

UK Data Archive SN 6614 – Understanding Society Waves 1-11: Revision September 2022

Date: 16th September 2022

This document describes revisions to the Understanding Society Waves 1-11 datasets, including Harmonised BHPS datasets, last fully revised in November 2021 with minor additions in March 2022. The User Guides and on-line documentation can be consulted for a study-wide perspective and variable level information, respectively.

Family Matrix (xhhrel) file

xhhrel is a new file that identifies all family relationships consistently across the study, previously referred to as the Family Matrix. The aim of it is to provide an individual level crosswave file of all sample members (those who were ever enumerated as part of the study) that contains familial relationship identifiers reported over the survey period for each sample member. It also contains an origin household identifier variable (osm_hh) which identifies the household they come from, so that sample members who are connected with each other can be identified (either because they were co-resident at some point or were co-resident with individuals who were co-resident with each other).

Full details are provided in the Family Matrix (xhhrel) User Guide (6614_family_matrix_xhhrel_user_guide.pdf).

New International Standard Classification of Education (ISCED) variables

The International Standard Classification of Education (ISCED) is a framework for comparing educational programmes and qualifications from around the world. Amongst other things, it supports international comparisons of attained qualifications. Constructing an ISCED measure for a given UK qualification required more detail than was collected in Waves 1 to 8 of Understanding Society and this was addressed in Wave 9. Consequently, two new ISCED variables have been added to the file w_indresp from Wave 9 onwards as follows: w_isced11_dv (Highest education qualification, short ISCED 2011) and w_nisced11_dv (Newest Highest education qualification, short ISCED 2011).

Further information is provided in 6614_ethnicity_immigration_guide_9thEdition.pdf.

Weighting changes

Weights have now been provided for the first time for datasets containing new-born information (w_newborn) and children information (w_child). Two types of weights are provided for these files: an issue weight *_li, which can be used to create tailored longitudinal weights with a starting point in a particular wave, or a particular age of a person; and a cross-sectional weight *_xw, which can be used in the current wave or by pooling information from children of a particular age or within an age range. The newborn dataset weights are called w_chdnb*_*_, and children weights specifically tailored to developmental questions are called w_chddv*_*_.

Additionally, all cross-sectional weights starting at Wave 2 in UKHLS datasets have been updated in two ways. Firstly, only cross-sectional weights with the highest sample size remain on the dataset – this is to avoid confusion and lead users to use the weights which give the most up-to-date new

immigrant and minority representation, and the best statistical power. Secondly, the weights have incorporated full mortality correction, matching to the longitudinal population distribution of age by gender and by country variables, as well as an additional elimination of zero weights (zero values are still present for TSM's from Wave 1 as per the sample design).

xwavedat corrections

Variables pertaining to first child and first partnership had previously not taken into consideration any new information reported in interviews for Waves 9 to 11. In addition, there was an issue with missing information regarding the date of birth of the first child in many cases. There were also two instances of the variable `birthy` containing the value 9898. These variables have now all been corrected.

COVID-19 Self-employment Income Support Scheme

Six COVID-19 Self-employment Income Support Scheme (SEISS) variables in Wave 11 were previously incorrectly named and have, therefore, been renamed from `k_jsseiss*` to `k_jseiss*`.

childpno dropped in indresp

The variable `w_childpno` has been dropped from `w_indresp` in all waves as it only identifies one child when there could be more children for the sample member for which there is no corresponding `w_childpno` variable.

Unnecessary Higher Education variables dropped from e_indresp

Eight variables, `e_anymore*` and `e_instok*`, previously released in `e_indresp`, have been dropped. These are associated with the Wave 5 Higher Education module, however, they were not intended for release and are of no use to researchers. All useful Higher Education variables remain available in the Special Licence dataset SN 8578.

PID Variable in Harmonised BHPS files reinstated

In the original release of Wave 11 the `pid` variable (Personal identifier (BHPS cohort)) was omitted from all harmonised BHPS wave files: `bw_indall` and `bw_indresp`. These have now been reinstated.

Pensionable age correction

Thirty-six females were identified in `k_pensioner_dv` (Whether R is of pensionable age) as being of pensionable age despite only being 60 years old. In addition, males aged 65 or older should not all be listed as being of pensionable age (those born after 6th December 1953). These have all been corrected.

Miscellaneous changes

Various minor data inconsistencies have been corrected.

Various labelling issues have been addressed.