

**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.**

## OBSERVATIONS ON THE DESIGN AND IMPLEMENTATION OF SAMPLE SURVEYS IN CHINA\*

*Donald J. Treiman, William M. Mason, Lu Yao,  
Pan Yi, Qi Yaqiang and Song Shige*

### ABSTRACT

Surveys in China conventionally sample from local area residential registers, which until recently have been of sufficient accuracy to function as *de facto* population registers. Due to a combination of large scale internal migration and massive replacement of housing in urban areas, a large fraction of the population currently does not live where registered. Consequently, recent surveys whose sampling procedures rely exclusively on residential registers provide a biased representation of the population due to exclusion of unregistered local residents. We report conclusions from, and observations related to, a pilot study designed to test the feasibility and effectiveness of a sampling method that does not depend on residential registers. In the pilot study we (1) purposely selected small areas; (2) enumerated the small areas; (3) randomly sampled individuals from the enumeration lists; and (4) interviewed sampled individuals. This approach substantially reduced the under-enumeration problem. As implemented, however, its point of departure required previously selected small areas. We describe an extension designed to achieve full coverage of the population of China through sampling of small areas as the penultimate stage of a multi-stage design.

### INTRODUCTION

Early in the post-Mao era in China, sample surveys became important sources of information for understanding Chinese society (Lavelly, Lee, and Wang 1990). Initially conducted under the auspices of ministerial level government organizations, by the late 1980s sample surveys were beginning to be carried out by Chinese and foreign scholars, often working together. In the early 1980s, low rates of internal geographic mobility and a government capable of mobilizing down to the lowest administrative level meant that surveys, when

---

\* This research was supported by U.S. NICHD grants 1R03HD40826 and R244HD41022. The first two authors share equal responsibility for this paper. A version of this paper was presented at the annual meeting of the Population Association of America, 2 April 2005, Philadelphia. We thank the anonymous referees for their thoughtful comments.

well designed, could achieve nearly complete coverage of the population, with response rates well over 90 percent. However, changes in Chinese society concomitant with and in consequence of the move toward a market economy have made it more difficult to conduct surveys than in the recent past, and have rendered results based on many contemporary surveys much more problematic.

This article reviews these developments. We first consider how increased internal geographic mobility has undercut the sampling method most often used in Chinese surveys. We then draw lessons and related observations from a recently completed pilot study for a national probability sample survey of Chinese adults. This multipurpose study was explicitly designed to test a sampling strategy that is new in the Chinese context. Because the pilot study used purposely selected small areas as first-stage units, the sampling design was incomplete. We thus end by describing an extension of the sampling strategy that is designed to achieve unbiased, full coverage of the population of China.

#### ELEVATED INTERNAL MIGRATION RATES AND SAMPLE SURVEYS

Since 1955 China has had an internal passport (*hukou*) system in which individuals are initially registered in their mother's locales. In rural areas, *hukou* locale is defined as a "village committee" (*cunweihui*). In urban areas, *hukou* locale is defined as a "neighborhood committee" (*juweihui*). Individuals are also assigned to agricultural or non-agricultural status, again following the mother's registration status. Except under special circumstances, mainly the achievement of secondary technical or tertiary schooling, it was and is extremely difficult to change from agricultural to non-agricultural status (Wu and Treiman 2004a). Even within the rural and urban sectors, changing registration from one place to another is not automatic and is sometimes difficult.

The sampling method conventionally used in past Chinese surveys, apart from any sampling process involving higher level aggregates, was to (i) select village and neighborhood committees, (ii) within each area, sample households from *hukou* lists—which functioned as *de facto* population registers—maintained by the village and neighborhood committees, and (iii) randomly select adults in chosen households. This method worked well when internal migration was limited,

since almost everyone lived where they were registered and also lived in family units. More recently, it has become less satisfactory.

Beginning in the early 1980s and continuing to the present, massive numbers of people have migrated to seek or take jobs, mainly but not entirely from rural to urban areas (Goldstein and Goldstein 1991; Davin 1999; Liang 2001; Liang and Ma 2004). Much of this migration currently is not, and has not been, state sanctioned. Except under limited circumstances, migrants are not permitted to change their place of registration (Yang 1993; Chan, Ta, and Yang 1999; Wang, Zuo, and Ruan 2002; Wu and Treiman 2004a). The problem is substantial. Some estimates put the current stock of informal migrants (those who have changed residence without changing registration), known in China as the “floating population”, as high as 144 million people, which amounts to approximately 12 percent of the population and an even larger fraction of the adult population (Liang and Ma 2004). Owing to the difficulty of tracing individuals who are not currently in residence at their place of registration, it follows that survey samples based on *hukou* lists are based on a subset of the population. Furthermore, this subset, which consists primarily of geographically stable individuals, is by definition anything but random.

It is true that internal migrants in China are encouraged to register as “temporary residents” at their place of destination (Chang and Zhang 1999; Solinger 1999a). Might it be possible to create registration-based sampling frames that achieve complete population coverage by augmenting permanent *hukou* lists with registers of temporary residents? Available evidence suggests otherwise. Apparently, a majority of migrants fail to register as temporary residents. For example, the Ministry of Public Security estimated that, in 1997, 62 percent of temporary migrants throughout China had failed to register (Chan and Zhang 1999). This is likely due both to variability in the vigor of local authorities and the reticence of informal migrants themselves, especially short-term migrants (see also Solinger 1999b; U.S. Consulate Guangzhou 2001).<sup>1</sup> Furthermore, it is probable that those who do

<sup>1</sup> Many cities do not have an established system for the registration of temporary residents and, for those that do, it cannot be assumed that the registers are up-to-date. In cities with established temporary migrant registration systems, migrants are thought to be reluctant to register, because they may be required to pay “fees” and may be monitored by local governments after registration.

register as temporary residents are a nonrandom subset of migrants. In sum, contemporary surveys whose sampling frames rely on lists of permanent and temporary residents will achieve incomplete and biased coverage of the population due to the under-sampling of migrants.

A second problem with the use of *hukou* lists is that urban China has experienced a massive building boom. Twenty-three percent of the urban population enumerated in the 2000 census lived in buildings constructed since 1996.<sup>2</sup> Because many people (the percentage is unknown) who move to new housing in the same city fail to change their registration from their old to their new neighborhood, these movers are excluded from *hukou*-based samples unless they can be traced, which is difficult and expensive.

Third, even apart from the failure of individuals to record their movements between neighborhoods, it can no longer be assumed that *hukou* lists are kept up-to-date with former administrative vigor. The sheer level of internal migration may be overwhelming, other forms of social control besides *hukou* registration probably have relaxed, and the anonymity that comes with increased size may also play a role. In particular, neighborhood committees have begun to be replaced by “community committees” (*sheweihui*) composed of several neighborhood committees or, in high rise buildings, by “residents’ committees” (*yezhuweiyuanhui*). With larger aggregations, it may be more difficult to keep the *hukou* lists up to date.

Fourth, based on our pilot study and discussions with others who have recently conducted surveys, it appears that, especially for residents’ committees in high-rise buildings, researchers are beginning to experience denial of access to *hukou* lists.

In sum, sample designs that rely on registration lists miss a substantial, nonrandom portion of the population and encounter related problems as well. It is unlikely that this conclusion will require modification in the near future.

#### PILOT STUDY

If *hukou* lists are incomplete, on what basis can a probability sample of the population be drawn? In societies without population regis-

---

<sup>2</sup> Estimates from the 2000 Census of China are from computations made by the authors from a 1:1,000 micro-data sample of households and individuals.

tration systems, the problem of sampling from a population can be addressed through the use of enumeration in small areas. For example, in the United States, national probability samples typically are created by using information from the most recent census to sample small areas known as block groups, which are subsets of census tracts. These areas typically contain several hundred households. When block groups have been selected for a particular survey, team members enumerate all households within the chosen block groups. The result of the canvassing is a complete list of the local area households, from which a sample of households is drawn at random. An interviewer then visits each sampled household, makes a list of eligible respondents, and on the spot samples one or more respondents to be interviewed.<sup>3</sup> Interviews are conducted either at the time of the first visit or during a return visit.

Although the Chinese National Bureau of Statistics employs enumeration in its decennial censuses and presumably in its mini-censuses, the use of enumeration in sample surveys conducted by academics or market research firms in China is unusual. For this reason, and given the inadequacy of samples based on *hukou* lists, we mounted a pilot study to determine the feasibility of enumerating small areas in (especially urban and peri-urban) China. The success of such a project could not be taken for granted, for several reasons:

1. Due to the large number of informal migrants, many neighborhoods in urban China have become extremely crowded, with apartment units subdivided and extra rooms built in courtyards, on rooftops, and in storage areas, to rent out to migrants. Enumeration in such areas is difficult.
2. Many people live in nonstandard spaces, such as sleeping spaces in shops, lofts over workshops, and tents and temporary dormitory buildings on construction sites. Enumeration in these places is also difficult.
3. The number of restricted-access residential sites is on the rise. Examples include high-rise apartment buildings, gated communities, and factory, construction-site, and other dormitories.
4. Anecdotaly, there is increasing suspicion of strangers, especially strangers asking questions.

---

<sup>3</sup> To insure the integrity of the probability sampling process, eligible respondents are randomly sampled using Kish-table or comparable procedures (Kish 1995).

In early 2003 we were poised to go into the field with a small survey designed to assess the practical difficulties of population enumeration for sample surveys in China. When the SARS epidemic flared, it became necessary to delay the study. We altered the design and content of the survey to allow study of social responses to the epidemic, while retaining the original intent to study the feasibility of enumeration. Then, in Fall 2003 and Spring 2004, we conducted a small survey ( $N = 1,059$ ).<sup>4</sup> We purposively chose two instances of six types of residential areas, for a total of 12 places.<sup>5</sup> An earlier pretest had suggested the need to consider various types of residential areas in order to address the full range of problems we were likely to encounter during enumeration. We identified the types of areas as (i) urban-fringe villages with workshops staffed mainly by migrants; (ii) urban-fringe “bedroom villages” housing migrants who work in cities; (iii) rural-to-urban transitional areas, possibly with overlapping civil jurisdictions; (iv) factory dormitories; (v) low-income urban areas; and (vi) middle-to-high-income urban areas. To minimize costs, all interviews were conducted in small areas in or close to four large cities: Beijing and Guangzhou (high SARS areas) and Chengdu and Suzhou (low SARS areas). Table 1 shows the distribution of respondents over places. We aimed for approximately 100 respondents in each place and achieved that goal except in medium-to-high SES places, a result that will receive attention later in the discussion. We next describe the derivation of the sample.

Complete enumeration of a small area is equivalent to conducting a perfect census of that area, and consists of knocking on every door and interviewing at least one household member about all those who live there, as well as probing for sub-households. It is difficult because it requires gaining access to semi-public interior spaces (apartment hallways, courtyards, etc.) and because it requires returning at a later time in cases where no household member is at home. Short of that gold standard, a range of fieldwork techniques can be cate-

<sup>4</sup> The fieldwork was carried out by a team from the Department of Sociology, Qinghua University, Beijing, headed by Professor Qiang Li.

<sup>5</sup> The design is quasi-experimental, with the distinction between high- and low-SARS areas regarded as the “treatment variable.” For each of the six neighborhood types, we chose one neighborhood from a high-SARS area and one neighborhood from a low-SARS area. Mason et al. (2005) report the results of an analysis of social responses to the SARS epidemic.

Table 1. *Distribution of Sampled Places*

Type of place	High SARS	N	Low SARS	N
<i>Rural (urban fringe)</i>				
Bedroom village	Beijing	97	Suzhou	102
In-migrant village	Guangzhou	101	Chengdu	99
Factory dormitory	Guangzhou	101	Chengdu	100
<i>Urban</i>				
Rural-to-urban transitional	Beijing	102	Suzhou	98
Low SES	Guangzhou	98	Chengdu	97
Medium-to-high SES	Beijing	6	Suzhou	58

Source: 2003–2004 Pilot Study for a National Probability Survey of China ( $N = 1,059$ ).

gorized as enumeration. The easiest and least invasive, counting doors without otherwise attempting to gain information on the number of inhabitants behind the doors (e.g., Landry and Shen 2005), is probably least accurate. The primary disadvantage of this approach is that it assigns a constant number of eligible respondents to each address. We sought to overcome this limitation. We reasoned that the properties of complete enumeration are well known, and that what remained to be discovered, in the Chinese context, is whether accurate estimates of the number of eligible respondents at each address can be obtained without actually knocking on each door. Toward that end, we used local informants who accompanied enumerators during neighborhood traversals, providing information for each doorway, with enumerators knocking on doors or interviewing neighbors only as needed.

To arrive at the sample, the steps were as follows. We first purposely identified village and neighborhood committees in which to carry out the enumeration. Suitable units were selected in consultation with local authorities. Upon beginning field work, our enumerators immediately discovered that several of the selected areas were too large to be enumerated by available staff, with populations estimated to be well over 1,000. In these instances the fieldwork team restricted enumeration to an area within the local administrative unit. We will return to the issue of determining the size of the area to be enumerated.

Second, at each selected place, enumerators sketched a map that identified the boundaries and main roads of the area. With the



assistance of a locally knowledgeable person (usually a village or neighborhood committee official), enumerators walked through the area drawing successively more detailed maps. The second level of detail identified small streets and alleyways within sectors defined by main roads. A third level displayed individual buildings in each small area. Additional maps were sketched for each large building, showing the layout of apartments within the building. Figures 1a and 1b provide examples of the maps.

As enumerators moved through an area, each address, or other household identifier if no household address was available, was recorded on an enumeration roster, together with an estimate of the number of eligible adults (people ages 20–60) residing in the household. To obtain estimates of the number of eligible adults in each household, enumerators first asked their local informants. If the informants did not know, enumerators sought information from the residents themselves or, when they were not available, from neighbors. For high-rise apartment buildings and in other instances where it was not possible to obtain informed estimates, it was assumed that households included two eligible adults.<sup>6</sup>

In addition to recording the number of eligible residents in each household, the enumerator noted for each household the lowest and highest sequence number in the cumulative sum of the number of eligible adults. Figure 2 is intended to clarify this point (figure 2a is the original, in Chinese; Figure 2b is the English translation). In that figure, note that the first household has four eligible adults, and thus the cumulative sum is recorded as having lowest sequence number 1 and highest sequence number 4; the second household has two eligible adults, and thus the cumulative sum has lowest sequence number 5 and highest sequence number 6. How the cumulative sums were used to draw the sample is described below.

Enumeration of households requires a definition of what constitutes a household. For our purposes, a “household” is a living space with a doorway opening into a common or public area. Thus, for

---

<sup>6</sup> As we demonstrate below, these estimates introduced error because the number of adults per household varies substantially. In addition, this procedure made it necessary to plan for post-weighting the data to correct for errors in estimates. A truly complete enumeration procedure, in which enumerators visit each household to determine the number of eligible residents, would obviate the need for post-survey weighting.

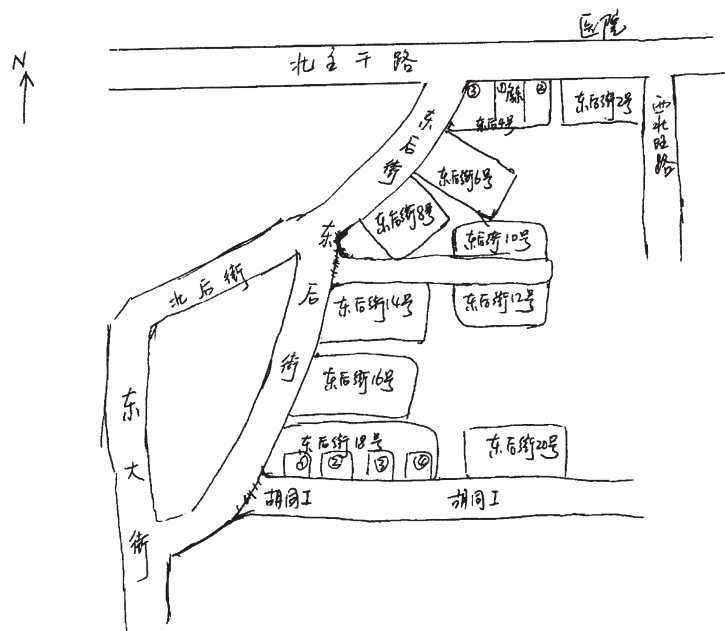


Figure 1a. Sketch Map of Area

东后街 42 号 户内详图.

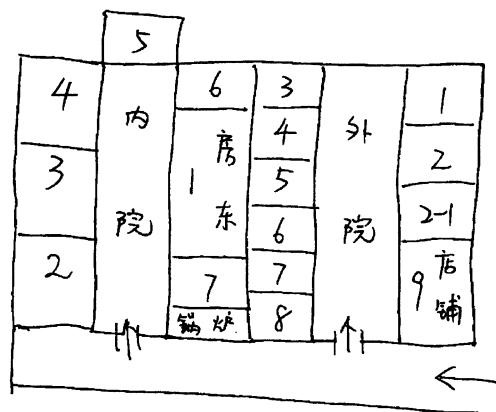


Figure 1b. Sketch Map of Building

抽样列举表

总页码: P

地图 页码	地图上的住户序号, 如有可能加上住址和其他描述信息	成人 数 20-60	序列范围	选出的 随机数	尝试 1	尝试 2	尝试 3
	1. <del>某某家</del> (1122)	4	1-4	1	1	2	
540	2. <del>某某家</del>	2	5-6		1	2	
	— 4' 52	10	7-16	14	14:5	14:5	X
	3. <del>某某家</del>	3	17-19				
	(— 4' 52)	12)					
	— 4' 52 63	2	20-21				
	50	2	22-23				
	40	2	24-25				
	2 在 11 田	2	26-27				
	2	2	28-29				
	3	2	30-31				
	4. <del>某某家</del> (1122)	3	32-34		14:5	14:5	
	4' 52 — 4' 52	2	35-36	35	14:5	14:5	
	— 4' 52 20	3	37-39				
	30	2	40-41				
	5. 七组 570 (1122)	4	42-45				
	4' 52 (1122)	5	46-50				
	— 4' 52 11 2 到 1	2	51-52				
18836	2	2	53-54	54	18:49	18:49	
	3	2	55-56				

列举员: — 11 日期: 2 10 4 0 3 2 3 分页码: P 1

Figure 2a. Page from Enumeration Roster, In-migrant Village

Map No.	Household no. on map (descriptive information)	No. age 20-60	Cum. sum	Selected random number	Disposition of interview		
					1	2	3
	1. xxx household (under) construction	4	1-4	1 4	x x	x x	
	2. xxx household —rental space	2 10	5-6 7-16	14	x		
	3. xxx household (—rental space 2nd floor, #6 #5 #4 3rd floor, #1 #2 #3	3 12) 2 2 2 2 2 2 2	17-19 20-21 22-23 24-25 26-27 28-29 30-31				
	4. xxx, 3rd floor —rental, 1st floor —rental, 2nd floor #2 —rental, 2nd floor #2	3 2 3 2	32-34 35-36 37-39 40-41	35	No	one	
	5. 7th group, #57 (xxx) —entrance —2nd floor (left to right)	4 5 #1 #2 #3	42-45 46-50 51-52 53-54 55-56	54	x		

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 onto a courtyard or passageway would be regarded as a separate household. This definition was relatively easy for enumerators to apply in a consistent way. However, enumerators often could not gain access to public or common areas onto which separate households opened. Examples include hallways in high-rise buildings and courtyards behind locked gates. In such cases, the number of households was determined by asking knowledgeable local people.

Once completed, we treated the enumeration roster as a census of all eligible individuals in the area. For example, the enumeration roster in Figure 2 lists 56 eligible adults who satisfied our selection criteria (ages 20–60). Using the enumeration rosters, for each of the 12 purposively selected places we drew random samples of somewhat more than 100 individuals in order to compensate for unoccupied units, never-at-homes, no one in the household meeting the age criteria, refusals, etc.—the exact number varied by site, depending on the judgment of the enumerators as to how difficult it would be to contact people. In the Chengdu factory enclave it took only 106 attempts to complete about 100 interviews. At the other extreme, the low SES neighborhood in Guangzhou, it took 180 attempts to complete about 100 interviews, due to large numbers of never-at-homes, lack of eligible respondents, and refusals. Sampling was performed by the fieldwork supervisor, using a random number table to choose the individuals to be interviewed. For example, on the enumeration page shown in Figure 2, random numbers 1, 4, 14, 35, and 54 were drawn. Interviewers were then assigned to interview two individuals in the first household (whose cumulative sum includes sequence numbers 1 and 4), one person in the rental space for the second household, which included 10 individuals (because 14 falls into the range of the cumulative sum 7–16), and so on.

When interviewers visited households, they knew the number of assigned interviews to be obtained at each household, but they did not know which individuals within a household were to be interviewed. For this task, at each household with sampled individuals, interviewers created a complete household roster. Then, using Kish-table procedures (Kish 1995), the required number of specific individuals to be interviewed was randomly selected—most often one person but sometimes more than one person (see, for example, household 1 on the roster page shown in Figure 2). The selected individuals were then either interviewed on the spot or appointments were made to return for the interviews.

For each of the 12 purposively selected places our goal was to complete 100 interviews. We did so in all but the middle-to-high-income areas, which consisted entirely of restricted-access buildings. As Table 1 demonstrates, in Beijing our interviewers had extreme difficulty gaining access, and obtained only six interviews.<sup>7</sup> In Suzhou

<sup>7</sup> Our medium-to-high SES site in Beijing is a high-rise apartment complex man-

they were more successful, but still obtained only 58 interviews.<sup>8</sup> Despite the best attempts of our enumerators and interviewers to secure the cooperation of building managers, or to enter buildings surreptitiously, we failed to achieve minimally acceptable completion rates. Because high-rise and restricted-access buildings are becoming increasingly common in urban China, unless reliable ways are found to enter such buildings, and to secure interviews in accordance with sampling procedures, it will become impossible to conduct adequate surveys of the urban population of China through use of face-to-face interviews.

Although the six-category place classification used in our pilot study is not standard and does not span the entire rural-urban range, it is nonetheless instructive to consider variation in social composition across the place types for what this can reveal about the problems surveys are increasingly likely to encounter in China. Table 2 presents distributions of various characteristics of the pilot study respondents.

In some areas, substantial fractions of the population lack any kind of local registration: 38 percent in in-migrant villages; 33 percent in dormitories, and 13 percent in low SES urban neighborhoods. All of these people would be missed in surveys based on local registration lists, even those that sample from the list of people with temporary residence permits.

Table 2 also shows that, even in the restricted age range of 20–60, the mean age of respondents in migrant communities is substantially lower than in communities with high proportions of permanent residents. In general, on an international comparative basis, communities with high proportions of migrants tend to be younger than those with low proportions of migrants (Massey et al. 1993).

---

aged by a single company, which refused to allow interviewers to enter the designated building. The company did allow our fieldwork team to place requests for interviews in 50 mailboxes, which resulted in one positive response. The company also phoned 50 respondents on behalf of the fieldworkers, which resulted in an additional five agreements to be interviewed. It is likely that had the interviewers been able to knock on doors, the response rate would have been somewhat higher. Moreover, had we used experienced professional interviewers, rather than inexperienced students, the probability of gaining permission to knock on doors would likely have been substantially greater.

<sup>8</sup> In both Suzhou and Beijing, many apartments were found to be vacant. In addition, a substantial fraction of respondents was never at home.

Table 2. *Selected Characteristics of Respondents in Various Community Types*

	Bedroom village	In- migrant village	Factory/ dormitory	Rural to urban transitional	Low SES urban	Medium/ high SES urban
<i>Registration Status (%)</i>						
Permanent local registration	26.1	22.0	9.0	39.0	72.8	87.5
Temporary local registration	9.6	1.0	2.0	0.5	0.0	4.7
Temporary residence permit	62.8	39.0	56.2	57.0	13.8	6.2
No local registration or permit	1.5	38.0	32.8	3.5	13.3	1.6
<i>Other Sociodemographic Characteristics</i>						
% with agricultural <i>hukou</i>	63.8	69.5	76.1	83.5	18.0	4.7
% female	43.7	46.0	60.2	47.5	47.2	53.1
Mean age	28.9	32.9	29.6	35.3	38.4	40.3
Mean years of schooling	10.7	8.5	9.4	7.8	11.1	12.6
% with at least some upper middle school	58.3	27.0	34.3	14.0	58.0	73.4
% employed	77.4	82.0	97.0	81.5	70.8	76.6
% unemployed	9.6	9.5	2.5	6.0	12.3	4.7
% self-employed without employees <sup>a</sup>	3.4	11.1	0.0	37.8	5.2	3.5
% small entrepreneur ( <i>getihu</i> ) <sup>a</sup>	5.7	18.1	1.0	16.1	16.9	5.3
% communist party member	5.5	1.0	2.5	2.0	10.8	25.0
<i>N</i>	199	200	201	200	195	64

Notes: <sup>a</sup> Among the employed.

Source: See Table 1.

Perhaps consistent with the pattern of layoffs from stagnating state factories in urban China (Lee 2000), the low SES urban areas in the pilot study have the highest unemployment rates, approximately 12 percent of respondents. Bedroom villages and in-migrant villages, which have high proportions of migrants, also have unemployment rates that are nearly as high. It is possible that urban fringe migrant

concentrations and low SES urban areas have high unemployment for different reasons.

Finally, the rural-to-urban transition areas in our sample are economically and socially marginal. On the fringes of Beijing and Suzhou cities, these are areas where settled rural villages have been overwhelmed by newcomers. These places are neither bedroom villages for migrants nor places with established workshops employing migrants. Rather, they appear to be places occupied by people without stable jobs. Eighty-six percent of respondents in the two rural-to-urban transition locales are migrants who took up residence after age 14, and 70 percent of the migrants arrived in the five years previous to the survey. Only 42 percent of the employed are employees, whereas 54 percent are self-employed workers without employees or very small scale entrepreneurs (*getihu*); the remaining four percent are still employed in agriculture. Reports from our field workers suggest that these may be areas of illegal or semi-legal activity. Certainly, the Beijing locale was a “tough” place, with a self-appointed local boss who challenged the enumerators and tried to discourage their activity.

#### LESSONS AND RELATED OBSERVATIONS

Experience gained during the fieldwork phase of our pilot study may be informative for others attempting surveys in urban China, including the urban fringe still administratively defined as rural. We next discuss lessons and related observations derived from our pilot study.

##### *Even imperfect enumeration is superior to reliance on registration lists*

We estimate that 17 percent of respondents in the pilot study would have been missed had we relied on registration lists for those with permanent residence as well as those with temporary residence permits. In factory dormitories and high in-migration villages, this figure rises to one-third or more. In the 2000 census, 30 percent of urban residents were informal migrants. Landry and Shen (2005), reporting on a Beijing area survey in which enumerators counted doors, report an even higher estimate. Forty-five percent of respondents in that survey would have been missed had the survey team relied on registration lists—20 percent had changed residence within Beijing, and 25 percent were unregistered in-migrants to Beijing. Thus, even for surveys of representative samples of the urban population, but



especially for surveys concerned with migrants, exclusive reliance on registration lists will result in substantial under-representation of important segments of the population.

*Enumeration of addresses or doors is not adequate*

Some data-collection agencies in China have begun to count addresses or doors rather than knocking on doors to ascertain the number of eligible respondents. This procedure introduces error because the number of adults per family household varies substantially. To illustrate the point, Table 3 displays distributions of the number of adults per household for the pilot study data and for a sample drawn from the 2000 census.

Although the averages in Table 3 are about the same for family households in the pilot study and the census samples (2.3 and 2.2 eligible adults per family household, respectively), and both are close to the mean of two eligible adults per family household that we assumed when no information was available, the dispersion around these averages cannot be ignored. Two-adult households constitute 59 percent of the family households in the pilot study. Family households with one or three adults are 30 percent of the distribution. Remaining numbers of adults per family household constitute 11 percent of the distribution in the pilot study. Thus, if our enumerators had simply counted doors, and had they known which doors led to family and which to collective households, our estimates would have been incorrect for 41 percent of the family households. The census sample suggests a similar conclusion for family households.

This magnitude of potential error may be a lower-bound estimate, since some of the time it is not obvious whether behind a door there is a family or a collective household. For collective households (dormitories, rooms shared by migrants, and other collective living arrangements), the problem is even more severe, because the distribution of household size is more variable, and because no specific number of adults constitutes even as much as one-third of the distribution. Because the distributions in Table 3 are unimodal but skewed to the right, counting doors is likely to under-represent people living in households with large numbers of eligible adults, and to over-represent people living in households with small numbers of eligible adults. A further problem with the practice of counting doors is that there often are doors behind doors, behind which are sepa-

Table 3. *Percentage Distribution of Number of Adults Age 20–60 Per Household, for Family and Collective Households, 2003–2004 Pilot Study and 2000 Census*

Number of residents age 20–60	Pilot Study		2000 Census of China	
	Family	Collective	Family	Collective
0	0.1	0.0	0.0	0.0
1	14.7	9.8	15.4	18.6
2	58.8	27.9	63.7	27.2
3	15.2	18.8	12.4	17.5
4	7.8	13.5	6.5	13.0
5	2.3	7.8	1.5	9.9
6	0.5	7.8	0.4	4.4
7	0.5	7.8	0.1	3.2
8	0.0	2.5	0.0	2.5
9	0.0	0.4	0.0	1.1
10 or more	0.0	3.7	0.0	2.6
Total	99.9%	100.0%	100.0%	100.0%
Mean	2.3	3.9	2.2	3.4
Standard deviation	1.0	2.8	.9	2.7
<i>N</i> (Households)	815	244	307,187	7,099

Sources: Pilot Study—see Table 1; 2000 Census of China—0.1 percent sample.

rate households that are completely invisible unless inquiry is made. In addition, sometimes the residential unit for which a doorway is visible is vacant, and the distribution of vacancies may not be random.

*Hybrid estimation of the number of eligible respondents per household performs only moderately better than door-counting*

If the hybrid enumeration strategy we used in the pilot study is superior to door-counting or random guesses, then, in a regression of the actual number of eligible adults per household on the estimated number of eligible adults per household, the coefficient should be positive and statistically significant. That is what we found using the 868 households for which the required cover-sheet information was available ( $b_{yx} = .41$ ,  $t = 12$ ,  $R^2 = 0.14$ ). The effectiveness of the hybrid enumeration strategy we used depends on the quality of the local informant's knowledge. We have no evidence to suggest that the

local informants who assisted during the enumeration were anything but among the most knowledgeable available individuals. For this reason, the 14 percent improvement over doorway-counting (or chance) indicated by the  $R^2$  may not be a lower bound. The evidence thus suggests that to obtain accurate information on the number of eligible respondents residing in each household, it is necessary to conduct a complete census of each local area. This needs to be done by conducting a mini-interview with a knowledgeable member of each household, with the enumerator recording the age, sex, and, if necessary, other identifying information for each eligible member of the household, and also aggressively inquiring about the possibility of sub-households within the household space. Obtaining cooperation where needed from local authorities, and then taking the time to enumerate thoroughly, will substantially increase the cost of conducting surveys that are representative of the population of China. Cost expectations based on past surveys are no longer valid, and must be revised upward.

*Criteria for the determination of households must be clear, precise,  
and exhaustive*

Enumerators must have clear rules to follow. Also, the quality of any realized sample depends in part on the precision and exhaustiveness of the criteria that enumerators are instructed to apply. As noted, our practice was to define as separate households those places that open into public space, but to include rooms rented to tenants that open into the living space of another household as part of that household. Members of primary families in households often neglect to report on people living in rented spaces, including rooms that open directly onto public spaces. Because this tendency results in under-counts of the number of household members as well as the number of households, enumerators and interviewers need to be trained to probe for renters, servants, and others who are not immediate family members.

*Unconventional living spaces must be enumerated*

In urban China, many people sleep in the shops where they work, in rooms behind restaurants, entryways to residential and commercial buildings, rooms in large commercial buildings, small houses on the edge of university campuses, lofts above garment workshops, rail-

way stations, factory dormitories, and in temporary dormitories erected on construction sites. For example, two surveys of migrants in Shanghai each found that about half of those surveyed lived at their work site (Roberts 1997: 274; Wang et al. 2002). In rural areas, unconventional dwelling arrangements include road construction crews living in tents, staff of power plants and dams, and boatmen and fishermen living on their boats. If these areas are not enumerated, an unknown but nontrivial fraction of the population will be missed. Thus, enumerators must be alert to the possibility of unconventional living arrangements, especially when enumerating areas that are not obviously residential. In these areas, enumerators will need to find knowledgeable individuals. For example, when enumerating commercial or industrial buildings, enumerators should attempt to locate building managers to inquire whether any people sleep in those buildings.

*It is important to adequately identify enumerated households so that interviewers can find the households chosen for interviewing*

In China, many houses have no obvious, standardized location identifiers. It helps to provide interviewers with as many of the following as is practical and appropriate: an address, where one exists; a location description; the name of the householder; a photograph of the doorway to the household; and coordinates supplied by handheld GPS (Global Positioning System) instruments.

*In the Chinese context, for practical reasons, samples of individuals may be preferable to samples of households*

Conditional on the adequacy of the enumeration roster, the procedure we used generates unbiased probability samples of individuals rather than of households that contain such individuals.<sup>9</sup> An alternative procedure used in some studies (e.g., the Indonesian Family Life Survey (Frankenberg et al. 1995)) is to sample households and then to collect information on each member of the household. With post-survey weighting it is possible to obtain unbiased probability samples of people using this procedure.

<sup>9</sup> The caveat of footnote 6 applies to this assertion. Had we been able to carry out a complete enumeration, no post-survey weighting would have been necessary.

Sampling households is most practical in situations in which it is possible to separately interview each eligible member of the household, or in which households consist of members of a single family who are so knowledgeable about each other that a single individual can serve as household informant. Interviewing everybody in the household is relatively impractical in contemporary urban China, where people are often not at home, and where the cost of securing interviews with each adult is thus prohibitively expensive. The use of key informants in Chinese surveys is also problematic where households include non-family members, because whoever is interviewed is unlikely to be adequately informed about the characteristics of all other adults in the household. In our data, almost all collective households, but only a small fraction of family households (two percent), include individuals who are not related to the household head. A tabulation from the 2000 census shows that about six percent of family households include non-family members. The census figure may be an underestimate, due to the way in which the census allocated individuals who had lived less than six months in a place (see the discussion below).

*Cooperation from local authorities is helpful; the value of cooperation from higher authorities seems less clear*

Cooperation of local authorities may improve access to restricted-access buildings and modulate harassment by police and local residents. The use of local guides often helps. In our experience, residents often became suspicious of strangers walking through the area drawing maps of buildings and notified either the police or local “bosses,” who challenged our enumerators to explain their purpose and tried to get them to desist. Official permission from local officials can be an effective counter to such challenges. It may not secure interviews, but it will protect enumerators and interviewers.

Higher-level administrative approval no longer guarantees lower-level cooperation. In the past, it was possible to secure the approval of a state ministry, which would then transmit the approval “down the line” to successively smaller units. Now, approval by a higher level of administration does not imply cooperation by local authorities. In one case, we had secured written approval from the township-level authority, which promised that a local policeman would be made available to accompany the enumerators without further

charge. When our enumerators tried to work out arrangements at the local level, they were refused with the question, "Who is going to pay for him [the policeman]?"

*In urban areas, and rural areas near cities, it is no longer feasible to sample administrative units below the street committee level*

Several changes in the administration of urban China make it impractical to specify a sampling stage below the street committee level (*jiedao*). First, in some parts of urban China, neighborhood committees have been replaced by community committees, and there is no longer a consistently defined unit below the street committee level. Second, in many rural-to-urban transitional areas, not all households within the area are regarded as part of the administrative unit. For example, in our pilot study we 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. It appeared that such individuals were under no local jurisdiction below the township level, in a manner akin to unincorporated areas of counties in the United States. Third, there appear to be areas (e.g., construction sites, camps for highway repair crews, power plants, and factories) in both urban and rural China that are similarly unincorporated, in the sense that they are not part of a specified village, but where people live nonetheless. All of China is covered by township-level units, but not all of China is included in village/neighborhood committees. Specifically, 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 and parks, are excluded. In a similar way, dams, power plants, rural factories, but also national parks and mountainous areas, are not necessarily included in the jurisdictions of specific villages.

*Township-level and even lower-level units are too large to enumerate*

There are approximately 51,000 township-level units in China. With an average population of 25,000 and an average number of households that is above 7,000, township-level units are too large to be enumerated by the typical survey organization. Thus, township-level units must be subdivided in a way that preserves the principle that each individual has a known probability of selection into a sample.

The same point holds at lower levels. Thus, for example, community committees may contain several thousand households. Furthermore, although until recently village and neighborhood committees had populations averaging about 1,000 (State Statistical Bureau 1996: 45–47), the locales selected for the pilot study tended to have larger populations. To put these figures in perspective, the National Bureau of Statistics defines its enumeration districts as areas that take an enumerator one week to fully enumerate and interview. On average, enumeration districts contain approximately 100 households and 300 people.

*A major unsolved problem is that of reliably and efficiently gaining access to restricted-access buildings*

As noted above, in 2000 nearly a quarter of the urban population of China lived in recently constructed housing, mainly high-rise apartment buildings. These buildings often restrict entry to residents and those invited by residents. Unlike the United States, where restricted access buildings and gated communities are uncommon, in China it appears that a large fraction of new high-rise apartment buildings, even those for people with relatively modest incomes, employ uniformed doormen to guard the entrance.<sup>10</sup> Moreover, as experienced in our pilot study, building management companies explicitly guard the privacy of their tenants against intruders, including academic researchers. Similar access problems occur for large factory or construction-site dormitories. The difficulty of enumerating and interviewing within restricted-access buildings has become a major threat to the representativeness of general population surveys conducted in China.

---

<sup>10</sup> About four percent of U.S. housing units are in restricted-access communities and about five percent are in restricted-access multi-unit buildings; but these two categories overlap, and somewhat less than nine percent of households have any kind of restricted entry (U.S. Census Bureau 2005: Table 2–8). Moreover, given that the probability of living in restricted-access housing is positively correlated with socioeconomic status and that household size is negatively correlated with socioeconomic status, the proportion of the population living in restricted-access housing is probably even smaller.

*The willingness of urban residents to cooperate with interviewers cannot be taken for granted*

On the basis of discussions with others who have conducted surveys in China over the past two decades, and relative to our own past surveys, refusal rates appear to have increased in urban areas. In the pilot study, refusal rates ranged from zero in the factory dormitories to 17 per cent in the low SES neighborhood in Guangzhou.<sup>11</sup> Refusals may be due to a combination of fear of crime, which makes people unwilling to open the door to strangers; an acceleration of urban life that is now more stimulating and option-filled than formerly; and an increased sense of personal autonomy coupled with a reduction in anxieties about offending authorities (Shambaugh 1993; Lau 2001). Moreover, although monetary payments are often used as incentives in Chinese surveys, the rising level of income inequality makes it difficult to provide large enough incentives to secure the cooperation of the new high income population without exceeding research budgets. Furthermore, graded incentives, calibrated to the presumed income of the respondents, are difficult to administer.

*Small area population estimates become increasingly out-of-date on a cycle driven by (decennial) censuses*

As the time since the last census increases, population estimates for small areas become less accurate due to population change. If that is true universally, it is especially true for China, due to the rapid economic growth of the nation, which has resulted in the upgrading of many villages to towns and of many towns to cities (Goldstein 1990). Moreover, areas of high in-migration tend to shift rapidly over time due to the development of new factories and workplaces as well as other changes in local conditions, such as the forced removal of migrants from certain areas (Fan 1996). In the pilot study, which was conducted no more than three-and-a-half years after the 2000 census, in the bedroom villages, in-migrant villages, and factory areas in our sample, more than half of all respondents arrived

---

<sup>11</sup> We had no refusals in either of our factory dormitories, although in one case there were some non-responses because dormitory rooms were occupied entirely by under-age respondents or because no occupant of the room was ever home. In the other case, respondents were called into an office for the interview during working hours and apparently were instructed by management to cooperate.



in 2001 or later, and hence would not have been included in the census counts. Had departures balanced new arrivals, the net change would have been small. But no way of estimating the net change is currently available. Although local authorities provide annual updates of the total size of the population, it is unclear how accurate such estimates are and, in any event, they do not generally include breakdowns by sociodemographic characteristics, and migrant status in particular.

*The 2000 census substantially undercounted migrants*

Demographic analyses (Lavelly 2001; Anderson 2004) suggest that the 2000 census substantially undercounted the migrant population. It is likely that many migrants were not located by census takers. In addition, the undercount was a function of an administrative decision regarding who should be counted as a migrant. Quite apart from the cyclical decay in small area estimates that depend on the decennial census, the undercount of migrants in the decennial census is problematic for those attempting to draw probability samples of the population. We return to this point after explicating the procedure used in the 2000 census.

In the census, migrants are those living in a place (village or neighborhood) other than where they are permanently registered. For 2000, an administrative decision was made to count as local residents only those migrants who had been living in the place they were enumerated for more than six months or who had left the place where their *hukou* was held more than six months before the census.<sup>12</sup> In our pilot study we found that 70 percent of informal migrants had arrived in the previous six months. Estimates from surveys of migrants conducted in Beijing and Shanghai in the 1990s (Joint Publications Research Service 1993; Wang and Zuo 1996; Roberts and Wei 1999) suggest that as many as half of all migrants had arrived less than six months earlier, which would double the

---

<sup>12</sup> Individuals who had been absent for less than six months from the places where they were registered were supposed to be counted as residents of those places. But it is probable that many such people were not counted, especially in situations in which the entire family moved away. Of course, even if they were completely enumerated, the result would be to undercount the *de facto* urban population and to overcount the *de facto* rural population.

number of migrants (defined as all those spending at least four nights a week in a place other than where they are permanently registered) in Beijing and Shanghai relative to the 2000 census estimates. If a roughly comparable ratio holds for other cities, then, to return to the implication of the migrant undercount, reliance on the 2000 census to establish a survey sampling frame will result in an underrepresentation of migrants and of cities with large proportions of migrants.

#### OUTLINE FOR A COMPLETE SAMPLING DESIGN

Because the pilot study focused on enumeration per se, it was unnecessary to use probability sampling methods to select local areas to be enumerated. We found in the pilot study that a form of enumeration is feasible, and that it detects individuals who do not appear in local *hukou* lists. There remains, however, the practical question of how to sample local areas when *hukou* lists are not used. We next provide an outline for a complete sampling design that includes enumeration and is responsive to those key difficulties discussed above that affect enumeration.

We propose a design for a national probability sample survey in China that combines (i) census-based sampling down to township-level units with (ii) the sampling of areal polygons within township-level units, and is followed by (iii) complete enumeration of individuals within each selected areal polygon, and (iv) probability sampling of individuals from the enumeration list.

Despite the undercount in the 2000 census, and the elapsed time since 2000, there is no viable alternative to that census for establishing a sampling frame for surveys to be carried out before data from the next decennial census become available. Use of the census affords advantages not otherwise available: it becomes possible to gain efficiency by incorporating stratification into the sample design, and unique and valuable contextual (e.g., township-level) data can be merged with individual data in the completed sample. We expand on these points later in the discussion.

In the proposed design, the first-stage sample is drawn by sampling counties and county-level urban units with probability proportional to size, using updated estimates of population size if available. Within each first-stage sample, a second-stage sample can be drawn in two steps: first contact the statistical office of every selected county-level

unit to obtain up-to-date administrative estimates of the population size of each township-level unit; and then use this information to sample township-level units with probability proportional to size. County- and township-level data files, with GIS information, are now commercially available.

For each sampled second-stage unit, (i) on a map or satellite photo for that unit impose a grid composed of squares small enough to be completely enumerated (e.g., 90 meter squares); (ii) randomly sample the squares or, if information regarding population density can be obtained, sample them with probability proportional to density; (iii) enumerate all individuals or all eligible respondents. GIS maps and satellite photos with adequate resolution (step (i)) are now available. Step (ii) can be carried out using GIS software. Step (iii) depends on the use of handheld GPS navigators to locate and delineate the actual squares to be enumerated. For a Beijing study, Landry and Shen (2005) successfully implemented the use of GPS technology for the location of squares or polygons to be enumerated.

The sampling of very small areas defined by geographic coordinates solves problems that would be intractable if survey designers attempted to delimit and then sample the areas administered by village, neighborhood, and community committees. First, the geographic coordinates approach sidesteps practical issues raised by post-2000 local area administrative, boundary, and population density changes (see the discussion above). If necessary, post-weighting can be carried out to correct for differences in the number of inhabitants in the enumerated polygons. Second, the geographic coordinates approach creates a systematic basis for sampling sub-areas within townships and street committees, which are the lowest level administrative units that are relatively time-invariant during the inter-censal decade. Third, it permits the specification of sub-areas small enough to be completely enumerated within a short period of time.

The proposed design is for a sample of the entire population of China. This is in deliberate contrast to the common practice of restricting a sample to urban or rural China. Partitioning the population in this way is inadvisable when the intent is to allow inferences that are applicable to the entire population. Rural-only and urban-only samples will typically be biased on factors related to the outcomes of analytic interest. Because there has been so much migration in recent years, a sample restricted to the urban population will represent the urban-origin population (since there is hardly any urban-

to-rural migration) but will not represent the rural-origin population. It is perhaps a universal fact that rural-to-urban migrants do not represent the entire rural-origin population; they are to some degree “selected” on characteristics that differentiate them from those who stay behind (Fan 1999; Yang and Guo 1999; Li and Zahniser 2002; Liang and Ma 2004; Wu and Treiman 2004a). Thus, with urban-only or rural-only samples, analyses of migration, social mobility, socioeconomic attainment, life-course transitions, health, or analyses of differences between urban-origin and rural-origin individuals, are likely to yield incorrect and misleading results (Wu and Treiman 2004b).

The proposed design is intentionally clustered and stratified. Surveys based on simple random samples are everywhere prohibitively costly. They are also impractical because equivalent efficiency can be obtained through a combination of clustering and stratification. Most surveys in China use clustered sampling because that reduces field costs. However, because the Chinese population is heterogeneous, clustering is inefficient relative to simple random samples. That is, clustered samples in China tend to produce standard errors equivalent to those yielded by smaller random samples. Stratification on variables likely to be associated with variables of analytic interest can substantially reduce and even eliminate the inefficiency induced by clustering (Kalton, Brick, and Lê 2005). Treiman et al. (1996) exemplify these observations in a series of sampling experiments based on a one percent sample of the 1990 Census of China. Of course, stratification is also important for its original purpose, which is to insure sample representation on key dimensions.

The proposed design is fully hierarchical in a useful way, because it is based on civil divisions down to the township/street committee level, and coordinate-defined small areas are sampled within townships or street committees. Consequently, it is possible to merge census-derived information about social, demographic, and economic contexts at the county and township levels with individual-level data. The resulting file is uniquely rich, and permits the use of hierarchical modeling. This is a clear advantage of combining the civil division-based and geographic coordinate-based sampling approaches. The purely geographic coordinate-based approach does not reliably provide for the nesting of small enumeration areas within low-level administrative units, and is not intended for that purpose.

## CONCLUSION

Long before we contemplated conducting the pilot study described here we began to discuss, with experienced researchers in China and elsewhere, the possibility that enumeration could be incorporated into the design for a major sample survey of the population of China, as a way of contending with inadequacies of *hukou* lists. The consensus among those we consulted was that enumeration would not succeed. It was claimed that officials would not cooperate, the populace would be suspicious and troublesome, and that fieldworkers would “get in trouble,” be warned off, or put in jail.

Recent evidence suggests otherwise. As Landry and Shen (2005) have shown, selection of small areas followed by the counting of doorways is feasible; this is a form of enumeration. As we have reported here, it is possible to take the enumeration a step further, using the hybrid estimation procedure we have described. Thus, either of these forms of enumeration is possible, and moreover, the evidence indicates that both procedures find substantial percentages of people who tell interviewers that they are not locally registered. These people will not be included in the *hukou* lists that apply to the enumerated areas.

Although the hybrid enumeration approach we piloted is better than doorway counting in the listing of local populations, the performance improvement is not major. This result strongly suggests the need for complete enumeration of small areas. Is this practical? It could be argued that decennial census enumerators perform effectively because they are government employees empowered by law to make inquiries. Academic and private sector fieldworkers, because they have no public mandate, must depend on the cooperation of officialdom, their own skills, the forbearance of the local populace, and the organizational skills of the survey groups that employ them. The experience of our fieldworkers during the pilot study suggests that, despite the legal advantage census enumerators have over academic and private sector enumerators, there is no fundamental obstacle to complete enumeration. The hybrid approach we employed in the pilot study included knocking on doors. Furthermore, although the estimates of household composition provided by local informants are not highly accurate on a household-by-household basis, these informants, who typically are local officials,

can accompany enumerators on their rounds to counter potential objections to the enumeration process.

Based on our fieldwork experience, we conclude that nothing less than a complete census of households in small areas will produce a sampling frame of sufficient accuracy to permit a true probability sample of individuals to be drawn. Complete enumeration will raise survey costs, but is not otherwise unfeasible.

The unanticipated level of new residential construction reported in the 2000 census, in combination with the unexpected prominence of restricted-access buildings in the urban and urban-fringe areas in which our pilot study was carried out, suggest that a profound change in the social organization of privacy is taking place. This change has major, not well understood, implications for the ways in which social surveys in China will need to be conducted.

#### REFERENCES

- Anderson, Barbara A. (2004). "Undercount in China's 2000 Census in Comparative Perspective." *PSC Research Report No. 04-565*. Ann Arbor: University of Michigan, Population Studies Center, September.
- Chan, Kam Wing, and Li Ziang. (1999). "The *Hukou* System and Rural-Urban Migration in China: Processes and Changes." *The China Quarterly* 160:818-855.
- Chan, Kam Wing, Liu Ta, and Yunyan Yang. (1999). "*Hukou* and Non-*hukou* Migration in China: Comparisons and Contrasts." *International Journal of Population Geography* 5:425-548.
- Davin, Delia. (1999). *Internal Migration in Contemporary China*. New York: St. Martin's Press.
- Fan, C. Cindy. (1996). "Economic Opportunities and Internal Migration: A Case Study of Guangdong Province, China." *Professional Geographer* 48:28-45.
- (1999). "Migration in a Socialist Transitional Economy: Heterogeneity, Socio-economic and Spatial Characteristics of Migrants in China and Guangdong Province." *International Migration Review* 33:954-987.
- Frankenberg, Elizabeth, Lynn A. Karoly, Paul Gertler, Sulistinah Ahmad, I. G. N. Agung, Sri Jarijati Hatmadji, and Paramita Sudharto. (1995). *The 1993 Indonesian Family Life Survey: Overview and Field Report*. Santa Monica, CA: Rand Corp.
- Goldstein, Sidney. (1990). "Urbanization in China, 1982-87: Effects of Migration and Reclassification." *Population and Development Review* 16:673-701.
- Goldstein, Alice, and Sidney Goldstein. (1991). "Temporary Migrants in Shanghai Households, 1984." *Demography* 28:275-291.
- Goodkind, Daniel, and Loraine A. West. (2002). "China's Floating Population: Definitions, Data and Recent Findings." *Urban Studies* 39:2237-2250.
- Joint Publications Research Service (JPRS). (1993). "Beijing's Mobile Population: Current Status, Policy." *Renko Yu Jingji* 79 (August 25). JPRS-CAR-93-091, December 29:44-47. [In Chinese.]
- Kalton, Graham, J. Michael Brick, and Than Lê. (2005). "Estimating Components of Design Effects for Use in Sample Design. Pp. 95-121 in *Household Sample Surveys in Developing and Transition Countries*. United Nations, Department of Economic and

- Social Affairs, Statistics Division. [Electronic version: ST/ESA/STAT/SER.F/96/WWW March 2005; accessed 25 April 2005 at [http://unstats.un.org/unsd/hhsurveys/pdf/Chapter\\_6.pdf](http://unstats.un.org/unsd/hhsurveys/pdf/Chapter_6.pdf)]
- Kish, Leslie. (1995). *Survey Sampling*. New York: John Wiley.
- 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* 13:1–22.
- Lau, Raymond W. K. (2001). "Socio-Political Control in Urban China: Changes and Crisis." *British Journal of Sociology* 52:605–620.
- Lavelly, William. (2001). "First Impressions from the 2000 Census of China." *Population and Development Review* 27:755–769.
- Lavelly, William, James Lee, and Wang Feng. (1990). "Chinese Demography: The State of the Field." *The Journal of Asian Studies* 49:807–834.
- Lee, Hong Yung. (2000). "Xiangang, the Chinese Style of Laying Off Workers." *Asian Survey* 40:914–937.
- Li, Haizheng, and Steven Zahniser. (2002). "The Determinants of Temporary Rural-to-Urban Migration in China." *Urban Studies* 39:2219–2235.
- Liang, Zai. (2001). "The Age of Migration in China." *Population and Development Review* 27:499–524.
- Liang, Zai, and Zhongdong Ma. (2004). "China's Floating Population: New Evidence from the 2000 Census." *Population and Development Review* 30:467–488.
- Mason, William M., Donald J. Treiman, Yaqiang Qi, Yao Lu, Yi Pan, Shige Song, and Wei Wang. (2005). "Behavioral Responses to China's 2002–2003 SARS Epidemic." Paper prepared for presentation at the annual meeting of the Population Association of America, Philadelphia, 30 March–3 April.
- 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* 19:431–466.
- 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* 31:249–293.
- Roberts, Kenneth D. and Jinsheng Wei. (1999). "The Floating Population of Shanghai in the Mid-1990s." *Asian and Pacific Migration Journal* 8:473–510.
- Shambaugh, David. (1993). "Losing Control: The Erosion of State Authority in China." *Current History* 92:253–259.
- Solinger, Dorothy J. (1999a). *Contesting Citizenship in Urban China*. Berkeley: University of California Press.
- (1999b). "Citizenship Issues in China's Internal Migration: Comparisons with Germany and Japan." *Political Science Quarterly* 114:455–478.
- State Statistical Bureau. (1996). *Zhongguo Nongcun Tongji Nianjian 1996 (Rural Statistical Yearbook of China 1996)*. Beijing: Zhongguo Tongji Chubanshe. [In Chinese.]
- 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/index/framei.htm>: click on "index," click on "Asia-China," click on "Life Histories and Social Change in Contemporary China, (1996)," click on "CLHS Site (non-UCLA researchers click here)," click on "app\_d\_public.pdf."]
- U.S. Census Bureau, Housing and Household Economic Statistics Division. (2005). *American Housing Survey National Tables: 2001* (last revised February 18 2005). [[http://www.census.gov/hhes/www/housing/ahs/ahs01\\_2000wts/ahs01\\_2000wts.html](http://www.census.gov/hhes/www/housing/ahs/ahs01_2000wts/ahs01_2000wts.html)] (Accessed 15 March 2005.)
- U.S. Consulate Guangzhou. (2001). "Township and Village Enterprises in Guangdong:

- Challenges for a New Decade.” Cable R 090849Z Jan. 1 (cited in Goodkind and West 2002:2248).
- Wang, Feng, and Xuejin Zuo. (1996). “Rural Migrants in Shanghai: Current Success and Future Promise.” Paper presented at the International Conference on the Flow of Rural Labor in China. Beijing, June 25–27.
- Wang, Feng, Xuejin Zuo, and Danching Ruan. (2002). “Rural Migrants in Shanghai: Living under the Shadow of Socialism.” *International Migration Review* 36:520–545.
- Wu, Xiaogang, and Donald J. Treiman. (2004a). “The Household Registration System and Social Stratification in China: 1955–1996.” *Demography* 41:363–384.
- (2004b). “Inequality and Equality Under Socialism: Occupational Mobility in Contemporary China.” Los Angeles: California Center for Population Research, UCLA. Working Paper CCPR-044–04. [[http://www.ccpr.ucla.edu/asp/papers\\_authors.asp](http://www.ccpr.ucla.edu/asp/papers_authors.asp)].
- Yang, Liushi, and Fei Guo. (1999). “Gender Differences in Determinants of Temporary Labor Migration in China: A Multilevel Analysis.” *International Migration Review* 33:929–953.
- Yang, Xiushi. (1993). “Household Registration, Economic Reform and Migration.” *International Migration Review* 27:796–818.