

# 7 Rural–Urban Migration in China: Survey Design and Implementation

**Sherry Tao Kong**

---

## 1 BACKGROUND AND OVERVIEW OF THE SURVEY DESIGN

The Rural–Urban Migration in China and Indonesia (RUMiCI) project was established in 2006 to study the patterns and effects of migration in China and Indonesia. The goal is to increase understanding of the urbanization process in both countries, and inform policy makers so that they can manage this process more effectively. Given that many existing studies on migration in China are plagued by data limitations, a substantial contribution of the RUMiCI project is to provide a rich, up-to-date dataset for migration studies in both countries.

The China component of the study, Rural–Urban Migration in China (RUMiC), is based on a sample of 5,000 urban households, 8,000 rural households and 5,000 migrant households. We intend to track these households over five years. Three independent surveys are being conducted to collect the data necessary for the research: the Rural Household Survey, the Urban Household Survey and the Urban Migrant Survey. The former two surveys are being conducted by China’s National Bureau of Statistics using a random sample from the standard annual household income and expenditure surveys the bureau carries out in cities and rural areas. The Urban Migrant Survey is being conducted by the RUMiC project team in collaboration with Datasea Marketing Research, a Shanghai-based professional survey company.

### **Background of Migration Surveys in China**

While many have articulated the desirability of longitudinal surveys for migration studies (see, for example, Stark 1976 and Biltsborrow 1998), collecting data on migrant populations in developing countries is known to be very difficult. The problems include complications in registering migrants, poor administrative procedures and the mobile nature of the migrant population. Biltsborrow,

Oberai and Standing (1984: 80) have described some of the difficulties typically associated with migrant surveys, especially in less industrialized countries such as Indonesia and China. In Indonesia, for example, the existence of seasonal and temporary workers makes it difficult to keep track of migrants. In addition, there are inconsistencies in government decision making, meaning, for example, that someone recorded as a temporary visitor in one area may be recorded as a new resident in another area.

In addition to the general challenges facing migration surveys, there are a number of specific issues complicating the investigation of migration patterns in China (Day and Ma 1994; Davin 1999). First, defining a migrant can be problematic, with duration, motivation and distance from home all needing to be specified. Differences in the definitions used by agencies have inevitably led to differences in the estimated size and composition of the migrant population. Second, the sheer size of the migrant population, already in the order of 120 million in 2004 (Sheng and Peng 2005), makes a nationwide survey of migrants a very challenging and expensive proposition. Third, migrants in general are a relatively vulnerable group, often treated as second-class citizens. This makes them particularly difficult to interview; out of a sense of self-preservation, they are naturally reluctant to provide detailed information on their livelihoods and social networks. Fourth, many migrants are missed by the usual dwelling-based surveys because they live at or near their workplaces in makeshift housing and factory dormitories, or in inaccessible locations such as shanty towns, squatter settlements or slums. To overcome difficulties such as these, the RUMiC project team developed a set of strategies embedded in the listing, sampling, interviewing and tracking process. These are discussed in sections 2–5 of this chapter.

### **Overview of the Survey Design**

The RUMiC survey is one of the largest of its kind and the first large-scale longitudinal survey to focus on patterns of migration in China. It aims to collect data on three populations: rural households both with and without migrants (through the Rural Household Survey); urban residents (through the Urban Household Survey); and rural-to-urban migrants (through the Urban Migrant Survey). The underlying consideration is to use rural households without migrants as a control group to study the effects of rural–urban migration on the living standards, socio-economic conditions and productivity of rural households; and to use urban residents as a control group to study patterns of social and economic assimilation among rural residents who have migrated to the city.

This chapter focuses on the design of the Urban Migrant Survey, which began collecting data in 2008. The target population is all rural–urban migrants,

that is, migrants who were registered in a rural area but living in an urban area in 2008.<sup>1</sup> The Urban Migrant Survey covers 15 cities in nine provinces or metropolitan areas, namely Shanghai, Guangdong, Jiangsu, Zhejiang, Anhui, Hubei, Sichuan, Chongqing and Henan. The first four provinces are China's largest migrant destinations; the remaining five are among the largest sources of migrants.<sup>2</sup> The distribution of the sample across the 15 cities is loosely associated with the overall population size of the city, with larger cities like Shanghai and Guangzhou being allocated larger sample sizes. The sample allocated to each city is shown in Table 7.1 (column 7). The location of the 15 cities is shown in Map 1.1 on page 4 of this book.

In designing the Urban Migrant Survey, the RUMiC researchers considered the need to (1) obtain an accurate listing of migrant workplaces; (2) institute reliable sampling procedures; (3) develop a comprehensive and meaningful questionnaire; and (4) track migrant households over the life of the study.

Because of the mobile and temporary nature of China's migrant population, there was no existing sampling frame that the RUMiC team could use for the Urban Migrant Survey. This meant that we could not follow the commonly used stratification sampling method. As a result, our first task was to establish a sampling frame that would provide a reasonable representation of the target population. Developing an unbiased, representative sampling frame required us to conduct a pre-survey 'census'—or listing—to gain an overall picture of that population. Using the listing results as a sampling frame, a total of 5,000 migrant households were selected randomly for face-to-face interviews. The information collected through these interviews provided the basis for subsequent research and analysis. The final step in the survey design was to develop a method to track respondents. We intend to interview as many of the respondent households as possible throughout the life of the survey, in order to capture changes over time. This will allow us to compile a genuine longitudinal dataset of Chinese migrant workers over a five-year span.

Throughout the design process, the RUMiC project team worked closely with Datasea to ensure that the survey was tested thoroughly in the field before being implemented. Together with Datasea, the RUMiC team tested the survey design, oversaw the training and supervision of enumerators and conducted quality control. Datasea and its subcontractors were responsible for the recruitment and training of the enumerators, actual data collection and data entry. The enumerators were mainly professional survey staff and university students.

The first questionnaire was drafted in 2006. Between 2006 and 2007, the draft questionnaire was tested twice in Beijing and once in the rural province of Hebei, each time involving 10–20 interviews. The complete listing and sampling procedures were tested between May and July 2007 in two major pilot studies covering 1,000 migrant households. Each involved one large and one medium-sized city: Shanghai and Wuxi in the first instance, and Guangzhou

and Shenzhen in the second. The formal listing procedure for the first wave of the survey was conducted between December 2007 and January 2008 in all 15 cities. The sampling and interview process took place between February and April 2008.

The rest of this chapter provides an overview of the Urban Migrant Survey's listing procedures, sampling strategy, questionnaire design and tracking methods. The listing and sampling methods employed in the survey have several innovative features that require some elaboration, so we focus mainly on them. The following description of sampling procedures and survey methodology should provide a useful framework for interpreting the data and research findings described elsewhere in this book.

## 2 LISTING OF MIGRANT WORKERS AND WORKPLACES

In China, there is a distinct gap in knowledge about the migrant population in any city at a particular point in time, because of problems with the sampling frame used for most household surveys. The existing surveys are generally based on an individual's current residential address and the place of domicile shown in a person's household register (*hukou*).<sup>3</sup> The *hukou* shows the place of registration of the household as a whole, as well as a range of personal details for each individual member of the household. In the case of permanent residents, whether urban or rural, the actual place of residence would usually be consistent with the *hukou*, meaning that there is an official record of the residential addresses of those people. For migrants, students and other temporary visitors, the residential address would be inconsistent with the address shown on the *hukou*. Such persons are required to register with the local authorities as temporary residents, so the authorities' official registers of temporary residents could be expected to provide information on these populations.

In the case of rural-to-urban migrants, however, the official register of temporary residents is largely incomplete. Many migrants do not register at all, and a large number live at their workplaces, without residential addresses as such. These migrants would not be captured by the existing surveys using administrative records of residential address as the basis for migrant survey sampling. The extent of undercoverage of this group of migrants, and how it would affect the results, is difficult to gauge, but it is safe to assume that many of the sampling frameworks used in China are not representative. This bias would in turn compromise the analysis and lead to an incorrect understanding of the migrant situation. The fundamental challenge was therefore to establish a random, representative sampling frame that could provide an accurate picture of the migrant population in all survey cities.

To overcome the problems associated with residential address-based sampling, we decided to conduct a listing covering a large number of workplaces in each city. We included all businesses in randomly selected enumeration areas within defined city boundaries. During the listing, we recorded the total number of staff and the total number of migrant workers in each workplace. This allowed us to estimate the total size of the migrant worker population in each city. The listing-based information on the size of the migrant population was designed to be representative of that city and to provide a sampling frame for subsequent random sampling and the interview process.

The listing procedure was conducted in three steps: we defined each city's boundaries; we divided each city into blocks and randomly selected enumeration areas; and we conducted a listing in each of the selected areas using in-built quality controls.

### **Defining a City's Boundaries**

The Urban Migrant Survey listing did not necessarily cover all areas within a city's administrative territory. Rather, the definition of a city's boundaries took into account both the feasibility of conducting a listing and the necessity to obtain a representative sampling frame. Accordingly, two general principles guided the identification of a city's boundaries: (1) the within-boundary area had to cover as many workplaces as possible where migrants might be employed; and (2) large empty spaces where little economic activity took place would not be included in the within-boundary area. To make sure the boundaries did not exclude any cluster of migrants, we imposed an additional rule: that there should not be any concentrated residential area or production site within 1 kilometre of the defined survey boundaries. If there was such an area, the boundaries would be expanded until this condition was met.

In practice, the areas where most economic activity takes place are far more concentrated than residential areas. As a result, most cities as defined by their survey boundaries proved to be smaller than their actual administrative territories. Although it was admittedly not possible to completely cover the urban city fringe or every business in a survey city, such omissions were minimized by defining the city boundaries to cover as large an area as possible. In Shanghai, for instance, nine of the city's 18 administrative districts were fully within the survey boundaries, and another four were partially within the boundaries.<sup>4</sup>

### **Dividing Cities into Blocks and Randomly Selecting Enumeration Areas**

The area within the survey boundaries of each city was divided into equal-sized blocks averaging 0.25 square kilometres in size, using up-to-date, equal-scale maps. These blocks were numbered in a consistent order.<sup>5</sup> The numbers

assigned to the blocks were subsequently used for the random selection of enumeration areas.

The number of blocks randomly selected to be included in the listing process was proportional to the sample size for each city. The actual ratio was 12 per cent of the sample size, of which 10 per cent would be used for the initial listing and 2 per cent would be kept in reserve in case there was a problem with the main sample. In Shanghai, for example, we started with more than 2,000 equal-sized blocks of 500 by 500 metres in size within the identified city boundary. As the sample size for Shanghai was to be 500 households, 60 blocks were selected randomly: 50 for the main sample and 10 for the reserve sample.

The way in which the blocks were selected was as follows. Three groups of random numbers were generated and the blocks with those numbers were marked on the map to indicate the distribution of the enumeration areas. Strictly speaking, each group of blocks had been chosen randomly, so the selection process was statistically sound. However, the businesses where migrants work are not distributed evenly across blocks. Considering this unevenness and the relatively small number of blocks selected for listing, we decided to discard the groups that excluded major clusters of migrant workplaces. For each city, we chose the single group of blocks that covered the most migrant workplaces. In doing so, we admit that we introduced a degree of arbitrariness into the selection of enumeration areas, and acknowledge the potential bias that may result.

Once the choice of enumeration areas was finalized, detailed maps were prepared for the next step in the listing process. These detailed maps allowed the supervisors to delineate the actual boundaries of each enumeration area and identify the main buildings within each block. This was an important step, because although the blocks were of equal size and shape on the map, the actual boundaries needed to take account of the presence of road and transport networks. As one would expect, the end result was irregularly shaped blocks whose boundaries were defined by roads in and around the area. Supervisors then decided the routes that enumerators would follow in conducting the listing.

### **Conducting the Listing and Ensuring Quality Control**

The listing procedure required enumerators to visit every workplace within a defined enumeration area, whether formal or informal, permanent or temporary—or even mobile. The enumerators first recorded the name and address of the business, then went on to enquire about staff. The information collected during this process included total staff numbers, the number of migrant workers employed, the contact details of the person providing the information and a range of basic descriptive indicators on the business, such as floor area and type of industry.

While the normal listing procedure required enumerators to obtain information on every business, we made an exception for large wholesale and retail markets (those with more than 50 stalls) on the basis that the stalls in such markets were likely to be relatively homogenous in size and industry type. In such cases, the enumerators were required to list the total number of businesses in the market and visit 10 per cent of them.

We recognized that taxi drivers, domestic servants, residential cleaners and migrants working in some other occupations would probably not be encountered by enumerators following the normal procedure of visiting workplaces. To ensure that they were not left out of the listing procedure, we developed a number of special procedures to contact such migrants. Where this was not possible, we developed statistical procedures to compensate for the missing information.<sup>6</sup>

An important extra step included in the Urban Migrant Survey listing was to ask enumerators, after they had recorded all the information they could obtain through an enquiry, to record their own subjective assessment of the reliability of the information provided. The enumerators were also instructed to provide their own estimates of the number of migrant workers at each workplace based on observation of such factors as floor area, industry type and the number of migrant workers in neighbouring shops of similar scale. Where business owners refused to provide information, these subjective measures proved particularly useful, although the subjectivity of this information required special treatment.

At the end of the listing process, supervisors compiled all of the information from every enumeration area to obtain a complete set of information that was representative at the city level. The Urban Migrant Survey listing covered more than 100,000 businesses and about 550,000 migrant workers in total. Table 7.1 replicates the summary information from Gong et al. (2008) and provides some basic statistics on the migrant population derived from the listing conducted in December 2007. The first column shows the number of enumeration areas in which a listing was conducted. The second shows the total number of blocks selected for enumeration, including reserve blocks. Column 3 indicates the total number of migrants listed in these enumeration areas. Column 4 presents the migrant population density per enumeration area. Based on the average number of migrants per enumeration area (column 4) and the total number of blocks within the defined city boundary (column 5), we estimated the aggregate number of migrants in each city (column 6).

Among the 15 cities, Dongguan in Guangdong province has the highest density of migrants (2,614 migrants per 0.25 square kilometre), followed by Shenzhen (2,295), Wuxi (1,691) and Guangzhou (1,171) (see column 4). Coinciding with anecdotal evidence, it is apparent that Dongguan, Shanghai, Guangzhou and Shenzhen are some of the most popular destinations for migrants (see column 6). Based on the listing results, the total number of migrants in the 15

Table 7.1 China: Basic Information on the Listing of Migrant Workers by City

Province	City	No. of Enumeration Areas in which Listing Was Conducted	Total No. of Enumeration Areas Selected (main + reserve)	Total No. of Migrants in Enumeration Area	Mean No. of Migrants per Enumeration Area	Total No. of Blocks within City Boundaries	Implied Total No. of Migrants	Total No. of Migrants to Be Sampled
		(1)	(2)	(3)	(4) = (3)/(1)	(5)	(6) = (5)*(4)	(7)
Guangdong	Guangzhou	40	48	46,822	1,171	1,048	1,226,736	400
	Shenzhen	30	36	68,851	2,295	474	1,087,846	300
	Dongguan	30	36	78,432	2,614	886	2,316,358	300
Shanghai	Shanghai	50	60	40,293	806	2,050	1,652,013	500
Jiangsu	Nanjing	40	48	30,352	759	893	677,608	400
	Wuxi	20	24	33,823	1,691	677	1,144,909	200
Zhejiang	Hangzhou	40	48	45,540	1,139	512	582,912	400
	Ningbo	20	24	18,637	932	249	232,031	200
Hubei	Wuhan	40	48	39,060	977	1,004	980,406	400
Chongqing	Chongqing	40	48	39,792	995	640	636,672	400
Sichuan	Chengdu	40	48	36,145	904	590	533,139	400
Anhui	Hefei	35	42	21,152	604	618	373,484	350
	Bengbu	20	24	5,689	284	121	34,418	200
Henan	Zhengzhou	35	42	28,604	817	655	535,303	350
	Luoyang	20	24	8,600	430	296	127,280	200

Source: Gong et al. (2008).



survey cities is estimated to be 12 million, or slightly less than 10 per cent of the official figure of 126 million for the entire country in 2005 (NBS 2006a).

During the listing period, the enumerators visited around 100,000 businesses. About 30 per cent of them were small groceries, restaurants or cafés and another 10 per cent were clothing wholesalers or retailers. These figures are reflected in the high proportion of migrants in Table 7.2 working in services (34 per cent) and wholesale and retail trade (33 per cent). To illustrate the significance of the sampling frame, Table 7.2 compares the Urban Migrant Survey listing results with those of three other urban-based datasets on migrant workers. The results for the Urban Migrant Survey find more migrant workers engaged in construction and manufacturing, and fewer in wholesale and retail trade.<sup>7</sup> It is possible that the choice of survey cities and the timing of the listing may have contributed to the differences in occupational distribution between the Urban Migrant Survey and the other surveys. However, a more important reason is the underlying sampling method. The three earlier surveys were all based on random sampling of residents in cities. As noted earlier, this approach effectively excludes from the sampling frame all migrants who live at their workplaces.

At the end of the listing process, a group of quality control officers carried out verification procedures to check the completeness of the records of businesses and confirm the accuracy of the information obtained. The officers covered 10 per cent of the enumeration areas in each city during this process, visiting one main street and one smaller street in each area.<sup>8</sup> A 10 per cent margin was allowed for the information completeness check, and a 30 per cent margin was allowed for the accuracy of the information obtained at a specific business. Where there were discrepancies greater than the allowed margins, the listing procedure was repeated for the rest of the enumeration area. If the information for an enumeration area was found to be grossly incomplete or inaccurate, the listing procedure was repeated for all enumeration areas conducted by the enumerator in question. The quality control procedure revealed that the quality of the listing was generally satisfactory. On average, less than 5 per cent of businesses had been omitted and discrepancies in the number of migrant workers ran at around 15 per cent.

### 3 SAMPLING

The next step was to select a simple random sample of migrants in each city for interview. This section first describes how the sample of migrants was drawn from the listing data. It then discusses how the sampling procedures were implemented in the field.

*Table 7.2 China: Listing and Sample Information on the Distribution of Migrant Workers by Industry*

City	Construc- tion (%)	Manu- facturing (%)	Educational & Govt Agencies (%)	Other Agencies (%)	Serv- ices (%)	Wholesale & Retail Trade (%)	Total No. of Migrants
Chengdu	10.7	3.1	0.6	6.4	45.8	33.5	36,145
Hangzhou	7.1	17.6	2.5	2.7	37.0	33.2	45,540
Nanjing	16.9	15.9	1.2	10.2	38.2	17.7	30,352
Ningbo	26.1	9.0	5.1	10.1	28.2	21.6	18,637
Shanghai	8.4	10.9	1.5	11.9	41.3	26.1	40,293
Wuhan	5.2	28.9	3.1	2.9	36.5	23.5	39,060
Wuxi	9.4	40.8	2.0	6.0	22.0	19.8	33,823
Bengbu	7.9	13.8	0.3	1.4	30.3	46.2	5,689
Chongqing	9.2	4.7	1.2	0.8	48.8	35.3	39,792
Dongguan	1.2	26.3	3.2	1.7	12.6	55.0	78,432
Guangzhou	1.3	23.2	2.9	6.1	40.1	26.4	46,822
Hefei	10.1	5.4	1.3	8.9	36.3	38.0	21,152
Luoyang	8.3	12.9	2.2	0.5	38.9	37.2	8,600
Shenzhen	7.2	21.8	1.0	2.8	36.3	30.9	68,851
Zhengzhou	21.5	8.2	1.3	1.6	34.5	33.0	28,604
<b>Total</b>	<b>8.4</b>	<b>18.2</b>	<b>2.0</b>	<b>4.7</b>	<b>34.0</b>	<b>32.7</b>	<b>541,792</b>
IOE, 1999 <sup>a</sup>	4.3	12.7	3.8	0.5	36.6	42.2	1,254
IOE, 2002 <sup>a</sup>	5.5	9.8	2.9	0.7	34.1	47.0	3,407
ULS, 2001 <sup>b</sup>	10.3	7.3	3.0	0.9	38.7	39.9	2,205
<b>Sample of respondents aged 15+</b>							
(no.)	1,089	1,391	558	285	2,052	2,065	7,440
(%)	10.9	18.8	6.4	4.3	31.9	27.7	100

a The Institute of Economics (IOE) at the Chinese Academy of Social Sciences conducted two urban-based migrant household surveys, the first covering 13 cities with 780 households (1,785 individuals) in 1999; and the second covering 12 provinces (27 cities) with a total of 2,000 households (5,327 individuals) in 2002. On average, the IOE surveyed less than 100 households in each city.

b The China Urban Labour Survey (ULS) was conducted by the Institute of Population and Labour Economics at the Chinese Academy of Social Sciences in 2001. Undertaken in five large cities (Shanghai, Wuhan, Shenyang, Fuzhou and Xian), it covered 340 migrant households and 2,365 individuals.

Source: Gong et al. (2008).

### **Random Selection of the Sample**

Theoretically, for each survey city, the total number of migrants working in all enumeration areas could be aggregated from the listing information for each business in each enumeration area. This aggregate number of migrants would then serve as the sampling frame for that city. In practice, however, the execution of this process needed to take account of two things. First, recall that the information collected in large markets did not cover every single business, so the businesses omitted from the listing process would have to be taken into account. Second, in reality it was not always possible to collect information on the number of migrants working in a particular business, because the owners of the business refused to provide the relevant information or the enumerators were refused entry to the premises. In such cases, the enumerators could still record descriptive information such as the industry category, location and estimated size of the business, but crucial information on the number of migrant workers employed in the business would be missing.

We addressed these two issues with the aid of a regression model. First, based on the assumption of homogeneity among businesses in large markets, we used the information collected on 10 per cent of the businesses in each market to estimate the total number of migrants working in all markets in each city. The second step in completing the sampling frame was to account for migrant workers who were not included in the listing. To do this, we constructed a regression model for the number of migrant workers employed in a given workplace, based on determinants such as total number of staff, floor area of the business, type of ownership, industry category and so on. We then used the information we had collected on these factors to predict the probable number of migrant workers in the businesses missed by the enumerators. This exercise was performed for each city to complete the listing information and provide a representative sampling frame for the subsequent sample selection.

To calculate the sampling factor for each city, we divided the sample size by the total number of migrant workers (including predications) in the sampling frame. This factor was assigned to each migrant worker in the sampling frame, giving all migrant workers in the frame an equal probability of being selected.<sup>9</sup> The next step was to randomly select the requisite number of migrant workers for each city, and for each workplace. Recall that the listing database contained business-level information on the number of migrant workers employed at each workplace. Therefore, at the end of this stage, each business was associated with a figure indicating the number of migrant workers to be selected.

### **Implementation of Random Sample Selection**

We now knew the number of migrants to be selected for interview at each workplace. However, to ensure an unbiased sample, we also needed to define

a procedure that would enable us to randomly select the designated number of migrants at each workplace. To describe how such a sample was selected in practice, three representative situations are described here.<sup>10</sup>

- 1 To choose one respondent from a workplace that employed only one migrant worker, sampling officers were instructed to identify the migrant worker and ask him or her to participate in the Urban Migrant Survey.
- 2 To choose one worker from a workplace that had more than one but less than 10 migrant workers, officers were asked to use the ‘random birth-month’ method—that is, to select the worker whose birth month was closest to a randomly chosen birth month selected previously.
- 3 To choose more than one respondent from a workplace with 10 or more migrant workers, sampling officers were instructed to apply a combined ‘random interception and random birth-month’ method, that is, to randomly intercept migrant workers and choose the ones whose birth months matched a randomly chosen birth month selected before the sampling process took place.<sup>11</sup>

The application of these three methods was supplemented by a set of rules to guide situations where the sampling process was unsuccessful, because of a lack of cooperation, for instance, or because of difficulties in gaining access to a selected workplace. These rules regulated when and how substitution of a selected migrant could take place. In practice, around 60 per cent of the total sample was chosen from businesses randomly selected by the original computer program; selection of the remaining 40 per cent involved substitution during the sampling process.<sup>12</sup> It is important to emphasize that sample substitution took place only between businesses of similar size in the same industry. Sampling officers were not permitted to substitute respondents from small-scale businesses with those from large-scale enterprises, or vice versa. Had such substitutions been allowed, the distribution of migrant workers by occupation and employment would have been biased.

Based on the sample selection, the Urban Migrant Survey collected a wide range of information on migrants and migrant households. The sample covered 5,003 households and 7,440 adult household members (defined as those aged 16 and above). The last two rows of Table 7.2 indicate that the occupational distribution of the sample is broadly consistent with the listing information.

## 4 QUESTIONNAIRE-BASED INTERVIEWS

The research focus of the RUMiCI project goes beyond the conventional topics of migration determinants and labour market discrimination to cover the

welfare status of migrants: their jobs, incomes and physical and mental health, their children's education and health, and the extent to which they assimilate into their city communities. This will allow researchers to explore a number of important aspects of migration that have so far been underresearched. This section briefly describes the structure of the questionnaire and the main survey instruments.<sup>13</sup>

A household-based questionnaire was employed for the Urban Migrant Survey. A household was defined as anyone who was living with the respondent at the time of the survey, sharing income and expenditure. The questionnaire had two main components: one to obtain individual-level information and the other to obtain household-level information. The individual-level component covered four areas: (1) household composition; (2) adult education; (3) adult employment; and (4) children.

Respondents were first asked to provide information to complete a household roster describing the household composition and basic characteristics of each household member: age, marital status, whether or not the individual was a member of an ethnic minority, work capability and so on. This section also included a set of general questions focusing on the mental health of adults.<sup>14</sup>

The sections on both education and employment were targeted at household members aged 16 and above who had completed their schooling. The questions on education focused on schooling and training. The questions on employment were designed to obtain detailed information on the labour market performance of migrant workers and other adults in the same household. This section of the survey first asked a set of general questions to identify the employment status of each individual. It then asked specific questions that differed depending on the respondent's employment status: wage worker, self-employed or not working at the time of the survey. This part of the questionnaire concluded with a set of questions on the migrant's first job in the city. The responses should allow researchers to extract historical information and observe how some labour market indicators, such as migrant workers' wages, change over time.

The final section of the individual component contained about 50 questions on migrant workers' children, defined as the sons or daughters of the household head (or other household members), and either aged under 16 or aged 16 and above and still at school at the time of the survey. Both children who were living in the city with their parents and those who had been left behind in the countryside were covered, to allow researchers to investigate the impact of migration on migrated children's health and education.

The household head (or that person's spouse) was asked to answer the household-level questions in the Urban Migrant Survey. This component covered (1) social networks; (2) lifecycle events; (3) household income and expenditure; (4) household assets; (5) housing conditions; and (6) information on the rural home village. The social networks section contained several sub-

sections covering spouses living separately, adult children living separately, the parents of both the household head and that person's spouse, and general information on the household's social networks, including information on up to five close associates.

All except the social networks section were relatively standard and self-explanatory. The questions on social networks were designed to capture the extent and depth of migrants' social connections, given the substantial literature on the sociological and economic role of social networks. It is well established that social networks have important implications for the intra-family transfer of income and wealth and the internal arrangement of household and social responsibilities (Montgomery 1992; Benabou 1993; Granovetter 2005). This part of the questionnaire therefore sought to gather data that would improve our understanding of the assimilation process.

In summary, the Urban Migrant Survey questionnaire is one of the most comprehensive ever developed for migration studies. It connects closely with the research questions of the RUMiCI project and has the potential to be used to investigate a wide range of questions of interest.

## 5 TRACKING

The RUMiCI project is designed to provide a longitudinal dataset covering a five-year time span. However, migration is an extremely dynamic process where migrants may move frequently or relocate across vast distances. We therefore designed a set of tracking strategies to enable us to maintain contact with respondents from year to year. The effectiveness of these strategies, and the attrition rate, will become clearer over time. We intend to modify and improve our tracking strategies to deal with problems as they arise, so this section will provide just a brief account of the current plan.

We intend to track respondents so long as they remain in the surveyed provinces and cities. While it would be valuable to follow migrants who return to their rural areas of origin, such an undertaking would not be feasible given the well-known problems and high cost of tracking returning migrants. To reduce the attrition rate, we developed a number of strategies to keep in touch with respondents between waves of the survey. On both the front and last pages of the questionnaire, household heads were asked to supply detailed contact information, including current address, address in the home village, telephone number, and names and contact details of three associates. We intend to contact respondents by telephone every three months and mail small presents to the home village at major festival times. This should help us to maintain an up-to-date list of respondents' contact details, in both the city and the home village, over the coming years. To provide an additional incentive for respondents to

stay in touch with the project team, we also set up a lottery scheme in which all respondents were entered at no cost to them.<sup>15</sup>

Of course, despite our efforts, we were aware that we would lose a number of respondents. When we conducted the first round of tracking and lottery activity about five months after the first wave of the survey, we found that we could not contact around 35 per cent of the respondent households. By February 2009, this proportion had reached almost 50 per cent. This high rate of attrition was associated with the severe economic downturn accompanying the global financial crisis. The crisis has had a particularly severe effect on the industries where migrant workers are concentrated, such as manufacturing, construction and services. To address the high rate of attrition, we plan to replace lost respondents with new respondents through a sampling process based on the 2007 listing information.

## 6 CONCLUSION

The Urban Migrant Survey is an important component of a large-scale study of rural–urban migrants in China and Indonesia. It addresses some of the problems with existing migrant surveys in China, which tend to be dwelling-based, one-off case studies. The project team devised a set of procedures and methodologies to generate a longitudinal dataset covering 5,000 migrant households across 15 Chinese cities. It is one of very few scientifically based, random-sample migration surveys in China. This chapter has highlighted the complex strategies adopted to address the current lack of information on the migrant population.

The first step in the survey design was to conduct a city-level listing to obtain workplace-based information on migrant workers in each city. This allowed us to capture most migrant workers within defined city boundaries. This new approach reduced the sample bias inherent in existing urban-based surveys that have used residential addresses as the basis for sampling. The project team then developed a set of protocols to ensure the randomness and representativeness of the selected sample. The third step was to conduct questionnaire-based, face-to-face interviews with respondent households, where a wide range of information was gathered. Subsequently, respondents have been contacted at regular intervals to ensure that as many as possible participate in future waves of the survey. With the availability of an extensive array of data covering a five-year span, the Urban Migrant Survey should contribute substantially to a better understanding of the patterns and effects of migration in China.

## NOTES

- 1 Unless stated otherwise, in the remainder of this chapter the terms 'migrant' and 'rural–urban migrant' are used interchangeably.
- 2 According to the 2000 census, Shanghai, Guangdong, Jiangsu and Zhejiang receive 66 per cent of all migrants, while Sichuan, Chongqing, Anhui, Hubei and Henan send 47 per cent of all migrants.
- 3 On the *hukou* system, see Cheng and Selden (1994), Chan, Liu and Yang (1999), Chan and Zhang (1999) and Liu (2005).
- 4 The four districts that were partially covered were Baoshang, Minhang, Jiading and New Pudong. The five districts that were excluded from the survey were Songjiang, Jinshan, Qingpu, Nanhui and Fengxian. The official website of the RUMiCI project provides detailed maps and information on the survey boundaries of all 15 cities; see <http://rumici.anu.edu.au>.
- 5 During the numbering process, blocks that contained mostly river or parks were excluded, as they would contain few businesses.
- 6 For a complete account of these procedures, see 'Urban Migrant Survey document: census manual' at <http://rumici.anu.edu.au>.
- 7 See Gong et al. (2008: 118–19) for a more detailed discussion of the aggregation of occupational categories as well as background information on previous urban-based surveys.
- 8 The tasks performed by quality control officers are described in more detail in the Urban Migrant Survey listing procedure; see 'Urban Migrant Survey document: census manual' at <http://rumici.anu.edu.au>.
- 9 Most statistical software packages make it easy to generate a random sample once a sampling factor has been specified. We used the STATA statistical package to randomly select the designated number of migrant workers for each city, plus an extra 30 per cent in case a reserve sample was needed.
- 10 For a detailed account of the procedures used for taxi drivers, street vendors, cleaners and other 'special situations', see 'Urban Migrant Survey document: sampling manual' at <http://rumici.anu.edu.au>.
- 11 In the second and third situations, sampling officers were also allowed to use an alternative method based on rosters of workers. In practice, however, this method was hardly used because complete rosters of migrant workers were rarely available.
- 12 Information on the extent of sampling substitution is based on estimates provided by the local supervisors who conducted the sampling process.
- 13 For a copy of the questionnaire, see 'Urban Migrant Survey document: interview' at <http://rumici.anu.edu.au>.
- 14 These questions were based on the 12-item General Health Questionnaire (GHQ12), which is widely used as a general measure of mental health status. All adult members who were present at the time of the survey were asked to answer this set of questions.
- 15 The lottery is set up in such a way that about 20 per cent of respondents receive prizes. We are hopeful that the relatively high chance of winning a prize will encourage respondents to stay in touch with the survey team, even though the prizes themselves are modest.