Chinese Sociological Review
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/mcsa20

An Introduction to the China Family Panel Studies (CFPS)

Yu Xie Distinguished Professor[a] & Jingwei Hu research associate[b]
[a] University of Michigan and Peking University
[b] Peking University
Published online: 27 Jan 2015.

To cite this article: Yu Xie Distinguished Professor & Jingwei Hu research associate (2014) An Introduction to the China Family Panel Studies (CFPS), Chinese Sociological Review, 47:1, 3-29

To link to this article: http://dx.doi.org/10.2753/CSA2162-0555470101.2014.11082908

Chinese Sociological Review, vol. 47, no. 1, Fall 2014, pp. 3–29.

ISSN 2162–0555 (print)/ISSN 2162–0563 (online)
DOI: 10.2753/CSA2162-0555470101

An Introduction to the China Family Panel Studies (CFPS)

Yu Xie, *University of Michigan and Peking University*
Jingwei Hu, *Peking University*

Abstract: *The China Family Panel Studies (CFPS), launched by Peking University, is a nearly nationwide, comprehensive, longitudinal social survey that is intended to serve research needs on a large variety of social phenomena in contemporary China. This article describes the background and characteristics of the CFPS, which was designed with the help of methods learned from the most influential survey projects in the world and their experiences. Extensive information is collected through computer-assisted person-to-person interviews of all family members. The questionnaires not only cover a wide range of topics but also consist of intergraded modules for rural and urban interviews, gathering information on family structure and family members, migrant mobility, event history (e.g., history of marriage, education, and employment), cognitive ability, and child development. The CFPS promises to provide the academic community with the most comprehensive and highest-quality survey data on contemporary China.*

China has been undergoing a rapid, large-scale, and irreversible social transformation since the late 1980s. The most significant social changes in this great transformation include rapid economic growth, a huge education expansion, and the completion of China's demographic transition (Xie 2011). These dramatic changes have been accompanied by others, such as increased income inequality (Xie and

We are grateful to Cindy Glovinsky, Yan Sun, Chunni Zhang, and Yuan Zhou, for their research assistance. The research is supported by the Center for Social Research at Peking University and the Population Studies Center (with support from the National Institute of Child Health and Human Development, R24HD041028), the Survey Research Center, and the Center for Chinese Studies of the University of Michigan. Direct all correspondence to Yu Xie, P.O. Box 1248, 426 Thompson Street, Population Studies Center, University of Michigan, Ann Arbor, MI 48106-1248; e-mail: yuxie@umich.edu.

Zhou 2014) and the erosion of traditional family and marriage values (Xie 2013b). The great transformation in contemporary China is, sociologically speaking, a uniquely Chinese phenomenon that invites scholarly studies that pay specific attention to this particular context, with its unique political, economic, historical, cultural, and social backgrounds and constraints. It would be foolhardy to treat China as simply another data point in a statistical data analysis. Its transformation today is no less significant than those events that are commonly considered historical watersheds, such as the Renaissance that began in fourteenth-century Italy, the Protestant Reformation in sixteenth-century Germany, or the Industrial Revolution in eighteenth-century Britain (Xie 2011). Hence, we social scientists today are in a privileged position to observe, document, and analyze the social changes now taking place in China but affecting the course of human history. Even highly trained, experienced social scientists, however, may fail to properly understand this transformation and its full ramifications, as individuals' perspectives are easily confounded by the complex and highly heterogeneous nature of what is happening in China today. Furthermore, all humans, even social scientists, are inevitably limited by their own experiences and values and are inclined to make subjective judgments and unsubstantiated deductions. How, then, can we best understand the social changes that have been taking place in China comprehensively and objectively? What is the best way to analyze and describe the social processes underlying those changes? The answer, we believe, is empirical research based on high-quality data that reflect Chinese reality and its dynamics.

This is not a truly new idea. In fact, an increasingly large number of social scientists in China have begun to take this approach. Since the 1980s, the social science community has conducted many important social surveys, the best-known of which are the Chinese Household Income Project (CHIP), designed by economists (Li, Sato, and Sicular 2013), and the China General Social Survey (CGSS), designed by sociologists (Bian and Li 2012). Treiman, Lu, and Qi (2012) provide a recent overview of many important Chinese surveys. Data from these surveys provide us with important evidence for research on contemporary Chinese society. Most surveys in China to date, however, have been cross-sectional in design and narrowly focused in subject matter. Researchers are in desperate need of high-quality panel data in order to conduct high-quality empirical research (Ren and Xie 2011).

The China Family Panel Studies (CFPS) is a general-purpose, nationally representative, longitudinal survey of Chinese society that was launched in 2010 by Peking University. Designed by an interdisciplinary group of scholars, the survey collects individual-, family-, and community-level longitudinal data in contemporary China in order to provide current and future researchers with comprehensive and objective data on Chinese society. The CFPS focuses on the economic as well as the noneconomic well-being of the Chinese population, gathering a wealth of information covering topics such as economic activities, education outcomes, family dynamics and relationships, migration, and health. Four waves of the CFPS (2010, 2011, 2012, and 2014) have been carried out thus far by the Institute of Social

Science Survey (ISSS) at Peking University. The data that have accumulated in the project can now be used to support both academic and policy research that demand high-quality longitudinal data.[1]

The main objective of this article is to provide an introduction to the CFPS. In the remainder of this introduction, we will cover design ideas, survey practice, and content characteristics of the CFPS project.

The Theoretical Foundation of the CFPS

Units of Analysis

Social science is a population science and, as such, studies units of analysis that are heterogeneous by nature (Xie 2012, 2013a). Different individual units within a population are inherently different from one another, although they belong to the same population and even possess a few of the same observed attributes. A focus on heterogeneity, observed as well as unobserved, is what distinguishes social science from natural science (Xie 2007, 2013a). Sources of individual-level heterogeneity are numerous, ranging from diversities in individuals' genotypes that affect attributes such as height, weight, appearance, and intelligence to variability in social environment and individuals' personal experiences that affect social outcomes. In brief, heterogeneity is affected by social structure and evolves over time. Especially when unobserved, it makes true causal analysis extremely difficult if not impossible, as it introduces elements of uncertainty and potential confounding into social research. One implication is that individuals are often selected into different situations and may react differentially to common social stimuli (Xie 2013a). Through social dynamics of accumulation, also called "path-dependency," and feedback, these differences evolve over time and lead to divergent social outcomes for different individuals. Due to individual-level heterogeneity, statistical results are meaningful only as population averages and do not apply to specific individuals (Xie 2007). This is why even researchers concerned only with population-level statistics cannot afford to overlook individuals at the micro level. Nor can we treat individuals in a population as if they were interchangeable with one another or invariant over time.

Hence, although the CFPS is designed to study social change, its design and data collection focus on specific individuals who were systematically selected in a sample. Possessing great variability, individuals are the most basic units of analysis in a human society. Many social phenomena, such as social inequality, population dynamics, public opinion, and culture, are meaningful only at the macro level, but they are ultimately based on individuals' outcomes, such as income, age, health, happiness, and attitudes through aggregation. Macro-level social changes in domains such as intergenerational mobility regime, population structure, family system, and education are all manifested in attributes measured at the individual level. To understand a society, therefore, we must understand the individuals who constitute

it, in terms of how they are distributed concerning roles, statuses, behaviors, and attitudes. This is why the CFPS treats individuals as their most important units for data collection and will follow up the same group of people over the long term.

Chinese society is a multilevel system in which individuals are embedded in larger social institutions, such as work units, communities, families, and governments (Xie 2010). In this nested structure, the family is the most direct and basic social institution that affects individuals' roles, status, behaviors, and attitudes. Thus, it is of primary importance, for several reasons.

First, the family is usually the first and most significant environment in which a person's socialization takes place. It endows an individual with his/her initial social status and teaches him/her social norms and behaviors in childhood and youth (Duncan et al. 1998; McLanahan and Percheski 2008). The family environment also has a lasting effect on individuals' attitudes, behaviors, and aspirations (Lareau 2011; Thornton and Camburn 1987). Therefore, to study how young people grow into mature adults, we want to know about the families in which they live.

Second, the family affects relations between generations (Blau and Duncan 1967; Sewell and Hauser 1975). Thus, knowledge about family structure and detailed information on all family members is a prerequisite for research involving intergenerational relationships, such as how parents' social statuses affect children, how family resources are distributed among different children, and the transfer of resources between adult children and their elderly parents.

Third, the family is an important social institution that shapes gender roles. Most Chinese adults are married (Xie 2013b). Men and women from different families of origin form new families through marriage. Social status and resources are redistributed and recomposed during this process (Kalmijn 1998). In marriage, there is a gender disparity in the division of household labor that disproportionately burdens women more than men, potentially disadvantaging women relative to men in the labor market (Yu and Xie 2011).

Fourth, the family is a general unit for carrying out most economic activities and social interactions, such as consumption, housing, raising children, support for the elderly, and so forth. For example, poverty is a family-level phenomenon that affects all members in a family living in poverty. To understand Chinese society, we need to understand activities and events at the family level.

Finally, as an institution, the family is extraordinarily meaningful in Chinese culture. Traditionally, Chinese people worshiped their ancestors and observed filial piety. They strove for achievements to glorify their family name. There was a strong norm about forming marriages between families of comparable social status. The desire to continue the family lineage, along the male line, was paramount. The relationship between a young member and his/her family was twofold. On the one hand, the family provided essential material and moral support to the young individual. On the other hand, the young individual was obligated to pay back the family when he/she grew up. This implicit contractual relationship was expected particularly of sons, for whom the dependency relationship was lifelong (Greenhalgh 1985).

The Chinese profamily ideological tradition has survived despite repeated assaults by a number of radical social movements, particularly the large-scale revolutions in the twentieth century: the Xinhai Revolution in 1911 that overthrew the last dynasty, the Communist Revolution that founded the People's Republic of China in 1949, and the Cultural Revolution of 1966–76 (Whyte 2003). Although specific traditional family values have eroded over time, the importance of the family in Chinese culture has not changed. We can see this importance in parents' heavy investments in their children, family-based social relations, and reliance on informal transfers of financial resources among family members.

In summary, when studying Chinese society, we have to pay close attention to the family. For this reason, the CFPS treats it as an important unit of analysis. The CFPS collects comprehensive and exhaustive information about family relations and structures using a novel method, to be discussed later. With such information, researchers can construct an accurate family network of relations and acquire basic social and demographic information on each member. In addition, the CFPS collects detailed information about economic and social life at the family level, which will be useful for research on Chinese society as well.

Research Perspectives

Chinese society is a huge, complicated, and dynamic system. Its complexity manifests itself through social processes at multiple levels of social structure. Social processes are not isolated but closely related both within and across levels. For example, national policies and the economic environment at a contextual level affect individuals' incomes at the micro level (Xie and Zhou 2014). Conversely, the aggregation of behaviors of individuals at the micro level leads to changes at the macro level. The complexity of the social system is further reflected in dynamic interactions across different domains for individuals across the life course. For example, a person's occupation affects his/her income and social network, his/her income and social network in turn affect his/her health and life satisfaction, and so on.

To properly understand the complicated nature of Chinese society, we designed the CFPS surveys to collect comprehensive data at multiple levels and across multiple domains. Community, family and individual are three important levels of the study. At each level, the CFPS gathers information across multiple domains. At the community level, the surveys collect data on the political environment, physical landscape, basic facilities, population, resources, transportation, medical/health facilities, fiscal conditions of the local government, and so on. At the family level, data collection covers family structure and relationships, living conditions, family social networks, income and expenditure, family assets, and so on. At the individual level, the CFPS developed elaborate survey instruments to measure all important life domains such as education, occupation, income, residence, marriage and cohabitation, and physical and mental conditions. In this way, the CFPS is able

to provide researchers with data on multiple domains and across multiple levels of analysis. It will enable researchers to better understand the causal pathways on which individuals evolve over time as they are affected by their family and community environments and their own experiences in the past.

One important feature of all social phenomena is that they are situated not only in space but also in time. Thus, temporality is extremely important. What happened in the past affects an individual's behaviors and outcomes in the present. What happens now will affect an individual's behaviors and outcomes in the future. In other words, social phenomena are accumulated results over time. Thus, time is an important concept in social science research, especially in research to understand causal mechanisms and social changes. From a methodological perspective, time is information used to identify the chronological order of events (Xie 2012). A panel survey collects data for a fixed group of sampled subjects and captures changes over time, which is a highly effective approach to studying the dynamics of social processes. By observing the same group (i.e., the same sample) of subjects at different time points, we are able to acquire detailed information about whether, when, and how the same subjects have changed in their various domains. With this information, researchers will be able to study causal processes temporally at the micro level as well as population trends. Thus, a panel survey is of great value in advancing research in important areas such as understanding population heterogeneity, causal inference, and status transitions in the social sciences (Ren and Xie 2011). Despite high cost, complicated designs, and difficult operations, panel surveys provide more valuable information than cross-sectional and trend studies, offering more rewards for scientific studies. Because of this, a panel design was initially applied in the CFPS to collect data on the target group at different time points. In the long term, the CFPS intends to track gene members who were captured in the CFPS 2010 baseline survey as well as their newborn children and their adopted children if adoption occurs before age ten.[2]

Practice from 2010 to 2012

The CFPS project is being conducted by the ISSS of Peking University. The baseline survey was officially launched in April 2010 and lasted until February 2011. The 2011 survey began in July 2011 and ended in February 2012. The 2011 survey was a small-scale follow-up interview survey and served as sample maintenance and rehearsal for the 2012 full-scale follow-up interview that lasted from July 2012 to March 2013. In the following sections, we will describe the sampling design, fieldwork, and follow-up strategy for the CFPS.

Sampling Design

The CFPS covers twenty-five provinces or their administrative equivalents[3] (municipalities and autonomous regions) in China, excluding Hong Kong, Macao, and

Taiwan, as well as Xinjiang, Tibet, Qinghai, Inner Mongolia, Ningxia, and Hainan. Our target population is all family members in households in the twenty-five provinces. A "household" in the CFPS survey refers to an economically independent dwelling unit with at least one family member of Chinese nationality (excluding Hong Kong, Macao, and Taiwan). "Family members" in the survey refers to (1) all immediate relatives who are economically interdependent; and (2) all non-immediate relatives who are economically related and have been living in the household continuously for three months or longer. Note that a key criterion we use to identify family relationship is economic rather than current residence: people who have left home for school or work but have a close economic relationship with other members of the household are treated as family members. The population of our study's twenty-five provinces represents 94.5 percent of the total population in Mainland China. Given this nearly national coverage of the Chinese population, we refer to the CFPS sample simply as a "nationally representative sample."

The target sample size was 16,000 households. Half of the households came from the five provinces (or their administrative equivalents) of Shanghai, Liaoning, Henan, Gansu, and Guangdong with oversampling, 1,600 households from each province. For convenience, we call these five provinces (or their administrative equivalents) "large" provinces. The other 8,000 households were from one independent sampling frame composed of the remaining twenty "small" provinces. The five large provinces were representative at the province level, which would allow province-level inferences and cross-province comparisons. Through proper weighting, the entire CFPS sample represents the targeted national population in the twenty-five provinces. The "nationally integrated sample," consisting of a subsample from the five large provinces and the sample from the twenty small provinces, is an analytical sample that is almost a self-weighted, representative sample of the national population.

The CFPS used multistage probability proportional to size sampling (PPS) with implicit stratification to reduce the operational cost of the survey and better represent Chinese society. All the subsamples were obtained through three stages: the primary sampling unit (PSU) was either an administrative district (in urban areas) or a county (in rural areas), the second-stage sampling unit was either a neighborhood community (in urban areas) or an administrative village (in rural areas), and the third-stage (final) sampling unit was the household.[4] Administrative units and measures of socioeconomic development were used as the main stratification variables. Within an administrative unit, local gross domestic product (GDP) per capita was used as the ordering index for socioeconomic development. If GDP per capita was not available, the proportion of nonagricultural population or population density was used.

It is important to note that in drawing the CFPS sample, we treated the Chinese population as a single entity instead of separating it into urban and rural populations, as has traditionally been done. Given China's rapid urbanization, the official rural–urban division can hardly reflect its reality dynamically. Thus, we gather

information about the urban versus rural settings through a number of survey instruments, such as the family member's *hukou* status, the urban and rural classification of the community, the agricultural and nonagricultural activities a family engages in, and so on, rather than solely by administrative divisions.

In order to evaluate the representativeness of the sample to the population, we compare the age–sex structure using the CFPS 2010 data (before weighting) and the Census 2010 data. In the population pyramids, shown in Figure 1, we present sex-specific relative sizes of age groups in five-year intervals, from age zero to age 100 (and above) for each sex. Part A is based on the CFPS 2010 data and part B is based on the Census 2010 data. The shapes of the two pyramids are almost identical.[5]

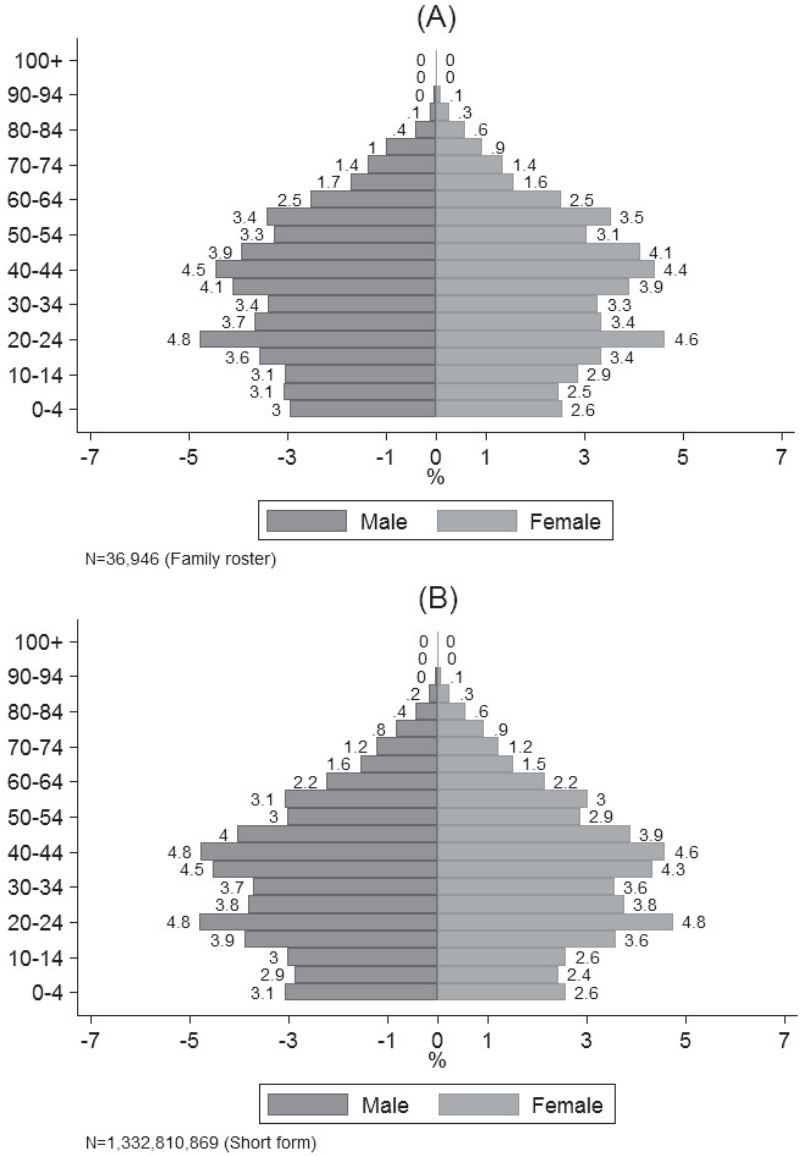
We calculated the weights for data sets respectively from the family, adult, and child questionnaires in 2010. We provide two sets of weights in each data set. One is for the complete national sample; another is for the nationally integrated sample. The process includes calculating for the design weight, the nonresponse weight, and the post-stratification weight, and applying trimming. The design weight is the reciprocal of the product of probabilities in three stages of sampling. The probability of subsampling counties is taken into account when calculating the sample weight for the nationally integrated sample. The nonresponse weight adjusts for differential response rates by observed covariates. The post-stratification weight aligns resulting statistics in demographic characteristics in the CPFS subsamples to be the same as those in the respective subpopulations.

The Fieldwork

The CFPS mainly conducts face-to-face interviews aided by computer-assisted personal interviewing (CAPI) technology. In situations where personal interviews are not feasible, telephone interviews using computer-assisted telephone interviewing (CATI) technology or Web interviews using computer-assisted Web interviewing (CAWI) technology are substituted. The CFPS developed five core questionnaires: the community questionnaire, the family roster questionnaire, the family questionnaire, the adult questionnaire, and the child questionnaire. The community questionnaire collects information at the village/community level and must be answered by a village/community administrator who knows the villages/neighborhoods well. The family roster questionnaire and family questionnaire collect information at the family level. The former is answered by a family member who knows the family relations best, and the latter is answered by one or several eligible family members who are most knowledgeable about the family's economic situation. The adult and child questionnaires collect information at the individual level and are answered by individual adults and children themselves along with children's caretakers.

In our effort to achieve the target sample size of 16,000 households, we enlarged the actual sample size to 19,986 households in the 2010 baseline survey

Figure 1. **Age-Sex Structure of the CFPS 2010 Baseline Survey and the Census 2010**



and successfully interviewed 14,960 households. These households were distributed over 635 urban neighborhoods or villages in 162 administrative counties of the twenty-five provinces. We identified 57,155 eligible family members in the contacted 14,960 households. Interviews with 42,590 of them were completed,

including 33,600 adults and 8,990 children. The 57,155 family members and any children born to or adopted by them before age ten in the future are treated as CFPS gene members and will be tracked throughout their lives. The response rate in 2010 is 81.3 percent at the household level and 84.1 percent at the individual level.

The 2011 survey was less ambitious for two reasons. First, it served to maintain the family sample. Second, we wanted to conduct this round as a rehearsal for the 2012 survey, which was going to be the first major round of follow-up interviews for the whole sample. Given that most youths still lived at home and their conditions, behaviors, and attitudes change quickly, a decision was made to interview all those young gene members who were successfully interviewed in 2010 and were no more than eighteen years old in 2011. Households that were successfully interviewed in 2010 were also followed up. Given our limited objectives, we had follow-up interviews neither with adult gene members who were over eighteen years old nor with the newborn and adopted gene members and the newly formed families since 2010. In the end, we successfully interviewed 13,130 households, and 8,803 gene individuals, including 1,279 adults (sixteen to eighteen years old) and 7,524 children (younger than sixteen years old). We achieved a success rate of 2010–11 reinterviewing at 89.1 percent at the family level.

In 2012, we followed up all members of the CFPS sample and the households in which they lived at the time of interviewing. The targeted group included not only all the old gene members from the 2010 survey but also all the new gene members born or adopted since that survey. The households could be the same ones as in 2010, but they could also be households newly formed or previously existing but first interviewed in 2012 by the CFPS due to gene members' marriages, divorces, or separations. We also collected extensive information about non-gene members in the same households. Among the 14,960 households that were successfully interviewed in the 2010 survey, 12,725 households were successfully interviewed in the 2012 survey. The successful-tracking rate was 85.1 percent at the family level. We contacted 52,336 gene members and 2,737 non-gene members from the 12,725 households and another 728 new households separating from them; 42,970 gene members and 1,714 non-gene members were successfully interviewed. Among the successfully interviewed 42,590 gene members in the 2010 survey, 33,956 members were successfully followed up in the 2012 survey. The successful-tracking rate was 80.6 percent at the individual level. Among the 52,336 gene members, 8,477 members left home—with or without financial relations with the households, 8,341 of them needed to be tracked, and 5,756 were successfully interviewed. Thus the successful-tracking rate for those gene members was 69.0 percent. Table 1 below summarizes basic information about the three surveys.

Follow-Up Strategy

As we discussed earlier, panel survey data can help researchers better understand causal processes over time. It is imperative to have a follow-up strategy in place

Table 1

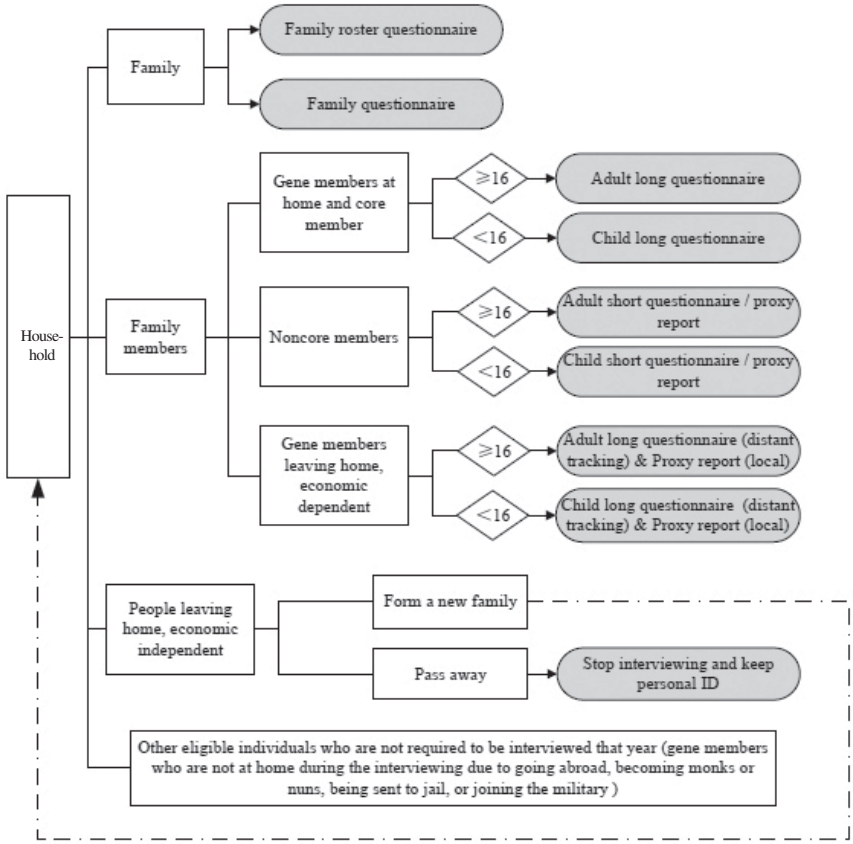
Basic Information on CFPS 2010, 2011, and 2012 Surveys

	2010	2011	2012
Objective			
Family level	19,986 households in random sampling	All households who completed interviews in 2010	Households of gene members
Individual level	All households members in sampled households	Gene members eighteen years olds and below who completed interviews in 2010	Gene members and other members from the same households
Community level	649 neighborhoods/ villages of sampled households	—	—
Sample size			
Community	635	—	—
Households	14,960	13,130	13,453
Gene members	42,590	8,803	42,970
Adults	33,600	1,279	34,447
Children	8,990	7,524	8,523
Interview mode	CAPI	CAPI+CATI+CAWI+ Mailing Questionnaire	CAPI+CATI
Contact results			
Family level	81.3% (response rate)	89.1% (reinterview rate)	85.1% (reinterview rate)
Individual level	84.1% (response rate)	88.8% (reinterview rate)	80.6% (reinterview rate)

about who is tracked and how to track them. Since the research subjects are families and individuals, complications arise when births and deaths occur and when families are split by individuals' moves across families. Any follow-up strategy should accommodate research needs while being constrained by limits in terms of time, funding, and technology.

The current plan is that a CFPS survey will be conducted biennially, with occasional annual surveys. Figure 2 shows the flow chart for tracking. All family

Figure 2. **the CFPS Flow of Tracking and Questionnaire Generation Rules**



members in the baseline survey and their new children born or adopted since the baseline are defined as CFPS gene members and will be tracked throughout their lives. All households in which at least one gene member lives will be interviewed, with interviewing discontinued whenever no gene member is living there (i.e., if all gene members in the household have moved out or have passed away). Over the long term, there will be both a depletion of old gene members due to natural deaths and a replenishment of new gene members through births in the CFPS sample. New CFPS families will appear for reasons such as marriage or divorce, and old CFPS families will disappear due to the deaths as well as movements of gene members. Thus, the CFPS sample is self-renewing. In the ideal situation of no attrition over time, it reflects the natural changes of the Chinese population and Chinese families.

In order to reduce costs and to increase efficiency while collecting valuable information about changing family dynamics, with the 2012 survey we began to

sort family members into three groups: gene members, core members, and noncore members. We now treat gene members, gene members' non-gene parents, non-gene children, and non-gene spouses living with them in the same household as core members, and other family members as noncore members. We require that core members answer the same individual (adult or child) questionnaires as gene members, which we call "long questionnaires." Noncore members answer the abbreviated 2012 individual questionnaire, which we call the "short questionnaire." We also use short questionnaires to collect key information through proxy reports by other family members when we are not able to conduct personal interviews. With this design, we can collect detailed information for the most important family members, as well as some vital information for the less important members in order to learn about the family background and relationships in the gene member's household. Core members and noncore members, however, will not be tracked permanently. We will stop interviewing them once they no longer live with any gene members. In addition, interviews with gene members who are not at home during the interviewing due to going abroad, becoming monks or nuns, being sent to jail, or joining the military are not required that year.

We pay particular attention to sample attrition when conducting a follow-up survey. In order to retain the sample, the following two measures are taken in the CFPS survey. On the one hand, for the family or respondent who could not be interviewed face-to-face, we substitute telephone or Web interviews whenever possible. On the other hand, for those gene members who cannot be reached by any means, we use proxy responses to the short questionnaire by other family members, discussed earlier.

Content Features

In the early stages of its work, the design team of the CFPS learned from the approaches and experiences of earlier successful research programs, such as the Panel Study of Income Dynamics (PSID), the National Longitudinal Surveys of Youth (NLSY), the Health and Retirement Study (HRS), and so on. As with those surveys, the content of the CFPS is thorough and comprehensive (see Table 2)—including important events over an individual's entire life course as well as some specific designs on family relationships, family economy, and communities—in order to meet various research needs from different academic fields. The CFPS not only covers common substantive subjects expected by many social science researchers but also has its own unique content.

Accurate Family Network and Complete Information on Family Members

Family, marriage, intergenerational relationships, and mobility are classical research topics in sociology, economics, and related fields that are still hotly studied

Table 2

Comparison Between 2010, 2011, and 2012 CFPS Questionnaire Contents

	2010	2011	2012		2010	2011	2012
Community Questionnaire				*Family Questionnaire*			
Facility, population, environment, and so on	√			Geography and transportation	√	√	
Family Roster Questionnaire				Living conditions	√		√
Family structure and relationship	√	√	√	Social activity	√		
Basin information of family member	√	√	√	Agricultural work	√	√	√
Adult Questionnaire				Nonagricultural business	√	√	√
Basic information	√	√	√	Individual income	√	√	
Migration	√	√	√	Transfer payments	√	√	√
Siblings	√			Family expenditure	√	√	√
Special experiences	√		√	Housing	√	√	√
Academic education	√	√	√	Land	√	√	√
Nonacademic education			√	Financial assets	√	√	√
Language ability	√	√	√	Debts	√	√	√
Marriage	√	√	√				
Relationship with children	√		√	*Child Questionnaire (answered by parents)*			
Work	√	√	√	Basic information	√	√	√
Retirement and pension	√		√	Migration	√		
Time usage	√	√		Parenting	√	√	√
Leisure activities	√	√		Growth	√		√

Cellphones and Internet	√	√		Health	√	√	√
Social network	√	√		Education expense	√	√	√
Religion			√	Views on parenting	√	√	√
Politics	√	√	√	Schooling	√	√	√
Attitude	√	√	√	Extracurricular/home tutoring	√	√	√
Health	√	√	√	Family interaction	√	√	√
Behavior	√	√	√	*Child Questionnaire (answered by children)*			
Mental status	√	√	√	Schooling history	√	√	√
Cognitive testing	√	√	√	Employment	√	√	√
Special physical activity testing	√	√	√	Time usage	√	√	
Parents' information			√	Language ability	√	√	√
				Social activity	√	√	√
				Cellphones and Internet	√	√	
				Health	√	√	√
				Attitude	√	√	√
				Behavior and mental status	√	√	√
				Cognitive testing	√	√	√

today. Research potential in these fields depends heavily on the extent to which complicated information on family members and family relationships can be adequately collected. Almost all surveys prior to the CFPS have selected only one respondent as the anchoring point for family relations and asked that respondent about relationships with other, coresiding family members. Because this traditional method can collect information only in a dyadic structure from one single anchoring member, typically the household owner or a randomly chosen family member, researchers can learn only about the relations between this anchoring respondent and other family members. The direct relations between family members/relatives other than those with the anchoring respondent remain unknown. Thus, information collected using this traditional method often fails to meet certain research needs. For example, information about in-laws is restricted if only one partner of a couple serves as the anchoring respondent. Similarly, research on intergenerational relations is restricted if the survey only bases information on either a parent or a child of the anchoring respondent. Moreover, with the traditional method, information on family members, except for the household owner or the key respondent, is always collected through proxies, even when they are important family members, such as spouses, parents or children.

The CFPS creatively developed T-tables to collect information on family members and family relations for the 2010 baseline survey. The T-tables are composed of three tables—T1, T2, and T3—used in the beginning section of the family roster questionnaire (see Figure 3). Table T1 records the basic social demographic features of all family members, including those who have left home but still keep close economic ties with the household. Table T2 records the names of mother, father, spouse, and children of each member in T1. The names listed in T2 but not shown in T1 are the fathers, mothers, spouses, and children of T1 members but themselves are not family members. Information on these nonfamily members' basic social demographic features is further collected and recorded in Table T3, structured similarly to T1. Therefore, Tables T1 and T3 provide researchers with information on the basic social and demographic characteristics of all family members and their parents, children, and spouse, regardless of whether they live in the household. Table T2 helps researchers identify the relations between all family members in T1 and the corresponding relations between T1 and T3 members. In addition, by matching the family roster data from the family roster questionnaire and the individual data from the adult or child questionnaire, we can obtain complete and symmetric information on many one-to-one pairs—parents and children, wife and husband, siblings living in the household, and so forth. For some families, researchers can even retrieve complete cross-generational data.

Compared to traditional domestic surveys, the design of the T-tables has the following advantages. First, traditional surveys focus family relations on those pertaining to one node—typically a randomly selected family member as the key respondent. No matter how the researcher identifies the key respondent, the assumption that we should be interested in family relationships around only one

Figure 3. **Illustration of T-Tables**

Table on Information of Regular Family Members (T1)

Personal code	Name	Date of birth/Year of Chinese Zodiac/Age	Gender	Marital status	Highest education	Primary occupation	Administrative/ managerial position	Current residential location
101								
102								
…								
301								
302								
…								

Table on Immediate Relatives (father, mother, spouse, and children) (T2)

Personal code	Name	Father	Mother	Spouse	Child 1	Child 2	…	Child 9	Child 10
101									
102									
...									

Table on Information of Immediate Relatives (father, mother, spouse, and children) not Listed in Table T1 (T3)

Personal code	Name	Date of birth/ Year of Chinese Zodiac/Age	Gender	Alive or not	Marital status	Highest education	Primary occupation	Administrative/ managerial position	Information on residence and hukou
301									
302									
…									

family member is an oversimplification. In fact, everyone in the family can be a node in family relations, resulting in networks with multiple nodes. However, the traditional surveys could only construct a radiating structure from a single anchoring family member (e.g., the household holder or a key respondent), which is only a small part of the complete family network, and of course can collect only partial information on family relations. The relations between family members/relatives other than those with the anchoring respondent are unknown. The T-table design allows each member to be an anchoring point in rotation. Researchers can learn not only about the direct relations but also about some indirect relations (e.g., stepparents and stepchildren), not only about relations across generations (e.g., grandparents and granddaughters) but also about relations between siblings of the same generation.

Second, in the CFPS survey, the interviewee is not restricted to one or two adults, as in many traditional surveys. Instead, the CFPS collects complete and symmetric information on all gene and core members, including children, by interviewing multiple family members. Personal interviews are conducted with all gene and core members age ten years or older. To collect information on children younger than age ten, the CFPS interviews the main caretaker, and it even conducts abbreviated or proxy interviews with less important, noncore family members. Moreover, the T3 table (used in 2010) collects some basic social demographic information on parents, spouse, and children who are not living in the household, important information that is typically missing in traditional surveys.

Third, many previous surveys only asked questions about the respondent's father, mother, and children without recording their names or coding for them. The naming and coding system of the CFPS makes it possible to measure direct relationships between all family members. The system also proves valuable for tracking in future waves.

The design of the T-tables fixes many problems that plagued previous surveys, such as inaccurate and incomplete information on family relations. It provides richer information on sampled CFPS families and useful information on all family members, enabling researchers to study these families with more and better data.

Integration of Rural and Urban Areas

As we discussed earlier, the CFPS treats rural China and urban China as an integrated population in sampling. Thus, we do not have two separate sets of questionnaires for rural and urban areas as many traditional surveys do. This is because, with the rapid development of urbanization, the official rural–urban division no longer reflects the actual reality of an area. Due to forces of modernization in rural areas and population migration, differences between the rural and urban areas are no longer as sharply drawn as before. In contrast, heterogeneity within each type of area has become more and more important. Operationally, interviewing with an

urban questionnaire in rural areas and a rural questionnaire in urban areas would lead to question inapplicability. Thus, it is necessary to have integrated questionnaires for rural and urban areas.

The CFPS's improvements over traditional practices are reflected in two ways. First, the CFPS collects a set of variables in the questionnaires that can be used to demarcate rural and urban areas. At the community level, it identifies whether a sampled community is an urban neighborhood or a village. At the family and individual level, the CFPS identifies individuals' *hukou* types and whether they engage in agricultural work. Researchers can analyze actual rural–urban differences according to research needs by using such information rather than simply relying on administrative division.

Second, the CFPS questionnaire design is modular. With the help of the CAPI/CATI/CAWI, the CFPS effectively resolves the potential problem of asking respondents inappropriate questions by constructing personalized questionnaires. For example, we launch the questionnaire's agricultural work module only for families and respondents participating in agricultural work.

Migrant Population

Population migration has been a significant social phenomenon in China since the 1990s. The sixth census in China revealed that the migrant population had reached 221.43 million in 2012 (Ma 2011). This migration not only changes the population structure and distribution between rural and urban areas but also affects the structure of the labor market and the social stratification order. New social problems emerge because of migration, such as migrant-related crimes, left-behind children, a rising divorce rate in rural areas, and unmet support needs among the elderly (Davin 1996; Lu 2012; Silverstein, Cong, and Li 2006; Wu 2013).

Researchers and policymakers alike are in need of high quality, reliable data on population migration. The available data from the many other sources on migrants typically suffer from three types of problems. First, traditional surveys on the migrant population (such as migrant workers) usually take samples from current migrants but cannot cover former migrants in Places of origin places or potential migrants who have not migrated. Research on migrants without information from the latter two groups is likely to be confounded by selection biases. Second, cross-sectional surveys, which focus only on the current situations of the migrants, do a poor job of collecting information on their premigration experiences. These shortcomings present major problems for studying the causal effects of migration on later life-course events. Although some cross-sectional surveys collect retrospective information to mitigate this problem, these retrospective items suffer from recall memory errors. Finally, the high frequency of migrants' mobility makes sampling and survey operation difficult, which in turn affects the representativeness and accuracy of data collected.

The CFPS resolves these problems. First of all, the CFPS survey collects data on all gene family members whether they are at home or have left home for study, work, marriage, or other reasons. This design meets the need for research on the selectivity of migrants and facilitates comparison between migrants and nonmigrants. By interviewing the same sample at different time points, the CFPS is able to capture all residential changes of respondents over time, tracking migration history more accurately and completely. Second, in each wave of the survey, the CFPS tracks all the gene members from their families in the last wave. Thus, if a person does not live in his/her original family, we can still learn where he/she now lives and obtain the person's contact information. Finally, the telephone and Web interviews help to reduce attrition among migrants, a very difficult group to follow over time. Even when we fail to complete either the face-to-face interview, the telephone interview, or the Web interview, proxy answers to a series of questions by other household members fill in some important values that otherwise would be completely missing.

Cognitive Assessments

Sociologists and economists have long been concerned with factors that affect individuals' attainment of socioeconomic status (e.g., income, occupation). Human capital—defined as a series of skills and capacities that can help to improve productivity—is an important factor in explaining income attainment as well as other social outcomes in the labor market (Mincer 1974; Sewell and Hauser 1975; Xie and Hannum 1996). Much prior research is based on the Mincer model, which measures human capital with education and work experience (Mincer 1974). However, schooling, training, and work experience are only part of human capital. Workers' ability, especially cognitive ability, affects their productivity in direct ways. Cognitive ability reflects an individual's intelligence as well as the effects of schooling and training. Ignoring this factor may lead to overestimation of regular education's effects on income (Griliches 1977). Cognitive ability tests are widely used abroad in social surveys, business management, and selection of armed forces. They also provide explaining variables for a substantial number of empirical research studies. Due to technology and cost constraints, however, tests of cognitive ability have seldom been conducted in social surveys in China.

The CFPS overcomes this limitation and plans to collect cognitive ability information over the long term. It has developed two sets of cognitive tests for respondents age ten years and older. One set of tests used in 2010 and 2011, also repeated in 2014, consists of literacy and mathematics questions. The other set of tests was used in 2012, consisting of memory and number-sequence questions, the latter of which are intended to assess respondents' mathematical reasoning abilities. The CFPS plans to rotate the two sets of tests by wave over the long term so that researchers can study the social and demographic determinants of cognitive ability over time and the influences it has on achievements, behavior, and attitudes.

Child Development

In recent years, social scientists have become increasingly concerned with children's development and its effects on their future social achievements, behaviors, and attitudes. However, few family surveys focus on children in China, and almost no panel data are available on Chinese children. To fill the gap, the CFPS collects longitudinal data on dependent children in the CFPS families. We first note that all gene children under sixteen years old are covered in the CFPS. Information for children under age ten is collected from their caretakers, while children age ten or older are interviewed directly, along with a questionnaire that caretakers answer. An apparent advantage of the CFPS survey is that researchers have access to information on newborn gene children from birth.

Further, the child questionnaire not only collects complete information on school education but also provides the researcher with a long-term assessment of children's cognitive ability. In addition, the CFPS especially implemented items to measure children's personality, mental and physical health, learning habits, and daily behaviors. All of these topics are of great value for research on children's growth and development, attainment of socioeconomic status, and contributions to their society and economy.

Life History

The individual life course is a reflection of social transformations (Elder 1985). Thus, by comparing the life courses and experiences of different age cohorts, researchers can better understand social transformations in history. Research on the life course requires life history data. Researchers need to know not only whether an event happened but also when it happened, how long it lasted, and what the sequence was between different events.

The CFPS collects detailed information on each gene family member's education and marriage history retrospectively and tracks later major life changes throughout a person's life. For education history, the CFPS asks about complete educational experiences in each stage, including the starting and ending dates, the field of study, whether the individual finished school in that stage, and so on. For marriage history, the CFPS is interested in respondents' first and current marriages and how these marriages have evolved over time. Questions are asked about the starting and ending time of each marriage, the spouse's age, education, and occupation, and how the couple knew each other. While life history data were already collected in earlier waves, from 2014 onward the CFPS began to collect detailed changes in residence, marriage status, and jobs through the instrument of the event history calendar.

The CFPS collects data on certain major events that may be turning points in an individual's life, such as joining the army, settling down in the countryside, and experiencing the Great Famine of 1959–62. It is notable that the CFPS gathers

information on cohabitation, which has sharply increased and become extremely important in recent years in China. However, past surveys did not cover this topic due to its sensitivity. The CFPS recognizes the importance of cohabitation and began to collect cohabitation data in its initial test phase in 2008.

Interview Observation Data

Interview quality is of the highest importance to the CFPS. To evaluate interview quality and to supplement interview data, the CFPS collects a set of observation data at the end of each questionnaire (shown in Figure 4). Interviewers are asked to answer a set of questions from their own observations. Questions in this module are mainly about the environment of the interview, the respondents' behaviors, attitudes and personal traits, and the interview experiences. These data can be used directly for research. As a procedure for data collection, these items are especially useful for evaluating the reliability of respondents' responses and providing evidence for evaluating interview data quality.

CFPS Data

The CFPS data consist of data sets pertaining to the community, family roster, family, adult family member, and child family member. As explained earlier, there are six sampling frames representing six subpopulations. In the data sets released, we marked different subpopulations with the indicator variable, "subpopulation." The values from 1 to 6 of the "subpopulation" respectively representing Shanghai, Liaoning, Henan, Gansu, Guangdong, and a combination of twenty other provinces. Also, we added a dichotomous indicator, "subsample." The value 1 of the "subsample" refers to the nationally integrated sample and the value 0 refers to the complete national sample.

The complete national sample includes the entire CFPS sample composed of the six subsamples representing the six subpopulations. After weighting,[6] the complete national sample represents the national population. The nationally integrated sample was constructed by resampling the oversampled subsamples in the five large provinces and combining them with the subsample in the other twenty small provinces so that they are proportional in the sampling ratio to the small provinces. The nationally integrated sample is directly representative on the national level (i.e., without weighting).

We created composite variables from original data. One reason for this is to create commonly used variables for users' convenience, such as "fdepression" and "depression," the factor score and the additive index score measuring the level of depression based on the mental health scale. "Wordtest" and "Mathtest" are scores of respondents' cognitive ability based on the cognitive tests. "Gap_fam" indicates the number of generations in a family. We also convert the occupational

Figure 4. **CFPS Interviewers Observations Questions**

Z1 Who was the main respondent that completed this questionnaire?
Z101 Who else answered this questionnaire in the family? [Limited to two respondents]
Z102 Who was present during the interview except the family members? [Select all that apply]
1. Relatives or friends
2. Interview supervisor
3. Neighbor
4. Cadre of the village or community
77. Other [Please specify his/her identity]
78. No one else
Z103 What was the main language used in the interview? [Single choice]
1. Mandarin (skip to Z201)
5. Dialect (continue to Z104)
Z104 What was the dialect used in the interview? _____
Z201 Respondent's comprehension of questions:
Very poor—1–2–3–4–5–6–7– > Very good
Z202 Respondent's physical condition:
Very poor—1–2–3–4–5–6–7– > Very good
Z203 Neatness/cleanliness of respondent's clothing:
Very poor—1–2–3–4–5–6–7– > Very good
Z204 Respondent's appearance:
Very poor—1–2–3–4–5–6–7– > Very good
Z205 Respondent's Mandarin fluency:
Very poor—1–2–3–4–5–6–7– > Very good
Z206 Respondents cooperation during the interview:
Very poor—1–2–3–4–5–6–7– > Very good
Z207 Respondent's intelligence:
Very low—1–2–3–4–5–6–7– > Very high
Z208 Respondent's courteousness:
Very poor—1–2–3–4–5–6–7– > Very good
Z209 Respondent's interest in the interview:
Very low—1–2–3–4–5–6–7– > Very high
Z210 Respondent's concern about the interview:
Very low—1–2–3–4–5–6–7– > Very high
Z211 Reliability of respondent's response:
Very low—1–2–3–4–5–6–7– > Very high
Z212 Respondents' ability to express themselves:
Very weak—1–2–3–4–5–6–7– > Very strong
Z5 Respondent's impatience with the interview:
No—1–2–3–4–5–6–7– > Yes

codes to International Standard Classification of Occupation codes (ISCO-88), International Socio-Economic Index of Occupational Status (ISEI), Treiman's Standard International Occupational Prestige Scale (Treiman's SIOPS), and Erikson and Goldthorpe's Class Categories (EGP), and create related composite variables, respectively.

The other reason for creating composite variables is to avoid confusion when there are multiple sources for the same variable. We have access to auxiliary information and thus are able to create better measures than ordinary data users, so we prepared such measures before data releases. The variable "qaly_best," best values of the year of birth, is a good example of this. In the baseline survey, we could acquire this information for family members from three sources: (1) the information in the family roster questionnaire provided by one family member's proxy report; (2) the information provided by the respondent him/herself in his/her individual questionnaire; and (3) the spouse's answer in the marriage module in the spouse's individual questionnaire, which we can obtain after spousal matching (if applicable). Through statistical analysis, we know that the information from these three sources was not always consistent. We derived the "best" year of birth both manually and with the help of a computer program. After checking the logic relations between the year of birth and several other life events and comparing birth year information between the baseline survey and the follow-up surveys, we eliminated the unreasonable values and gave the best values to "qa1y_best."[7] We created a set of such "best" variables: "qe1_best" is the most reasonable marital status; "qe605y_best" is the most reasonable year of the first marriage; "qe606y_best" is the most reasonable year of birth of the first spouse. More information about composite variables can be obtained from the CFPS users' manual and from technical reports.

Conclusion

The CFPS is currently the largest and most comprehensive social panel survey in China. To help researchers learn more about the project and the data, this paper introduced the design, the practice, and the content of the survey project in great detail. We also briefly introduced the data.

In terms of design, the CFPS treats both individuals and families as its research subjects and carries out its design from multiple levels and a longitudinal perspective, giving full consideration to the heterogeneity, nesting structures, complexity, and temporality of social phenomena. In practice, the survey obtains its nationally representative sample by integrating rural and urban areas. It has improved interview quality with the support of advanced technology and minimizes refusal and attrition rates by including telephone and Web interviews. Its follow-up strategy has improved the practicability and success rate of tracking respondents over time. In terms of content, the survey has not only learned from earlier successful research programs in the United States but also developed its own unique features, such as the T-tables, the integration questionnaire for rural and urban areas, life history,

cognitive tests, child development, and interviewer-observation data.

China has been undergoing an unprecedented social transformation. It is historically imperative that social scientists objectively record the process of this transformation and carefully study its mechanisms and outcomes. Only by making this effort can social scientists contribute to the accumulation of knowledge, the very basis for formulating effective policies that create positive social changes. Of course, collecting high quality data is only a beginning. For the CFPS project to realize its full value, the world scientific community must fully exploit its data in conducting meaningful and significant social research on family life in China in the years ahead.

Notes

1. The CFPS Web site is www.isss.edu.cn/cfps/.

2. "Gene member" refers to all family members in the CFPS 2010 baseline survey and their newborn/adopted children thereafter.

3. For simplicity, they will all be called "provinces" in this paper.

4. Shanghai is different from other "large" provinces; therefore, the sampling procedure was slightly different. More information can be found in Xie, Qiu, and Lu (2012).

5. We calculated the sampling bias rate of each age–sex group compared to the Census 2010 (see Xie 2013c).

6. For details on weights, see Lu and Xie (2013).

7. We did not delete the original values along with this variable, although there is a discrepancy between them. There were several reasons for this. First, they were the most reasonable values after our careful considerations but we were not sure if they were 100 percent correct. Second, the age of the respondent is directly related to the type of questionnaires he/she had answered. Their answers were either correct or incorrect, the questionnaires of individuals being automatically created according to the year of birth they claimed. The original values provided evidence for the questionnaires. For this reason, we must keep them; otherwise it might confuse data users about the rules of creating the questionnaires. But the "qaly_best" variable is the best one for researchers to use in measuring age.

References

Bian, Yanjie, and Lulu Li. 2012. "The Chinese General Social Survey (2003–2008): Sample Designs and Data Evaluation." *Chinese Sociological Review* 45(1): 70–97.

Blau, Peter, and Otis Dudley Duncan. 1967. *The American Occupational Structure.* New York: Wiley.

Davin, Delia. 1996. "Gender and Rural–Urban Migration in China." *Gender and Development* 4: 24–30.

Duncan, Greg J., W. Jean Yeung, Jeanne Brooks-Gunn, and Judith R. Smith. 1998. "How Much Does Childhood Poverty Affect the Life Chances of Children?" *American Sociological Review* 63(3): 406–423.

Elder, Glen H. Jr. 1985. "Perspectives on the Life Course." In *Life Course Dynamics: Trajectories and Transitions, 1968–1980,* edited by Glen H. Elder Jr., pp. 23–49. Ithaca, NY: Cornell University Press.

Greenhalgh, Susan. 1985. "Sexual Stratification: The Other Side of 'Growth with Equity' in East Asia." *Population and Development Review* 11: 265–314.

Griliches, Zvi. 1977. "Estimating the Returns to Schooling: Some Econometric Problems." *Econometrica* 45: 1–22.
Kalmijn, Matthijs. 1998. "Intermarriage and Homogamy: Causes, Patterns, Trends." *Annual Review of Sociology* 24: 395–421.
Lareau, Annette. 2011. *Unequal Childhoods: Class, Race, and Family Life, with an Update a Decade Later.* Berkeley: University of California Press.
Li, Shi, Hiroshi Sato, and Terry Sicular. 2013. *Rising Inequality in China.* Cambridge: Cambridge University Press.
Lu, Ping, and Yu Xie. 2013. "Weight Calculation (2nd ed.) of the CFPS 2010 Baseline Survey ." *Technical Report: CFPS-17*. Institute of Social Science Survey (in Chinese). Available at www.isss.edu.cn/cfps/d/file/p/a88086f58ae9d67eb1a8a1a5bf24563c.pdf.
Lu, Yao. 2012. "Education of Children Left Behind in Rural China." *Journal of Marriage and Family* 74: 328–341.
Ma, Jiantang. 2011. "Data Report from 2010 Chinese Census." National Bureau of Statistics. Available at www.stats.gov.cn/tjfx/jdfx/t20110428_402722238.htm (in Chinese).
McLanahan, Sara, and Christine Percheski. 2008. "Family Structure and the Reproduction of Inequalities." *Annual Review of Sociology* 34: 257–276.
Mincer, Jacob. 1974. *Schooling, Experience, and Earnings.* New York: Columbia University Press.
Ren, Qiang, and Yu Xie. 2011. "Statistical Analysis of Longitudinal Data ." *Population Research* 35(6): 3–12 (in Chinese).
Sewell, William H., and Robert M. Hauser. 1975. *Education, Occupation, and Earnings: Achievement in the Early Career.* New York: Academic Press.
Silverstein, Merril, Zhen Cong, and Shuzhuo Li. 2006. "Intergenerational Transfers and Living Arrangements of Older People in Rural China: Consequences for Psychological Well-Being." *Journal of Gerontology* 61B: S256–S266.
Thornton, Arland, and Donald Camburn. 1987. "The Influence of the Family on Premarital Sexual Attitudes and Behavior." *Demography* 24(3): 323–340.
Treiman, Donald J., Yao Lu, and Yaqiang Qi. 2012. "New Approaches to Demographic Data Collection." *Chinese Sociological Review* 44(3): 56–92.
Whyte, Martin K. 2003. "China's Revolutions and Intergenerational Relations. " In *China's Revolutions and Intergenerational Relations,* edited by Martin Whyte, pp. 3–23. Ann Arbor: Center for Chinese Studies, University of Michigan.
Wu, Yaowu. 2013. "One-Child Policy and the Elderly Migration." *Sociological Research* 4 (in Chinese).
Xie, Yu. 2007. "Otis Dudley Duncan's Legacy: the Demographic Approach to Quantitative Reasoning in Social Science." *Research in Social Stratification and Mobility* 25: 141–156.
———. 2010. "Understanding Inequality in China." *Society* 30(3): 1–20 (in Chinese).
———. 2011. "Evidence-Based Research on China: A Historical Imperative." *Chinese Sociological Review* 44(1): 14–25.
———. 2012. *Sociological Methodology and Quantitative Research*. Beijing: Social Sciences Academic Press (in Chinese).
———. 2013a. "Population Heterogeneity and Causal Inference." *Proceedings of the National Academy* of Sciences (PNAS) 110(16): 6262–6268.
———. 2013b. "Gender and Family in Contemporary China." *Research Report 13-808.* Population Studies Center, University of Michigan. Available at www.psc.isr.umich.edu/pubs/pdf/rr13-808.pdf.
———. 2013c. *China Family Panel Studies User's Manual*. Institute of Social Science Survey. Available at www.isss.edu.cn/cfps/d/file/wd/jsbg/2010jsbg/a16476476881bc46c90086fa1508880b.pdf (in Chinese).

Xie, Yu, and Emily Hannum. 1996. "Regional Variation in Earnings Inequality in Reform-Era Urban China." *American Journal of Sociology* 101: 950–992.
Xie, Yu, and Xiang Zhou. 2014. "Income Inequality in Today's China." *Proceedings of the National Academy of Sciences* (PNAS) 111: 6928–6933.
Xie, Yu, Zeqi Qiu, and Ping Lu. 2012. "Sampling Design." *CFPS-1*. Institute of Social Science Survey (in Chinese). Available at www.isss.edu.cn/cfps/d/file/wd/jsbg/2010jsbg/c70c703752f7f6f24e27b5a8a85c1cba.pdf.
Yu, Jia, and Yu Xie. 2011. "The Varying Display of 'Gender Display.'" *Chinese Sociological Review* 44 (2): 5–30.

About the Authors

Yu Xie is Otis Dudley Duncan University Distinguished Professor of Sociology, Statistics, and Public Policy and a research professor at the Institute for Social Research, University of Michigan, and Visiting Chair Professor at Peking University. His main areas of interest are social stratification, demography, statistical methods, Chinese studies, and sociology of science. His recently published works include: *Marriage and Cohabitation* (University of Chicago Press 2007) with Arland Thornton and William Axinn, *Statistical Methods for Categorical Data Analysis* with Daniel Powers (Emerald, 2008, 2nd ed.), and *Is American Science in Decline?* (Harvard University Press, 2012) with Alexandra Killewald.

Jingwei Hu is a research associate at the Center for Social Research, Peking University, where she works on questionnaire design, data cleaning, coordination, and documentation for CFPS. She has a bachelor of law degree in politics and a master of management degree in public administration from the School of Government, Peking University.