China Health and Retirement Longitudinal Study

Wave 5 (2020) User Guide

Yaohui Zhao Xinxin Chen Yafeng Wang Qinqin Meng
Hai Bo Chuan Chen Yanfeng Chen Jinquan Gong Yuan Jia
Gewei Wang Xiaoyang Wu Li Yan Peng Yang Hongyan Zhou

VersionID: 20231106

Nov. 2023

How to cite this document: Yaohui Zhao, Xinxin Chen, Yafeng Wang, Qinqin Meng, Hai Bo, Chuan Chen, Yanfeng Chen, Jinquan Gong, Yuan Jia, Gewei Wang, Xiaoyang Wu, Li Yan, Peng Yang, Hongyan Zhou (2023), China Health and Retirement Longitudinal Study Wave 5 (2020) User Guide, *National School of Development, Peking University*

National School of Development
Institute of Social Science Survey
Peking University

This page intentionally left blank

How to Cite the CHARLS Data

When you write a paper using the CHARLS data, in data description part, making proper citations of authoritative papers can enhance the credibility of your research.

1. How to cite this user guide?
    - Yaohui Zhao, Xinxin Chen, Yafeng Wang, Qinqin Meng, Hai Bo, Chuan Chen, Yanfeng Chen, Jinquan Gong, Yuan Jia, Gewei Wang, Xiaoyang Wu, Li Yan, Peng Yang, Hongyan Zhou (2023), China Health and Retirement Longitudinal Study Wave 5 (2020) User Guide, *National School of Development, Peking University*.
2. When using the CHARLS data, for general introduction of CHARLS, please cite the following paper:
    - Zhao, Yaohui, Yisong Hu, James P Smith, John Strauss, Gonghuan Yang. (2014). Cohort Profile: The China Health and Retirement Longitudinal Study (CHARLS), *International Journal of Epidemiology*, 43 (1): 61–68.
3. When using the baseline data or introducing sampling information, please cite the following paper:
    - Zhao, Yaohui, John Strauss, Gonghuan Yang, John Giles, Peifeng (Perry) Hu, Yisong Hu, Xiaoyan Lei, Man Liu, Albert Park, James P. Smith, Yafeng Wang. (2013). China Health and Retirement Longitudinal Study: 2011-2012 National Baseline User's Guide, *National School of Development, Peking University*.
4. When using the CHARLS data of venous blood sample, please cite the following paper:
    - Chen, Xinxin, Eileen Crimmins, Perry Hu, Jung Ki Kim, Qinqin Meng, John Strauss, Yafeng Wang, Yuan Zhang, Yaohui Zhao, 2019, Venous Blood-based Biomarkers in the China Health and Retirement Longitudinal Study: Rationale, Design, and Results of the 2015 Wave, *American Journal of Epidemiology*, 188(11): 1871–1877

This page intentionally left blank

Preface

This document describes the overall process, including the design, implementation, and data release, of the China Health and Retirement Longitudinal Study (CHARLS) national survey of wave five in 2020. This manual aims to enhance the user's understanding and application of the survey data.

CHARLS is a longitudinal survey that aims to be representative of the residents in mainland China aged 45 and older, with no upper age limit. It attempts to set up a high-quality public micro-database, which can provide a wide range of information from socioeconomic status to health conditions, to serve the needs of scientific research on the elderly.

To ensure the adoption of best practices and international comparability or results, CHARLS is harmonized with leading international research studies in the Health and Retirement Study (HRS) model. The national baseline survey was conducted in 2011-12, with wave 2 in 2013, wave 3 in 2015, wave 4 in 2018, and wave 5 in 2020. In addition, the life history survey was implemented in 2014. To ensure sample representativeness, the CHARLS baseline survey covered 150 countries/districts, 450 villages/urban communities, across the country, involving 17,708 individuals in 10,257 households, reflecting the mid-aged and older Chinese population collectively. In late 2019 and early 2020, there was an outbreak of Covid-19 in China. In order to document the impact of the pandemic on the lives and health of middle-aged and elderly people in China, the information related to Covid-19 was additionally collected in wave 5.

All data collected in CHARLS are maintained at the Institute of Social Science Survey of Peking University, Beijing, China. The first four waves of CHARLS data plus the Life History wave have all been released publicly, on the CHARLS website (https://charls.pku.edu.cn/en/). By the end of October, 2023, more than 88,000 users have registered and downloaded the data. Publications based on CHARLS have grown rapidly. Using an online search, we found 4,587 articles as of the end of September 2023, of which 2,079 are English journal articles, 1,639 are Chinese journal articles, and 697 are university theses.

This page intentionally left blank

Acknowledgements

The China Health and Retirement Longitudinal Study (CHARLS) is an enormous project that required the efforts of many people. We want to express our gratitude to the CHARLS research team, the field team, every respondent and all the supporters for the survey.

The design of the fifth wave of CHARLS was overseen by Professor Zhao, Yaohui, with Dr. Wang, Yafeng serving as the head of questionnaire and data work, and Dr. Chen, Xinxin taking charge of on-site execution. The on-site supervision team, consisting of members of the execution department such as Yuan Jia, Yongjie Wang, Qianyu Hu, and Yu Ma, recruited, trained, and managed a team comprising 549 interviewers, 130 quality checkers, and 15 supervisors. The data team, including Qinqin Meng, Hai Bo, Yanfeng Chen, Jinquan Gong, Gewei Wang, Xiaoyang Wu, Li Yan, Peng Yang, and Hongyan Zhou, was responsible for designing the questionnaire, conducting real-time data quality checks using parallel data, cleaning data, and constructing sampling weights. Chen, Chuan and Yinxia Zhao provided information technology support for on-site visits, developing a CAPI system for the Android platform and corresponding support systems for interviewer recruitment, training, on-site management, and multi-method quality control checks. Song, Na handled all logistical matters during on-site execution, and Jin, Haiyu managed the financial affairs of CHARLS. Dr. Wenjin Wang participated in the translation of the questionnaire. Over ten students participated in post-survey data checks, cleaning, phone callbacks, open-ended question coding, and preparations for data public release.

The survey could not have taken place without the understanding and support of all households participating in the CHARLS project. The data provided not only lays the foundation for academic studies on the Chinese aging issues but also throws light on the future development of the social welfare system for our government. We extend our sincerest thanks.

This page intentionally left blank

Contents

This page intentionally left blank

1 General Introduction to CHARLS

1.1 Background and Significance

China has the largest aging population in the world, and also one of the highest aging rates in the world today. By the end of 2021, the older adults aged 65 and above had accounted for 14.2% of the total population, which implies that China doubled the share of its older adults aged 65 and over from 7% to 14% in 21 years. On the contrast, the developed countries took about half a century to double the share (Kinsella, He, and Way, 2009). By 2065, China will be the most seriously aging country among the world's 20 most populous countries. By 2050, the number of people aged 65 and over in China will reach 395 million, equivalent to 1.2 times the current population of the United States. The number of senior citizens (those aged 80 and above) will reach 135 million, more than the current population of Japan (United Nations, 2022).

Related to the aging process, China has been undergoing a rapid health transition in which the nature of health problems changes from infectious diseases, which affect mainly the young, to chronic diseases affecting the elderly (Yang et al., 2013). Moreover, China is undergoing the aging process at much lower income levels than was the experience in industrial countries. Compared to most other countries with Health and Retirement Studies, China is much more rural, with lower levels of schooling among the elderly, lower levels of public services available, and the traditional family support system is facing more serious challenges with rapid urbanization. How to deal with problems of support for the well-being of the elderly is one of the most significant challenges to the fast-booming Chinese society in the decades to come.

In response to this challenge, the Chinese government has taken robust actions to solve the problem. In recent decades, a series of social safety nets have been put into place. Such policies include the Minimum Living Standard Guarantee System, the New Cooperative Medical Insurance System, the Urban and Rural Resident Medical Insurance System, the New Rural Social Pension Program, the Urban Resident Social Pension Program, etc. Although some of these policies are not specifically designed for the elderly, the aged population is undoubtedly one of the most important beneficiary groups. Similar to many other policies, they are initiated by the central government, but the local governments maintain certain autonomy in the process of implementation. The local governments may decide on the schedule for pilot tests and promotion, and they may have different implementation plans. CHARLS is measuring the existence of these social safety nets at both the household and community levels and will allow analysis that hopefully will provide a more scientific

basis for the government to further revise and amend the existing policies.

Prior to the CHARLS's baseline survey of 2011-12, scientific studies of China's aging-related issues were at an early stage, the greatest obstacle being a lack of sufficient micro, longitudinal data. The existing data tended to be small scale in parts of China, not collecting the breadth of data necessary for good social scientific analysis of the health of the older population. For instance, there exist some health data sets that centered on health measures, with indicators of socioeconomic status largely neglected; on the other hand, data sets collected by social science scholars tend to be insufficient in health-related measures. Since the welfare of the elderly is closely associated with their health and socioeconomic status, and also because health and socioeconomic levels are themselves interrelated, micro-data that is of extensive coverage and high accuracy is highly needed for research on Chinese aging problems. CHARLS is an attempt to fill this gap.

CHARLS is the first nationally representative survey of the older population that enables the study of the health of the older population in China patterned after US Health and Retirement Study (HRS) and related aging surveys around the world (e.g., the English Longitudinal Survey of Aging, ELSA, and the Survey of Health, Aging and Retirement in Europe, SHARE, Japanese Study of Aging and Retirement (JSTAR), the Longitudinal Aging Survey of India (LASI), the Indonesia Family Life Survey (IFLS) and the Korean Longitudinal Survey of Aging (KLoSA), etc.). More importantly, this study has closely tracked the implementation of social security policies introduced in China in recent years and the impact of major public health events (like Covid-19) on the health of middle-aged and elderly people in China.

1.2 Ethical Approval

Ethical approval for all the CHARLS waves was granted from the Institutional Review Board at Peking University. The protocol of fieldwork of the main household survey was approved `IRB00001052-11015`. For those who require this approval letter for publication, please contact CHARLS team by sending an email to charls_info@pku.edu.cn, with specific information about your needs. During the fieldwork, each respondent who agreed to participate in the survey was asked to sign two copies of the informed consent. One was kept by the respondent and the other was kept in the CHARLS office, which was also scanned and saved in PDF format.

1.3 Organization of this Document

Section 2 of this manual documents the sample design, focusing on the baseline sampling procedures, the refresher samples, and proxy interviews. In section 3, we introduce the contents of the survey, including a brief description of the survey contents of wave 5. Section 4 describes how the fieldwork was organized and how we tracked the respondents. Section 5 describes how sampling weights were constructed, and section 6 provides some basic information of dataset.

This page intentionally left blank

2 Sampling

The CHARLS national baseline survey was conducted in 28 provinces, 150 countries/districts, 450 villages/urban communities across the country in 2011-2012, with wave 2 in 2013, wave 3 in 2015, wave 4 in 2018 and wave 5 in 2020. The CHARLS sample is representative of people aged 45 and over, living in households; institutionalized elderly is not sampled, but wave 1 respondents who later move to an institution are followed. In Wave 5, due to covid-19 prevention and control policies, we fail to visit three sample villages in Xinjiang and some sample villages in Inner Mongolia. This section describes how CHARLS samples originated.

2.1 Baseline Sampling

A implicit stratified (by area, urban districts and rural counties, and GDP per capita) multi-stage (county/district-village/community-household) PPS random sampling strategy was adopted. At the first stage, all county-level units were sorted (stratified) by region, within the region by urban district or rural county, and by GDP per capita (Tibet was the only province not included). The region was a categorical variable based on the NBS division of provinces. After this sorting (stratification), 150 counties or urban districts were chosen with probability proportional to population size (Zhao et al., 2013). For each county-level unit, 3 PSUs (villages and urban neighborhoods) are randomly chosen with probability proportional to population (Zhao et al., 2013). Hence CHARLS has national regional representativeness. 450 villages/urban neighborhoods, 150 Counties and districts in 28 provinces are included in the CHARLS sample.

In light of the outdated household listings at the village/community level due to population migration, CHARLS designed a mapping/listing software (`CharlsGIS`) that makes use of Google-earth map images to list all dwelling units in all residential buildings to create sampling frames. In each sampled household, a short screening form was used to identify whether the household had a member meeting the age eligibility requirements. If a household had persons older than 45 and meeting the residence criterion, one of them was randomly selected. If the chosen person was 45 or older, then he/she became the main respondent, and his or her spouse was interviewed. As a result, the CHARLS's baseline survey includes one person per household aged 45 years of age or older and their spouse, totaling 17,708 individuals, living in 10,257 households in 450 villages/urban communities (Table 1).

After applying sampling weights created using the sampling procedure, the CHARLS base-

line sample demographics mimics very closely that of the population census in 2010 (Zhao et al., 2014), indicating it is well representative of middle-aged and older Chinese adults.

2.2 Refreshment Samples

CHARLS is a study of people aged 45 and over. As the study progresses, the sample respondents get older, leaving the youngest ages unrepresented unless new sample members are recruited to fill the gap. For example, those aged 45-46 at the baseline will be 47-48 in the next wave two years later, leaving the sample aged 45-46 in 2013 unrepresentative. Therefore, to consistently maintain representativeness of people aged 45 and over, CHARLS reserved the refreshment samples aged between 40 and 44 in the baseline survey for future waves of survey. That is, if a household had persons older than 40 in the baseline, we randomly selected one of them as a reserved sample. when he/she is 45 or older in the following waves, he/she becomes a main respondent and will be interviewed with his or her spouse. Thus, in wave 2, respondents who were aged 43-44 in wave 1 (plus their spouses) were added from the refresher sample, the same for waves in 2015 and 2018, out of those aged 41-42 and 40 in the baseline. With the depletion of the refresher sample, the CHARLS sample is no longer nationally representative of the younger cohort in the following waves after 2018.

Counting refresher samples and age-eligible respondents who failed to be found in the baseline but successfully contacted in the follow-up waves, the total number of individuals (main respondents plus spouses) has increased from 17,708 in wave 1 to 19,395 in wave 5 (Table 1).

Table 1 describes the age/sex composition of the CHARLS during wave1 and wave 5. In the baseline wave, we have data on 17,708 individuals, of which 52.1% are female, and 58% of the respondents are under 60 years while in wave 5 only 38% are under 60 years as there are no younger refreshment samples.

2.3 Proxies

A proxy interview was pursued under some special circumstances when the respondent could not complete the survey, for example, when the respondent was physically or cognitively impaired, in a hospital, or couldn't be tracked by the field teams during the fieldwork period. Some of those who refused to take part in person but someone else could do the interview on their behalf could also have a proxy interview. Before the proxy interview, the interviewer was asked to identify a proxy informant who knew the respondent well and could provide enough information about him. In most cases, a close family member, such as a spouse or child, assumed this role.

In order to limit the use of proxies during interviews, the interviewer was required to call the team leader to apply for a proxy code who checked by asking some specific reasons for the proxy request. If the request was approved, the interviewer received a proxy code and proceeded to the ”complete proxy” mode. In the previous four waves, after that, the CAPI system would automatically switch to the proxy mode before entering the first module. However, this was not the case in Wave 5. Considering that respondents might be able to answer some questions in some specific modules, we made some changes in the CAPI system. That is, after the interviewer enter this proxy code, in each module the CAPI system will first confirm whether the interview will be done in a proxy mode. If the answer is no, then questions in that module will be exactly the same as those in the regular questionnaire; if the answer is yes, then a proxy will be chosen and less questions (see the logic diagram for each module for details) will be asked. During Wave 5 a total of 1,837 proxy codes were applied for.

This page intentionally left blank

3 Survey Content

In wave 5, each household was first invited to complete the module of cover screen (CV). Then, for respondents who passed away between the two waves, we will invite those who were familiar with the respondents to complete the exit questionnaires (EX); for all other living respondents, regular survey questionnaires will be applied.

The core regular questionnaire consists of the following modules: (B) demographics, (C) family structure/transfer, (D) health, (F) work and retirement, (G) Income and expenditure (including housing characteristics), and (V) Covid. Table 2 summarizes the main questionnaire contents in Wave 5.

In addition to the wealth of individual social, economic and behavioral data, CHARLS is characterized by the rich information on the respondent's health. The section of health begins with the self-reports, including the respondent's self-assessment of general health, questions asking about diagnoses by doctors of a set of chronic diseases, question on activities of daily living (ADLs), instrumental activities of daily living (IADLs), and physical functioning. Sections on depressive symptoms and cognition follow. Furthermore, information is collected on several health behaviors. These include detailed information on smoking, drinking, and physical activities.

CHARLS used preloads in designing the questionnaire in CAPI in the four follow-up waves, 2-5. Starting in Wave 5 we converted all questions into logic graphs (which will be released with the questionnaire) to show the routing patterns for each possible answer from both cross-section and preloaded variables, which helped to avoid programming errors.

In order to understand the impact of covid-19 on the health of middle-aged and older adults, a new specific module focusing on covid-19 was added to collect respondents' precautionary behaviors during the outbreak, their disease precautionary awareness, personal infection and quarantine experiences, personal activities and local control initiatives. In addition, information on whether the health care utilization is postponed or cancelled, income change, and that on whether they are unemployed during the outbreak is collected in the health module, income module, and work module separately.

What needs to be pointed out is that, due to the extra interview time added by the new COVID module, less information was collected in some other modules in Wave 5 compared to previous waves. For example, information on siblings was not collected in the family module; Additionally, there was less information on health status, health care utilization, health insurance, pension and assets in this survey.

This page intentionally left blank

4 Fieldwork and Response

The successful completion of this project requires implementing a set of core activities: questionnaire designing, CAPI programming, field staff recruitment and training, fieldwork organization, and quality control. It usually takes one year to prepare the fieldwork of each wave. The CHARLS research team started revising the questionnaire in August 2019 and organized several pilots to test and improve the questionnaire and the survey procedure. In April 2020, we conducted several formal pretests before finalizing the questionnaire using CAPI. From the experience of the pretests, the research team revised the questionnaires and procedures for the field survey. In March 2020, the project team started to recruit interviewers, and prepare training materials. From April to July in 2020, 673 university students took the training courses for about 60 hours and 549 students finally passed the training examinations and went to the field. Most of the fieldwork was completed by the end of September of 2020. All these activities of the field work, including the tracking strategies are discussed in this section.

4.1 Questionnaire Design

Because Wave 5 is a longitudinal survey, which has complicated rules about preloading answers from previous waves and probing when detecting a change in status between waves, the research team worked very hard to streamline the questionnaire. We painstakingly drew electronic flow charts of each questionnaire module to sort out the logic of each question. In the process, we corrected previous CAPI errors and logic mistakes that existed in earlier waves.

In addition, we added a specific module to collect the respondent's experience of infection, quarantine and activity during the outbreak of covid-19. In the process of producing the final questionnaire, several formal pretests took place in some communities in Beijing. The questionnaires were finalized following these pretests.

4.2 Improvements in the CAPI System

In order to meet the requirements of the resurvey, a substantial amount of programming was carried out, particularly to accommodate the needs of follow-up procedure of the newly age-eligible respondents and non-response sample in the previous waves. Since Wave 4 in 2018, we have continually used tablets instead of laptops, which facilitates collecting GPS, photos and voice recording and running much faster. The updated system not only helps designing questionnaires based on logic diagrams and generate questionnaires automatically, but also

facilitates automatic questionnaire logic testing, in addition to providing data of multiple types for quality control of the interviews.

4.3 Interviewer Recruitment and Training

Due to the prevention and control policy of covid-19, the recruitment and training of interviewers in Wave 5 was held mainly online. In order to facilitate recruiting more than 500 interviewers in a limited time, we employed a web-based recruitment system where job applicants filled in relevant information online, and initial screening was conducted. The positions were advertised at CHARLS Wechat public account. In many cases, we relied on colleagues in local universities to recruit their students as interviewers. The criteria used in selecting interviewers were their stated work attitudes, any previous field experience, and communication skills. We targeted interviewers who can speak local dialects. In order to minimize the movement of interviewer across regions, we chose student interviewers whose hometowns were the sample districts/counties as much as possible. In this wave, 332 of 549 interviewers came from universities outside Beijing, accounting for 61%. As a result, most of them were able to speak the local dialect, which facilitated more efficient communication with the respondents.

Most of our interviewers are undergraduate and graduate students, which has been true for all waves. Due to the tight covid-19 containment policy, student movement among universities was not encouraged. Therefore, the training was mainly hold online. In order to overcome the disadvantage of less classroom interaction in the online training, we added online Role Modeling Exercises in classes for up to 25 students; developed a self-study and testing system for the interviewers to check and improve the effectiveness of the training; arranged Live interviews with convenience samples during training and a 2-3 day rehearsal right before the fieldwork observed by supervisors. The training course lasted for about 60 hours and ensured that students were familiarized with the CAPI system, mastered the questionnaire, developed interviewing skills, and understood the fieldwork protocols and their respective responsibilities. Students who passed the final training assessment were identified as interviewers.

4.4 Field Organization

In Wave 5, 549 interviewers were organized into 76 teams. Mostly, a team was responsible for interviewing all respondents in 6 villages/communities in 2 counties, including those who had moved from elsewhere into the area. Two people of each team played the role of an advance unit that updated the contact information of respondents so that movers could

be identified early on and cases reassigned. They also organized the logistics for the team and conducted community-level and county-level surveys. The rest of the team consisted of a team leader and interviewers who were in charge of carrying out all the follow-up procedures and conducting individual interviews.

During the fieldwork, each team was assigned a supervisor who had experience in fieldwork. The field supervisors report to the Field Director at the CHARLS headquarter. Each supervisor manages 2-5 teams, and the CHARLS headquarters responded to all problems that could not be solved by the supervisors.

In each of previous waves, all interviews were conducted face-to-face CAPI interviews. However, in wave 5, in some cases when the respondent fear of being infected by covid-19, or when he/she lives in a nursing home where visits are not allowed, the video interview was conducted after it was approved by CHARLS headquarter. In total, 267 interviews were done by video in wave 5.

4.5 Languages Used in the Field

The great majority (92%) of Chinese are Han and use the same written language. Some of the large ethnic minorities have adopted the Han written language, too, such as Manchurian, Hui and Zhuang. Therefore, we estimate that less than 4% of the Chinese population use a different written language. Despite having the same written language, various dialects are used in different parts of China.

To minimize communication difficulties, when we recruit interviewers, we prioritize those with dialectic or language skills. All of our students are college students. In Wave 5 there are 564 student interviewers from universities or colleges throughout China, and every province is well represented in students, so we have had no difficulty recruiting students from places of our survey, even those from Xinjiang or Tibet (we have a Tibetan county in Sichuan province in our sample). In rare cases in which none of the interviewers speak the respondent's dialect, or if a minority language other than Mandarin is needed, another staff or a local resident who speaks that dialect or language is used as a translator.

4.6 Quality Control in the Field

Quality control measures that previously worked well continued to be employed in Wave 5, including quality control sessions at the time of training and self-checklist upon completion. Conventional measures were also used to produce accurate assessment of the data quality, such as data checking and audio recording playback (Zhao et al., 2013).

The use of Computer-Assisted Personal Interview (CAPI) dramatically increases our ability to catch and correct errors made by the interviewers in the field. While they are still in the field, CAPI informs the interviewer immediately when a section has been improperly skipped, was incomplete, or was found to take too little time. We train interviewers in these procedures.

Other data checking measures used include comparing respondent photos between different waves to ensure that the same persons are interviewed, checking audio recordings, and short phone interviews with respondents. The CAPI system allows the teams to send the data back through the internet to the head office at the end of each day, uploading the data onto a secure website, which allows for checking in real-time. We check the first interview of each interviewer, plus 10% of subsequent interviews on average. For those whose interview is inferior to their peers, their following interviews are checked more often. Usually, the interviewer will get the feedback from the quality control team within 48 hours after they finish the interview and send back the required data for quality control. By doing so, some errors are corrected before the team leaves for the next village/community.

In Wave 5, 130 members of the quality control team provided detailed feedback to interviewers on a day-to-day basis. Such communication proved to be a useful means to supplement the pre-field training and complement the in-field supervision.

4.7 Sample Tracking

For respondents surveyed in the baseline wave, they were re-contacted in each of the follow-up waves. In addition, we did not give up households or individuals we could not find in the baseline and contacted them in follow-up waves, so are respondents who failed to respond in any one or more waves.

Respondents and spouses are tracked if they exit the original household. Main respondents and spouses in the baseline survey are followed throughout the life of CHARLS, or until they die. If the main respondent or spouse re-marries, the new spouse is interviewed so long as they are still married to the baseline respondent at the time of the specific wave.

CHARLS aims to interview every respondent face to face, no matter where they moved to. To achieve this, we borrowed tracking procedures from the very successful Indonesia Family Life Survey (IFLS) and innovated by computerizing the procedure so that samples that moved can be quickly transferred to the team in the destination area. From the previous wave, we have contact information such as the current land and cell telephone numbers for each respondent. We also have the name, address and telephone numbers of relatives or

friends who would likely know where they were, should they move. We use this information if we do not find a respondent where we found them previously.

Through these efforts, we obtained high rates of follow-up of our respondents. The response rate for the baseline survey was 80.5% (94% in rural areas and 69% in urban areas, lower in urban areas as is common in most surveys undertaken in developing countries (Zhao et al., 2013)). A description of the sample and response rates in waves 2, 3, 4 and 5 are provided in Table 3.

The response rate of the tracked sample (panel sample) remains at higher than 86% in any of the follow-up waves. Specifically, among those households which were interviewed in the baseline survey, about 88% of them completed at least one module in Wave 2 in 2013 (92% in rural areas and 83% in urban areas). In Waves 5, about 87% of the tracked households completed at least one module (Table 3). The success rates are high compared to many HRS-type surveys.

This page intentionally left blank

5 Sampling Weights

Weights are recommended when making inferences at the popluation level in order to minimize bias resulting from different non-response rates among key subgroups. Based on sampling weights for the baseline wave data (Zhao et al., 2013), we construct cross-sectional sample weights directly from the sampling probabilities for households and individuals in Wave 5, taking account of the death and divorce. We do not provide panel weights from wave 3, since panel weights depend on analysis purpose, users may use the panel of all waves, or only a subset of waves. Users can construct panel weights according to any appropriate sample attrition adjustment method.

We provide two sets of cross-sectional household weights, one with corrections for non-response (`HH_weight_ad1`) and one without (`HH_weight`). We provide two sets of cross-sectional individual weights, one with corrections for household and individual non-response corrections (`INDV_weight_ad2`), and one without any non-response corrections (`INDV_weight`).

Weights are calculated for the representative sample only, including proxy and partial interviews. All other non-sample individuals that were interviewed, such as wrong baseline respondents (interviewer went to the wrong address, for example, which is confirmed in follow-up waves) receive no weights. The variable `crosssection` in the `Sample_Infor` dataset can be used to identify the cross-sectional sample.

This page intentionally left blank

6 Dataset Information

In this released version (versionID: 20231106) we released *eight* main datasets, the associated *two* datasets (sample information and cross-sectional weights). The PSU information is the same as previous waves. The following table provides detailed information about these *ten* datasets.

Module in Questionnaire	**Dataset**	**Information**
B. Demographic Backgrounds	Demographic_Background	Demographic information for main respondent and spouse
C. Family Information	Family_Information	Information of household and family members
D. Health Status and Functioning	Health_Status_and_Functioning	Health status, health behaviour and cognition
F. Work and Retirement	Work_Retirement	Work status and retirement
G1. Household Income and Expenditure	Household_Income	Household income, expenditure and housing
G2. Individual Income	Individual_Income	Individual income
V. COVID	COVID_Module	Information during COVID
EX. Exit	Exit_Module	Deceased respondents information
	Weights.dta	Cross-sectional weights
	Sample_Infor.dta	Responsed Samples, whether cross-sectional, whether died, and interview date

All the data sets are stored in Stata 14 format, users can also find the summary information of variables from the released "codebook".

The IDs (ID, householdID and communityID) can be matched with there counterparts in the previous waves. Users need to adjust the householdID and ID in the baseline wave, as noted in the release note of wave 2.

This page intentionally left blank

7 Table Appendix

Table 1: Number and Age/Sex Structure of Individuals: 2011 - 2020

Age Group	Baseline, 2011			Wave 2, 2013			Wave 3, 2015			Wave 4, 2018			Wave 5, 2020		
	Total	Male	Female	Total	Male	Female	Total	Male	Female	Total	Male	Female	Total	Male	Female
-44	483	74	409	435	76	359	718	137	581	255	31	224	87	11	76
45-49	3,575	1,643	1,932	3,153	1,398	1,755	3,175	1,503	1,672	1,960	820	1,140	526	120	406
50-54	2,707	1,310	1,397	2,827	1,348	1,479	3,551	1,694	1,857	3,500	1,664	1,836	2,727	1,255	1,472
55-59	3,520	1,721	1,799	3,406	1,655	1,751	3,095	1,532	1,563	3,045	1,429	1,616	3,934	1,856	2,078
60-64	2,823	1,432	1,391	3,152	1,581	1,571	3,594	1,723	1,871	3,375	1,665	1,710	2,687	1,303	1,384
65-69	1,836	928	908	2,084	1,051	1,033	2,537	1,297	1,240	3,162	1,512	1,650	3,495	1,698	1,797
70-74	1,291	681	610	1,466	756	710	1,679	823	856	1,996	1,002	994	2,701	1,325	1,376
75-79	850	427	423	981	511	470	1,083	577	506	1,330	656	674	1,598	779	819
80+	612	260	352	750	331	419	841	368	473	1,193	561	632	1,640	746	894
Obs.	17,697	8,476	9,221	18,254	8,707	9,547	20,273	9,654	10,619	19,816	9,340	10,476	19,395	9,093	10,302

There are 11 individuals in 2011, 10 individuals in 2013, and 11 individuals in 2015 lacking age information.

Table 2: Summary of Wave 5 Data Collected in Household Questionnaire

B. Demographic Backgrounds
- Birthdate
- Residence
- Hukou, education and marital information
- Pension and health insurance

C. Family Information
- Children
  - Demographics
  - Contact and transfer with parents
  - The Impact of COVID on income and relationships with parents
- Household members
- Family reunion during the Spring Festival

D. Health Status and Functioning
- Self-reported general health
- Doctor diagnosed chronic Pain, accidents, fall and fracture
- Sleep, physical activity, social activities, smoking and drinking
- Functional limitations and helpers
- Cognition and depression
- Utilization of medical services and its impact by COVID

F. Work and Retirement
- Current job status
- Detailed information on the current main job
- Unemployment and job search activities
- Work during COVID
- Retirement

G. Income and Expenditure
- Household income and expenditure
- Individual income
- Housing

V. COVID
- Awareness
- Infection and quarantine
- Activities
- Restrictions in the residence area

Table 3: Response Rates: 2011-2020

		Baseline	Wave 2, 2013		Wave 3, 2015		Wave 4, 2018		Wave 5, 2020	
		2011[a]	Cross Section[b]	Panel[c]	Cross Section[b]	Panel[c]	Cross Section[b]	Panel[c]	Cross Section[b]	Panel[c]
Response Rate (%)	Total	80.51	82.63	88.30	82.13	87.15	83.84	86.46	84.31	86.81
	Rural	94.15	91.74	92.18	91.32	93.13	91.40	92.79	92.54	93.80
	Urban	68.63	72.20	82.61	71.64	78.45	74.55	77.24	74.45	76.76
No. of Households	Total	10,257	10,629	9,022	11,797	8,715	10,524	8,288	10,204	8,046
	Rural	6,033	6,340	5,547	6,993	5,483	6,456	5,226	6,303	5,089
	Urban	4,224	4,289	3,475	4,804	3,232	4,068	3,062	3,901	2,957
No. of Respondents	Total	17,708	18,264	15,196	20,284	14,522	17,970	13,567	17,364	13,095
	Rural	10,537	10,950	9,439	12,075	9,200	11,017	8,622	10,739	8,352
	Urban	7,171	7,314	5,757	8,209	5,322	6,953	4,945	6,625	4,743

[a] The response rate in the baseline is computed as the number of households that completed at least one main module divided by the number of implied age-eligible households.

[b] The cross-sectional response rate in the follow-up waves (2013-2020) is computed in the same way as that in the baseline.

[c] The panel response rate in the follow-up waves (2013-2020) is computed as the number of respondents which were interviewed in the baseline and completed at least one main module in the current wave divided by the number of respondents that were interviewed in the baseline. Since CHARLS not only tracks the response samples in previous waves, but also include the non-response samples and refresh samples, so the cross-sectional sample is not necessarily the same as the panel sample.

References

Kinsella, Kevin, Wan He, and Peter-O. Way. 2009. *An aging world: 2008, International Population Reports*. U.S. Governement Printing Office.

United Nations, Population Division, Department of Economic and Social Affairs. 2022. *World Population Prospects: The 2022 Revision*. UN.

Yang, Gonghuan, Yu Wang, Yixin Zeng, George F. Gao, Xiaofeng Liang, Maigeng Zhou, Xia Wan, Shicheng Yu, Yuhong Jiang, Mohsen Naghavi, Theo Vos, Haidong Wang, Alan D. Lopez, and Christopher JL Murray. 2013. "Rapid health transition in China, 1990–2010: findings from the Global Burden of Disease Study 2010." *The Lancet* 381 (9882):1987–2015.

Zhao, Yaohui, Yisong Hu, James P Smith, John Strauss, and Gonghuan Yang. 2014. "Cohort Profile: The China Health and Retirement Longitudinal Study (CHARLS)." *Int J Epidemiol* 43 (1):61–8.

Zhao, Yaohui, John Strauss, Gonghuan Yang, John Giles, Peifeng (Perry) Hu, Yisong Hu, Xiaoyan Lei, Man Liu, Albert Park, James P. Smith, and Yafeng Wang. 2013. "China Health and Retirement Longitudinal Study: 2011-2012 National Baseline User's Guide." Tech. rep., National School of Development, Peking University.