

UK Data Archive SN 6614 – Understanding Society Waves 1-8: Revision November 2019

Date: 30th October 2019

This document describes revisions to the Understanding Society Waves 1-8 datasets, including Harmonised BHPS datasets, last fully revised in November 2018. The User Guides and on-line documentation can be consulted for a study-wide perspective and variable level information, respectively.

Wave 5 Higher Education Establishments

In the Wave 5 Higher Education Consent module adult respondents who finished higher education in 1995 or after at a UK institution were asked about the Higher Education establishment that they attended. Nine variables have now been released in a new special licence dataset, Higher Education Codes, SN 8578. The variables are: e_heinst, 1st higher education institute where received a degree; e_heiname, name of 1st higher education institute; together with three more pairs of variables for 2nd, 3rd and 4th institutes respectively. e_heinst* is the institution code which can be matched to publically available institution codes. pidp is also included to enable linkage to the main understanding Society data files. It should be noted that there are instances where e_heiname* contains an establishment name but the corresponding e_heinst* contains -1 “Don’t know”. This occurs when a name has been provided by a respondent but it has not been possible to identify a code, typically due to there being multiple options available.

Distance Moved by Respondents

Two variables have been added to w_indresp.dta relating to distance moved by respondents since the last wave. In i_indresp.dta this will relate to the distance moved since wave h, in h_indresp.dta this will relate to the distance moved since wave g and so on. The new variables are: w_distmov_dv, Distance participant moved since last wave (km), and w_addrmov_dv, Participant changes address postcode since last wave. Distances calculated are in kilometres “as the crow flies” and the centre location of the postcode area is used, not the actual location of an individual property. The distances are calculated from postcodes so these variables are only applicable for moves between different postcodes. Moves within the same postcode cannot be identified using these variables. Where one or other postcode is not known or is not valid then both variables will be set to -1, Don’t know.

Council Tax Bands

Two new variables have been made available in w_hhresp.dta relating to council tax bands. w_ctband_dv, identifies the council tax band for each household and has been derived using reports, external data and imputes. The variable w_ctband_if identifies the data source used for each instance of w_ctband_dv.

Income Changes

The variable containing “total household net income net of household deductions” (w_fihhmnnet3_dv) and “council tax liability” (w_ficount_dv) have been revised to incorporate the new council tax band data (see Council Tax Bands above).

Weighting Changes

The following weights and probabilities have been added to the crosswave file `xwavedat.dta`:

- `psnenu_xd` – UKHLS design weight
- `psnengp_xd` – GPS design weight
- `psnenub_xd` – UKHLS+BHPS design weight
- `prob91e` – BHPS91 England selection probability
- `prob91w` – BHPS91 Wales selection probability
- `prob91s` – BHPS91 Scotland selection probability
- `prob99w` – BHPS99 Wales selection probability
- `prob99s` – BHPS99 Scotland selection probability
- `prob01ni` – BHPS01 Northern Ireland selection probability
- `prob09ni` – GPS09 Northern Ireland selection probability
- `prob09e` – GPS09 England selection probability
- `prob09w` – GPS09 Wales selection probability
- `prob09s` – GPS09 Scotland selection probability
- `prob09emb` – EMB09 selection probability

Please refer to the User Guide for further details.

Geography Datasets Changes

The ONS Postcode Directory (ONSPD) used to assign geographies for Wave 9 is the May 2018 version. This version, however, does not support Local Education Authority (LEA) which is released in SN 6671. The previous version of the ONSPD, May 2017, has been used to assign LEA's for Wave 9. 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 for other waves and then to the relevant ONSPD documentation for full details of their definition.

School Codes Dataset (SN 7182) Changes

There are examples where there can be duplicated school codes across countries. In order to be able to identify which country the school is in the variable `w_countrycode` has been released in both `w_schcode.dta` and `w_schcodeest.dta`.

Interviewer Characteristics Crosswave Changes

In previous releases the crosswave files `xivdata.dta` and `xivdata_bh.dta`, which contain details on the interviewers, were supplied with the EUL (SN 6614) and Special Licence (SN 6931) datasets. From this release onwards these files are made available to researchers in a new separate Special Licence dataset, Interviewer Characteristics, SN 8579.

The interviewer ethnic groups in the file `xivdata.dta` is only supplied by the fieldwork agency in five basic groups for some interviewers from Wave 8 onwards. As a consequence a new variable `racel_5_dv` has been created containing those five ethnic groups. This variable is valid for all interviewers in the data file and is therefore suitable for use with all Waves. The pre-existing variable `racel_dv`, covering a larger group of ethnicities, has been retained but is only suitable for use for interviewers with the more detailed ethnic detail. Interviewers for which only the restricted set (in `racel_5_dv`) is available are coded as -8, inapplicable, in `racel_dv`.

The data required to calculate the variable veteran in the file xivdata.dta has not been supplied by the fieldwork agency in Wave 9. This variable is judged to be of minimal use, with only 15 entries that are not equal to “None or unknown”, and consequently it has been dropped.

xwavedat Crosswave Changes

The variable psmwave_bh, BHPS Wave when sample status changes from TSM to PSM, has been harmonised with psmwave and is therefore no longer available.

The variable dcsedw_dv_bh, BHPS wave by which R reported deceased, has been harmonised with dcsedw_dv and is therefore no longer available.

The variables ynlp14, agelh, lvag14 and lvag16 have been harmonised across BHPS and UKHLS.

Corrections have been made to the first job, first marriage, first cohabitation and first child variables: j1semp; j1boss; j1mngr; ynlp; agelh; lvag14; lvag16; school_dv; feend_dv; anychild_dv; ch1bm; ch1by; evermar_dv; lmar1m_dv; lmar1y_dv; evercoh_dv; coh1m_dv; and coh1y_dv. Note that not all changes are due to corrections. Values of these event based variables will change if the event occurs for the first time by the latest wave (e.g., if a person becomes a parent for the first time or gets married for the first time by the latest wave).

Some information is collected when a 16+ year old person is interviewed for the first time. So, for those sample members who turn 16 and complete the adult interview for the first time, values of these variables will also change. For example, information on country of birth, ethnic group, and parents’ countries of birth.

The variable j1none, Still in full-time education / never had a job, is no longer available as this was only available for 2% of the BHPS sample and not for any UKHLS sample member. Researchers interested in first job should look at the first job variables and if interested in age left FE education should use feedn_dv.

Three BHPS sample members had duplicate PIDP assigned and the duplicate cases have now been dropped.

Wave 1 SOC Variables

The following Wave 1 soc variables: a_j2soc00_cc; a_j2soc10_cc, a_j2soc90_cc and a_j2nssec8_dv have now been released in both EUL and SL versions of a_indresp.dta. The following variables: a_j2soc00; j2soc10 and j2soc90 have been released in the SL version of a_indresp.dta. These variables were already previously available in all other waves.

Timing Files Changes

The data file h_hhgrid_timings_start_end.csv was released in error at Wave 8 and has now been removed.

Improvements to Key Variables

The following youth related variables in c_youth.dta all had problems due to errors in the original coding from the paper questionnaires which have now all been corrected:

- c_yppay – how many hours paid work did you do last week? (note: this label has been corrected from that in previous releases)
- c_yppkmp – how much money did you receive last week to spend on yourself?
- c_ypsave – which of the following describes what you usually do with your money?
- c_ypwhrs – how many hours paid work did you do last week?
- c_ypwklw – did you do any paid work last week?
- c_ypsocweb – do you belong to a social web-site such as bebo, facebook or myspace?
- c_ypnetcht - hours using social media on weekdays

There have been ongoing efforts to improve the consistency of the variable w_sampst, sample status, i.e. whether the respondent is an osm, psm or tsm.

The following variables relating to the end times of household interviews for Wave 6 which were all set to -9 for non-IEMB sample members have now been corrected: f_hgenddaty, f_hgenddatm, f_hgenddatd, f_hgenddathh, f_hgenddatmm, f_hgenddatss, f_hgstrtdaty, f_hgstrtdatm, f_hgstrtdatd, f_hgstrdathh, f_hgstrtdatmm, and f_hgstrtdatss.

Miscellaneous Added Variables

The variable h_ideviceused, Device used for first accessing individual interview, has been released in h_indresp. In addition, to avoid confusion, the variable h_deviceused, Device used for first accessing the household grid, already released in h_hhsamp, has been renamed h_hideviceused.

The variable w_absparnam, Absent parent identifier, has been released in e_chmain.dta and g_chmain.dta. This variable is an indicator which shows which children have the same absent parent.

Miscellaneous Dropped Variables

The variable d_brloop in b_indresp has been dropped as it is superfluous.

The variables pidp1, pidp2, pidp3 and pidp4 in g_indresp have been dropped as they were released in error.

Miscellaneous Changes

Various minor data inconsistencies have been corrected.

Various labelling issues have been addressed.