



User Guide for China Family Panel Studies 2018

Qiong Wu

Lihong Dai

Qi Zhen

Liping Gu

Yirui Wang

August, 2021

This user guide draws heavily from documentation from prior years written by numerous CFPS staff members. This version was edited by Yiwen Zhao.

This document may be cited as follows:

Wu, Q., Dai, L., Zhen, Q., Gu, L., Wang, Y. User Guide for China Family Panel Studies 2018. Institute of Social Science Survey, Peking University, August 2021.

User Guide for China Family Panel Studies 2018

The earlier waves of the China Family Panel Studies are described in the user's manual. This report includes new features that the CFPS users may find helpful when analyzing data from its more recent wave of 2018.

1. General introduction

Field work for CFPS2018 began in June, 2018. Face-to-face interviews were completed by March, 2019. Telephone interviews continued till May, 2019. A total of 15,051 families completed the family roster questionnaires. At the respondent level, we completed a total of nearly 44,000 interviews including both self reports and proxy reports. About 400 interviewers participated in data collection. At the household level, the CFPS2018 had a response rate of 69.3% and cross-wave retention rate of 86.6%. At the individual level, the cross-sectional response rate was 67.4% and cross-wave retention rate was 80.8%. If we focus on gene members from the baseline wave of 2010, the response rate in 2018 was 64.5%.

1.1 New features in questionnaire designs

The CFPS2018 survey included a family roster questionnaire focusing on basic sociodemographic information of all family members and their relationships to one another; a family questionnaire focusing on family income, expenditure and assets; an individual self-report questionnaire for all respondents aged 10 and above; an individual proxy questionnaire which was an abbreviated version of individual self report; and a child proxy questionnaire for respondents aged 0-15. All questionnaires are posted on the project website under "Documentation" in the "Questionnaires" section.

As a longitudinal survey, CFPS is intended to keep most parts of its questionnaire design consistent across waves. Only a few changes were made in CFPS2018. The first change is related to harmonizing individual self reports for respondents between 10-15 and those above 15. They were administered different questionnaires in years prior to 2018. In CFPS2018, we harmonized them into an integrated self report. Within the individual self report questionnaire, individuals would be screened into appropriate modules based on age and other indicator variables. Second, all individuals who were not residing in the home address at the time of the household interview would be eligible for proxy report, and proxy reports would be completed by a family respondent. In previous waves, proxy reports were available only for

those who were financially connected within the original family, but not for those in split households. Starting from the 2012 survey, proxy reports were also administered when respondents were unable to complete the questionnaires due to physical or mental conditions (e.g., to ill to complete the questionnaires). Compared with the full scope individual self report questionnaire, proxy report only asked a small number of questions. In the family roster questionnaires, we confirmed with family respondents about missing or inconsistent information in basic sociodemographic variables from earlier waves.

In addition, CFPS2018 added a few new measures, mostly in the individual self report questionnaire. The “first job (GD)” module collected the first job of the respondent, including job description, educational requirements and how the respondent got the job. Charity behaviors and financial literacy assessments were administered. A brief big five personality scale was introduced first time in CFPS for all respondents aged 16 and above. For those between 10-15, CFPS collected ratings on internalizing and externalizing problem behaviors. Next we provided more detailed related to the above two measures. *Big Five Personality Scale:* In the CFPS 2018 individual self report questionnaire, the CFPS collected personality data for the first time for respondents aged 15 and above. We adopted a brief 15-item version of the big five personality scale, which was used from household surveys in other countries (e.g., PSID, GSOEP and BHPS). The scale had three items in each of the following five dimensions: conscientiousness, extraversion, agreeableness, openness, and neuroticism. Table 1 lists the 15 items and their corresponding dimensions. Note that four out of the five dimensions contain one item that is reverse coded, and those items are noted in Table 1. If users want to form summation scores for each scale, they should reverse code the items before summation. We also found that the presence of reverse coded items lowered the internal consistency of the scales. Users may consider creating composite variables by removing the reverse coded items.

Table 1. Big five personality scale in CFPS 2018 respondent questionnaire

| Variable Name | Items | Dimensions |
|---------------|--|-------------------|
| QM201 | I do a thorough job. | Conscientiousness |
| QM202 | I am talkative. | Extraversion |
| QM203 | I am sometimes rude to others. (reverse coded) | Agreeableness |
| QM204 | I am original and come up with new ideas. | Openness |
| QM205 | I worry a lot. | Neuroticism |
| QM206 | I have a forgiving nature. | Agreeableness |

| | | |
|-------|---|-------------------|
| QM207 | I tend to be lazy. (reverse coded) | Conscientiousness |
| QM208 | I am outgoing and sociable. | Extraversion |
| QM209 | I value artistic experiences. | Openness |
| QM210 | I get nervous easily. | Neuroticism |
| QM211 | I do things efficiently. | Conscientiousness |
| QM212 | I am reserved. (reversed coded) | Extraversion |
| QM213 | I am considerate and kind to almost everyone. | Agreeableness |
| QM214 | I have an active imagination. | Openness |
| QM215 | I am relaxed and handle stress well. (reversed coded) | Neuroticism |

Problem behaviors for adolescents. In the CFPS 2018 individual self report questionnaire, the CFPS collected data on problem behaviors among adolescents aged 10 to 15 for the first time. Both internalizing problem behaviors and externalizing problem behaviors were measured. We adopted those items from the Early Childhood Longitudinal Survey, including 8 items for internalizing problem behaviors (variable names prefixed with Qint) and 6 items for externalizing problem behaviors (variable names prefixed with Qext), as detailed in Table 2. We assessed the reliability and validity of the scale. We first used factor analysis to assess the unidimensionality of the 8-item internalizing problem behaviors and the 6-item externalizing problem behaviors respectively. Initial results indicated that the unidimensional model did not fit the empirical data well.¹ Subsequent analyses revealed that a pair of variables within each scale (QINT005 and QINT007 for internalizing problem behaviors and QEXT004 and QEXT006 for externalizing problem behaviors) showed common factors beyond those associated with the general factors. A closer look at those pair of variables revealed that QEXT004 and QEXT006 may appear redundant to the Chinese respondents (hard to pay attention vs get distracted easily). If we remove one of the variables from the scale (e.g., Qint007 for internalizing deviance and QEXT006 for externalizing deviance), data from the remaining variables presented reasonable fit to the unidimensional model. ² The Internal consistency coefficient Cronbach's alpha for the internalizing question dimension was 0.65, and the Cronbach's alpha coefficient for the externalizing question dimension was 0.64.

¹Goodness-of-fit statistics for an unidimensional model based on data from the 8 internalizing problem behavior questions: chi-square=338.44, $p < .001$, RMSEA=.081, CFI=.868, TLI=.816, SRMR=.044. For the externalization bias behavior items, the corresponding goodness-of-fit indices are as follows: chi-square=357.14, $p < .001$, RMSEA=.126, CFI=.834, TLI=.723, SRMR=.057

²The 7 questions after removing QINT007 were analyzed using a unidimensional model for the dimension of internalized deviant behavior, resulting in the following fit indices: the chi-square=104.653, $p < .001$, RMSEA=.051, CFI=.950, TLI=.926, SRMR=.027. For the externalization problem behavior, the fit indices based on the analysis of 6 questions after removing QEXT006 are as follows: chi-square=71.091, $p < .001$, RMSEA=.073, CFI=.955, TLI=.909, SRMR=.030.

Table 2. Items for the internalizing problem behaviors and externalizing problem behaviors for adolescents aged 10 to 15 in CFPS2018

| Variable Name | Items | Dimensions |
|---------------|--|---------------|
| QInt001 | I feel angry when I have trouble learning. | Internalizing |
| QExt002 | I often argue with other kids. | Externalizing |
| QInt003 | I worry about taking tests. | Internalizing |
| QExt004 | It is hard for me to pay attention. | Externalizing |
| QInt005 | I often feel lonely. | Internalizing |
| QExt006 | I get distracted easily | Externalizing |
| QInt007 | I feel sad a lot of the time. | Internalizing |
| QExt008 | It is hard for me to finish my school work. | Externalizing |
| QInt009 | I worry about doing well in school. | Internalizing |
| QInt010 | I worry about finishing my homework. | Internalizing |
| QInt011 | I worry about having someone to play with at school. | Internalizing |
| QExt012 | I get in trouble for talking and disturbing others. | Externalizing |
| QExt013 | I get in trouble for fighting with other kids. | Externalizing |
| QInt014 | I feel ashamed when I make mistakes at school. | Internalizing |

1.2 Modifications in data structure

Prior to CFPS2018, individual level data were presented in two datasets by respondents' age. All those above 15 were in the adult datasets, and those between 0 and 15 were in the child datasets. The child dataset used to contain both proxy reports from the guardians as well as self reports of the children if they were between 10 and 15. In CFPS2018, the datasets were restructured to reflect the harmonization of self reports for all individuals aged 10 and above. Self reports from those between 10 and 15 and those above 15 were in the same datasets in CFPS2018, facilitating analysis that involved respondents from different age groups. Consequently, a separate dataset that included only proxy reports from the main guardians was available for children below age 16.

1.3 Modifications in weighting variables

Two changes were made to the weighting variables in CFPS2018. Prior to CFPS2018, two sets of weights were available, one applied to the total sample, and the other one applied to the national resampling sample. As the two sets were supposed to generate comparable estimates but often caused confusion to the users, we only retained the set applicable to the total sample in CFPS2018.

The second change was a conversion from population weights to normalized weights. At baseline, we were able to generate population weights based on the benchmark population from 25 provinces covered by the baseline sampling framework. As years passed by, a proportion of those originally in the 25 provinces moved to other provinces. We lack a good official benchmark of its current population size. Therefore, we made this transition from population weights to normalized weights. The basic algorithm was similar to those introduced in the user's manual, but we derived the normalized weight by dividing the population version of weight variable by its mean. We recommend the use of proportions instead of population size estimates. If needed, users may estimate population sizes of subgroups based on those proportion estimates, with the caveats that the actual size of the population is different from the one we can use (e.g., the population of the whole country).

2. Descriptions of the public-released datasets

We harmonized data from individual self-report questionnaires and individual proxy questionnaires to form a single dataset for all respondents aged 10 and above. Data from the remaining questionnaires (family roster questionnaire, family questionnaire, and child proxy questionnaire) were each in a separate dataset. Table 3 displays the sample sizes and number of variables for each dataset. Codebooks for each dataset are posted in the "Data Description" section under "Documentation" on the project website.

Table 3. Basic information of CFPS2018 public-use datasets

| Questionnaires | Dataset name | n | Number of variables |
|--|--------------|--------|---------------------|
| Family Roster | Famconf | 58,504 | 296 |
| Family | Famecon | 14,218 | 321 |
| Individual self report+individual proxy report | Person | 37,354 | 1371 |
| Proxy report of child guardians | childproxy | 8,735 | 289 |

2.1 Family Roster Database (famconf)

The released family roster dataset in CFPS2018 includes 58,504 individual observations from 15,051 households. Each row in the family roster dataset represents an individual member of families that had completed the family roster questionnaire in CFPS2018. All individual members were uniquely identified

by the individual identifier “pid” and the family id in 2018 was “fid18”. The dataset includes the basic sociodemographic information (birth year, gender, education) of each individual as well as their father (the _f series of variables), mother (the _m series of variables), spouse (the _s series of variables), and up to 10 children (the _c1-_c10 series of variables).

The main purpose of the family roster questionnaire in the follow-up survey is to identify whether there are any changes in the family composition. Such changes could be related to family sizes, e.g., addition of new family members through marriage or birth, or the death in the family members. Such change could also be related to family splitting processes. For example, adult children become financially independent of the original family and would be assigned a new family id in CFPS.

During the family roster questionnaire interview, a family respondent would report whether any member is not staying in the family residence. If not, the family respondent would be further asked about whether the non-coresiding family member was maintaining financial relationship with the original family. If yes, the non-coresiding member would still be considered a member of the original family. If not, the non-coresiding member would form a new family and assigned a new family id at the time of the interview. However, the response by the family respondent of the original family regarding financial relationship did not determine the final status of the non-coresident members. All non-coresident members would receive a family roster questionnaires³, and they would be asked whether they were financially connected with the original family. The final status of the non-coresident members would be determined by their own responses. For example, if family respondent in the original family considered the non-coresident financially independent of the original family, during the time of the interview, the non-coresident unit would not be considered a member of the original family and they would be assigned a new family id. However, if the non-coresident member claimed themselves to be still financially connected with the original family, their family id would be changed back to the original family id.

All eligible members of a family was indicated by a variable co_a18_p. When the value is equal to 1, the individual is financially connected with the corresponding fid18 and thus considered an eligible family member. If the value is equal to 0, the individual is not financially connected with the corresponding fid18 and thus not considered an eligible family member. Familysize18 was computed by counting all members with co_a18_p=1 within the same fid18. Instead of financial connection, another

³ All those residing at the same address would receive one family roster questionnaire interview.

variable `tb6_a18_p` indicates whether the individual was a coresiding member of the family. For members who were financial connected with the family but lived in a different residence, their `co_a18_p` would be 1, but `tb6_a18_p` was equal to 0.

In addition to `fid18`, the family roster dataset also provides the family ids of all individuals in previous waves, namely `fid16`, `fid14`, `fid12`, and `fid10`. For the same observation, if the value of the family ids changes from one wave to another, that means that household has experienced some splitting processes and some members from previous waves are now financially independent from other members.

Similar to previous waves, the existing series of variables (`TB601_A18_*`: Reason for leaving home) were updated in 2018. `TB601` was based on question A3 of the family roster questionnaire, in which the open-ended text information collected for the category "77" (77.Other reasons [please record the respondent's words]) was coded. The following categories were added to the original categories: leaving the country, going to work, going to school, getting divorced, getting married, visiting relatives, separating from family, moving, and going for medical treatment.

For the ease of use, we added a few more composite variables in the CFPS2018 family roster dataset.

[Appendix 1](#) lists the variable names and relevant information.

2.2 Family Dataset (famecon)

Unlike the family roster dataset with observations at the individual level, the family dataset contains 14,218 observations at the family level. It focuses on family income, expenditure, and assets. Each family is identified by a unique `fid18`. Since families may experience splitting, family ids may change across waves. Their family ids from earlier waves are identified by `fid10` to `fid16`. Even if the family ids remain the same across waves, we cannot be certain that the family structure remain exactly the same. An accurate assessment of whether a household is identical to its status in the previous survey requires comparing family members across different waves.

All observations in the family dataset have corresponding family observations from the family roster dataset, however, not every family that appeared in the family roster dataset is present in the family dataset. A small number of families did not complete the family questionnaire after completing the family roster questionnaire.

In the CFPS2018 family dataset, there are two variables related to family size: `familysize18` and `fml_count`. `fml_count` in the dataset reflects family sizes determined by the family id assigned at the time of the interview, while `familysize18` reflects family sizes determined by the corrected family id after data processing. Values of the two variables are mostly the same, but in occasions when responses regarding financial dependence are inconsistent between family respondent of the original family and the non-coresident unit, the two values may differ.

Many questions in the family datasets are related to family finances, e.g., incomes, expenditures, and assets. Two issues are prominent with those variables. First, compared with other items, item-level non-responses for finance questions are relatively high. For some questions, the CFPS arranged unfolding brackets questions following a non-response. For example, in case of missing data in the total expense question, the family respondent would be asked the following question:

*FEXPUB In the past 12 months, was your family's total expenditure higher/lower than (10,000/25,000/**50,000**/100,000/250,000yuan)?*

The series of questions would start with the bolded value in the middle, if the respondent answered yes (which means that the total expenditure is higher than 50,000), the series would proceed to the next higher value (100,000) combined with “higher”. If the respondent answers no, the series would proceed to the next lower value (25,000) combined with “lower”. A brackets with a higher and lower bound would be formed with such a process, or the boundary values would be reached (i.e., higher than 250,000 or lower than 10,000). The released datasets contains information on the bounds.

The second issue with the finance questions is the confusion between different units, especially those between Yuan and tens of thousands of Yuan. Such confusion would lead to either reporting error (when respondents mistook one unit for the other), or recording error (when interviewers misheard the units of the reported values). In order to minimize such error, we took a series of steps on variables that involved money values. We first identified outliers both based on their distributions within the same wave, and comparison with previous waves. Then, we tried to verify their credibility by using information from other auxiliary variables (e.g., location information to verify house values, income and expenditure to verify assets etc.) Additionally, we drew upon para-data from audio recording of the interview process to minimize possible recording error. Variables that went through such processes included the following: 1)

Asset related: FM401 (total assets of family business); FQ5 (the purchasing value of the current residence); FQ6 (the current market value of the current residence); FR2 (the market value of other real-estate properties); FS6V (total value of consumer durables), FS7V (the total value of agricultural machinery), FT1 (cash and deposits), FT101 (savings), FT201 (the total value of financial products); FT501 (amount of loans to be paid to the banks), FT601 (amount of loans to be paid to friends and relatives), FT901 (amount individuals or institutions owe your family). 2) Income-related variables FM4 (net profit from family business), FN301 (pension income), FR501 (rental income).

2.3 Individual dataset (person)

The individual dataset includes data from 37,354 respondents aged 10 and above. The individual sample is uniquely identified by the variable “pid”. Pid remains constant across datasets within or across years. In addition to the individual identifier, the respondent dataset also contains family id to identify the family that the individual belongs to based on the CFPS definition.

The 2018 individual dataset includes both data from self reports (identified by selfrpt = 1) and proxy reports (identified by proxyrpt = 1). Questionnaires for proxy reports are abbreviated versions of the self reports. Interview modes are either face to face (identified by self_iwmode=1 or proxy_iwmode=1), or telephone (identified by self_iwmode=2 or proxy_iwmode=2). Face-to-face questionnaires are very similar to those for telephone interviewers, except that cognitive tests were only administered to those participating in face-to-face interviews. When we combined data from different forms of questionnaires, we assigned them the same variable names if the questions and response options were the same, and we kept their variable names different if either the questions or the response options were different. In the case of duplicate records of the same individuals from the self report and proxy report, we would retain the self report data and discard the proxy report.

We note the following variables that the user may need to pay attention to.

Individual wage income. The raw version of wage income is generated by the interview system during the time of the interview. “incomeB” includes wage income from the main job in the past 12 months, while “incomeA” includes wage income from all other jobs in the past 12 months. Combined together, they formed the total wage income in the past 12 months, indicated by the variable “income”. It is important to note that all the above variables were applicable to respondents who had a hired job and paid by others.

If the respondent was unemployed or engaged in family business or family farm work, they would not have a valid value in the above variables.

Cognition function. The CFPS2018 cognitive test follows the design of the CFPS2014 questionnaire and consists of two parts: a literacy test and a math test. We calculated respondents' total scores in literacy and mathematics (giving one point for each correct answer), represented by the variables `wordtest18_sc2` and `mathtest18_sc2`. Meanwhile, we generated an additional set of possible scores for respondents assuming a fixed starting point to ensure comparability with the CFPS2010 cognitive scores, represented by the variables `wordtest18` and `mathtest18`. The correlation between the scores of the two algorithms was very high, with a correlation coefficient close to 1 for the two sets of scores for the literacy test and over .97 for the math test. We recommend using the `_sc2` family of variables if users are using 2014 or/and 2018 cognitive test data, but if data from 2010 data included in the analysis, we recommend using the 2010 comparable variables, i.e., `wordtest18` and `mathtest18`. The cognitive tests were collected only in the individual self-report interview mode; no data were available for individuals interviewed by telephone.

More information about the designs and scoring algorithms can be found in the user's manual.

Depression. In CFPS2018, the Center for Epidemiologic Studies Depression Scale (CES-D) was used to measure individuals' depression levels. In CFPS2012, we used the CESD20, a 20-question scale, which was not well accepted by respondents based on field interviewer feedback. So in CFPS2016, we selected a random sample to use the abbreviated 8-item version (similar to that used in HRS), and rest still used CES-D20 in preparation for the gradual transition to the streamlined version.

In the CFPS2018 survey, we provided two composite variables for CESD. One is `CESD8`, which was computed by summing up scores from the 8 items, taking into account reverse coding. In addition, we provided `CESD20sc`, using equipercntile equating based on the CFPS2016 random design. `CESD20sc` was comparable to the scores from `CESD20`. With data from CFPS2016, users can also generate their equivalent scores as they deem more appropriate. We recommend that users use the 8-item version of the total score if they only use data from CFPS2016 or/and CFPS2018 data, but use the `CESD20sc` instead if they need to use the CFPS2012 data for analysis.

2.4 Child Proxy dataset (childproxy)

The child proxy dataset contains proxy reports from guardians of children aged 0 to 15. Each line represents a child, uniquely identified by pid. Proxy respondent is identified by respc1pid. The proxy respondent is the guardian of the child, often the child's mother or father, and sometimes grandparents or other family members. Compared with child datasets from previous waves, child proxy dataset in CFPS2018 contains only proxy data from guardians, whereas the child dataset in previous waves additionally contains self report data for children between 10 and 15. The self report data for children between 10 and 15 in CFPS2018 is part of the individual dataset.

2.5 Cross-wave individual core variable dataset (crossyearid)

For all individual samples that ever entered the CFPS through family roster questionnaire, the cross year ID dataset provides their basic information across all waves from baseline. A total of 74,130 individuals were in the cross year id dataset, including 64,208 gene members, 6,919 members who were core members in at least one wave, and 3,183 non-core members in at least one wave.

Variables in the cross-wave individual dataset are mainly in three categories. The first category is the time-constant demographic variables, including pid, birth year, gender, ethnicity, sampling information at baseline. They are considered constant in the study design across waves. The second category is the time-varying sociodemographic variables, including marital status, education, residence permit status (hukou). They may change across waves, and thus have a different variable for each wave. The third category refers to the interview status, including the entry year of the individual sample, fid at each wave, whether the individual was financially connected with the corresponding fid of the particular wave, whether the individual was living in the same address of the family residence, whether the individual completed an individual survey, and whether the individual survey was a self report.

Appendix 1. Variables in the CFPS2018 family roster datasets and their corresponding items in the questionnaire

| Variable name | Variable label | Questionnaire Item | Notes |
|--|--|--------------------|--|
| Existing variables from previous waves | | | |
| FID_PROVCD18 | Province ID 2018 | | |
| FID_COUNTYID18 | County ID 2018 | | |
| FID_CID18 | Community ID 2018 | | |
| FID_URBAN18 | Urban area (Census Bureau's definition) | | |
| SUBSAMPLE | Is it in the national resampling sample? | | Defined by subsample status linked with fid_base (defined later) |
| SUBPOPULATION | Sampling subpopulation | | Defined by subpopulation status linked with fid_base (defined later) |
| GENETYPE18 | Type of gene member in 2018 | | Recode based on RTYPE_END18 and gene |
| fid1* | Family ID 2018/6/4/2/0 | | |
| FAMILYSIZE18 | Number of Family members (defined by T1) | | Counting number of individual within co_a18_p=1 with the same fid18 |
| TB2_A_P | Gender | BC2、E1、D105 | Synthesizing information from family roster questionnaire, individual self-report, and information from previous waves |
| TB1Y_A_P | Date of birth (year) | BC3、E2、D104 | Synthesizing information from family roster questionnaire, individual self-report, and information from previous waves |
| TB1M_A_P | Date of birth (month) | BC3、E2、D104 | Synthesizing information from family roster questionnaire, individual self-report, and information from previous waves |

| | | | |
|----------------------|---|-----------------------------|--|
| TB3_A18_P | Marital status | BC4、E3 | Synthesizing information from family roster questionnaire, individual self-report, and information from previous waves |
| TB4_A18_P | Highest degree of education | BC5、E4 | Synthesizing information from family roster questionnaire, individual self-report, and information from previous waves |
| HUKOU_A18_P | Hukou status | BC6、E5、D106 | Synthesizing information from family roster questionnaire, individual self-report, and information from previous waves |
| TB6_A18_P | Currently live in this family | A2、A201 | |
| CO_A18_P | Whether respondent was financially connected with fid18 | F102、B1 | Using reports by non-coresident members first, supplemented by reports from family respondents in the original family |
| OUTPERS_R_WHER E18_P | Residence of left-home person in 2018 | G1、H1 | Synthesizing information on single-person households and multiple-person households |
| TB602ACODE_A18_P | Province code (Residence of left-home person) | G101、H101 | Synthesizing information on single-person households and multiple-person households Values 1 to 6 reflect the original options from the questionnaire, values 10 to 18 were coded based on the text information for option 77 |
| TB601_A18_P | Reason for 'pid's leaving home by 2018 | A103 | |
| OUTUNIT18 | Serial number of resident unit leaving original household | F1 | Divide non-coresident members into different units, depending on the number of addresses reported |
| COREMEMBER18 | Whether core family member in 2018 | Type of member in CFPS 2018 | |
| CFPS2018_INTERV_P | Status of individual survey | | |

| | | | |
|------------------------------------|---|-----------------|---|
| ALIVE_A18_P | Still alive | A3 | Synthesizing newly reported death and information from previous waves |
| TA4Y_A18_P | Year of death | A4 | Synthesizing newly reported death and information from previous waves |
| TA4M_A18_P | Month of death | A4 | Synthesizing newly reported death and information from previous waves |
| TA401_A18_P | Cause of death | A401 | Synthesizing newly reported death and information from previous waves |
| pid_a_* | Father: Personal ID Mother: Personal ID Spouse: Personal ID Child: Personal ID | C2、C3、C4、 C5 | |
| Newly added variables for CFPS2018 | | | |
| TB602CCODE_A18_P | County id (Residence of left-home person) | G102、H102 | |
| TB602CCODE_A18_* | Father: county id Mother: county id Spouse: county id Child: county id | G102、H102 | |
| C105_A18_P | Reason for 'pid's joining into the family by 2018 | C105 | |
| RTYPE_END18 | Meaning of rtype in the questionnaire | | Type of member in CFPS 2018 |
| PSU | Primary sampling unit | | Defined by PSU linked with fid_base |
| ADS1_18 | Migrated or not | ADS1 | |
| KZ103_18 | Main language used during the interview | Z103 | |
| FID_BASE | Baseline Family ID | | Related family id at the baseline year of 2010 |
| INTERVIEWERID18 | Interviewer id | | |