

UK Data Archive SN 6614 – Understanding Society Waves 1-6: Revision 2017

Date: 06th November 2017

This document describes revisions to the Understanding Society Waves 1-6 datasets last fully revised in October 2016 with further between-edition adjustments made in February and March 2017. The User Guide and on-line documentation can be consulted for a study-wide perspective and variable level information, respectively.

Changes to data files. The cross-wave file XWAVEDAT has been fully revised to reflect the inclusion of the harmonised BHPS data in Understanding Society. This file now includes all sample members ever enumerated in either of the studies. New derived variables included in XWAVEDAT are: ch1bm_dv, ch1by_dv, anychild_dv, lmar1m_dv, lmar1y_dv, evermar_dv, coh1m_dv, coh1y_dv, evercoh_dv, dcsedfl_dv, dcsedw_dv, dcsedw_dv_bh, lwenum_dv, lwenum_dv_bh, fwenum_dv, fwenum_dv_bh, lwintvd_dv, lwintvd_dv_bh, fwintvd_dv, fwintvd_dv_bh, school_dv, scend_dv, feend_dv, xwdat_dv, w6osmflag, bornuk_dv. Some additional initial questions have been included: paid, maid, j1semp, j1none, j1boss, j1mngr, lvag16, agelh, lvag14 and ynlp14.

The XWAVEID file has not been harmonised, so XWAVEID pertains to the UKHLS survey and XWAVEID_BH pertains to the BHPS survey, but a few harmonised variables have been included in both. New derived variables included in XWAVEID are: dcsedfl_dv, dcsedw_dv, lwenum_dv, fwenum_dv, lwintvd_dv, fwintvd_dv and w6osmflag. The quarter variable is considered to be stable across waves and so now one quarter variable without the wave prefix is included. Additional variables included are: sex, birthy, birthm, strata, psu, sampst and psmwave.

New derived variables. The variables w_age_dv (age at the time of interview) and w_sex_dv (cross-wave consistent sex) have been computed and added to the INDALL, INDRESP, CHILD and YOUTH data files. This uses the most up-to-date information available in the sample administration data base. The variables w_age_cr and w_sex_cr have been withdrawn. Interview dates used to compute w_age_dv have been added (w_inddaty_dv, w_intdatm_dv, w_intdatd_dv, w_doby_if, w_dobm_if and w_intdatd_if) as have derived date of birth variables: w_doby_dv, w_dobm_dv (SL only), w_dobd_dv (Secure only) and w_doby_if w_dobd_if.

For consistency across files, the PIDP for co-resident grandparents (w_grfpid and w_grmpid) have been added to complement the already existing PNOs, in all waves. The pointers to natural parents (mpid and fpid) have been renamed to w_mnpid and w_fnpid respectively, in order to harmonise with the BHPS.

The variable w_yprelig_dv is now provided in the Wave 5 and Wave 7 YOUTH data files.

More complete socio-economic classifications in INDRESP. An error in the jbsoc00 values fed forward from the BHPS has been corrected leading to a larger number of valid jbsoc00 codes. Data quality was further improved by a mop-up coding exercise of jbsoc00 values for respondents who had missing jbsoc00 in the previous release data. The variable w_jbsocgr_dv, a recode of current job occupation to 1 or 2 digits, has been added.

Revisions to within-household relationships. Inconsistencies within household relationships have been investigated and corrected, see variable `w_relationship_dv` on data file EGOALT. The checking code now relies on the derived variable `w_age_dv`. Changes affect derived variables based on `w_relationship_dv`. The variable `w_rel_dv` has been added to EGOALT. It resembles the code frame of the BHPS relationship variable but users should be aware of the different data collection methods. The next wave status identifiers `w_anwstat` and `w_enwstat` have been added to F_EGOALT.

IEMB Migration Module. In wave 6 (IEMB only), all country variables associated with the module, e.g. `f_mlivedc1` and `f_mabroadc_all`, where the country value is over 103, have been changed to the correct country code.

Income and investment. Variable names and labels have been renamed in order to harmonise them with UKHLS naming conventions as follows:

Old Variable Name	New Variable Name:	File
<code>w_jb2pay_if</code>	<code>w_j2pay_if</code>	INDRESP
<code>w_inc1clabj2</code>	<code>w_j2paynet_dv</code>	INDRESP
<code>w_inc2misc</code>	<code>w_fimnmisc_dv</code>	INDRESP
<code>w_inc3prben</code>	<code>w_fimnprben_dv</code>	INDRESP
<code>w_inc5inv</code>	<code>w_fimninvnet_dv</code>	INDRESP
<code>w_inc6pen</code>	<code>w_fimnpen_dv</code>	INDRESP
<code>w_inc7sben</code>	<code>w_fimnben_dv</code>	INDRESP
<code>w_netinc1</code>	<code>w_fimnnet_dv</code>	INDRESP
<code>w_jb2pay_dv</code>	<code>w_j2pay_dv</code>	INDRESP
<code>w_frjt_mod</code>	<code>w_frjtmod_dv</code>	INDRESP
<code>w_frjtpn_mod</code>	<code>w_frjtpnmod_dv</code>	INDRESP
<code>w_jtkeep</code>	<code>w_frjtkeep_dv</code>	INDRESP
<code>w_hhnetinc1</code>	<code>w_fihhmnnet1_dv</code>	INDRESP
<code>w_dep9ctax</code>	<code>w_ficountax_dv</code>	INDRESP
<code>w_hhnetinc3</code>	<code>w_fihhmnnet3_dv</code>	INDRESP

The following income and investment variables have been added to HHRESP: `w_fihhmnlabgrs_dv`, `w_fihhmnlabnet_dv`, `w_fihhmnmisc_dv`, `w_fihhmnprben_dv`, `w_fihhmninv_dv`, `w_fihhmnpen_dv` and `w_fihhmnsben_dv`.

Values for income variables are likely to have changed since the last release for waves 1 to 6 for cases where missing data made imputation necessary because the imputation uses data from multiple waves, including the latest wave.

Weights. The following previously-released weights have been revised to correct some errors: `b_ythscbh_xw`, `f_psnenui_li`, `f_psnenui_xw`, `f_hhdenub_xw`, `f_indpxui_xw` and `f_indinui_xw`. The weight `f_hhdenub_xw` has been added to `f_hhresp`, for cross-wave consistency. The two household weights `f_hhdenub_xw` and `f_hhdenub_xw` have been added to F_HHSAMP for the first time (in addition to remaining on F_HHRESP).

Other Miscellaneous Changes.

Some missing values have been recoded. In particular, derived variables that rely on `w_dvage` now use the full range of interview outcomes codes applicable to children under the age of 16, as appropriate, if a child's age is missing.

An inaccurate reference to long-term unemployed and never worked has been removed from the NSSEC5 value label for parental and other occupations. Value labels for the Wave 4 investment variables have been added.

Errors in the Wave 6 twin module have been corrected.

A review of variable and value labels across all files has resulted in a number of improvements in this area.

Five instances of `f_mabroadc_all` being out of the valid range have been corrected.

The variable `a_pmarstat` has been removed as it does not add any value to the study.

Various minor data inconsistencies have been corrected.