other family members were asked for the U.S. addresses of these migrants, many refused, probably because in many cases they were in the U.S. illegally. However, when the study team sent an especially trained and highly skilled interviewer to revisit the households, he was able to obtain U.S. cell phone numbers and addresses for almost all of the emigrants; then another highly skilled female interviewer successfully contacted and interviewed a large fraction of this sample (personal communication from Elizabeth Frankenberg).

Computer-assisted personal interview (CAPI) methods

With the low-cost availability of portable computing devices ("personal digital assistants" (PDAs), notebook computers, and netbook computers), it has become possible to carry out computer-assisted personal interviews (CAPIs) in field settings. This is done by programming the computing device with the survey protocol. The interviewer moves from screen to screen in much the same way as s/he would move through a paper version of a questionnaire, reading questions and recording answers.

There are several advantages to CAPI surveys. First, response errors are reduced. The CAPI protocol can be programmed to require a response (including “Don’t know” or “Refuse to answer”) before moving to the next question, reducing inadvertent skipping of questions or even blocks of questions (e.g., by turning two pages at once). Moreover, complex filters and skip patterns, in which respondents are asked or not asked certain sets of questions, depending on their responses to previous questions (e.g., only those who say they have ever been married are asked questions pertaining to their spouse), can be managed much better in a CAPI survey than in a PAPI (“paper and pencil interview”) survey, where errors on the part of the interviewer are frequent.

Second, the CAPI protocol can be programmed to generate an analysis-ready computer file, eliminating the need for key entry and radically reducing the amount of “cleaning” (post-interview editing) of the completed questionnaires. This is an extremely important advantage, since many PAPI surveys are not available for analysis until several months after completion of data collection and even then often contain many errors and inconsistencies.

Third, it often is possible to arrange for completed questionnaires to be transmitted via the internet to headquarters, either in real time or every evening, rather than after fieldwork is completed. This permits greater management control of the fieldwork process and may even permit “mid course corrections” of the kind discussed above in the section on **Adaptive surveys**.

Real time transmissions also may reduce cheating by interviewers, who have been known to complete questionnaires on their own kitchen tables. For example, if GPS coordinates have been identified as part of the enumeration process (see the discussion above), the management team can determine whether the interview is actually being conducted at a selected address.

The obvious downside to CAPI procedures is the cost, both the expense of programming the CAPI protocol and the cost of computing devices for each interviewer. However, as the cost of computers has declined, with quite powerful GPS-equipped netbook computers (which are, by definition, internet-capable) available in China for about rmb 1,500 as this is written, CAPI procedures have become financially viable.

A "net café" procedure

A CAPI variant that has been used in some recent Chinese surveys is to rent computers in an internet café in each neighborhood or village where interviews are being conducted. Interviewers visit households and conduct screening in the usual way. But then individuals chosen for interviewing are invited to visit the internet café to be interviewed. The respondent and interviewer sit together and the interviewer conducts the interview on the basis of the CAPI screens. There are several advantages to this procedure. The most important is that it may reduce refusals if people have time to reflect on the request that they be interviewed and to weigh the value of any incentive offered against inconvenience to them, rather than being required to respond on the spot. Moreover, it makes it possible for them to check out the legitimacy of the survey operation before agreeing to an interview. Given the low level of trust in contemporary China, this may be very helpful. A second advantage is that it obviates the technical issues associated with the use of netbook or other small computers in field settings (battery life, visibility of the screen in bright light, the lack of adequate workspace in crowded and poorly furnished houses, the difficulty of formatting complex lists for a small screen, etc.). The obvious disadvantage is that internet cafes may not be readily available, especially in remote villages, although they appear to be increasingly ubiquitous in China. However, this problem can be overcome by using a "survey van," a specially equipped internet-connected van, which can travel to remote places where net cafes are unavailable.

Handling sensitive questions in the absence of privacy

In many Chinese households it is difficult to conduct interviews in private. Often space is limited and even when it is not it is common and culturally acceptable for family members and also neighbors to gather around to watch the entertainment afforded by the interview. In such circumstances, candid and accurate responses to sensitive questions, such as extra-marital affairs, drug use, homosexual activity, or even whether people have had various diseases, whether they are happy, depressed, or have particular attitudes or opinions, are unlikely. People may refuse to answer the question or may give socially desirable responses. In such circumstances, a useful procedure is Audio Computer-Assisted Self-Interviewing (ACASI). Here is a description of the procedure from the Research Triangle Institute (RTI), a leading U.S. survey research organization:

> Using this tool, the respondent uses headphones connected to a laptop computer to listen to questions that have been digitally recorded to an ACASI program, then keys his/her responses directly into the computer. This methodology has proven to be a highly successful means of gathering information on sensitive or personal topics, due to the

added measure of privacy that results from the absence of an interviewer and the use of headphones (http://www.rti.org/).

RTI reports having conducted more than 225,000 interviews using ACASI, including a survey of a national sample of women on such topics as abortion, sexual partners, and HIV risk behaviors. The procedure may be used in conjunction with standard CAPI procedures. For example, the *National Survey on Drug Use and Health (NSDUH)*, conducted by RTI, includes within a CAPI protocol an extensive ACASI component that involves the use of tobacco, alcohol, illicit drugs, and nonmedical use of prescription drugs.

Sometimes researchers are interested in establishing incidence or prevalence rates of behaviors that are so socially unacceptable that it is likely that respondents would be unwilling to admit to them under any circumstances in which they could be connected to their responses, such as criminal activity, domestic violence, rape, molestation of children, and so on. To collect such information, Warner (1965) some years ago proposed a procedure known as the "randomized response technique" (RRT), which has since been used in many studies; for good overviews see Fox and Tracy (1986) and Chaudhuri and Mukerjee (1988); for a recent meta-analysis of the validity of responses, which arrives at a favorable evaluation, see Lensvelt-Mulders et al. (2005).

The essence of the technique is that respondents are asked to flip a coin, without revealing the outcome to the interviewer, and then are asked to give a positive response to a sensitive question such as "have you ever hit your wife" if *either the answer is yes or if the coin is "heads."* Since the interviewer does not know the outcome of the coin flip, s/he has no way of knowing whether the respondent has admitted to wife beating or not. Now, suppose 60 per cent of respondents answer yes to the question. Since the known (expected) percentage of "heads" in coin flips is 50 per cent, the researchers can conclude that 10 per cent of respondents have ever hit their wives. Of course, which respondents did this is unknown. But the technique has been shown to more accurately establish incidence and prevalence rates than direct questions. Moreover, it is possible to establish subgroup variations. For example, suppose that the percentage "positive" is 53 per cent for urban men and 63 per cent for rural men. We could then conclude that 3 per cent of urban men and 13 per cent of rural men have ever hit their wives.

Panel maintenance: keeping track of respondents between waves

In panel surveys, where the same respondents are periodically re-interviewed (often approximately every three years, but sometimes much more frequently), it is crucial to minimize sample attrition because the ability to compare responses over time is limited to cases for which there are interviews at each time point in the comparison. Obviously, if a survey has an excellent response rate in the first wave of interviewing, say, 80 per cent, but then loses half of the wave 1 respondents through failure to keep track of those who have moved or to retain the motivation to participate of those who have not, the effective response rate for wave 2 is 40 per cent.

For this reason, high quality panel surveys devote considerable effort to "panel maintenance," which includes periodically re-contacting respondents between waves, often with a small gift and a brief phone call or letter reminding them of the importance of the survey, and tracing those who have moved between waves. In order to do this, it is necessary to collect extensive information from each respondent at wave 1, including names, addresses, landline and cell phone numbers, and also, crucially, the names, addresses, and phone numbers of relatives, friends, and neighbors who are likely to know the whereabouts of respondents if they have moved. Doing this in such a way as to avoid refusals and the provision of deliberately incorrect information by respondents requires considerable interviewer skill. For a useful discussion of the panel maintenance procedures used in the *British Household Panel Survey (BHPS)* see Laurie, Smith, and Scott (1999).

INFORMATION COLLECTED

In addition to methods of data collection, there have been many new developments in the kind of data collected by population scientists. Three areas have seen major expansion in recent years: the integration of data on individuals or families with data on characteristics of their social contexts (neighborhoods and communities but also classrooms, schools, hospitals, and other institutions); the integration of survey data with geographic information obtained from satellite-generated data and images; and the addition of biological data to population surveys. In addition, there have been some new developments in scaling and measurement procedures, including the use of anchor points to calibrate self-reports of health status and of the method of "unfolding brackets" to obtain information on income and wealth.

Contextual variables and multilevel designs

It often has been hypothesized that characteristics of communities affect individual and household outcomes and behaviors (e.g., Brooks-Gunn et al. 1993; Sampson, Morenoff, and Earls 1999). One strategy for studying such possibilities using data from sample surveys is to construct summary indicators of community characteristics by averaging data from all respondents residing in each community. However, this strategy is sub-optimal both because not all community characteristics are aggregates of individual characteristics and, even for those that can be reasonably represented as aggregates, autocorrelation difficulties are likely to arise when both individual-level and aggregate-level variables derived from the same information are included in the same model. For this reason, an increasing number of surveys of households and individuals also include a "community characteristics" component, often obtained by interviewing community officials and staff representatives from schools, health, and other community facilities; compiling records; and using the interviewer as an informant to report on the type and condition of housing, the presence of infrastructural facilities, e.g., the location of the nearest school, hospital, library, bank etc.

The usual goal is to gather information that captures community infrastructures and levels of socioeconomic development. The measures commonly collected include community educational, health, and sanitation facilities, means of transportation, availability of electricity, and characteristics of the local economy. Most often the focus is on community characteristics at the time of the survey, but sometimes historical information is collected as well. From an analytic standpoint, data about the nature of the community in the past are often extremely useful since population researchers often want to know about the nature of the social environment at the time respondents were growing up. Thus, for example, in a study of adult outcomes it probably is more important to know whether a lower middle school existed in the community at the time respondents were age 14 than whether a lower middle school existed at the time of the survey.

Good examples of sample surveys that systematically collect community characteristics via separate "community questionnaires" include the *CHNS*, the *IFLS*, and the *Living Standards Measurement Study (LSMS)*, a collection of national and sub-national surveys conducted in over

30 developing nations. For good introductions to statistical procedures for simultaneously handling individual and community level data, see DiPrete and Forristal 1994; Mason 2001.

Spatial data

The increasing, and now widespread, availability of two types of geographic information has made it possible to incorporate spatial measurements of various kinds into population studies (see also the chapter in this volume by Logan, Zhang, and Xu). The first are measurements of latitude and longitude via the "global positioning system" (GPS). The second are satellite-based environmental measurements and images of the earth's surface, which can be tagged with geo-coordinates.

Geo-coordinates

The GPS is a satellite-based navigation system made up of a network of 24 satellites placed into orbit by the U.S. Department of Defense. The GPS originally was intended for military applications, but in the 1980s, the U.S. government made the system available without cost for civilian use. Each GPS satellite circles the earth twice a day in a very precise orbit and transmits signal information to earth. GPS receivers record the signals from three or more satellites and use triangulation, based on the elapsed time between transmission and receipt, to calculate the user's exact location. Relatively inexpensive GPS receivers (less than rmb 700) are accurate to within about 15 meters and various enhancements can substantially increase accuracy. In addition, training of users is necessary to ensure accurate measurements for research purposes (Spencer et al. 2003; Montana and Spencer 2004).

Location and distance are key components of many processes population scientists study. For example, distance to market affects the choices by peasants of what crop to produce; the availability of local schooling is an important component of decisions about whether to keep children in school (Fafchamps and Wahba 2006); the degree of isolation from transport hubs may affect decisions to migrate (Treiman and Lu 2009); and so on. In the past, researchers usually have relied on subjective reports by respondents about distances, but they are known to be highly inaccurate, with error often correlated with outcomes of interest (Gibson and McKenzie 2007). Using GPS technology, it is possible to precisely measure the location of each household and each other point of interest (schools, roads, markets, clinics, or other public services, nearest cities, etc.). For useful examples of GPS-based studies of distance as a barrier to services, see Entwisle et al. (1997); Rosero-Bixby (2004); Hong, Montana, and Mishra (2006); and Gibson et al. (2007).

In a different kind of application, Rosero-Bixby and Palloni (1998) studied the relationship between population growth and deforestation in Costa Rica, which has experienced an increase in population and a decline in forest cover over a 40 year period. Their paper addresses a central debate in research on population and environment—the extent to which rapid population growth is associated with the massive deforestation currently underway in many tropical nations. They compiled geocoded data from two population censuses and a series of land cover maps and used these to model the net impact of population growth on the 1973-83 probability of deforestation in about 31,000 parcels of 750 meters square, which were covered with forest at the beginning of the period. They show a strong connection between population potential (the size of the population in surrounding areas) at the beginning of the period and subsequent deforestation.

Other satellite data

Satellite-based data, including both images and measurements of properties of the earth's surface, are now available without cost to researchers for essentially all locations in the world, albeit with higher resolution in some places than others. Population scientists have made a variety of uses of such data, often combining satellite measurements with data collected through population surveys. For example, Frankenberg, McKee, and Thomas (2005) studied the impact of forest fires that blanketed the Indonesian islands of Kalimantan and Sumatra in 1997 on health outcomes by combining satellite measurements of aerosol levels in the atmosphere with longitudinal data for individuals from the *IFLS*. They showed that individuals who were exposed to haze from the fires experienced greater increases in difficulty with activities of daily living than did people living in nonhaze areas.

In another study, Frankenberg and her colleagues (Frankenberg et al. 2008) studied the effect of the 26 December 2004 tsunami on the mental health of residents of the Indonesian island of Sumatra, where 130,000 people died. Of interest here, they combined longitudinal sample survey data collected before and after the tsunami with community-level assessments of the extent of devastation caused by the tsunami. Devastation was measured by comparing satellite images from 17 December and 29 December for each of their 585 study sites. For each site, images for a .6 kilometer square centered on the study site were classified into three categories based on the degree to which the pre-tsunami ground cover visible in the first image had been replaced by bare earth in the second image. About 15 per cent of the areas were classified as heavily damaged (at least 20 per cent of pre-tsunami ground cover replaced by bare earth); about 35 per cent were classified as undamaged (no change in ground cover); and the remaining 45 per cent were classified as moderately damaged. They then showed that mental health was strongly affected by the level of damage, net of other factors.

A different application of remote sensing technology is to use night-light measurements to study population density. Zhuo and his colleagues (Zhuo et al. 2009) generated population density measurements for one kilometer squares covering all of China based on 1998 night-light data. Because these data are geo-coded, they can be integrated with ground-based data from sample surveys and, in principle, from the population census (provided that census data are made publicly available with sufficiently fine-grained geographic detail, e.g., at the township level or for even smaller units such as are available for the U.S.). See also the U.S. Department of Energy's global population database, LandScan, which provides annually updated population size estimates for 30-second by 30-second polygons (less than one kilometer square), using a regionally-specific model based on provincial-level census data and four geospatial input datasets (land cover, roads, degree of slope, and extent of night-time light) (see Dobson et al. 2000; Bhaduri et al. 2002; and http://www.ornl.gov/sci/landscan/).

Finally, there has been increasing interest in combining microdata (from surveys and censuses) with remotely sensed data on land cover and land use to study the impact of population on the environment, particularly global warming, with attendant changes in weather patterns; increased vulnerability to natural disasters; and the loss of biodiversity. The collection of papers edited by Fox et al. (2003) provides a good introduction to this topic. See also Liverman et al. (1998), Matthews (2003), Walsh and Welsh (2003), and Walsh, Evans, and Turner (2004).

Biological data

As population scientists have become increasingly interested in health, and health researchers have increasingly seen the advantage of population-based health studies, there has been a move to integrate questions about social and economic circumstances, health histories, and anthropometric and biometric measurements. One development, in particular, has been the collection of blood samples in general population surveys in order to do bioassays for the presence of various diseases and also of buccal cells to carry out genetic assays (see the chapter in this volume by Guo and Hu; for an introduction to the genetics underlying population processes, see the chapter by Gage and Fang). However, the collection of venous blood in field conditions is difficult and expensive. This difficulty has led to the development of procedures for conducting bioassays and genetic assays from dried blood spots (DBS), which are much easier to collect and manage (McDade, Williams, and Snodgrass 2007). The *IMHC* is an example of a Chinese study that has successfully collected DBS.

The basic procedure is that a respondent's finger is sterilized and then pricked using a spring-loaded device. The resulting blood droplets (usually four or five) are then collected on filter paper, dried, stored in plasticine bags, and within seven days shipped to the laboratory that will do the bioassays, where they are stored in medical freezers (-20^{o}C) until ready for analysis. In China, by law the person doing the blood collection must be medically trained. Thus, a nurse, doctor, or other health worker must accompany the interviewer to do the blood collection. This, of course, adds considerably to the cost and complexity of the fieldwork procedures; but it has the advantage of improving other biometric and anthropometric measurements because nurses and doctors are used to touching and physically manipulating patients and thus are able to put respondents at ease when they are taking body measurements, such as hip and waist circumference, procedures that may be awkward for interviewers, especially when dealing with the opposite sex.

Apart from blood spots, many population-based studies now include measurements of blood pressure, heart rate, lung capacity, hip, waist, and upper arm circumference, height and weight, and brief physical ability tests, e.g., the “timed chair stand” (how long it takes for a person to stand up and sit down five times), a grip strength test, and so on. Sometimes other biological specimens, such as buccal cell samples for DNA analysis and urine samples to study sexually transmitted diseases are collected as well. Examples of studies that include biological measurements are the U.S. *National Longitudinal Study of Adolescent Health (Add Health)*, the U.S. *National Health and Nutrition Examination Survey (NHANES),* the U.S. *Health and Retirement Study (HRS)*, *IFLS*, *MxFLS,* and *IMHC.* All these studies except for *IMHC* currently are available for analysis by the research community and the *IMHC* data will be made available for public use by the end of 2010. In addition, biological data will be collected by the *Chinese Health and Retirement Longitudinal Study (CHARLS)*, a survey currently in progress. These data, too, will be made available for analysis by the research community.

Two particularly interesting examples of the use of biological data in population research are a field experiment on the efficacy of iron supplements in Indonesia (Thomas et al. 2006) and a study of gene-environment interactions in the U.S. (Guo, Roettger, and Cai 2008).

Iron-deficiency anemia is a serious problem in many developing nations, increasing the risk of reduced immune response, delayed cognitive and physical development, and fatigue and reduced work capacity. In conjunction with their on-going panel study of the Indonesian population (*IFLS*), Thomas and his colleagues carried out an experiment in Java involving approximately 17,000 participants in which weekly iron supplements were administered to all members of half of the participating households selected at random (note that this is an example of a cluster randomized trial since clusters (households) rather than individuals are randomly assigned to the treatment or control condition). Measurements of hemoglobin, a marker for anemia, were taken before the supplements began and again after the first six months of intervention. In addition, changes in employment status and income were assessed. The results were striking. Men who were iron deficient prior to the intervention and who were provided supplements substantially increased their hemoglobin levels, increased the probability of working, reduced their hours of sleep, lost less time to illness, became more able to conduct physically arduous activities, and improved their emotional health. Among the self-employed, hourly earnings, and therefore monthly earnings, increased. Benefits for women were in the same direction but the effects were more muted.

Guo and his colleagues carried out an analysis of data collected as part of *Add Health*, which extracted DNA from buccal cells for a subsample of the 20,000 adolescents in the initial sample. They used data from about 1,100 males for whom measures of both DNA and delinquency were available to study the interaction of genetic and environmental influences on delinquency. They were able to show that three genetic polymorphisms were significant predictors of serious and violent delinquency when added to a social-control model of delinquency. But two of them interacted strongly with family processes, school processes, and friendship networks. Specifically, the risk of delinquency only increased among those who both had the genetic propensity and suffered from weak social control. This is one of the first studies to show genetic-environment interactions on the kinds of behaviors ordinarily studied by social scientists.

Improved measurement

Here we touch briefly on the value added by longitudinal data, discuss the advantages of multiple measurement of concepts, and conclude by considering two developments in scaling procedures, to improve the measurement of self-reported health (and other self-reports), and to improve the measurement of income and wealth.

Longitudinal data

We already have referred a number of times to longitudinal, or panel, data, in which the same individuals are queried at two or more points in time. Such data have many advantages, including in particular their ability to fix people's circumstances, conditions, behavior, and psychological states in time and hence to improve the possibility of being able rigorously to establish causal effects. But, as noted above, it is particularly important in panel studies to minimize panel attrition.

Even when it is not possible to mount panel studies, it sometimes is possible to generate over-time measurements by including retrospective questions in questionnaires, although in this case it is important to be sensitive to the possibility of recall error (Smith and Thomas 2003). One useful way to ask such questions is to collect "event history" rosters, for example, asking people to report each time they changed residence, each time they changed jobs, each time they married, and so on (for an example of the use of such rosters, see the questionnaires for *LHSCC* and *IMHC*).

Multiple measurement

It is well known that responses to survey questions, even factual questions such as the level of education completed by the respondent's father, are sometimes erroneous. Even random error has serious consequences for statistical analysis, generally reducing the strength of the association between variables and, if the degree of error differs between variables in a model, distorting the relative strength of different effects. Systematic error, for example a tendency to overstate the similarity between mothers and fathers or to "yea-say," giving positive responses regardless of the content of the question, is even worse, seriously biasing statistical estimates. A variety of methods for dealing with measurement error, particularly random error, are well summarized in Alwin's monograph on measurement error (Alwin 2007). Most require either *multiple measurements* or *multiple indicators*.

Multiple measurements are independent measurements of a single underlying dimension either in the same survey (for example, by asking attitude questions using a 7-point scale near the beginning of a questionnaire and an 11-point "feeling thermometer" near the end of the questionnaire, or by asking about both the years of school completed and the highest level of schooling attained, etc.) or, if the phenomenon can be assumed to be constant over time, as father's education usually is, by asking the same question in two or more waves of a panel study (see Smith and Stephenson 1979). Ganzeboom, Treiman, and Ultee (1991:292) discuss the advantages of multiple measurement and Ganzeboom (2005) provides an example of the advantage gained by measuring occupations in two different ways.

Multiple indicators are items that, at least in part reflect some underlying dimension, for example, asking people whether in the past year they had experienced unfair treatment in a variety of circumstances—by the police, by shopkeepers, etc.—in order to construct a “harassment” scale—see the questionnaire for the *IMHC* study.

Multiple measurements make it possible to compute measures of measurement reliability, which can then be used to correct multivariate models using “errors-in-variables regression” (Kmenta 1997:352-357; Draper and Smith 1998:89-91; Treiman 2009:258-261). Multiple indicators permit utilization of “multitrait-multimethod” models via confirmatory factor analysis procedures (Alwin 2007).

For variables that must be coded, such as narrative reports on occupational position, an alternative to multiple measurement is multiple coding by independent coders—although building multiple measurements into the questionnaire is preferable. If a set of narrative reports is independently assigned codes in the Chinese Standard Occupational Classification (CSCO) by two different coders, the codes can be converted into status scores (Ganzeboom, De Graaf, and Treiman 1992; Ganzeboom and Treiman 1996) and reliability estimates can then be computed. It is important to recognize that in order to fully exploit the power of either multiple measurement or multiple coding, it is necessary to apply these procedures to all the variables entering into one’s model. For example, to study intergenerational occupational mobility, it doesn’t help much to have multiple measurements of respondent’s occupation but not of father’s occupation.

Anchor points to improve scaling

As a reflection of the growing interest in health among population scientists, many recent social surveys have collected information on health status. However, due to cost and space constraints, the health measurement often is restricted to self-ratings of health status (e.g., “How would you rate your health—as excellent, good, fair, or poor?"). Although this simple indicator has been shown to be a powerful predictor of subsequent mortality (Idler and Benyamini 1997; Benyamini and Idler 1999), its utility is compromised by differences among population subgroups in the way they calibrate the scale (Murray et al. 2002; Sadana et al. 2002; Sen 2002; Thomas and Frankenberg 2002; Shmueli 2003; King et al. 2004; Lindeboom and van Doorslaer 2004). Self-perceptions may be different from true health status due to different definitions of health, different expectations for health, and different cognitive processes. Moreover, the standard for "good health" may differ dramatically between old and young respondents, between males and females, and between ethnic groups. Consequently, the cut points relating the observed categorical measure to a true underlying continuum of health status may vary systematically across social groups within a population or across populations, making direct comparisons of the observed distribution of self-reported health between groups problematic since they may not accord well with population differences in the underlying dimension. As a concrete example, the first author had to abandon a study of racial differences in health in South Africa when it turned out that Blacks reported better health than Whites in response to a self-rating question, despite the objectively poorer health of Blacks. Presumably, the life circumstances of Blacks predisposed them to expect a level of chronic health difficulties that Whites found unacceptable. A similar problem might arise in China in comparisons between peasants and urban residents.

The importance of establishing measurements of self-reported health (and other self-ratings) that permit valid cross-national and sub-national comparisons has led to the development of a technique for "anchoring" self-reports by using vignettes to calibrate responses (Tandon et al. 2002). The essence of the technique is, for example, to ask the respondent to rate his/her own health and then to rate the health of a set of vignettes, presented in random order, that are clearly ordered with respect to the level of health they represent. Tandon et al. (2002) illustrate this approach with an example concerned with the physical mobility of an elderly population. The respondent was asked “Overall in the past 30 days, how much difficulty did you have in moving around?” and five response categories were provided: “Extreme/cannot do,” “Severe difficulty,” “Moderate difficulty,” “Mild difficulty,” and “No difficulty.” The respondent was then asked to rate a set of vignettes on the same scale. The vignettes were written so as to be clearly ordered with respect the degree of mobility difficulty they imply but were presented to the respondent in random order. Here are three of five vignettes:

Vignette 1: [Paul] is an active athlete who runs long distance races of 20 kilometers twice a week and engages in soccer with no problems.

...

Vignette 3: [Rob] is able to walk distances of up to 200 meters without any problems but feels breathless after walking one kilometer or climbing up more than one flight of stairs. He has no problem with day-to-day physical activities, such as carrying food from the market.

...

Vignette 5: [Louis] is able to move his arms and legs, but requires assistance in standing up from a chair or walking around the house. Any bending is painful and lifting is impossible.
[PERHAPS CHANGE THE NAMES TO CHINESE NAMES WHEN TRANSLATING.]

Since each vignette describes a fixed state of health, any differences in responses to the vignette questions can be attributed to reporting heterogeneity by respondents. Thus, the problem becomes one of adjusting each individual's responses using information from their scoring of the vignettes. There are two ways to do this.

The simplest approach is to rank the vignettes and then to assign to the respondent's self-rating a score at the midpoint between the ranks for the vignettes just above and just below the respondent's self rating. For example, if the respondent gave him/herself a score higher than the score s/he gave Vignette 4 and lower than the score s/he gave Vignette 5, s/he would be scored 4.5 (Salomon, Tandon, and Murray 2004; King et al. 2004). A more efficient approach is the hierarchical ordered probit (HOPIT) model proposed by Tandon et al. (2002) and discussed in detail in King et al. (2004) and Jones et al. (2007). This model use the vignette questions to identify the reporting style/cut points of each respondent and then analyzes self-reported health after adjusting for cut point variations. This technique has been shown to largely eliminate effects of differences in the way people use rating scales.

Salomon, Tandon, and Murray (2004) provide a cross-national study of self-reported mobility levels using anchoring vignettes. They demonstrate that there was significant reporting heterogeneity with respect to age and country of residence, so that direct comparisons of the observed values of self-reported health would be quite misleading. Kapteyn, Smith, and van Soest (2007) use the same approach to compare self-reported work disability between the Netherlands and the United States, and find that the cross-national difference in reported work disability is largely due to cut point variations between the two populations. Finally, the anchoring-vignette approach has been used to study other subjective phenomena, for example job satisfaction (Kristensen and Johannson 2008) and political efficacy (King et al. 2004); see King's web site (http://gking.harvard.edu/vign/) for additional applications.

Unfolding brackets

A different sort of problem occurs with respect to the measurement of income and wealth. Many respondents simply refuse to provide values when asked. For example, about 30 per cent of respondents to the *HRS* refused to give estimates of the value of their assets, and those who refused differed in systematic ways from those who cooperated, yielding biased results. To cope with this problem, and a similar problem in the U.S. *Panel Study of Income Dynamics (PSID)*, researchers at the University of Michigan developed the technique of "unfolding brackets" (Juster et al. 2007:4-33). If the respondent fails to answer a question about his/her assets, either saying that s/he doesn't know or simply refusing, s/he is asked "Is the amount more than X, less than X, or what?" (where X is some designated amount); if s/he says, "more than X," s/he is asked, "Is it more than Y? (where Y is some amount greater than X); and so on. In this way, the amount of missing data can be greatly reduced and estimates can be obtained of the same order of precision as asking people to specify which category of a range of asset levels fits.

DATA SHARING

Earlier in this paper we several times have mentioned data "in the public domain," that is, freely available for use by the research community. The sharing of data is highly desirable, both because no one researcher or research group will have the imagination or energy to fully exploit the large and very expensive data sets that constitute the data base for much of contemporary demographic research and because independent validation and replication of results is the way science advances and error is corrected (Fienberg 1994). In the U.S., the two leading funders of scientific research, the National Science Foundation and the National Institutes of Health, require that all sample surveys and other systematic data collection efforts paid for by them must be deposited in an archive or otherwise made available[8] within a reasonable period after the data are collected (the convention is two years, but many studies make their data available as soon as they are analysis-ready). The Organization for Economic Co-operation and Development (OECD), a consortium of about 30 democratic nations, mainly in Europe, has adopted a similar policy (OECD 2007).

Unfortunately, China is a laggard in this respect. To our knowledge, there currently are only three data archives in China: the Chinese Humanistic and Social Science Data Archive at People's University (http://www.cssod.org/index.php); the China Survey Data Network at the China Center for Economic Research, Peking University (http://www.chinasurveycenter.org/CSDN_EN/default.aspx); and the Databank for China Studies (DBS) at the Chinese University of Hong Kong (http://www.usc.cuhk.edu.hk/databank.asp). But the holdings of each of these archives are quite limited. Moreover DBS charges a usage fee. Nor are there strong norms encouraging researchers to share their data. As a result, most of the studies of the Chinese population freely available for research use either were conducted by or in conjunction with foreign researchers or have been made available in foreign archives with English-language documentation, but are not readily available from Chinese sources or with Chinese-language documentation. It is our hope that as China increasingly participates in the international research community, this situation will change.

---

[8] The Interuniversity Consortium for Political and Social Research (ICPSR) at the University of Michigan is the leading U.S. archive for large sample surveys, but there also are many other archives, in the U.S. and elsewhere. For additional information on the holdings and terms of use of social science data archives, see http://www.sociosite.net/databases.php and http://www.sscnet.ucla.edu/issr/da/Home.Other%20Archives.htm. In addition, major studies increasingly have established their own web sites with provisions for downloading data; see the URLs listed in the References and in the Appendix.

REFERENCES

Books and articles

AAPOR. 2006. *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys.* 4th Edition. Lenexa, KS: American Association for Public Opinion Research.

The ALLHAT Officers and Coordinators for the ALLHAT Collaborative Research Group. 2002. "Major Outcomes in Moderately Hypercholesterolemic, Hypertensive Patients Randomized to Pravastatin vs Usual Care: the Antihypertensive and Lipid-Lowering Treatment to Prevent Heart Attack Trial (ALLHAT-LLT)." *Journal of the American Medical Association* 288(23):2998-3007.

Allison, Paul D. 2005. *Fixed Effects Regression Methods for Longitudinal Data Using SAS.* Cary, NC: SAS Institute Inc.

Alwin, Duane F. 2007. *Margins of Error: A Study of Reliability in Survey Measurement.* Hoboken, NJ: Wiley-Interscience.

Benyamini, Yael, and Ellen L. Idler. 1999. Community Studies Reporting Association between Self-Rated Health and Mortality." *Research on Aging* 21:392-401.

Bhaduri, Budhendra, Edward Bright, Phillip Coleman, and Jerome Dobson. 2002. "LandScan: Locating People is What Matters." *Geoinformatics* 5(2):34-37.

Brooks-Gunn, Jeanne, Greg J. Duncan, Pamela Kato Klebanov, and Naomi Sealand. 1993. "Do Neighborhoods Influence Child and Adolescent Development?" *American Journal of Sociology* 99(2):353-395.

Chaudhuri, Arijit, and Rahul Mukerjee. 1988. *Randomized Response: Theory and Techniques*. New York: Marcel Dekker.

Chaudhuri, A., M. Bose, and J. K. Ghosh. 2004. "An Application of Adaptive Sampling to Estimate Highly Localized Population Segments." *Journal of Statistical Planning and Inference* 121:175–189.

Cumming, R. G. 1990. "Is Probability Sampling Always Better? A Comparison of Results from a Quota and a Probability Sample Survey." *Community Health Studies* 14(2):132-7.

Darroch, John N., Stephen E. Fienberg, Gary F. V. Glonek, and Brian W. Junker. 1993. "A Three-Sample Multiple-Recapture Approach to Census Population Estimation with Heterogeneous Catchability." *Journal of the American Statistical Association* 88(423):1137-1148.

Davis, James Allan, Tom W. Smith, and Peter V. Marsden. 2009. *General Social Surveys, 1972-2008: Cumulative Codebook.* Chicago: National Opinion Research Center.

de Leeuw, Edith, and Wim de Heer. 2002. "Trends in Household Survey Nonresponse: A Longitudinal and International Comparison." Pp. 41-54 in *Survey Nonresponse*, edited by Robert M. Groves, Don A. Dillman, John L. Eltinge, and Roderick J. A. Little. New York: Wiley.

DiPrete, Thomas A., and Jerry D. Forristal. 1994. "Multilevel Models: Methods and Substance." *Annual Review of Sociology* 20:331-357.

Dobson, Jerome E., Edward A. Bright, Phillip R. Coleman, Richard C. Durfee, and Brian A. Worley. 2000. "LandScan: A Global Population Database for Estimating Populations at Risk." *Photogrammetric Engineering and Remote Sensing* 66(7):849-857.

Draper, Norman R., and Harry Smith, Jr. 1998. *Applied Regression Analysis*, 3rd ed. New York: Wiley.

Duncan, Greg J. 2008. "When to Promote, and When to Avoid, a Population Perspective." *Demography* 45(4):763–784.

Entwisle, Barbara, and William M. Mason. 1985. "Multilevel Effects of Socioeconomic Development and Family Planning Programs on Children Ever Born." *American Journal of Sociology* 919(3):616-649.

Entwisle, Barbara, Ronald R. Rindfuss, Stephen J. Walsh, Tom P. Evans, and Sara R. Curran. 1997. "Geographic Information Systems, Spatial Network Analysis, and Contraceptive Choice." *Demography* 34(2):171-187.

Erickson, Bonnie H. 1979. "Some Problems of Inference from Chain Data." *Sociological Methodology* 10:276-302.

Fafchamps, Marcel, and Jackline Wahba. 2006. "Child Labor, Urban Proximity, and Household Composition." *Journal of Development Economics* 79(2):374-397.

Feng, Xiaotian. 2007. "Gao Huishou Lv Geng Hao Ma? Dui Diaocha Huishou Lv de Ling Yizhong Renshi (Is High Response Rate Better? Another Understanding on Response Rates for Social Survey)." *Shehuixue Yanjiu* (*Sociological Research*) 3:121-135.

Fienberg, Stephen E. 1992. "Bibliography on Capture-recapture Modeling with Application to Census Undercount Adjustment." *Survey Methodology* 18:143-154.

Fienberg, Stephen E. 1994. "Sharing Statistical Data in the Biomedical and Health Sciences: Ethical, Institutional, Legal, and Professional Dimensions." *Annual Review of Public Health* 15:1-18.

Fisher, N., S. W. Turner, R. Pugh, and C. Taylor. 1994. "Estimated Numbers of Homeless and Homeless Mentally Ill People in North East Westminster by Using Capture-recapture Analysis." *British Medical Journal* 308(6920):27-30.

Fox, James A., and P. E. Tracy. 1986. *Randomized Response: A Method for Sensitive Surveys*. Quantitative Applications in the Social Sciences, No. 07-058. Beverly Hills, CA: Sage.

Fox, Jefferson, Ronald R. Rindfuss, Stephen J. Walsh, and Vinod Mishra (eds.). 2003. *People and the Environment: Approaches for Linking Household and Community Surveys to Remote Sensing and GIS*. Norwell, MA: Kluwer Academic Publishers.

Frankenberg, Elizabeth, Douglas McKee, and Duncan Thomas. 2005. "Health Consequences of Forest Fires in Indonesia." *Demography* 42(1):109-129.

Frankenberg, Elizabeth, Jed Friedman, Thomas Gillespie, Nicholas Ingwersen, Robert Pynoos, Iip Umar Rifai, Bondan Sikoki, Alan Steinberg, Cecep Sumantri, Wayan Suriastini, and Duncan Thomas. 2008. "Mental Health in Sumatra After the Tsunami." *American Journal of Public Health* 98(9):1671-1677.

Ganzeboom, Harry B. G. 2005. "On The Cost Of Being Crude: A Comparison Of Detailed And Coarse Occupational Coding." In *Methodological Aspects of Cross-National Research*, edited by Jürgen H.P. Hoffmeyer-Zlotnik. Mannheim: ZUMA-Nachrichten (Special Issue #10).

Ganzeboom, Harry B. G., and Donald J. Treiman. 1996. "Internationally Comparable Measures of Occupational Status for the 1988 International Standard Classification of Occupations." *Social Science Research* 25:201-239.

Ganzeboom, Harry B. G., Paul de Graaf, and Donald J. Treiman. 1992. "An International Scale of Occupational Status." *Social Science Research* 21:1-56.

Ganzeboom, Harry B. G., Donald J. Treiman, and Wout Ultee. 1991. "Comparative Intergenerational Stratification Research: Three Generations and Beyond." *Annual Review of Sociology* 17:277-302. Reprinted, 2002, as "Duibi Daijijian Fenceng Yanjiu: Qiansandai Jiqi Zhihou De Fenceng Yanjiu" in *Qing Hua She Hui Xue Ping Lun* (*Tsinghua Sociological Review*), Beijing, 3:151-194.

Gelman, Andrew. 2007. "Struggles with Survey Weighting and Regression Modeling." *Statistical Science* 22(2):153-164. [See also five comments and Gelman's rejoinder in the same issue.]

Gibson, John, and David McKenzie. 2007. "Using Global Positioning Systems in Household Surveys for Better Economics and Better Policy." *The World Bank Research Observer* 22(2):217-241.

Gibson, John, Geua Boe-Gibson, Halahingano Rohorua, and David McKenzie**.** 2007. "Efficient Remittance Services for Development in the Pacific." *Asia-Pacific Development Journal* 14(2):55-74.

Gile, Krista J., and Mark S. Handcock. 2008. "Respondent-Driven Sampling: An Assessment of Current Methodology." Paper presented at the Sunbelt XXVIII International Sunbelt Social Network Conference. St. Pete Beach, FL January 22-27.

Goyder, John. 1985. "Face-to-Face Interviews and Mail Questionnaires: The Net Difference in Response Rate." *Public Opinion Quarterly* 49:234-252.

Groves, Robert M. 2006. "Nonresponse Rates and Nonresponse Bias in Household Surveys." *Public Opinion Quarterly* 70(5, Special Issue):646–675.

Groves, Robert M., and Mick P. Couper. 1998. *Nonresponse in Household Interview Surveys.* New York: Wiley.

Groves, Robert M., Don A. Dillman, John L. Eltinge, and Roderick J. A. Little, eds. 2002. *Survey Nonresponse.* New York: Wiley.

Gu, Dongfeng, Kristi Reynolds, Xigui Wu, Jing Chen, Xiufang Duan, Robert F. Reynolds, Paul K. Whelton, and Jiang He. 2005. "Prevalence of the Metabolic Syndrome and Overweight among Adults in China." *Lancet* 365:1398-1405.

Guo, Guang, Michael E. Roettger, and Tianji Cai. 2008. "The Integration of Genetic Propensities into Social-Control Models of Delinquency and Violence among Male Youths." *American Sociological Review* 73(4):543-568.

Hansen, Morris H., William N. Hurwitz, and William G. Madow. 1993 (1953). *Sample Survey Methods and Theory. Volume I: Methods and Applications*. New York: Wiley.

Hao, Dahai. 2007. "Yingda Lv de Yiyi Ji Qita: Dui Zhongguo 'Gao' Diaocha Huishou Lv de Ling Yizhong Jiedu (The Significance of Response Rates and Others: Another Explanation to 'High' Response Rates for Chinese Social Survey)." *Shehuixue Yanjiu* (*Sociological Research*) 6:180-194.

Harris, Donald R., Mavis MacLean, Hazel Glenn, Sally Lloyd-Bostock, Paul Fenn, Peter Corfield, and Yvonne Brittan. 1984. *Compensation and Support for Illness and Injury.* Oxford: Clarendon.

Hayes, Richard J., and Lawrence H. Moulton. 2009. *Cluster Randomised Trials.* Boca Raton, FL: CRC Press.

Heckathorn, Douglas D. 1997. "Respondent-Driven Sampling: A New Approach to the Study of Hidden Populations." *Social Problems* 44(2):174–199.

Heckathorn, Douglas D. 2002. "Respondent-Driven Sampling II: Deriving Valid Estimates from Chain-Referral Samples of Hidden Populations." *Social Problems* 49(1):11–34.

Heckathorn, Douglas D. 2007. "Extensions of Respondent-Driven Sampling: Analyzing Continuous Variables and Controlling for Differential Recruitment." *Sociological Methodology* 37:151-208.

Heimer, Robert. 2005. "Critical Issues and Further Questions About Respondent-Driven Sampling: Comment on Ramirez-Valles, *et al*. (2005)." *AIDS and Behavior* 9(4):403-408.

Hong, Rathavuth, Livia Montana, and Vinod Mishra. 2006. "Family Planning Services Quality as a Determinant of Use of IUD in Egypt." *BMC Health Services Research* 6(1):79.

Idler, Ellen L., and Yael Benyamini. 1997. "Self-rated Health and Mortality: a Review of Twenty-seven Community Studies." *Journal of Health and Social Behavior* 38:21-37.

Kmenta, Jan 1997. *Elements of Econometrics*. 2nd ed. Ann Arbor: University of Michigan Press.

Jenkins, Paul, Charles Scheim, Jen-Ting Wang, Roberta Reed, and Allan Green. 2004. "Assessment of Coverage Rates and Bias Using Double Sampling Methodology." *Journal of Clinical Epidemiology* 57(2):123-130.

Johnston, L. G., M. Malekinejad, C. Kendall, I.M. Iuppa, and G.W. Rutherford. 2008. "Implementation Challenges to Using Respondent-driven Sampling Methodology for HIV Biological and Behavioral Surveillance: Field Experiences in International Settings." *AIDS and Behavior* 12:131–141.

Jones, Andrew M., Nigel Rice, Teresa Bago d'Uva, and Silvia Balia. 2007. *Applied Health Economics.* Routledge: London and New York.

Juster, F. Thomas, Honggao Cao, Mick P. Couper, Daniel Hill, Michael D. Hurd, Joseph P. Lupton, Michael Perry, and James P. Smith. 2007. "Enhancing the Quality of Data on the Measurement of Income and Wealth." Michigan Retirement Research Center Research Paper No. WP 2007-151, January 1. Available at SSRN: http://ssrn.com/abstract=1095815.

Kalton, Graham. 1983. "Models in the Practice of Survey Sampling." *International Statistics Review* 51:175-188.

Kapteyn, Arie, James P. Smith, and Arthur van Soest. 2007. "Vignettes and Self-Reports of Work Disability in the United States and the Netherlands." *American Economic Review* 97:461-473.

King, Gary, Christopher J. L. Murray, Joshua A. Salomon, and Ajay Tandon. 2004. "Enhancing the Validity and Cross-Cultural Comparability of Measurement in Survey Research." *American Political Science Review* 98(1):567-583.

Kristensen, Nicolai, and Edvard Johansson. 2008. "New Evidence on Cross-Country Differences in Job Satisfaction Using Anchoring Vignettes." *Labour Economics* 15:96-117.

Kumar, Naresh. 2007. "Spatial Sampling Design for a Demographic and Health Survey." *Population Research and Policy Review* 26:581–599.

Landry, Pierre F., and Mingming Shen. 2005. "Reaching Migrants in Survey Research: The Use of the Global Positioning System to Reduce Coverage Bias in China." *Political Analysis* 13(1):1-22.

Laurie, Heather, Rachel Smith, and Lynne Scott. 1999. "Strategies for Reducing Nonresponse in a Longitudinal Panel Survey." *Journal of Official Statistics* 15(2):269-282.

Lee, Chanam, Anne Vernez Moudon, and Jean-yves Pip Courbois. 2006. "Built Environment and Behavior: Spatial Sampling Using Parcel Data." *Annals of Epidemiology* 16(5):387–394.

Lensvelt-Mulders, Gerty J. L. M., Joop J. Hox, Peter G. M. van der Heijden, and Cora J. M. Maas. 2005. "Meta-Analysis of Randomized Response Research: Thirty-Five Years of Validation." *Sociological Methods and Research* 33(3):319-348.

Levy, Paul S., and Stanley Lemeshow. 2008. *Sampling of populations: Methods and Applications*. 4th edition. Hoboken, NJ: Wiley.

Lindeboom, Maarten, and Eddy van Doorslaer. 2004. "Cut-Point Shift and Index Shift in Self-Reported Health." *Journal of Health Economics* 23:1083-1099.

Liverman, Diana., Emilio F. Moran, Ronald R. Rindfuss, and Paul C. Stern (eds.). 1998. *People and Pixels: Linking Remote Sensing and Social Science*. (Committee on the Human Dimensions of Global Change, Commission on Behavioral and Social Sciences and Education, National Research Council.) Washington D.C.: National Academy Press.

Lu, Yao. 2007. "Educational Status of Temporary Migrant Children in China: Determinants and Regional Variations." *Asian and Pacific Migration Journal* 16:29-55.

Malekinejad, M., L. Johnston, C. Kendall, L. Kerr, M. Rifkin, and G. Rutherford. 2008. "Using Respondent-Driven Sampling Methodology for HIV Biological and Behavioral Surveillance in International Settings: A Systematic Review." *AIDS and Behavior* 12:105–130.

Mare, Robert D., and Vida Maralani. 2006. "The Intergenerational Effects of Changes in Women's Educational Attainments." *American Sociological Review* 71(4):542-564.

Mason, William M. 2001. "Multilevel Methods of Statistical Analysis." pp. 14,998-14,994 in *International Encyclopedia of the Social and Behavioral Sciences*, edited by Neil J. Smelser and Paul B. Baltes. Amsterdam: Elsevier Science.

Massey, Douglas S, Joaquin Arango, Graeme Hugo, Ali Kouaouci, Adela Pellegrino, and J. Edward Taylor. 1993. "Theories of International Migration: a Review and Appraisal." *Population and Development Review* 19(3):431-466.

Massey, Douglas S, Joaquin Arango, Graeme Hugo, Ali Kouaouci, Adela Pellegrino, and J. Edward Taylor. 1998. *Worlds in Motion: Understanding International Migration at the End of the Millennium*. Oxford: Clarendon Press.

Matthews, Stephen A. 2003. "GIS and Spatial Demography." GIS Resource Document 03-63 (GIS_RD_03-63), October, Population Research Institute, Pennsylvania State University [http://www.pop.psu.edu/gia-core/pdfs/gis_rd_03-63.pdf]

McCarty, Daniel J., Eugene S. Tull, Claudia S. Moy, C. Kent Kwoh, and Ronald E. LaPorte. 1993. "Ascertainment Corrected Rates: Applications of Capture-Recapture Methods." *International Journal of Epidemiology* 22(3):559-565.

McDade, Thomas W., Sharon Williams, and J. Josh Snodgrass. 2007. "What a Drop Can Do: Dried Blood Spots as a Minimally Invasive Method for Integrating Biomarkers into Population-based Research." *Demography* 44(4):899–925.

Merli, M. Giovanna, Jing Li, William Whipple Neely, and Rick Nordheim. 2008a. "Sensitivity of Respondent Driven Sampling to Assumptions." Unpublished paper, Center for Demography and Ecology, University of Wisconsin. (Last revised 8/8/08.)

Merli, M. Giovanna, William Whipple Neely, Feng Tian, Gao Ersheng, Tu Xiaowen, Gu Weiming, and Yang Yang. 2008b. "Sampling Female Sex Workers in Shanghai Using Respondent Driven Sampling." Paper presented at the Symposium on Health, Equity and Development, Shanghai Academy of Social Sciences, Shanghai, December 19-20.

Miguel, Edward, and Michael Kremer. 2004. "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities." *Econometrica* 72(1):159–217.

Montana, Livia, and John Spencer. 2004. "Incorporating Geographic Information into MEASURE Surveys: A Field Guide to GPS Data Collection." Chapel Hill: University of North Carolina, Carolina Population Center. [http://www.measuredhs.com/basicdoc/gps/DHS_GPS_Manual.pdf].

Moser, C. A. and G. Kalton. 1971. *Survey Methods in Social Investigation*. London: Heinemann Educational Books.

Muhib, Farzana B., Lillian S. Lin, Robin L. Miller, Wesley L. Ford, Wayne D. Johnson, and Philip J. Smith. 2001. "A Venue-based Method for Sampling Hard-to-reach Populations." *Public Health Reports* 116(Supplement 1):216-222.

Murray, Christopher J. L., Ajay Tandon, Joshua A. Salomon, Colin D. Mathers, and Ritu Sadana. 2002. "New Approaches to Enhance Cross-Population Comparability of Survey Results." Pp. 421-431 in *Summary Measures of Population Health: Concepts, Ethics, Measurement and Applications*, edited by Christopher J. L. Murray, Joshua A. Salomon, Colin D. Mathers and Alan D. Lopez. Geneva: World Health Organization: Geneva.

National Center for Health Statistics. 1987. *Linking the National Survey of Family Growth with the National Health Interview Survey* (Vital and Health Statistics, Series 2, No. 103). Washington, DC: U. S. Government Printing Office.

OECD [Organization for Economic Co-operation and Development]. 2007. *OECD Principles and Guidelines for Access to Research Data from Public Funding*. Paris: OECD.

Owen, L., A. McNeil, and C. Callum. 1998. "Trends in Smoking During Pregnancy in England, 1992-7: Quota Sampling Surveys." *British Medical Journal* 317(7160):728-730.

Peterson, James A., Heather Schacht Reisinger, Robert P. Schwartz, Shannon Gwin Mitchell, Sharon M. Kelly, Barry S. Brown, and Michael H. Agar. 2008. " Targeted Sampling in Drug Abuse Research: A Review and Case Study." *Field Methods* 20(2):155-170.

Ramirez-Valles, J., D. D. Heckathorn, R. Vasquez, R. M. Diaz, and R. T. Campbell. 2005. "From Networks to Populations: The Development and Application of Respondent-driven Sampling among IDUs and Latino Gay Men." *AIDS and Behavior* 9(4):387–402.

Roberts, Kenneth D. 1997. "China's Tidal Wave of Migrant Labor: What Can We Learn from Mexican Undocumented Migration to the United States?" *International Migration Review* 31(3):249-293.

Robinson, W. T., J. M. Risser, S. McGoy, A. B. Becker, H. Rehman, M. Jefferson, V. Griffin, M. Wolvertwon, and S. Tortu. 2006. "Recruiting Injection Drug Users: A Three-site Comparison of Results and Experiences with Respondent-drive and Targeted Sampling Procedures." *Journal of Urban Health* 83(7):29–38.

Rosero-Bixby, Luis. 2004. “Spatial Access to Health Care in Costa Rica and Its Equity: A GIS-Based Study.” *Social Science and Medicine* 58(7):1271–84.

Rosero-Bixby, Luis, and Alberto Palloni. 1998. “Population and Deforestation in Costa Rica.” *Population and Environment* 20(2):149-185.

Sadana, Ritu, Colin D. Mathers, Alan D. Lopez, Christopher J. L. Murray, and Kim Moesgarrd Iburg. 2002. “Comparative Analysis of More Than 50 Household Surveys on Health Status.” Pp. 370-386 in *Summary Measures of Population Health: Concepts, Ethics, Measurement and Applications*, edited by Christopher J. L. Murray, Joshua A. Salomon, Colin D. Mathers and Alan D. Lopez. Geneva: World Health Organization.

Salganik, Matthew J., and Douglas D. Heckathorn. 2004. “Sampling and Estimation in Hidden Populations Using Respondent-Driven Sampling.” *Sociological Methodology* 34:193–239.

Salomon, Joshua A., Ajay Tandon, and Christopher J. L. Murray. 2004. “Comparability of Self Rated Health: Cross Sectional Multi-country Survey Using Anchoring Vignettes.” *British Medical Journal* doi:10.1136/bmj.37963.691632.44 (published 23 January).

Sampson, Robert J., Jeffrey D. Morenoff, and Felton Earls. 1999. “Beyond Social Capital: Spatial Dynamics of Collective Efficacy for Children.” *American Sociological Review* 64(5):633-60.

Sen, Amartya. 2002. “Health: Perception versus Observation.” *British Medical Journal* 324:860-861.

Shmueli, Amir. 2003. “Socio-Economic and Demographic Variation in Health and in its Measures: The Issue of Reporting Heterogeneity.” *Social Science and Medicine* 57:125-134.

Singer, Eleanor. 2002. “The Use of Incentives to Reduce Nonresponse in Household Surveys.” Survey Methodology Program No. 051. Ann Arbor: University of Michigan, Institute for Social Research, Survey Research Center.

Singer, Eleanor, Nancy Gebler, Trivellore Raghunathan, John Van Hoewyk, and Katherine McGonagle. 1999. “The Effect of Incentives in Interviewer-Mediated Surveys.” *Journal of Official Statistics* 15:217-230.

Smith, James P., and Duncan Thomas. 2003. “Remembrances of Things Past: Test-Retest Reliability of Retrospective Migration Histories.” *Journal of the Royal Statistical Society (A)* 166(1):23-49.

Smith, Tom W., and C. Bruce Stephenson. 1979. "An Analysis of Test/Retest Experiments on the 1972, 1973, 1974, and 1978 General Social Surveys." GSS Methodological Report No. 8. Chicago: National Opinion Research Center, December. [http://publicdata.norc.org:41000/gss/DOCUMENTS/REPORTS/Methodological_Reports/MR008.pdf].

Song, Shige, Wei Wang, and Peifeng Hu. 2009. "Famine, Death, and Madness: Schizophrenia in Early Adulthood after Prenatal Exposure to the Chinese Great Leap Forward Famine." *Social Science and Medicine* 68(7):315-1321.

Spencer, John, Brian Frizzelle, Philip Page, and John Vogler. 2003. *Global Positioning System: A Field Guide for the Social Sciences*. Oxford: Blackwell.

Stark, Oded, 1992. *The Migration of Labour*. Cambridge, Mass: Blackwell, Oxford.

St Clair, D., M. Xu, P. Wang, Y. Yu, Y. Fang, and F. Zhang. 2005. "Rates of Adult Schizophrenia Following Prenatal Exposure to the Chinese Famine of 1959–1961." *Journal of the American Medical Association* 294(5):557-562.

Sudman, Seymour, Monroe G. Sirken, and Charles D. Cowan. 1988. "Sampling Rare and Elusive Populations." *Science,* New Series 240, No. 4855 (May 20):991-996.

Tandon, A., C., J. L. Murray, J. A. Salomon, and G. King. 2002. "Statistical Models for Enhancing Cross-Population Comparability." Global Programme on Evidence for Health Policy Discussion Paper No. 42. Geneva: World Health Organization.

Thomas, Duncan, and Elizabeth Frankenberg. 2002. "The Measurement and Interpretation of Health in Social Surveys." PP. 387-420 in *Summary Measures of Population Health: Concepts, Ethics, Measurement and Applications*, edited by Christopher J. L. Murray, Joshua A. Salomon, Colin D. Mathers, and Alan D. Lopez. Genvea: World Health Organization.

Thomas, Duncan, Elizabeth Frankenberg, Jed Friedman, Jean-Pierre Habicht, Mohammed Hakimi, Nicholas Ingwersen, Jaswadi, Nathan Jones, Christopher McKelvey, Gretel Pelto, Bondan Sikoki, Teresa Seeman, James P. Smith, Cecep Sumantri, Wayan Suriastini, and Siswanto Wilopo. 2006. "Causal Effect of Health on Labor Market Outcomes: Experimental Evidence." Population Working Paper PWP-CCPR-2006-070. California Center for Population Research, University of California, Los Angeles.

Thompson, Steven K. 1997. "Adaptive Sampling in Behavioral Surveys." Pp 296-319 in *The Validity of Self-Reported Drug Use: Improving the Accuracy of Survey Estimates*, edited by L. Harrison and A. Hughes. NIDA Research Monograph 167. Rockville, MD: National Institute of Drug Abuse.

Thompson, Steven K. and Linda M. Collins. 2002. "Adaptive Sampling in Research on Risk-related Behaviors." *Drug and Alcohol Dependence* 68:57-67.

Thompson, S. K., and O. Frank. 2000. " Model-based Estimation with Link-tracing Sampling Designs." *Survey Methodology* 26:87–98.

Thompson, Steven K., and George A. F. Seber. 1996. *Adaptive Sampling*. New York: Wiley.

Thompson, S. K., F. L. Ramsey, and G. A. Seber. 1992. "An Adaptive Procedure for Sampling Animal Populations." *Biometrics* 48:1195-1199.

Treiman, Donald J. 2009. *Quantitative Data Analysis: Doing Social Research to Test Ideas.* San Francisco: Jossey-Bass.

Treiman, Donald J., and Yao Lu. 2009. "The Effect of Internal Migration in China on Socioeconomic Outcomes and the Level of Living." Paper presented at the Annual Meeting of the Population Association of America. Detroit 29 April-2 May.

Treiman, Donald J., William Cumberland, Xilai Shi, Zhongdong Ma, and Shaoling Zhu. 1996. "A Sample Design for the Chinese Life History Survey." Appendix D.1.a of the documentation for the survey, "Life Histories and Social Change in Contemporary China." [Downloadable from the UCLA Social Science Data Archive: http://www.sscnet.ucla.edu/issr/da/].

Treiman, Donald J., William M. Mason, Yao Lu, Yi Pan, Yaqiang Qi, and Shige Song. 2006. "Observations on the Design and Implementation of Sample Surveys in China." *Social Transformations in Chinese Societies* 1:81-111.

Volz, E., and D.D. Heckathorn. 2008. "Probability Based Estimation Theory for Respondent Driven Sampling." *Journal of Official Statistics* 24 (1):79–97.

Walsh, Stephen J., and William F. Welsh. 2003. "Approaches for Linking People, Place, and Environment for Human Dimensions Research." *Geocarto International* 18(3):51-61.

Walsh, Stephen J., Tom P. Evans, and Billie L. Turner II. 2004. "Population-environment Interactions with an Emphasis on Land-use/Land-cover Dynamics and the Role of Technology." Pp. 491-519 in *Geography and Technology*, edited by Stanley D. Brunn, Susan L. Cutter, and J. W. Harrington, Jr. Norwell, MA: Kluwer Academic Publishers.

Warner, Stanley L. 1965. "Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias." *Journal of the American Statistical Association* 60:63-9.

Watters, John K., and Patrick Biernacki. 1989. "Targeted Sampling: Options for the Study of Hidden Populations." *Social Problems* 36(4):416-430.

Whyte, Martin King, and William L. Parish. 1984. *Urban Life in Contemporary China*. Chicago: University of Chicago Press.

Wooldridge, Jeffrey M. 2006. *Introductory Econometrics: A Modern Approach.* 3rd edition. Mason, OH: Thomson South-Western.

Wu, Xiaogang, and Donald J. Treiman. 2007. "Inequality and Equality under Chinese Socialism: The *Hukou* System and Intergenerational Occupational Mobility." *American Journal of Sociology* 113(2):415-445.

Zhuo, L., T. Ichinose, J. Zheng, J. Chen, P. J. Shi, and X. Li. 2009. "Modelling the Population Density of China at the Pixel Level Based on DMSP/OLS Non-radiance-calibrated Night-time Light Images." *International Journal of Remote Sensing* 30(4):1003-1018.

Non-Chinese Surveys Referred to in the Text (Chinese studies are described in the Appendix)

Add Health: [U.S.] National Longitudinal Study of Adolescent Health
http://www.cpc.unc.edu/projects/addhealth

BHPS: British Household Panel Survey
http://www.iser.essex.ac.uk/survey/bhps

GSS: [U.S.] General Social Survey
http://www.norc.org/GSS+Website/

HRS: [U.S.] Health and Retirement Study
http://hrsonline.isr.umich.edu/

IFLS: Indonesian Family Life Survey
http://www.rand.org/labor/FLS/IFLS/

LSMS: The Living Standards Measurement Study
http://go.worldbank.org/IFS9WG7EO0

MxFLS: Mexican Family Life Survey
http://www.ennvih-mxfls.org/

NHANES: [U.S.] National Health and Nutrition Examination Survey
http://www.cdc.gov/nchs/nhanes.htm

NHIS: [U.S.] National Health Interview Survey
http://www.cdc.gov/nchs/nhis.htm

NSDUH: [U.S.] National Survey on Drug Use and Health
https://nsduhweb.rti.org/

NSFG: [U.S.] National Survey of Family Growth
http://www.cdc.gov/nchs/NSFG.htm

PSID: [U.S.] Panel Study of Income Dynamics
http://psidonline.isr.umich.edu/

APPENDIX

Chinese Sample Surveys of Interest to Population Scientists

[Listed in the order of the year of (initial) data collection][9]

Date:	1982
Title:	*China Census of Population*
Sample size:	10,039,191
Sample specifications:	1% public use sample drawn from census records
Design:	Single cross section
Principle investigator(s):	National Bureau of Statistics
Available from:	IPUMS International (https://international.ipums.org/international/)
Language of documentation:	English
Additional comments:	Chinese language documentation is available from DCS.

Date:	1985
Title:	*China In-Depth Fertility Survey -- Phase I*
Sample size:	13,307
Sample specifications:	Married women < age 50 who are permanent household members in Hebei, Shaanxi and Shanghai Provinces.
Design:	Single cross-section. Individual, household, and community questionnaires.
Principle investigator(s):	National Bureau of Statistics
Available from:	DCS
Language of documentation:	English

[9]The surveys listed here are available without charge to members of the academic research community. We list the major surveys known to us. Additional surveys, often restricted to specific regions, cities, or subgroups of the population, may be obtained from the Chinese Humanistic and Social Science Data Archive (CSSOD) at People's University (http://www.cssod.org/index.php); the China Survey Data Network (CSDN) at the China Center for Economic Research, Peking University (http://www.chinasurveycenter.org/CSDN_EN/default.aspx); and the Databank for China Studies (DCS) at the Chinese University of Hong Kong (http://www.usc.cuhk.edu.hk/databank.asp). Note that the last of these archives charges user fees. The Chinese Population Information Network Statistical Database ("China PopIn") website (http://www.cpirc.org.cn/tjsj/tjsj_cy_3.asp) of the China Population and Development Research Center, a center within the National Population and Family Planning Commission, lists a number of surveys, mostly large scale surveys conducted by government agencies, but has no standardized provision for making them publicly available, although usage can be negotiated for a fee. The China Data Center at the University of Michigan (http://www.umich.edu/~iinet/chinadata/) makes available a variety of aggregate data files constructed from various Chinese censuses. However, this center also charges fees, which often are beyond the capabilities of academic researchers. Finally, the National Bureau of Statistics (NBS) maintains a website, the NBS Statistical Data Link (http://www.stats.gov.cn/tjsj/), which shows various aggregate statistics drawn from the censuses and surveys conducted by NBS.

Date:	1987
Title:	*China In-Depth Fertility Survey -- Phase II*
Sample size:	49,458 households; 39,210 women.
Sample specifications:	Married women < age 50 who are permanent household members in Gansu, Guangdong, Guizhou, Liaoning, and Shandong Provinces.
Design:	Single cross-section. Individual, household, and community questionnaires.
Principle investigator(s):	National Bureau of Statistics
Available from:	DCS
Language of documentation:	English

Date:	1988
Title:	*CHIP: Chinese Household Income Project*
Sample size:	10,258 (51,352 people) rural and 9,009 (31,827 people) urban households
Sample specifications:	National probability sample of households and individuals in rural and urban China
Design:	Cross section
Principle investigator(s):	Keith Griffin, and ZHAO Renwei
Available from:	ICPSR (http://www.icpsr.umich.edu/icpsrweb/ICPSR/), Study # 9836
Language:	English
Additional comments:	See also *CHIP 1995*.

Date:	1989, 1991, 1993, 1997, 2000, 2006
Title:	*CHNS: China Health and Nutrition Survey*
Sample size:	Approximately 4,400 households, yielding approximately 19,000 individuals
Sample specifications:	Multistage semi-probability sample of 9 provinces
Design:	Panel (rotating panel). Also community questionnaires.
Principle investigator(s):	Barry Popkin and ZHAI Fengying
Available from:	Carolina Population Center (http://www.cpc.unc.edu/projects/china)
Language of documentation:	English; Chinese

Date:	1990
Title:	*China Census of Population*
Sample size:	11,835,947
Sample specifications:	1% public use sample drawn from census records
Design:	Single cross-section
Principle investigator(s):	National Bureau of Statistics
Available from:	IPUMS International (https://international.ipums.org/international/)
Language of documentation:	English

Date:	1993
Title:	*China Housing Survey*
Sample size:	2,096
Sample specifications:	Probability sample of households in Shanghai and Tianjin
Design:	Single cross-section.
Principle investigator(s):	John R. Logan and BIAN Yanjie
Available from:	ICPSR (http://www.icpsr.umich.edu/icpsrweb/ICPSR/)
Language of documentation:	English

Date:	1994
Title:	*The State and Life Chances in Urban China 1949-1994*
Sample size:	4,074
Sample specifications:	National probability sample of people age 25-65 residing in urban areas in China.
Design:	Single cross-section
Principle investigator(s):	ZHOU Xueguang and Phyllis Moen
Available from:	ICPSR (http://www.icpsr.umich.edu/icpsrweb/ICPSR/)
Language of documentation:	English

Date:	1995
Title:	*CHIP: Chinese Household Income Project*
Sample size:	7,998 (34,739) rural and 6,931 (21,694) urban households
Sample specifications:	National probability sample of households and individuals in rural and urban China
Design:	Cross section
Principle investigator(s):	Carl Riskin, Renwei Zhao, and Shi Li
Available from:	ICPSR (http://www.icpsr.umich.edu/icpsrweb/ICPSR/), Study # 3012
Language:	English
Additional comments:	See also *CHIP 1988.*

Date:	1996
Title:	*LHSCCC*: *Life Histories and Social Change in Contemporary China*
Sample size:	6,090
Sample specifications:	National probability sample of people age 20-69
Design:	Single cross-section.
Principle investigator(s):	Donald J. Treiman, Andrew Walder, and LI Qiang
Available from:	UCLA SSDA (http://www.sscnet.ucla.edu/issr/da/).
Language of documentation:	English, Chinese.
Additional comments:	Also available from the Chinese Humanistic and Social Science Data Archive [http://www.cssod.org/index.php]. The documentation is not as complete, but is in Chinese.

Date:	1998, 2000, 2002, 2005, 2008
Title:	*CLHLS: Chinese Longitudinal Healthy Longevity Survey*
Sample size:	Initial wave: 9,073 people age 80+; see project web site for other sample sizes.
Sample specifications:	People age 80 and older in 1998. Also, a comparison group age 65-79 and other special samples.
Design:	Panel, with replacements for deceased
Principle investigator(s):	ZENG Yi
Available from:	Duke Aging Center (http://www.geri.duke.edu/china_study/index.htm)
Language of documentation:	English

Date:	1999-2000
Title:	*CHFLS: Chinese Health and Family Life Survey*
Sample size:	3,426
Sample specifications:	National probability sample of people age 20-64, with an oversample of high STD areas.
Design:	Single cross-section
Principle investigator(s):	William Parish and Edward O. Laumann
Available from:	http://www.src.uchicago.edu/prc/chfls.php
Language of documentation:	English; Chinese

Date:	2003, 2004, 2005, 2006, 2008
Title:	*CGSS: Chinese General Social Survey*
Sample size:	2003: 5,894 (urban only); remaining years, approximately 10,000 each.
Sample specifications:	National probability sample of people age 18-69
Design:	Repeated cross sections
Principle investigator(s):	BIAN Yanjie, LI Lulu
Language of documentation:	Chinese and English
Available from:	CSSOD
Additional comments:	Currently only the 2003, 2005, and 2006 data are available for public distribution.

Date:	2008
Title:	*IMHC: Internal Migration and Health in China*
Sample size:	3,000
Design:	Single cross-section
Sample specifications:	National probability sample of people age 18-64, with an over-sample of high in-migration areas.
Principle investigator(s):	Donald J. Treiman, William M. Mason, SONG Shige
Available from:	UCLA SSDA (http://www.sscnet.ucla.edu/issr/da/)
Language of documentation:	English and Chinese. Currently, the Chinese documentation is limited.
Additional comments:	(1) Currently only the documentation is available. The data will be released for public use in late 2010. (2) Conditional on securing funding, this survey will be converted to a panel design.

Date:	2010-2020
Title:	*CFPS: Chinese Family Panel Studies*
Sample size:	Approximately 16,000
Design:	Panel, with a new wave of interviews in the spring of each year.
Sample specifications:	Nearly national probability sample (excluding Tibet, Qinghai, Xinjiang, Ningxia, Inner Mongolia, and Hainan).
Principle investigator(s):	QIU Zeqi
Available from:	Institute of Social Science Survey, Peking University (http://www.isss.edu.cn/)
Language of documentation:	Chinese and English.
Additional comments:	A pilot study (N = 2,400) was conducted in 2009 in Beijing, Guangdong, and Shanghai.

Date:	2011-
Title:	*CHARLS: China Health and Retirement Longitudinal Study*
Sample size:	Approximately 10,000 households, 17,000 individuals.
Design:	Panel, with a new wave of interviews every two years.
Sample specifications:	National probability sample of adults age 45 and older.
Principle investigator(s):	ZHAO Yaohui, LIN Justin Yifu, John Strauss, Albert Park
Available from:	National School of Development (China Center for Economic Research) at Peking University (http://charls.ccer.edu.cn/charls/)
Language of documentation:	English and Chinese.
Additional comments:	A pilot study (N = 1,570 households, 2,685 individuals) was conducted in 2008 in Zhejiang and Gansu provinces.