World Values Survey, China 2018

Sampling and Fieldwork Report

Center for Public Opinion Research

Shanghai Jiao Tong University (SJTU)

January 2019

Principal Investigator	Zhong Yang, Professor, SJTU
Sampling Specialists	Yan Jie, Associate Professor, Peking University
Project Director	Han Guanghua, Associate Professor, SJTU
Project Assistant	Ji Chengyuan, Assistant Professor, SJTU

Table of Contents

I. Overview

1.1 Length of Fieldwork

July 7, 2018—August19, 2018

September 26, 2018—October12, 2018 (call-back due to partially completed interviews)

1.2 General Outturn

Planned sample size：5450

Planned valid interview size：3000

Eligible Sample drawn in the field：4921

Completed and valid interviews：3036

Response Rate：61.7%

II. Sampling

2.1 Sample Population Overview

The target population covers Chinese nationals between the ages of 18 and 70, living in the sampled district and county for more than 6 months in all 31 provinces of the Chinese Mainland (Hong Kong, Macao and Taiwan were not included).

2.2 Sample Population Exclusions

Individuals who reside in the places listed below were not included in the study:
● Military residential complexes
● Residential units in compounds of Central Ministries
● Embassies and consulates
● Infrastructural buildings（i.e. Power Stations, Water Stations etc.）
● Prisons
● Tourist destinations and religious sites

Individuals with the following characteristics who were residing in valid residential complexes were not included in the study: residents of Hong Kong, Macau, Taiwan, and non-Chinese citizens.

2.3 Sampling Method

The sampling plan for the general public uses the *GPS Assisted Area Sampling Method*[1], which incorporates population as a measure of size, stratification and multi-stage PPS (Probabilities Proportional to Size).

2.4 Sampling Unit

Primary Sampling Units (PSUs):
> County level administrative units (municipal districts, county-level cities, counties)

Secondary Sampling Units (SSUs):
> Half-square minutes (HSM) of latitude and longitude

Tertiary Sampling Units (TSUs):
> Spatial square seconds (SSS), approximately 90m*90m

Basic Sampling Units:
> Dwellings in the sampled units

2.5 Stratification Method

For the purpose of allocating PSUs across all large regions with different levels of economic development, stratification was based on internationalization level, urbanization level, and economic development level.

Strata 1: Municipal districts of the international cities (国际化城市的市辖区). According to the data of the international influence of the world cities released by the Kearney Management Consulting Company in 2014, cities enter the top 100 cities in the world is selected as the first strata, including Beijing, Shanghai, Guangzhou, Shenzhen and Chongqing. The municipal districts of these cities serve as the primary sampling units for the first strata.

Strata 2: Municipal districts of sub-provincial cities (副省级城市的市辖区).

Strata 3: Municipal districts of the prefecture-level cities (地级市的市辖区).

Strata 4: Counties/county-level cities/suburban counties of prefecture-level cities and above with higher level of economic development (经济发展水平高的县/县级市/地级市及以上城市的郊区县).

---

[1] Landry, P. F., & Shen, M. (2005). Reaching migrants in survey research: The use of the global positioning system to reduce coverage bias in China. *Political Analysis*, *13* (1), 1-22.

Strata 5: Counties/county-level cities/suburban counties of prefecture-level cities and above with middle level of economic development (经济发展水平高的县/县级市/地级市及以上城市的郊区县).

Strata 6: Counties/county-level cities/suburban counties of prefecture-level cities and above with lower level of economic development (经济发展水平高的县/县级市/地级市及以上城市的郊区县).

Levels of economic development are based on per capita GDP, and are divided into three levels equally.

Table 1. The population size of each stratum and their respective PSU

Strata	Total PSUs in each strata	2010 census population	Propor-tion	PSUs that should be allocated according to the size proportion	PSUs allocated finally	FPC_PSU
1. Municipal districts of the international cities	64	77,632,475	0.06	2.91	3	0.047
2. Municipal districts of sub-provincial cities	104	72,279,486	0.05	2.71	3	0.029
3. Municipal districts of prefecture-level cities	694	328,509,934	0.25	12.32	12	0.017
4. Counties/county-level cities/suburban counties of prefecture cites and above with higher level of economic development	668	304,392,723	0.23	11.42	12	0.018
5. Counties/county-level cities/suburban counties of prefecture cities and above with middle level of economic development	668	282,773,148	0.21	10.61	10	0.015
6. Counties/county-level cities/suburban counties of prefecture cities and above with lower level of economic development	667	267,222,659	0.20	10.02	10	0.015
Total	2865	1,332,810,425	1	50	50	

2.6 Selection Method in Each Stage

PSU: Out of 2,855counties in China, 50 counties were chosen by stratified PPS.

SSU: Two HSMs were selected by PPS within each of the selected county.

The measures of size (HSM) used at this stage is 2012 DMSP/OLS nighttime lighting data.

Selected SSUs in a district or county:
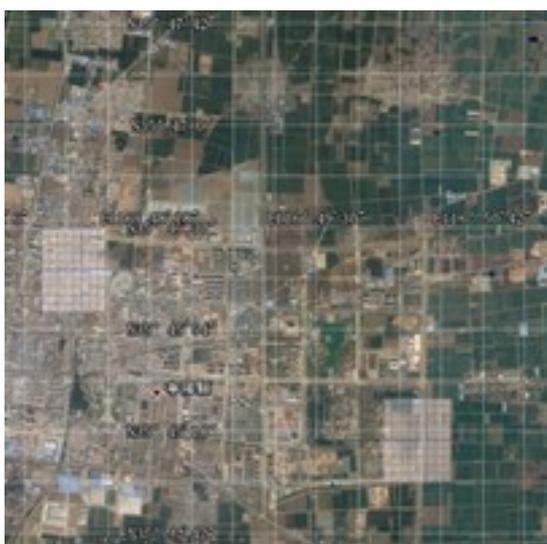
TSU: Divide the HSM into 8 columns and 10 rows, and obtain 80 small grids. According to the population density in the HSM and the number of respondents that should be completed, the number of grids that should be sampled was calculated, and then grids by simple random sampling were selected.

QSU: Trained interviewers equipped with GPS receivers were then sent to locate and enumerate the sampled "spatial square seconds" (SSS). For maintaining equal probabilities of selection across households, all dwellings enumerated in the SSSs were included in the sample.

The field address map:

Respondents: Qualified respondents (Chinese nationals between the ages of 18 and 70 who had lived in the urban district or county over 6 months) were selected from dwellings using the Kish Grid method[2].

2.7 Sample Size

To satisfy a confidence level of 95%, with a permissible error of 5%, and taking into consideration factors such as the outcomes of multi-stage sampling (design effect), non- responses (caused by reasons such as unqualified individuals, vacant residential units, interview refusals, language barriers etc.), a total of 5,450 residential units were planned to be selected, with an effective sample size of 3,000.

In actuality, 4,921 valid addresses were selected eventually, containing an effective sample size of 3,036, with an effective response rate of 61.7%.

III. Fieldwork Report

This survey project was carried out by a consortium of 12 universities in China from different geographical locations. It should be mentioned that all the 12 consortium members had the experience of conducting survey research in China. Faculty and students of social science disciplines from these universities served as field survey supervisors and field survey interviewers.

3.1 The Pre-test

A small scale pre-test was conducted in Beijing, Shanghai, Qingdao and Xi'an in May. 2018 with 50 complete interviews. Statistics showed that the sampling method and the designed field procedures were feasible, effective and representative in the field. Based on the results of the pre-test, the questionnaire was revised and finalized. Then the Question-by-Question Book was developed for the team to use in the training sessions, which is also to be used to clearly specify how to deal with possible questions raised by respondents in the field interviewing.

3.2 Supervisor Training

3.2.1 Supervisor

Supervisors of this project were faculty members and doctoral students of social science disciplines from the 12 consortium members. A joint training session for the supervisors was held at Shanghai Jiao Tong University in May 2018.

---

[2] Kish, L. (1965). *Survey sampling*. New York: John Wiley & Sons.

3.2.2 Areas Included in Supervisor Training

- Project Background
- Basic Interview Techniques
- Specific Requirements for the Project
- How to use GPS and the Sample Area Selection Process
- Address and Interviewee Selection Process
- Overview of the Questionnaire in the Survey
- Classroom Exercises
- Project Implementation Procedures
- Quality Control Procedures
- Code of Conduct and Safety Protocols

3.2 Interviewer Training

3.2.1 Interviewer

The interviewers for the project were all college students in the surveyed area and were trained, according to the Interviewer Manual, by their supervisors.

3.2.2 Areas included in Interviewer Training

- Project Background
- Basic Interviewing Techniques
- Specific Requirements for the Project
- Interviewee Selection Process
- Overview of Questionnaire in the Survey
- Classroom Exercises
- Home Interview Procedures
- Quality Control Procedures
- Code of Conduct and Safety Protocols

3.3 Official Implementation

3.3.1 Project Team

Principal Investigator: Zhong Yang
Co-principal Investigators: Yan Jie, Tang Wenfang and Yang Dali

Sampling Specialists: Yan Jie

Project Operation Director: Han Guanghua

Project Assistant: Ji Chengyuan

Regional Project Managers (alphabetically listed):

Chen Chao, Chen Shaowei, Hao Heling, He Yanling, Lei Xuchuan, Liu Lin, Lu Chunlong, Shi Yaojiang, Xiao Tangbiao, Xu Guihong, Yan Jie, Zheng Zhenqing, and Zhong Yang

Supervisors：Approximately 40 from the consortium

Interviewers：Approximately 290 college students (both undergraduate and graduate students)

Quality Inspectors：Project Operation Director and Supervisors

3.3.2 Implementation Process

a. Interviewee Address Sampling

First, supervisors proceeded to the half-square minutes as determined by longitudes and latitudes prescribed. Within the half-square minutes, supervisors were given the relevant longitudes and latitudes to identify and approach the targeted small-grid cluster. For maintaining equal probabilities of selection across households, supervisors then enumerated all dwellings in the SSSs （"spatial square seconds"）.

b. Interviewee Sampling

After interviewers entered a valid address, they identified all individuals who had resided in the address for more than 6 months, and recorded them into the Kish grid. All individuals between ages of 18 and 70 were then separated by gender, and then ordered according to age, from oldest to youngest so that the Kish method can be used to select one interviewee.

c. Field Interviewing

The field interviewing was launched on July 7 2018. The interviewers entered each target SSS by small groups, led by field supervisors. Individual interviewers were responsible to find the designated address, identify the respondent, conduct paper-pencil and face-to-face interview and complete the questionnaire (sometimes assisted by people from the neighborhood committee/village committee).

If the designated respondent was not available, the interviewer would try to make an appointment through a member of the household for a later visit. If no one at home, the interviewer would make a second or third visit some other time. If the dwelling or respondent refused to be interviewed, the supervisor would assign another interviewer to pay a visit. At least five more visit attempts were made by the interviewers before declaring a particular case as non-response.

The completed questionnaires were collected, checked, and signed by field supervisors on location.

d. How to Handle Refusals

Interviewers were instructed not to give up on any initial refusal until it was decided as a real refusal. Interviewers were required to write up a detailed record of every attempt. The record had to show clearly why the respondents refused to participate in the survey so that the supervisor could use such information to assign an appropriate interviewer for subsequent attempts. In the case of refusals either by the respondent or at the household level, the supervisor sent another interviewer or went there himself/herself until there had been 5 refusals from the chosen interviewee before deciding to give up on this address or this particular respondent.

e. Supervisors' Daily Responsibilities

Arranging interviews and ensuring quality

Leading teams into the targeted communities to perform interviews, collecting surveys, checking questionnaire completion rate.

Completed surveys must be checked daily so that potential problems could be discovered to prevent similar problems from surfacing again in the future. Supervisors must sign their initials on valid surveys. Names of interviewees should be recorded on the back of each survey questionnaire, while exact address of the interviewee must be recorded on the front of the questionnaire. After the interview, interviewers were expected to hand their completed surveys and interview records to the supervisor within 24 hours.

During the interview, supervisors must take note of the progress of the interview so as to properly fill out daily work journals and interview summaries. Forms such as interview progress forms and completed survey tables must be copied with clear handwriting and error-free.

Report to Center for Public Opinion Research (CPOR) the progress of fieldwork. Any problems that may arise must be reported to CPOR to ask for further instructions.

3.3.3 Quality Control

The overarching goal of the data collection quality control procedures was to ensure that the data collection activities yield the highest quality data possible.

Every completed questionnaire went through three rounds of checking: by the interviewer herself/himself right after leaving the interview location, by his/her field supervisor in the field, and by data managers at CPOR. All of them had to sign their names on the cover sheet of the questionnaires.

To make sure interviews were conducted according to procedures, a minimum of 20 percent of each interviewer's completed interviews were verified by the supervisor on location. Verification of interviews was done in random fashion. Verification

methods include phone calls to the respondents and/or paying visit to the respondents by supervisors, asking, for example, a set of selected factual questions with regard to the finished interviews. The random checking continued throughout the entire data collection period.

CPOR's data manager carried out the third round checking after receiving the first batch of questionnaires from the field supervisors. All completed and uncompleted questionnaires were required to be verified.

Verification included:

- Whether the address visited was correct
- Whether the respondent was properly selected within chosen household
- Whether the respondent was actually interviewed
- Whether the interviewing time was reasonable
- Whether all should-be-asked questions were asked
- Whether clear logic errors occurred between question and answer on the questionnaire
- Whether gift for the respondent was given
- Other questionable issues

The verification followed the specifications below:

- The supervisor read through all completed questionnaires submitted by the interviewers
- All questionnaires with suspected problem(s) must be verified, regardless whether they were selected into the 20% random checking pool
- Random checking covered all interviewers' work
- Any interviewer who was suspected of interview falsification should have 100 percent of his or her work verified. In situations like this, all suspected interviews were re-taken

3.4 Reasons for Unsuccessful Visits

Table 3. Reasons for unsuccessful visits

Status of visit	Amount	Percentage
1 Completed interview	**3,036**	**61.7**
2 Refusal by chosen household	537	10.9
3 Refusal by chosen interviewee	429	8.7
4 Failure to access chosen interviewee	248	5.0
5 Physical and language barriers/inability to understand questions due to old age	16	0.3
6 Unable to get in touch with people living in the chosen household	632	12.8
9 Missing data or kish sampling problem	23	0.5
Total Valid Addresses	**4921**	**100.0**
7 Vacant residential unit	332	
8 No occupants meet requirements	154	
77 Others	71	
Total Invalid Addresses	557	
Pre-planned Addresses	**5478**	

IV. Database Creation

4.1 Data Entry

Entry of raw data was performed using EPIData, beginning on November 15, 2018. To improve the accuracy of entered data, double data entry method was used to ensure accuracy. Data entry work was completed on November 28, 2018.

4.2 Sampling Verification

Three Shanghai Jiao Tong University graduate students were responsible for verifying the validity of both completed and incomplete questionnaires. Specifically, they checked whether the interviewees were properly selected and whether the interviews were indeed conducted on the chosen respondents.

4.3 Data Cleaning

Data cleansing began on January 4, 2019, headed by experienced research assistants at CPOR. The process was divided into 3 steps. The first step was verifying whether data was properly entered into the system through the double data entry method. The second step was to identify incorrect data and logical fallacies within the responses. The final step was double-checking of the variables of the entire dataset by

different individuals for any logical fallacies after correction of all the previous mistakes. These tasks were completed on January 24, 2019.

4.4 Database Creation

On January 24, 2019, after finishing data cleansing for the entire survey, database creation process was initiated. Steps include changing variable tags and descriptions as well as entering sampling information such as stratification and PSU information. The final database and coding manual were completed on February 5, 2019.

4.5 Weighting

4.5.1 Base Weight（Weighting Design）

Two weighting methods were used on the database. One of them, base weighting, is shown below, where the reciprocal of each interviewee's selection probability (including weight of empty responses) is calculated and given the variable name "wt_base".

First, with each half-square minutes (SSU) considered one unit; the valid sampling addresses and valid addresses of each half-square minutes were tabulated. Then using the formula below, each address's selection probability was calculated.

$$f_{address} = \frac{a_h Mos_{h\alpha}}{\frac{1}{c} \bullet \sum_{\alpha=1}^{\alpha_h} Mos_{ha}} \bullet \frac{d_{h\alpha} Mos_{h\alpha\beta}}{Mos_{ha}} \bullet \frac{b^*_h}{Mos_{h\alpha\beta}}$$

h represents strata number (6 stratum total)

a_h represents the number of PSUs in each strata

$d_{h\alpha}$ represents the number of SSUs sampled in each PSU. (For this project, 2 SSUs are sampled in each PSU)

Mos_{ha} represents each PSU's size measurement (that is the population in each county level unit in 2010)

$Mos_{h\alpha\beta}$ represents each SSU's size measurement，(that is the number of residential addresses in each half-square minutes)

b^*_h represents the number of valid sampled addresses in each chosen SSU

c represents the average household size in 2010.

b^*_h represents the number of addresses

$\sum_{\alpha=1}^{\alpha_h} Mos_{ha}$ represents the population at each strata

Because this project selected only inhabited residential addresses, we needed to divide population at each strata by the average household size c to obtain the number of residential addresses and thus be able to calculate each address's selection probability. The average household size in 2010 was obtained from China's 2010 census.

Next, using the number of valid participants in each address, the probability of selection for each individual is calculated.

$$f = f_{address} \bullet \frac{1}{q2}$$

q2 is a variable in the database representing the number of valid participants in each address.

Then，the base weight for each individual is calculated

$$wt = \frac{1}{f}$$

Subsequently, the empty response weight is calculated.

$$wt\_respo = \frac{Valid\ Sampling\ Size}{Valid\ Response\ Sample\ Size}$$

Lastly，the base weight is calculated

$$wt\_base = wt \bullet wt\_respo$$

The three variables, wt, wt_respo and wt_base are all present in the database, and their usage depends on how empty responses are accounted for. If weighting is used to account for empty responses, then wt_base can be used directly to add weighting to the variables. If other methods are used, then further adjustments can be performed using wt as a basis.

4.5.2 Weight Calibration

Weight calibration is based on Table 4-1a to Table 4-1c of the *1% National Population Sample Survey Data 2015* published by the Department of Population and Employment Statistics of the National Bureau of Statistics of China. Gender, age group (18-30, 31-40, 41-50, 51-60, 61-70, education level (primary school and below, junior high school, senior high school, college and above), and residency (urban and rural) were the target variables. Weight calibration is completed based on the base weight (wt_base), and the calibration weight range is limited to the 1% quantile and the 99% quantile of the base weight. The calibration weight is the variable "wt_final".
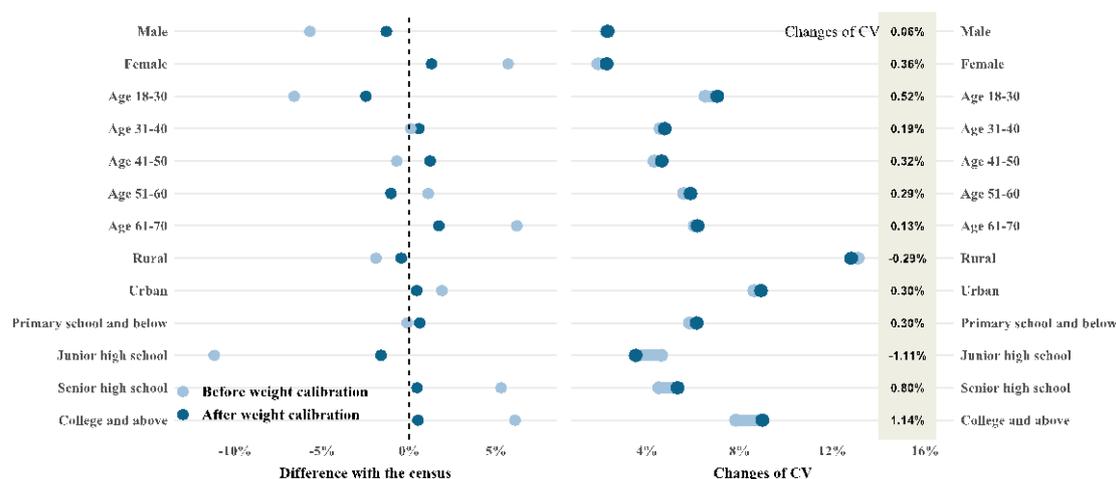
V. Data Usage Instructions

5.1 Documents and Data File

Following material are provided:
(1) Database, SPSS version
(2) Codebook
(3) Sampling and fieldwork report
(4) Questionnaire in Chinese (PDF version)

5.2 Variable Naming Instructions

(1) Basically consistent with the questionnaire;
(2) Questions with open-end opinion is naming with "**k".