
CHINA HEALTH AND RETIREMENT
LONGITUDINAL STUDY

FOLLOWUP 2015

RELEASE NOTE

VERSIONID: 20170531

MAY 2017



CHINA CENTER FOR ECONOMIC RESEARCH

INSTITUTE OF SOCIAL SCIENCE SURVEY

PEKING UNIVERSITY

1 Released Modules

In this released version (versionID: **20170531**) we released *eight* main modules, and the associated *two* datasets (sample information and cross-sectional weights). The PSU information is the same as the 2013 wave. Table 1 provides detailed information about these *ten* datasets.

Table 1: Released Datasets

Module in Questionnaire	Dataset	Information
B. Demographic Background	Demographic_Background.dta	Demographic information for main respondent and spouse
D. Health Status and Functioning	Health_Status_and_Functioning.dta	Health behaviour and Status
E. Health Care and Insurance	Health_Care_and_Insurance.dta	Health care utilization, health care costs and medical insurance
F. Work Retirement and Pension	Work_Retirement_and_Pension.dta	Work history and current status, pension
G&H. Income, Expenditures and Assets (G2, HA)	Household_Income.dta	Household income, expenditure and assets
G&H. Income, Expenditures and Assets (G1, HB)	Individual_Income.dta	Individual income and assets
I. Housing Characteristics	Housing_Characteristics.dta	Construction materials and home facilities
Biomarker	Biomarker.dta	Anthropometric measurements
	Weights.dta	Cross-sectional weights
	Sample_Info.dta	Responded Samples, whether cross-sectional, whether died, and interview date

All the data sets are stored in Stata 13 format, you can also find the summary information of variables from the released codebook.

2 Weights

We have only released cross-sectional weights from this wave, since panel weights depend on analysis purpose, you may use all the three waves panel (balance), or particular two waves. Users can construct panel weights according to any appropriate sample attrition adjustment method.

There are two sets of cross-section weights, one without non-response adjustment, and another with, the construction method is similar to the baseline weights but taking account of the death and divorce.

Since we not only tracking the previous (2011, 2013 and 2014 wave) respondent sample in this wave, but also the non-response sample and refresh sample, so the cross-section sample is not necessarily the longitudinal sample, as we also find several errors in the

previous waves interview, so the longitudinal sample is not necessarily the cross-section sample. Please be careful in identifying samples.

We release the cross-sectional biomarker weight, which is calculated based on the individual cross-sectional weights and a logit regression of whether the individual responses in biomarker, the biomarker level inverse probability weight factor is calculated by the inverse of predicted probability for each individual (cap at the 99th percentile).

3 IDs

The IDs (ID, householdID and communityID) can be matched with their counterparts in the previous waves. You need to adjust the householdID and ID in the baseline, as noted in the release note of 2013 wave.

You may also want to match household members, parents and children between waves, this is simple since we do not change their IDs (or rather, the order, which is *i* in the variable like ca001_i_).

4 Variable Naming Rule

_w3 in the variable name means that this is a new variable. Variable name with prefix "z" means this is a preload variable, for example, "zba001" contains the value of ba001 in the 2013 wave. Variable name with prefix "x" mean this is a generated variable by our CAPI, for example, "xchildage" is generated from the birth year of that child.