

UK Data Archive SN 6614 – Understanding Society Waves 1-7: Revision November 2018

Date: 1st November 2018

This document describes revisions to the Understanding Society Waves 1-7 datasets, including Harmonised BHPS datasets, last fully revised in November 2017 with a significant between-editions release in August 2018. The User Guides and on-line documentation can be consulted for a study-wide perspective and variable level information, respectively.

Improvements to Key Variables

There have been ongoing efforts to improve the quality of the data in `age_dv`, `sex_dv`, `relationship_dv` and `jbsoc00` variables by utilising the best available data from different sources. These improvements impact a wide range of derived variables as well as all income variables.

Income Changes

The component `w_fimnsben_dv`, amount income component 7: social benefit income, in `w_indresp`, also now includes the sources “universal credit” and “personal independence payments”, both from Wave 4 onwards.

The `w_indresp.dta` variables `w_inc1alabem`, amount income component 1a: net earnings main job; and `w_inc1blabse`, amount income component 1b: net self employ income, have been dropped from all waves as they are working variables not intended for researchers.

Weighting Changes

The longitudinal response weights for proxy and main interview (`g_indpxui_lw`), main interview (`g_indinui_lw`), and self-completion interview (`g_indscui_lw`), all in `g_indresp.dta`, have been recalculated to now include rising 16-year olds.

The weight, `b_ythscub_xw`, BHPS+UKHLS cross-sectional youth interview weight, has been added to `b_youth.dta`.

Derived Variables, Pointers and Imputation Changes

The variable `w_ethn_dv`, ethnic group (derived from multiple sources), has been included in `w_indall.dta` and crosswave files (see Crosswave Files Changes below).

The pointers to household reference persons (`w_hrpno` / `w_hrpno`), grandfathers (`w_grfpno` / `w_grfpno`) and grandmothers (`w_grmpno` / `w_grmpno`) have been added to all `w_child.dta` and `w_youth.dta` files.

Gender-neutral pointers to biological parents 1 (`w_pn1pno` / `w_pn1pid`) and 2 (`w_pn2pno` / `w_pn2pid`) and biological / step / adoptive parent1 (`w_pns1pno` / `w_pns1pid`) and 2 (`w_pns2pno` / `w_pns2pid`) have been added to all individual-level enumeration and response files. The variables `w_pn*sex` and `w_pns*sex` report the respective parent’s gender.

A number of counts of specific relationships have been added to all individual-level enumeration and response files: number of respondent’s grandparents (`w_ngrp_dv`), biological parents (`w_npn_dv`), biological / step / adoptive parents (`w_npns_dv`), biological siblings (`w_nnsib_dv`), biological / half / step / adoptive siblings (`w_nnssib_dv`) has in the current-wave household.

The variable `w_ppsex`, partner’s sex, is now included in all individual-level enumeration and response files.

The variable `w_nbrsnci_dv`, Buckner's Neighbourhood Cohesion Instrument, short ($\alpha = .88$), has been added to `w_indresp.dta` for waves 1, 3 and 6.

The variables `w_nbrcoh_dv`, Neighbourhood Social Cohesion, ($\alpha = .78$), and `w_nbrcohdk_dv`, neighbourhood social cohesion: Total number of don't knows, have been added to `w_indresp.dta` for waves 3 and 6.

The variables:

- `w_agegr5_dv`, age group (DVAGE): 5 year intervals
- `w_agegr10_dv`, age group (DVAGE): 10 year intervals
- `w_agegr13_dv`, age group (DVAGE): 13 categories
- `w_buno_dv`, benefit unit number
- `w_adresp15_dv`, PNO of person responsible for child under 16
- `w_rach16_dv`, whether responsible for child under 16
- `w_hhresp_dv`, household response status
- `w_sex_dv`, sex, derived

have been added to `w_youth` files.

The variables:

- `w_agegr5_dv`, age group (DVAGE): 5 year intervals
- `w_agegr10_dv`, age group (DVAGE): 10 year intervals
- `w_agegr13_dv`, age group (DVAGE): 13 categories
- `w_buno_dv`, benefit unit number
- `w_depchl_dv`, whether dependent child: DWP definition
- `w_sex_dv`, sex, derived

have been added to `w_child` files for waves 1, 3, 4 and 5.

The variables:

- `w_intdatd_if`, interview date: Day, imputation flag
- `w_intdatm_if`, interview date: Month, imputation flag
- `w_intdaty_if`, interview date: Year, imputation flag
- `w_doby_if`, DOB Year imputation flag
- `w_age_if`, imputation flag for `age_dv`

have been added to `w_child` files for waves 1, 4 and 5.

The variable `w_racel_dv`, ethnic group incorp. all waves, codings, modes and bhps, has been removed from `w_indall.dta` as its correct location is `w_indresp.dta`.

The derived interview date variables in `w_indresp.dta`: `w_intdatd_dv`, Interview date: Day, derived; `w_intdatm_dv`, Interview date: Month, derived; and `w_intdaty_dv`, Interview date: Year, derived, have been corrected, where appropriate, for proxy respondents.

Further BHPS Harmonisation Changes

The variable `bw_highqual_dv` has been added to all `bw_indresp.dta` files. It is fully harmonisable with `w_highqual_dv` in the `w_indresp.dta` files.

Secure Access Dataset (SN 6676) Changes

In previous releases work has been undertaken to improve the quality of the dates of birth across Understanding Society. The full date of birth file `ukhls_fulldob_m1_8_control.dta` has been updated to utilise the derived date of birth variables: `dobd_dv`, `dobm_dv` and `doby_dv`.

Geography Datasets Changes

The Travel To Work Areas (TTWA's) (SN 6675) have changed due to ONS implementing a large scale revision of them in 2015. The old UKHLS variable `w_ttwa` has been replaced with `w_ttwa07` (the ONS 2007 scheme) and `w_ttwa15` (the ONS 2015 scheme). The equivalent variable in the harmonised BHPS waves has also been renamed to match. Please refer to the document `understanding_society_mainstage_panel_geographical_lookup_tables` to identify which ONS Postcode Directory (ONSPD) has been used to extract these variables and then to the relevant ONSPD documentation for full details of their definition.

The Output Area Classifications (OAC's) have been augmented with additional variables matching the individual elements of the structure. For OAC 2001 (SN 6674), for continuity reasons, the pre-existing string variable `w_oac01` has been retained unchanged from the last release. The additional variables, all of which are numeric, are: `w_oac01spgrp` (super-group), `w_oac01grp` (group) and `w_oac01sbgrp` (sub-group). The equivalent variables have been created for OAC 2011 (SN 7639). The equivalent harmonised BHPS variables in `bw_oac01` have also been created.

In addition, postcodes in Waves 1 to 7 have been corrected following a review process resulting in more matches.

School Codes Dataset (SN 7182) Changes

Due to a scripting error in the Wave 7 parents and children module, questions `STILLPRIV` to `SCHTOWNCPT` were not asked from the start of fieldwork until the script was corrected in November 2015 midway through Q4. These questions were asked at wave 8 just for people that missed them at wave 7, and were made available with the release of wave 8. The adult school code files, `w_schcodest.dta`, which are normally only present for alternate waves, is therefore present for Wave 8.

Crosswave Files Changes

`ethn_dv`, ethnic group (derived from multiple sources), is a new ethnic group variable derived using: the ethnic group reported in the main adult questionnaire; the youth questionnaire; Wave 1 and Wave 6 enumeration grids and by matching parents information for children as this was not collected for most 0-9 year olds. A new flag variable, `ethn_dv_source`, source for `ethn_dv`, identifies the information source to compute `ethn_dv` used for each person. This is an improvement on `racel_dv`, ethnic group incorp. all waves, codings, modes and bhps, which only combined ethnic group information collected in adult interviews across all waves and was thus missing for children (<16 year olds). These are available in `xwavedat.dta` and `indresp.dta` files in all eight UKHLS waves.

Computation for the variable `bornuk_dv` in `xwavedat.dta` has been improved by including information collected in the Wave 6 IEMB enumeration grid about whether the household member was born in the UK or not.

The variables genetics, genome wide snp data available, and epigenetics, dna methylation data available, have been added to xwavedat.dta and xwaveid.dta to allow researchers to identify whether genetics and epigenetics data is available for survey members.

Miscellaneous Added Variables

The variable g_ringfence , CAPI ringfenced household, has been newly included in g_indall.dta, g_hhresp.dta, g_indresp.dta, g_youth.dta and g_child.dta.

The variable w_ff_nonprod, non productive at preceding wave, has been added to w_hhsamp.dta so that it is present from Wave 2 onwards.

The variables w_hhorig, sample origin (household), w_psu, primary sampling unit, w_strata, strata, w_sampst, sample status, w_memorig, sample origin (individual), w_ivfio, individual interview outcome, w_sex, sex, and w_ivfho, final household interview outcome, have been added to a variety of files to aid access to commonly used variables.

Miscellaneous Dropped Variables

The variable f_issued2w in f_hhsamp.dta has been dropped as it superfluous. It is correctly retained in w_indsamp.dta.

The variable b_plboth_code in b_indresp.dta, previously only available in the Special Licence version, has been dropped as it does not appear in any other wave as the variable that should be used is w_plbornc_all.

The variables a_intdatem and a_intdatey have been dropped from a_indresp.dta as a_intdatd_dv, a_intdatm_dv and a_intdaty_dv should be used for interview dates. Wave a has been made consistent in this area.

The variables f_lcse* and IEMB were previously incorrectly released in f_indresp.dta and have consequently been dropped.

Miscellaneous Issues Noted

There was an error in the scanning of the responses to c_yppkmp, how much money did you receive last week to spend on yourself?, and c_yppay, what was your total pay last week?, in the Wave 3 Youth questionnaire. It is planned to include the correct responses in the November 2019 release.

Miscellaneous Changes

Various minor data inconsistencies have been corrected.

Various labelling issues have been addressed.