

The BHPS/Understanding society harmonisation project

Laura Fumagalli (ISER, University of Essex)

April 19, 2018

Harmonised BHPS/Understanding Society dataset

- A revised version of BHPS data from wave 1 (1991) to wave 18 (2008)
- Understanding Society data from wave 1 (2009-2010) onwards (updated each year with data from Understanding Society).

BHPS is modified to match Understanding Society.
Understanding Society is left untouched!

The resulting dataset covers more than 25 years!

Existing research

Existing research combining BHPS and Understanding Society is:

- New
- Cutting edge
- Interdisciplinary

But it was also extremely **time-consuming** before the release of the **Harmonised BHPS/Understanding Society dataset**

Type of research questions

1. Studying neighborhood effects where neighborhood characteristics change slowly over time [▶ topic 1](#)
2. Studying relationships which may vary with the economic cycle [▶ topic 2](#)
3. Studying long run impact of behaviours and policies [▶ topic 3](#)
4. Studying cumulative effects (e.g., long term socio-economic status) [▶ topic 4](#)
5. Studying a phenomenon longitudinal in nature, by comparing England with other countries with long panel data [▶ topic 5](#)
6. Studying long run trends [▶ topic 6](#)
7. (Epi)genetics [▶ topic 7](#)
8. Studying small sub samples of the population [▶ topic 8](#)

General principle

The Harmonised BHPS-Understanding Society follows a **variable first** approach

1. Potential matches between a/few BHPS and a Understanding Society variables are searched for
2. Comparability of selected matches is assessed
3. Only variables meeting pre-set standard of comparability (i.e., have the same -or highly similar- question wording and routing) are considered harmonisable

Consequences

- The **number** of harmonised variables is likely to **increase over time** (new matches can be found or made possible)
- The quality of the harmonisation is **roughly constant**
- The criteria for harmonisation are quite restrictive
- Data recoding is minimal
- Variable renaming is minimal. When possible, variable root remains the same as in the BHPS
- Both harmonised and not-harmonised variables are released
- The variables are released in separate files (e.g., bw_indresp, bw_hhresp)

What is harmonised at the moment

- The structure of egoalt ▶ Egoalt
- xwavedat: there is now a unique file for BHPS and Understanding Society
- windresp, windsamp, windall, whhresp, wincome, wyouth:
 - Waves naming conventions
 - Variables names
 - Variables response categories
 - Variables existing in Understanding Society only, which can be derived on BHPS data by combining two or more variables

Problems → Solutions

- Different naming conventions
 - Naming conventions have been harmonised (BHPS datasets and variables have been given the prefix bw_)
- Variables with the same content had different names
 - BHPS variables have been given the Understanding Society name.
 - ▶ Example
- Variables with the same name had different content or coding frame
 - BHPS variables have been given the suffix _bh.
 - When possible, new variables -with the name and the structure of Understanding Society- have been created by recoding the original BHPS variables
 - ▶ Example
- Batteries of questions are only partially carried over into Understanding Society
 - The original battery of questions is kept with the original name
 - The questions carried into Understanding Society are duplicated and given the Understanding Society name.
 - ▶ Example

What is scheduled for releases 7.1 and 8

- 7.1 Release (June 2018)
 - whhsamp
 - Corrections of mistakes from wave 7 release (e.g., removal of the the suffix _bh when not necessary)
 - More individual questionnaire variables
 - More derived variables
- Post 7.1 Release
 - Harmonisation of pointers and identifiers
 - More harmonisation of egoalt
 - More harmonisation of income variables (e.g., net income)
 - Harmonisation of labels
 - More derived variables
 - Value added data sets (e.g., partnership histories)

Documentation

A harmonised **documentation** to navigate within and between datasets.

▶ **User Guide:** <http://repository.essex.ac.uk/21094/1/bhps-harmonised-user-guide.pdf>

▶ **New website:** <https://www.understandingsociety.ac.uk>

Can you help?

You can help:

- By providing feedback on the current state of the project
- By letting us know the variables you have already harmonised for your research
- By suggesting areas where there might be some scope for additional harmonisation
- By suggesting improvements in the documentation
- By collaborating with us in substantive/methodological work exploiting the features of the dataset.

Questions?

Feel free to contact me: lfumag@essex.ac.uk

Neighborhood effects

- Impact of ethnic diversity on attitudes towards immigration in Britain (Kaufmann, 2015, Comparative Political Studies) ▶ [Kaufmann, 2015](#)
- Impact of ethnic diversity on satisfaction about the neighborhood (Langella and Manning, 2016, LSE WP) ▶ [Langella and Manning, 2016](#)

▶ [Back](#)

Relationships varying with the economic cycle

- The labour market impacts of leaving education when unemployment is high (Taylor, 2013 ISER WP) [▶ Taylor, 2013](#)
- Non-employment, age, and the economic cycle (Jenkins and Taylor, 2012, Longitudinal and Life courses Studies) [▶ Jenkins and Taylor, 2012](#)
- Scarring effect of Unemployment, from the early 90 to the Great Recession (Tumino, 2015, ISER WP) [▶ Tumino, 2015](#)
- Protective role of savings against future financial hardship in different macroeconomic environments (Brown et al., 2017, IZA WP) [▶ Brown et al., 2017](#)
- Relationship between expectations on future financial hardship and savings in different macroeconomic environments (Brown and Taylor, mimeo)

[▶ Back](#)

Long run impact of behaviours/policies

- Impact of saving behaviour as a child on saving in adulthood (Brown and Taylor, 2016, J. Bank. Finance) [▶ Brown and Taylor, 2016](#)

▶ [Back](#)

Cumulative effects

- The income-health gradient (Davillas et al., 2017, HEDG WP)

▶ [Davillas et al., 2017](#)

▶ [Back](#)

Comparison with other long panels

- Within-couple inequalities and women's labour market outcomes in Britain and Germany (Dieckhoff et al., 2016, Research in Social Stratification and Mobility) [▶ Dieckhoff et al., 2016](#)
- House ownership and well-being in Britain and Australia (Smith et al., 2017, Environment and Planning: A) [▶ Smith et al., 2017](#)
- Transition from parental home to home ownership for Britain, Germany and Australia (Coulter et al., mimeo)
- Job Polarization and the Declining Quality of Knowledge Workers for the UK and Germany (Cavaglia and Etheridge, mimeo)

[▶ Back](#)

Long term trends

- Demography of mid-life, from the 1980s to the 2000s (Demey et al, 2011, Population Trends) [▶ Demey et al, 2011](#)
- Transition into adulthood (Schoon and Lyons-Amosb, 2016, Research in Social Stratification and Mobility) [▶ Schoon and Lyons-Amosb, 2016](#)
- Labour market segmentation in the UK from the late 90s until the late 2000s (Yoon and Chung, 2016, Social Indicators Research) [▶ Yoon and Chung, 2016](#)

[▶ Back](#)

(Epi)genetics

- Socioeconomic position and accelerated DNA methylation age (Hughes et al., 2018; R&R at the American Journal of Epidemiology)

▶ [Back](#)

Sub-populations

- Impact of retirement on well-being (Kesavayuth et al., 2016, Economic Inquiry) ▶ [Kesavayuth et al., 2016](#)

▶ [Back](#)

Egoalt

- Original structure (BHPS): alter to ego vs Original structure (Understanding Society): ego to alter
- Changes made
 - spouse → spouse
 - partner → partner
 - natural brother, sister → natural brother, sister
 - aunt, uncle → nephew, niece
 - natural parent → natural son, daughter
 - employee → employer

NOTE: `relationship_bh` is not exactly the same as `relationship_dv`.

Example (1)

- Straightforward cases:
 - `wwlsha` → `bw_wlsh1`
 - `wwlshb` → `bw_wlsh2`
 - `wwlshc` → `bw_wlsh3`
 - `wwlshd` → `bw_wlsh4`
 - `wwlshe` → `bw_wlsh96`
- But also:
 - `wdoi4` → `bw_istrtdaty`
 - `wdoim` → `bw_istrtdatm`
 - `wdoid` → `bw_istrtdatd`
 - `wfisit` → `bw_finnow`

NOTE: the original names of BHPS variables disappear → documentation will facilitate the link between old and new variable names.

Example (2)

- Cases without Understanding Society equivalent
 - $wjbsat \rightarrow bw_jbsat_bh$
 - $wjbsec \rightarrow bw_jbsec_bh$
 - $wm(p)aage \rightarrow bw_m(p)aag_bh$
 - $wm(p)afar \rightarrow bw_m(p)afar_bh$
 - $wracel \rightarrow bw_racel_bh$
 - $wrelkid \rightarrow bw_relkid_bh$
- Cases with Understanding Society equivalent
 - $wjbsect \rightarrow bw_jbsect_bh$ (original) & bw_jbsect (recoded)
 - $wmlstat \rightarrow bw_mlstat_bh$ (original) & bw_mlstat (recoded)
 - $wsampst \rightarrow bw_sampst_bh$ (original) & bw_sampst (recoded)

Example (3)

Series: wnvesta wnvestb wnvestc wnvestd wnveste wnvestf wnvestg
wnvesth wnvesti wnvestj wnvestk

- Variables carried over into Understanding Society
 - wnvestb → bw_nvestb & duplicated as bw_svacts5
 - wnvesth → bw_nvesth & duplicated as bw_svacts1
 - wnvesti → bw_nvesti & duplicated as bw_svacts2
- Variables not carried over into Understanding Society
 - wnvesta → bw_nvesta
 - wnvestc → bw_nvestc
 - wnvestd → bw_nvestd
 - wnveste → bw_nveste
 - wnvestf → bw_nvestf
 - wnvestg → bw_nvestg
 - wnvestj → bw_nvestj
 - wnvestk → bw_nvestk