Lessons from a Pilot Study for a National Probability Sample Survey of Chinese Adults Focusing on Internal Migration

Donald J. Treiman, Yao Lu, Yi Pan, Yaqiang Qi, Shige Song, and William Mason
(all California Center for Population Research and Dept. of Sociology, UCLA, except Song, who is at the Chinese Academy of Social Sciences)

EXTENDED ABSTRACT

Problematic aspects of doing survey research in China

Doing survey research in China has become increasingly problematic, for several reasons.

- Sampling from population registers, the conventional method in China, is no longer viable, for two reasons. First, there has been massive internal migration, particularly from rural to urban areas, with some estimates placing the size of the "floating population" (people living in locales other than where they are formally registered) as high as 150 million, more than 12% of the population. Many of these migrants (perhaps half) fail to temporarily register where they are living and working. Second, urban China has experienced a building boom (23% of the urban population enumerated in the 2000 census lived in buildings constructed since 1996). But many people who move to new housing in the same city fail to register at their new locale.

- Many new residential buildings restrict access (akin to gated communities and doorman-controlled buildings in the U.S.). Given the high percentage of the population living in new residential buildings, this is a non-trivial problem.

- The urban population is becoming increasingly unwilling to cooperate with interviewers, because of a combination of fear of crime, which makes people unwilling to open the door to strangers, and a breakdown in civility, with people—especially the *nouveau riche* and new middle class—increasingly inclined to ask, "what's in it for me?".

- There has been a breakdown in the power of local authorities, who in the past could be relied upon to secure the cooperation of residents.

- There has been a breakdown in the willingness of authorities at each level to do the bidding of higher-level authorities, so that securing the cooperation of high level authorities is no guarantee of local cooperation, as it used to be until fairly recently.

- There have been changes in small-area geography since 2000, which make multi-stage probability sampling difficult.

- The 2000 census, which substantially under-counted migrants, is rapidly becoming severely deficient as a sampling frame because of the massive internal migration.

Our pilot study

As an alternative to sampling households or individuals from population registers, we carried out a small survey (N=1,059) in Fall 2003 and Spring 2004 designed to test the feasibility of doing complete enumeration of the adult population, sampling individuals from the enumeration list, and conducting interviews with the sampled individuals. We purposively chose 12 small areas—two areas corresponding to each of six neighborhood types. (Because another aspect of our analysis, reported in a separate paper, was to study social responses to the SARS epidemic, for each of the six neighborhood types we chose a matched pair of neighborhoods from a high- and low-SARS area.) The neighborhood types were specifically chosen to cover the range of problems we were likely to encounter in carrying out complete enumerations. They consisted of urban-fringe villages with workshops staffed mainly by in-migrants, urban-fringe "bedroom villages" housing migrants who worked in cities, rural-to-urban transitional areas, factory dormitories, low-income urban areas, and middle-to-high-income urban areas. To minimize costs, we conducted all of our interviews in small areas in or close to four large cities: Beijing and Guangzhou (the high SARS areas) and Chengdu and Suzhou (the low SARS areas). For each of the 12 areas our target was 100 completed interviews. We achieved our target in all but the middle-to-high-income areas, which consisted entirely of restricted-access buildings—in Beijing our interviewers more or less failed completely, obtaining only six interviews; in Suzhou we were more successful, but still obtained only 58 interviews.

The lowest administrative unit in China is "villages" on the rural side and "neighborhoods" on the urban side (often described as "village committees" and "neighborhood committees"). Subsequent to the 2000 census, some urban areas were reorganized into "community committees" approximately three times as large as "neighborhood committees." Because we had a small enumeration team for the pilot study, we enumerated only portions of each selected village or neighborhood.

To conduct the enumeration, our enumerators first drew a sketch map of the area and then, with the assistance of a locally-knowledgeable person (usually a neighborhood committee official), walked through the area drawing successively more detailed maps (e.g., one identifying buildings and another set showing the layout of units in each building), and recording each "household" on an enumeration roster together with the number of adults residing in the household, estimated as "2" if was not possible to obtain the information from a resident. The enumerator also recorded the cumulative sum of the number of adults in each household. After the enumeration was completed, the fieldwork supervisor drew from the enumeration roster a random sample of 100 individuals to be interviewed and also a reserve sample to substitute for "never-at-homes," refusals, etc., with the goal of achieving 100 interviews per area. The final step was for an interviewers to go to each of the selected households, do a complete household enumeration, randomly select one adult to be interviewed (or more, if the household fell into the sample more than once), and complete the interview, returning if necessary if the chosen individual(s) were not available at the moment of selection.

Lessons learned

- Enumeration is clearly superior to relying on registration lists. We estimate that overall 19% of residents of households in which we conducted interviews would have been missed if we had relied on registration lists (that is, lists both of permanent residents and of those holding temporary residence permits), but in our factory dormitories and high in-migration villages this figure rises above one-third. Landry and Shen (2004), reporting on an enumeration-based probability sample survey of Beijing, show an even higher estimate—only 55% of their respondents would be have been captured from registration lists. Another 20% had changed residence within Beijing and 25% were unregistered in-migrants to Beijing.

- Enumeration of "addresses" or "doors" is inadequate. The conventional practice of those data collection agencies in China that have begun to do enumeration is to count "addresses" or "doors" rather than knocking on doors to ascertain the number of eligible residents (e.g., the number of adults). But this creates two problems. First, the number of persons per household is highly variable in China, especially when "collective households" (dormitories, rooms shared by migrants, etc.) are included. Second, there often are "doors behind doors," that is, separate households that are completely invisible unless an inquiry is made. Getting good information on the number of adults per household will increase the cost of enumeration by a factor of two to three, but in our judgement is worth the added cost.

- These problems raise the vexing question of what is a household. "Households" need to be carefully defined so that enumerators have clear rules to follow. Our practice was to define as separate households any places that opened into public space, but to include rooms rented out to tenants that open into the living space of another household as part of that household.

- It is crucial to enumerate "unconventional" living spaces. In urban China, many people sleep in their shops, in rooms behind restaurants, in entryways to residential and commercial buildings, in rooms in large commercial buildings, in small houses on the edge of university campuses, in lofts above garment workshops, in railway stations, in factory dormitories, in temporary dormitories erected on construction sites, etc. In rural areas there are similar "unconventional" dwellings: road construction crews living in tents, staff of power plants and dams, boatmen and fishermen living on their boats, etc. If these areas are not enumerated, an unknown but nontrivial faction of the population of China will be missed.

- It is important to adequately identify enumerated households so that interviewers can find those households chosen for interviewing. In China, many houses have no "address" in the U.S. sense. Thus, a location description is necessary. It may help to provide GPS (Global Positioning System) coordinates in addition to a location description and (as our enumerators often did) the name of the householder. The interviewer can then confirm that s/he has the correct location by checking the GPS coordinates. Small handheld GPS locators are becoming increasingly inexpensive and increasingly accurate.

- Cooperation from authorities from the top to the bottom of the administrative hierarchy may help to penetrate restricted-access buildings and certainly will minimize harassment by the police and/or local residents. Being accompanied by a local guide often helps.

- The shift from "neighborhood committees" to much larger "community committees" (and the growth in size of many villages near cities) make enumeration of small areas defined by the census impractical. Procedures need to be established to sample in a probabilistic way areas smaller than these administrative units.

A proposed design

On the basis of the experience gained in our pilot study, we now believe that the optimal design for a national probability sample survey in China is a hybrid of census-based sampling down to local administrative areas and GPS sampling of subunits of local administrative areas, followed by complete enumeration of individuals within each subunit and random sample selection from the enumeration list.

China is divided into four administrative levels (ignoring the seldom-used category, "prefecture," which stands just below the highest level). On the rural side, these consist of provinces, counties, townships, and villages, with each level subdivided into units at the next lower level. The corresponding categories on the urban side are provinces, county-level cities and districts of larger cities, "street committees," and "neighborhood committees." As noted above, in some parts of China neighborhood committees have been replaced over the past two years by larger "community committees." A crucial point is that all of urban China is divided into "street committees" but not all territory within "street committees" is covered by "neighborhood committees;" commercial buildings, factories, hospitals, universities, and other institutions, as well as roadways, parks, etc. are excluded. We believe the same distinction holds in rural China as well.

Despite having coverage gaps and becoming increasingly out-of-date, there is no viable alternative to the census for establishing a sampling frame. We therefore propose to sample counties and county-level urban units with probability proportionate to size (on the basis of updated estimates of their population size if such are available); to then visit the statistical office of every selected county-level unit to obtain up-to-date administrative estimates of the population size of each township and street committee; to sample these units with probability proportionate to size; to obtain maps of each selected township and street committee and divide these maps into grids small enough to be completely enumerated (e.g., 100 meter squares); to randomly sample these grids or, if sufficient information regarding population density can be obtained, to sample them with probability proportionate to size; to do a complete enumeration of eligible respondents (e.g., adults) within each selected grid; and to randomly sample individuals to be interviewed from the completed enumeration roster.

The selection of grids to be enumerated and their location in space relies on GPS technology, which has been successfully implemented in China (see Landry and Shen 2004). The advantage of a GPS-based approach to identify very small areas within townships and "street committees" is threefold:

First, it completely bypasses the problem of post-2000 boundary changes and changes in population density (if necessary, post-weighting can be carried out to correct for differences in the number of inhabitants in the enumerated grids). Second, it creates a systematic basis for sampling sub-areas within the lowest set of (more or less) time-invariant administrative units: townships and street committees. Third, it permits the specification of sub-areas small enough to be completely enumerated within a short period of time. By saturating a very small area with enumerators, exploring *every* building and contacting *every* household, researchers can minimize the risk of missing households (e.g., rooms built on roofs, in courtyards, people sleeping in commercial space or institutional buildings, etc.) and can minimize bias in the selection of respondents by sampling *individuals* at random rather than sampling households, which vary widely in population size.

Reference

Landry, Pierre F., and Mingming Shen. 2004. "Reaching Migrants in Survey Research: the Use of the Global Positioning System to Reduce Coverage Bias in China." New Haven: Yale University, Dept. of Political Science, unpublished paper.