CHINA HEALTH AND RETIREMENT LONGITUDINAL STUDY

FOLLOWUP 2013

RELEASE NOTE

VERSIONID: 20150131

JAN 2015

CHINA CENTER FOR ECONOMIC RESEARCH

PEKING UNIVERSITY

1 Released Modules

In this released version (versionID: **20150131**) we released ten main modules and the associated two datasets (weights and community information). Table 1 provides detailed information about these twelve datasets.

Table 1: Released Datasets

Module in Questionnaire	Dataset	Information
B. Demographic Background	Demographic_Background.dta	Demographic information for main respondent and spouse
C. Family (CM, C1, A)	Family_Information.dta	Information for household members and family members
C. Family (C2)	Family_Transfer.dta	Transfer among family members
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
J. Interviewer Observation	Interviewer_Observation.dta	Attitude and comprehensive ability of respondent
	Weights.dta	Cross-section and longitudinal weights
	PSU.dta	Province, city name and area type

You can also find the summary information of variables from the released codebook.

2 Weights

We have released two sets of weights, one for the cross-section sample, and another for the longitudinal sample. The variable **crosssection** identifies whether the sample is a cross section sample, and the variable **longitudinal** identifies the longitudinal sample. 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. The longitudinal weights are based on the baseline weights (or rather, the one with non-response adjustment), an inverse probability weighting factor is constructed from a logit regression of whether a respondent participates in the second wave conditional on the participation in the baseline, note that in constructing the longitudinal weights, we treat dead as response.

Since we not only tracking the baseline 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 baseline interview, so the longitudinal sample is not necessarily the cross-section sample. Please be careful in identifying samples.

3 IDs

The IDs (individualID, householdID and communityID) can be matched with there counterparts in the baseline sample. The community IDs are the same, but the household IDs are different, the householdID in the baseline is an 9 digit number, but in this wave, it changed, now it is an 10 digit number, the 10th digit is an indicator for splitting household due to divorce, you can safely adjust the householdID in baseline sample to 10 digit by adding "0" in the 10th digit, then the modified householdID is comparable to that in the second wave. As the householdID changed in the second wave, the inidividualID also changed, now it is an 12 digit number, where the last two digit is the identifier for individuals within households which is comparable to those in baseline individualID, so you can simply use the adjusted householdID with the last two digit of individualID as made-up numbers to generate the new individualID which is comparable to that in the second wave.

In Stata, you can use the command "replace householdID = householdID + "0"; replace individualID = householdID + substr(individualID,-2,2)" to adjust the IDs in the baseline sample.

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

_w2 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 baseline sample. 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.