**The Study of Family Life in Urban China**

**Principal Investigators:**

Yu Xie

Population Studies Center

The University of Michigan

and

Zhongdang Pan

Communication Science

University of Wisconsin-Madison

**Methodology Description**

1. Introduction

The Study of Family Life in Urban China, also referred to simply as the “Three-City Survey,” is a 1999 survey of urban residents about their family lives in three large Chinese cities: Shanghai, Wuhan, and Xi’an. The study was a result of close collaboration between Professor Yu Xie at the University of Michigan and Professor Zhongdang Pan, then at the Chinese University of Hong Kong. Xie was interested in social and economic aspects of the family, while Pan was interested in cultural practices and civic values. Pooling resources from the National Science Foundation, the National Institute for Aging, the Population Studies Center at the University of Michigan, and the Chinese University of Hong Kong, we (Xie and Pan) conducted the survey in Shanghai, Wuhan, Xi’an, three of the largest cities in China, during May-August 1999.

The survey contains a main sample as well as an intergenerationally matched sample. At each research site of the Three-City Survey, we initially targeted a probability sample of about 1,300 adults (18 or older) as the main sample. Whenever possible, we also interviewed the respondent’s adult child (for respondent 61 years of age or above) or elder parent (for respondent 60 years of age or younger) who also lived in the same city. To avoid errors in implementation, we designed distinct questionnaires for the four groups of respondents. If the initially selected person was 60 years old or younger, we used Questionnaire A, and then interviewed the selected parent with Questionnaire A+. If the main sample respondent was age 61 or older, we used Questionnaire B, and the selected adult child was interviewed with Questionnaire B+. Forms A and B+ are identical except that Form A contains additional questions for paired respondent selection. Forms B and A+ are identical except that Form B has additional questions for selecting an adult child as a paired respondent. Table 1 shows the designation of questionnaire form in the main sample and the paired sample. All four questionnaires are provided to users in both English and Chinese.

**Table 1. Designation of Questionnaire Form**

|  |  |  |
| --- | --- | --- |
| **Respondent’s Age** | **Main Sample** | **Paired Sample** |
| **18 – 60** | A | A+ |
| **61 or above** | B | B+ |

In each city, the main sample was drawn using a two-stage probability sampling method. At the first stage, we drew 50 neighborhood communities at random. Within each selected neighborhood community, 26 households were randomly chosen. A total of roughly 1,300 households were thus targeted in each city. Next, a Kish table was used to select an adult respondent within each selected household. Table 2 shows the distribution of the main sample by city and the response rates.

Table 2. Distribution of the Main Sample by City

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **City** | | | **Total** |
| **Shanghai** | **Wuhan** | **Xi’an** |
| **Household Selected** | 1,300 | 1,319 | 1,294 | 3,913 |
| **Interview Completed** | 1,022 | 1,140 | 1,019 | 3,181 |
| **Response Rate** | 78.62% | 86.43% | 78.75% | 81.29% |

Face-to-face interviews were conducted to collect the data. Each interview lasted about one hour. Including respondents from the paired sample, a total of 4,561 interviews were successfully completed in the three cities. After data cleaning, it was determined that there were altogether 4,444 usable records. Table 3 presents the distribution of the usable records by city and questionnaire form.

Table 3. Completed Interviews by City, and by Questionnaire Form

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Sample Types** | **Questionnaire Form** | **City** | | |
| **Shanghai** | **Wuhan** | **Xi’an** |
| **Main Sample** | A | 731 | 890 | 705 |
| B | 291 | 250 | 314 |
| **Total** | **1,022** | **1,140** | **1,019** |
| **Paired Sample** | A+ | 197 | 373 | 272 |
| B+ | 119 | 136 | 166 |
| **Total** | **316** | **509** | **438** |
| **Total by City** | | **1,338** | **1,649** | **1,457** |
| **Total** | | **4,444** | | |

The structure of the data file is rectangular with one record for each respondent. It consists of 924 variables and 4,444 unit records. The data are generally organized in the order of the questionnaire items, and the constructed variables are located after the variables obtained directly from the questionnaires. The data have been cleaned to the best available information. Special error codes are assigned for unreasonable or invalid values. (See Section 5 for details.) Should users of this dataset identify and report problems to us, we will double check and provide corrections where appropriate.

The data file is released in STATA format (with the extension name “.dta”) with two companion documents:

1. *Three-City Survey Variable Reference List*. The list indexes names and labels of all the 924 variables in the data file. The user can also find the input variable name for each variable, which indicates the question (and option) on which the variable is based and created.
2. *Three-City Survey Constructed Variables Documentation*. Approximately half of the variables in the data file were subsequently constructed based on existing variables from the questionnaires. This document provides construction details of all these variables.

These two documents, together with the four questionnaire forms and this methodology description, provide all the necessary information for users to work with the data.

For publications using the data, see Xie, Jiang, and Greenman (2008), Xie, Lai, and Wu (2009), Xie and Wu (2008), and Xie and Zhu (2009).

The suggested citation when using this data set is: Xie, Yu, and Zhongdang Pan. 2009. *The Study of Family Life in Urban China*. (http:/yuxie.com/3city).

1. Research Team

Yu Xie and Zhongdang Pan collaborated in designing the survey. With assistance from other collaborators listed below, Zhongdang Pan was primarily responsible for implementing the survey.

Besides the University of Michigan and the Chinese University of Hong Kong, the project involved the cooperation of the Center for Population Information and Research (CPIR) and the Institute for Marketing Information (IMI) at the Beijing Broadcasting Institute in the fieldwork phase of the study. Listed below by institution are members of the research team who were involved in the various phases of the project: (Unless they are noted as professors, they were either students or staff.)

The University of Michigan

Yu Xie (Professor, PI)

Meichu Chen

Seth Hauser

Qing Lai

Zhen Zeng

Haiyan Zhu

The Chinese University of Hong Kong

Zhongdang Pan (Professor, PI)

Yanhong Li

Linchuan Qiu (Jack)

Kai Xu

Wenhong Wang

Jing Zhang

Yong Zhang

The Center for Population Information and Research

Xuejun Yu (Research Fellow and Director)

Wenli Gao

Chunjuan Zhou

The Institute for Marketing Information

Shengmin Huang (Professor)

Jinghua Huang (Professor)

Bing Wang (Professor)

Jin Kang

Chunying Wen

Lin Lin

Ning Fang

Lingli Huang

We would also like to thank Professor Donald J. Treiman of the University of California, Los Angeles, who provided help with the occupation coding and its conversion to prestige score and socioeconomic index.

1. Sample Design

At each research site of the Three-City Survey, we first targeted about 1,300 respondents as the main sample with a two-stage probability sampling method. Based on the main sample, a paired sample was then generated through intergenerational matching.

1. *Main Sample*

For the main sample, two sampling schemes were adopted depending on the availability of neighborhood information. In Shanghai, where the number of households within each neighborhood was known, neighborhood selection was done in proportion to size (Scheme 1). In Wuhan and Xi’an, however, no reliable information was available at the neighborhood level. Therefore, we conducted Simple Two-Stage Cluster Sampling (Scheme 2) based on estimated average neighborhood sizes.

Scheme 1: Proportional-to-Size Two-Stage Cluster Sampling (Shanghai)

Step 1: List all neighborhood committees in Shanghai in a single column, with all neighborhood committees within a district in the same block. This allows implicit stratification by district through systematic sampling. Each neighborhood committee governs and represents a neighborhood. Let us label the neighborhood as j, j = 1, 2… M, where M is the total number of neighborhood committees (i.e. the total number of neighborhoods) in Shanghai.

Step 2: List the number of households of each neighborhood (Nj) next to its neighborhood committee in the next column.

Step 3: Calculate the cumulative number of households Cj = k=1 j Nk. CM=N, the total number of households in the city.

Step 4: Sample 50 neighborhoods. Draw 50 random numbers (R’s) between 1 and N. The jth neighborhood is chosen if an R is greater than Cj-1 and smaller than Cj. If a random number R falls within a neighborhood that has already been selected, a new random number R is drawn to replace it.

Step 5: Calculate the household sampling interval for each selected neighborhood from Step 4. For each selected neighborhood, divide the number of households Nj by 26, where Nj is the number of households in the jth neighborhood and 26 is the number of households to be selected from each neighborhood. The rounded integer of Nj/26, Ij, is the sampling interval for the jth neighborhood.

Step 6: Select 26 households from each selected neighborhood. Array all households in the jth neighborhood and assign sequential numbers from 1 to Nj for all the households. Randomly select a number between 1 and Ij as the random start number (Fj) for the jth neighborhood. The first household selected is the Fjth household. Then select every Ijth household after the Fjth. The actual number of households selected for each neighborhood might not be 26 due to the rounding error of the sampling interval.

Step 7: Use a Kish table to select one adult respondent from each selected household.

Scheme 2: Simple Two-Stage Cluster Sampling (Wuhan and Xi’an)

Step 1: For each city, list all neighborhood committees in the city in a single column, with all neighborhood committees within a district in the same block and assign sequential numbers for all neighborhood committees as 1, 2, …, M, where M is the total number of neighborhood committees in the city. This allows implicit stratification by district through systematic sampling. Each neighborhood committee governs and represents a neighborhood.

Step 2: Compute the sampling interval for selecting neighborhoods (In) as the rounded integer M/50, where M is the total number of neighborhoods in the city and 50 is the number of neighborhoods to be selected (i.e., the sample size of neighborhoods).

Step 3: Sample 50 neighborhoods through a systematic sampling at the interval In. First, select the first neighborhood by randomly drawing a number between 1 and In, the random start number. Then, select every Inth neighborhood.

Step 4: Calculate the household sampling interval for a selected neighborhood (I) as the rounded integer 50N/1300M, where N is the total number of households in the city, M is the total number of neighborhoods in the city, 50 is the sample size of neighborhoods, and 1,300 is the sample size of households. In other words, the household sampling interval for a selected neighborhood (I) is the average number of households in a neighborhood (N/M) divided by the sample size (26) in a neighborhood. Note that the household sampling interval (I) is the same for all neighborhoods.

Step 5: Select households from each selected neighborhood from Step 3. For each selected neighborhood, array all households in the neighborhood and assign sequential numbers from 1 to Nj for all the households. Randomly select a number between 1 and I as the random start number (Fj). The first household selected is the Fjth household. Then select every Ith household after the Fjth from the neighborhood. Note that the number of households selected in each neighborhood is not necessarily 26 by this design.

Step 6: If Step 5 does not yield a sample of 1,300 households, repeat Step 5 with additional randomly selected backup neighborhoods until the target of 1,300 households is reached.

Step 7: Use a Kish table to select one adult respondent from each selected household.

1. ***Paired Sample***

The paired sample was determined when main sample respondents were interviewed. For each respondent age 60 or younger with at least one surviving parent, the sole surviving parent -- or, randomly selected, one of the respondent’s parents --, was interviewed if he/she lived in the same city. Alternatively, if the primary respondent was 61 or older with at least one surviving adult child, one of the respondent’s adult children was randomly selected for interview if he/she lived in the same city. Although the design called for a “random” selection when an elderly parent was first interviewed and multiple adult children were possible candidates, actual field operation allowed some interviewers to take the shortcut of interviewing the coresidential adult child if the elderly person was in a coresidential household. Here, we define coresidential families as those in which an elderly person (60 years or older) lives with his/her adult child. Thus, there may be an upward bias in the incidence rate of coresidence in our data.

1. **Fieldwork**

The Three-City Survey was fielded in Shanghai, Wuhan, and Xi’an in the summer of 1999. The operation in Shanghai lasted from May 13 to June 28 and covered thirteen urban districts: Baoshan, Changning, Hongkou, Huangpu, Jing’an, Luwan, Minhang, Nanshi, Pudong, Putuo, Xuhui, Yangpu, and Zhabei. Wuhan’s fieldwork was conducted from May 13 to June 18 in seven urban districts: Hanyang, Hongshan, Jiang’an, Jianghan, Qiaokou, Qingshan, and Wuchang. The fieldwork in Xi’an was conducted from July 16 to August 24. The sampling area included four urban districts: Beilin, Lianhu, Xincheng, and Yanta. Trained supervisors were sent to each city. With assistance of local collaborators, the supervisors recruited and trained samplers and interviewers in each city, and oversaw the entire field operation.

1. *Sampling*

In all three cities, the sampling was implemented through the following procedures:

First, construct sampling frames. In each city, we created a list of neighborhoods and, whenever possible, a list of the households within each neighborhood. In Shanghai, an official document released by the Municipal Civil Affairs Bureau detailed the neighborhood information including the number of households within each neighborhood. In Wuhan and Xi’an, however, only a list of neighborhoods was available, and we had to estimate the number of households within each neighborhood. All sampling frames went through rigorous cross-validation (i.e., published literature, governmental officials, researchers, and field investigation) before being put into use.

Second, select neighborhoods. As documented earlier, the neighborhood selection in Shanghai was proportional to size while the selection in Wuhan and Xi’an was based on estimated average neighborhood sizes. In later phases of the fieldwork, several selected neighborhoods were found to be nonexistent due to rapid urban changes, and some neighborhood committees refused to cooperate. Therefore, back-up neighborhoods were randomly chosen for replacement. Eventually, data were collected from 50 neighborhoods in Shanghai, 56 in Wuhan, and 58 in Xi’an.

Third, select households. After training, samplers were sent to assigned neighborhoods for household selection. Whenever possible, resident registration cards kept at the neighborhood committees were used to select households. Otherwise, samplers were instructed to produce neighborhood maps with necessary information for household selection. In roughly two-thirds of the neighborhoods in Shanghai and most of the neighborhoods in Wuhan and Xi’an, where we gained official support, resident registration cards were used.

Finally, select respondents. Samplers visited all the selected households, and recorded information for each adult member. Nonexistent and unoccupied households, after reasonable search efforts were made, were replaced or abandoned following a standard procedure. If the household was occupied by other people than those shown on the residence registration card, the current occupants’ information was recorded for respondent selection, as long as they planned to stay until the end of our field operation. After information about all adult members of the selected households was collected, a Kish table was used to select respondents. If the selected respondent was recorded as physically incapacitated or otherwise unfit for interviewing (e.g. deaf, absent from the household during the entire period of the field operation), a replacement would be sought in the same household among members of the same generation, in the order of spouse, the oldest sibling, the second oldest sibling, and so on.

1. *Interviewing*

As a measure of quality control, the interviewing phase followed sampling as a separate stage. Interviewers were mostly recruited from among new applicants. A few samplers stayed with our project and continued to work as interviewers. They were assigned different neighborhoods in which the sampling was done by others. All interviewers went through a probationary period. During the probation, if an interviewer’s work was twice evaluated as unsatisfactory, he/she would be dismissed and the work re-done.

Local official support varied. In Shanghai, about one-third of the neighborhood committees refused to cooperate, which made it difficult to gain the trust of selected respondents. The main sample response rate in Shanghai was 78.62%. Fortunately, most neighborhood committees in Wuhan provided our field workers with effective assistance, yielding a response rate of 86.43% for the main sample. In Xi’an, the local administrations were cooperative, but the field operation was affected by some unexpected complications. For example, part of the city was undergoing a replacement of street number plates. On July 22, one week after the fieldwork started, Beijing launched a national campaign against Falungong (a religious sect banned by the Chinese government). In the end, we achieved a main sample response rate of 78.75% in Xi’an.

When interviewing the main respondents, we also tried to randomly select one of their elderly parents or adult children to construct the matched sample. However, many primary respondents either did not know or refused to provide the contact information of the paired respondent. To ensure an adequate paired sample size, the field operation introduced the following convenience measures: a) If the primary respondent lived with one and only one adult child or elderly parent, that person would be the paired respondent; b) If the primary respondent lived with two or more adult children, the oldest child would be the paired respondent; c) If the primary respondent lived with both parents, the paired respondent would be the father if the serial number on the questionnaire was odd, or mother if the serial number was even; d) If the primary respondent had multiple adult children or both parents living in the same city but did not live with any of them, the person who lived at the closest address would be the paired respondent; if the primary respondent could not tell which relative lived the closest, a Kish table would be used to select one at random. Consequently, we successfully interviewed 316 paired respondents in Shanghai, 509 in Wuhan, and 438 in Xi’an.

Each questionnaire, after completion, went through three rounds of inspection. Immediately after each completed questionnaire was submitted, a supervisor would check through it item by item, together with the interviewer. If logical inconsistencies or recording errors were identified, the interviewer was responsible for clarification and/or correction. Whenever necessary, the respondent would be contacted or even re-interviewed. A different supervisor would then conduct the second-round inspection. In case of any problems, the questionnaire would be returned to the first-round inspector, who would, in turn, contact the responsible interviewer to fix the problems. Finally, two to three supervisors would work on the final inspection collaboratively.

1. *Data Input*

All questionnaires, after three rounds of inspections, were sent to Beijing for data input. The field supervisors continued to supervise the data input, for which thirty-two university students were recruited and trained. They worked in sixteen pairs, with one reading from the questionnaire and the other entering the data on the keyboard. Each questionnaire was entered twice by different pairs of students. Any inconsistency between the two independent inputs would be detected and refused by the program, and thus a correction was called for.

1. **Data Cleaning and Editing**

The data was cleaned at the Chinese University of Hong Kong and the University of Michigan. We relied heavily on the principle of internal logical consistency. All the variables were checked individually, especially if they were contingency questions, either explicit or implicit. When suspicious data were detected, the original questionnaires were double-checked. If problem persisted, special error codes were assigned.

1. ***Special codes***

In the data file, all variables directly obtained from questionnaires were initially coded and labeled as designated on the questionnaires. However, the special codes (i.e., “Not Applicable”, “Don’t Know”, missing data, and erroneous data) specified on the questionnaires were not consistent across items. For instance, “Don’t Know” is indicated by 9, 99, 999, and -9 in different questions; and 98, 8, -8, and -6 are used for “Not Applicable”. To avoid mistakes in data use, we have standardized the special codes as the following:

* -99, -999, -9999, or -999999 for blanks when responses are expected.
* -98, -998, -9998, or -999998 for “Don’t Know”.
* -97, -997, -9997, or -999997 for constructed variables that miss crucial information when responses are expected.
* -88, -888, -8888, or -888888 for “Not Applicable” due to the particular questionnaire form that was used for the interview. For example, the question for the variable DISEASA was only applicable for elderly respondents (i.e., FORM=A+ or B). Therefore this variable was coded -88 for all young respondents (i.e., FORM=A or B+). Similarly, the variable NUMSONS was only applicable for young respondents. Therefore, it was coded -88 for all elderly respondents.
* -87, -86, -85, and onward for “Not Applicable” due to skipped question by design or logic. For the first skip -87 was used, and the second skip was coded -86, and the third -85, etc.
* -77, -777, -7777, or -777777 were used for erroneous data. When a value exceeded the possible range or was otherwise unreasonable and could not be fixed by double-checking the completed questionnaires, an error code was assigned.

1. ***Occupation Coding***

The occupation variables were coded using the Chinese Standard Classifications of Occupations (CSCO) developed by Professor Donald J. Treiman. Each occupation variable was converted to three other versions: the International Standard Classification of Occupations 1968 Edition (ISCO68), the Standard International Occupations Prestige Scores (SIOPS), and the International Socioeconomic Index (ISEI).

References

Xie, Yu, Yang Jiang, and Emily Greenman. 2008. “Did Send-Down Experience Benefit Youth? A Reevaluation of the Social Consequences of Forced Urban-Rural Migration during China’s Cultural Revolution.” *Social Science Research* 37: 686-700.

Xie, Yu, Qing Lai, andXiaogang Wu. 2009. “*Danwei* and Social Inequality in Contemporary Urban China.” *Sociology of Work* 19:283-306.

Xie, Yu, and Xiaogang Wu. 2008. “*Danwei* Profitability and Earnings Inequality in Urban China.” *The China Quarterly* 195:558-581.

Xie, Yu, and Haiyan Zhu. 2009. “Do Sons or Daughters Give More Money to Parents in Urban China?” *Journal of Marriage and Family.* 71:**174-186.**