Understanding Society: Marital and Cohabitation Histories, 1991-2018 User Guide

Alita Nandi Seetha Menon Robert Smith

Institute for Social and Economic Research

University of Essex

CITATION FOR THE DATA:
University of Essex, Institute for Social and Economic Research. (2020). Understanding Society: Marital and Cohabitation Histories, 1991-2018. [data collection]. 2nd Edition. UK Data Service. SN: 8473, http://doi.org/10.5255/UKDA-SN-8473-2

CITATION FOR THE USER GUIDE:
Nandi, Alita; Menon, Seetha and Smith, Robert. (2020). Understanding Society: Marital and Cohabitation Histories, 1991-2018 User Guide. Colchester: University of Essex

Table of Contents

1. Introduction

The *Understanding Society* partnership history file, "*Understanding Society: Marital and Cohabitation Histories, 1991-2018*" (SN 8473) contains information about partnership spells reported by adult respondents in all *Understanding Society* (UKHLS) and British Household Panel Survey (BHPS) samples up to Wave 9.

For UKHLS samples, initial partnership history and current partnership status information was collected in Wave 1 and after that information was collected on changes since last interview. For new entrants after Wave 1, only partial information about their past partnership history was collected.

For BHPS sample members information about their partnership history until 2008 has been extracted from the single partnership history file created by Pronzato (2009). For those BHPS sample members who were interviewed as part of Understanding Society (from onwards 2010), their prior information was combined with that collected during these interviews.

The *Understanding Society: Marital and Cohabitation Histories, 1991-2018* dataset is based on the following datasets which are available from the UK Data Service:

***SN 6614**: Understanding Society: Waves 1-9, 2009-2018 and Harmonised BHPS: Waves 1-18, 1991-2009*
***SN 5151**: British Household Panel Survey: Waves 1-18, 1991-2009*
***SN 5629**: British Household Panel Survey Consolidated Marital, Cohabitation and Fertility Histories, 1991-2009*

Data users are encouraged to read the user manuals for the datasets used to produce this data file to know more about the underlying surveys and datasets (Knies 2018, Taylor et al 2009, Pronzato 2009). All individuals in this file can be linked to their *Understanding Society* data, **SN 6614**, using the unique cross-wave identifier **pidp**. The unique cross-wave identifier in **SN 5629** is **pp** which is the same as **pid,** the unique cross-wave identifier for non-harmonised BHPS files (**SN 5151)**. So, **pid** is also

included[1]**.** The different sample members can be identified in this file by the variable **hhorig**.

2. File structure and variable descriptions

SN 8473 includes two files. One of them**, phistory_long.dta**, is in long format, i.e., each row represents one spell for each person in the dataset. In this section the long format file and its content is described. The other file, **phistory_wide.dta** is in wide format, that is, each row represents one person. In this wide format file, variable names pertaining to spells have a number suffix which represents the spell number. Both files only include individuals who have provided information about at least one union or partnership and so all persons in the dataset have at least one partnership spell. The following variables are available in this dataset:

- **pidp:** This is the unique cross-wave person identifier available within the **SN 6614** data files. This identifier can be used to link information from **SN 6614** data files.
- **pid:** This is the unique cross-wave person identifier available within the **SN 5151** data files. This variable is also available in **SN 6614** data files (except for in any of the first wave files as these files do not include BHPS sample members). This variable equals -8 for all UKHLS sample members and new BHPS sample members (that is, individuals who joined the BHPS sample members' households for the first time during the UKHLS survey period).
- **hhorig:** This variable identifies the sample origin that the person belongs to. The following is a summary of the frequency of sample members within each sample origin.
- **spellno:** This is the spell number of the union of each individual where the first spell is the earliest partnership.
- **spellnoR:** This is the spell number of the union of each individual in reverse order, that is, where the first spell is the most recent partnership.
- **status:** This variable identifies whether the partnership is a marriage, civil partnership or cohabitation (living together as a couple). It takes on the following values; 2 "Married", 3 "Civil Partner" and 10 "Living together as a couple". Note the coding frame is the same as the marital status derived variable, **w_maststt_dv** in **w_indall** and **w_indresp** files.

[1] Note: Those who joined the households of BHPS sample members after the BHPS ended, that is, as part of Understanding Society, do not have a valid existing PID. So, their PID equals -8.

- **partner:** This variable records the **pidp** of the partner in spells where this information was available.
- **starty & startm:** This is the year and month of the start date of the union.
- **endy & endm:** This is the year and month of the end date of the union.
- **divorcey & divorcem:** When marriage separation and divorce dates are available, end date represents the separation date and the divorce date is recorded here. If only separation date is available this is a valid skip (-8).
- **start_if:** This variable identifies if the start date has been imputed. It takes on the following values: 0 is no imputation, 1 is month imputed 2 month and year is imputed.
- **end_if:** This variable identifies if the end date has been imputed. It takes on the following values: 0 is no imputation, 1 is month imputed 2 month and year is imputed.
- **divorce_if:** This variable identifies if the divorce date has been imputed. It takes on the following values: 0 is no imputation, 1 is month imputed 2 month and year is imputed.
- **spell_if:** This takes on the value 2 if the whole spell has been imputed. When the marital or partnership status changed between two consecutive wave interviews but no spell was reported in the interview the entire spell had to be imputed.
- **ongoing:** This binary variable identifies whether a spell is ongoing at the time of the survey. It takes on the value of 0 and 1 with 1 indicating an ongoing spell.
- **lastinty & lastintm:** This is the year and month of the last interview date.
- **startdate, enddate, divorcedate, lastintdate:** these are the start, end, divorce and last interview dates in Stata date format, that is, these are measured in months since January 1960. As Stata data formatting has been attached to the variables, these will appear as the correct date. For example, if the start date is January 1960 then **startdate** will have the value 2 and it will appear on screen as 1960m1.
- **ttl_married, ttl_civil_partnership, ttl_cohabit**: These variables measure the total number of marriage, civil partnership and cohabitation spells for each person.
- **ever_married**, **ever_civil_partnership, ever_cohabit:** These variables are indicator variables which show if a person has ever married, been in a civil partnership or been in a non-marital cohabitation.

3. Variable creation, missing data and imputations

In Wave 1, all adult respondents (aged 16+) were asked about their (i) current cohabitation, (ii) current marriage and the cohabitation spell preceding that with the same partner, (iii) past cohabitation and marriage history if applicable. If the current marriage spell was preceded by a cohabitation spell with the same partner then the cohabitation end date is not asked and set to the month prior to the marriage start date. From the second wave onwards, they were asked about changes to their marital status and information about these changes. They were also asked about any new cohabitation spells since the last interview.

From the second wave onwards the new household members who join were asked about their first marriage and total number of marriages when they were interviewed for the first time. So, the start date of their first marriage is available. If currently married then their marriage end date has been set to the current interview date (ongoing spell). All other marriage dates are set to -9.

For ongoing spells the end date has been set to the date of current interview. The **ongoing** flag can be used to identify such cases. Note as respondents do not give interviews every year, the current interview date varies across respondents. The variables **lastintdate** show the date that they were last interviewed. The month and year of this date are also available as separate variables: **lastinty**, **lastintm.**

Imputed dates:

- If the start or end month of a spell was missing but a valid start or end year was reported, the month was imputed by setting it to June.
- If marriages followed cohabitations with the same partner then the end date of the cohabitation spell was set to the month before the marriage start date, if available. If the start date was missing it was imputed as mid-point from the cohabitation start date to the current interview date or marriage end date.
- If a change in marital status between waves was noted but the change since last interview questions were not asked then the start date was set to the last interview date.
- There were some missing dates that could not be imputed and were left as is.
- Overlapping spells as reported were not changed to make them consistent.

- In cases, where the current marital status (i.e., at the time of interview) changed between two consecutive interviews but no change was reported in the annual events history (i.e., where changes since the last interview are recorded), then the dates were imputed. Specifically, the year of last interview was taken as the start date and June to be the month, while the end date was taken to be the mid-point of the imputed start date and the current interview date. If this was the final spell then (as with other spells), the current interview was taken as the end date and the ongoing spell indicator was set to 1. Start dates which had been imputed to greater than their end dates were later corrected to follow on from the previous end date. These new spells that were added, were done in a coherent manner which aligned roughly with the rest of the data, and didn't disrupt the original data. If the spell thus created overlapped another reported spell, both the start and end date of this new imputed spell was imputed to be the mid-point between the two interview dates. This was done to signpost this spell but the dates were made to be equal as there was no additional information to help impute the dates. Such spells can be identified by spells with the same start and end date and for which the variable **spell_if** equals 2. Users can decide whether to keep such spells in the datafile before using it. For example, a cohabitation spell was reported to start in January 2000 and end in January 2005 and a marriage spell was imputed with start and end dates June 2004 and June 2005. To avoid the overlap the start and end dates of this new imputed marriage spell were changed to January 2005 and **spell_if** set to 2.
- When these imputations were all put together the spells were merged with one another to create more coherent spells, using information on the status, partner and logical timing.
- For proxy interviews, if any information was provided then that was used, if not, the status was noted as changing and then the imputation method discussed above was used.
- Those who had never responded, but their marital status was married or cohabiting as a couple, could not be included in this partnership history file due to lack of data on partnership spells. Similarly, in a few cases, adult respondents reported invalid dates for their current marriage and no past history and again could not be included in this partnership history file.

Missing values:

The same missing value codes as in the BHPS and Understanding Society datafiles has been used, that is:

-8: valid skip
-9: missing

However, as **startdate**, **enddate** and **divorcedate** are in Stata date format, -8 and -9 values would be interpreted by Stata as 8 months and 9 months before January 1960, which could be valid dates so, a value of -720 has been assigned if the date is missing (-9). This translates to January 1900 which is earlier than the earliest start date for any partnership spell in the dataset. For valid skip (-8), a value of -721 has been assigned which is interpreted by Stata as December 1899.

4. Data Description

A brief description of the dataset – names, value labels, variable labels of variables included in the dataset is as follows.

```
  obs:       130,076                          Substantive data for responding adults (16+), incl. proxies
 vars:            33                          29 Apr 2020 10:28
 size:    18,210,640                          (_dta has notes)
```

variable name	storage type	display format	value label	variable label
pidp	long	%12.0g		cross-wave person identifier (public release)
pid	long	%12.0g	pid	person number identification
hhorig	float	%27.0g	f_hhorig	Sample origin, household
spellno	float	%9.0g		spell no, earliest first
spellnoR	float	%9.0g		spell no, recent first
status	float	%42.0g	status	partnership status
partner	long	%12.0g	partner	partner pidp
startdate	double	%tm	startdate	start date
starty	float	%10.0g	starty	start year
startm	float	%10.0g	startm	start month
enddate	double	%tm	enddate	end date
endy	float	%10.0g	endy	end year
endm	float	%10.0g	endm	end month
divorcedate	float	%tm	divorcedate	Divorce date if available
divorcey	float	%15.0g	divorcey	Divorce year if available
divorcem	float	%16.0g	divorcem	Divorce month if available
mrgend	float	%10.0g	mrgend	how marriage/civil partnership ended
cohend	float	%9.0g	cohend	how cohabitation ended
ongoing	float	%13.0g	ongoing	ongoing spell indicator
start_if	float	%18.0g	start_if	start date imputation flag
end_if	float	%18.0g	end_if	end date imputation flag
divorce_if	float	%18.0g	divorce_if	divorce date imputation flag
spell_if	float	%17.0g	spell_if	full spell was imputed
ttl_spells	float	%9.0g		total number of any partnership spells
ttl_married	float	%9.0g		total number of marriage spells
ttl_civil_par~p	float	%9.0g		total number of civil partnership spells
ttl_cohabit	float	%9.0g		total number of cohabitation spells
ever_married	float	%9.0g	ever_married	ever married
ever_civil_pa~p	float	%9.0g	ever_civil_partnership	ever in civil partnership
ever_cohabit	float	%9.0g	ever_cohabit	ever cohabit
lastintdate	float	%tm	lastintdate	last interview date
lastinty	float	%9.0g	lastinty	last interview year
lastintm	float	%9.0g	lastintm	last interview month

Summary statistics of all the variables included in the dataset is as follows.

Variable	Obs	Mean	Std. Dev.	Min	Max
pidp	130,076	6.93e+08	4.95e+08	2727	1.65e+09
pid	130,076	1.66e+07	3.72e+07	-8	1.89e+08
hhorig	130,076	2.567791	2.438668	1	16
spellno	130,076	1.661413	1.005057	1	12
spellnoR	130,076	1.661413	1.005057	1	12
status	130,076	5.245264	3.924857	2	10
partner	130,076	4.45e+08	5.12e+08	-9	1.65e+09
startdate	130,076	360.167	271.0161	-720	714
starty	130,076	1940.373	317.5553	-9	2019
startm	130,076	6.183254	3.917059	-9	12
enddate	130,076	499.581	266.6567	-720	725
endy	130,076	1940.899	352.3491	-9	2020
endm	130,076	6.042583	4.176473	-9	12
divorcedate	130,076	-628.0667	311.3895	-721	713
divorcey	130,076	180.5356	582.6167	-9	2019
divorcem	130,076	-6.566707	4.547056	-9	12
mrgend	130,076	-3.134245	4.486582	-8	4
cohend	130,076	-3.810741	4.667685	-8	2
ongoing	130,076	.401373	.4901781	0	1
start_if	130,076	.2106461	.5728585	0	2
end_if	130,076	.2428426	.607703	0	2
divorce_if	130,076	-6.73251	3.140232	-8	2
spell_if	130,076	.1193302	.4737325	0	2
ttl_spells	130,076	2.322827	1.389795	1	12
ttl_married	130,076	1.202428	.7905986	0	8
ttl_civil_~p	130,076	.0104477	.1121077	0	3
ttl_cohabit	130,076	1.109951	1.123322	0	11
ever_married	130,076	.8513946	.3557005	0	1
ever_civil~p	130,076	.0093945	.0964693	0	1
ever_cohabit	130,076	.6697239	.4703143	0	1
lastintdate	130,076	626.4066	62.26003	-9	712
lastinty	130,076	2008.419	82.98599	-9	2019
lastintm	130,076	6.591654	3.566951	-9	12

The data file only includes adult respondents who have reported at least one partnership spell. It includes 74,534 adult respondents and 130.076 spells. Table 1 shows the distribution of spells across these respondents.

Table 1: Distribution of partnership spells

Number of spells	Number of adult respondents with these types of spells			
	Any partnership	Marriage	Civil partnership	Living as a couple
0		13,739	77,094	39,890
1	46,771	52,642	399	26,751
2	16,795	9,428	39	7,940
3	8,719	1,550	2	2,136
4	3,571	148		560
5	1,096	18		166
6	382	7		60
7	133	1		20
8	44	1		5
9	15			4
10	5			1
11	2			1
12	1			
Total	77,534	77,534	77,534	77,534

Table 2 shows the distribution of respondents with at least one marriage, civil partnership or cohabitation (living as a couple) spell across the different samples.

Table 2: Distribution of respondents with at least one marriage, civil partnership or cohabitation (living as a couple) spell across the different samples

Sample	Number of adult respondents
UKHLS GP-GB sample	40,932
UKHLS GP-NI sample	1,903
UKHLS EMB sample	6,456
UKHLS IEMB sample	3,439
BHPS GB sample	14,341
BHPS Scottish Boost sample	2,894
BHPS Welsh Boost sample	3,131
BHPS Northern Irish Boost sample	3,221
ECHP – SCPR	556
ECHP – ONS	509
ECHP - NI	152
Total number of adult respondents	77,534

5. Data Quality and reporting errors

There are 289 spells where the end date is earlier than the start date, 19 spells where the end date is later than the divorce date and 2,554 spells where the start date is earlier than the end date of the previous spell (overlapping spells).

Table 3: Distribution of spells with missing start and/or end dates

	Spell end date missing?		
Spell start date missing?	No	Yes	Total
No	123,307	3,386	126,693
Yes	2,658	725	3,383
Total	125,965	4,111	130,076

Table 4: Distribution of spells with imputed start and/or end dates

	Spell end date imputed?			
Spell start date imputed?	No	Yes, month imputed	Yes, whole date imputed	Total
No	105,336	4,733	3,136	113,205
Yes, month imputed	3,646	2,617	79	6,342
Yes, whole date imputed	1,566	118	8,845	10,529
Total	110,548	7,468	12,060	130,076

Comparing the partnership status and partnership PIDP in this dataset with that reported in the individual adult respondent files, **w_indresp**, no mismatch was found in partner PIDP (**partner vs w_ppid**), but there were some mismatches in marital status (**status vs w_mastat_dv**).

If you have any questions or would like to provide feedback please email usersupport@understandingsociety.ac.uk

6. References

Fumagalli, Laura, Knies, Gundi and Buck, Nick (2017): *Understanding Society*: The UK Household Longitudinal Study harmonised British Household Panel Survey (BHPS) User Guide. Colchester: University of Essex.

Pronzato, C. (2011). *British Household Panel Survey Consolidated Marital, Cohabitation and Fertility Histories, 1991-2009*. [data collection]. *3rd Edition.* University of Essex. Institute for Social and Economic Research, [original data producer(s)]. UK Data Service. SN: 5629, http://dx.doi.org/10.5255/UKDA-SN-5629-1.

Taylor, M., Brice, J., Buck, N., Prentice-Laine, E. (ed.) (2009). British Household Panel Survey User Manual Volume A: Introduction, Technical Report and appendices.

University of Essex, Institute for Social and Economic Research, NatCen Social Research, Kantar Public. (2019). Understanding Society: Waves 1-9, 2009-2018 and Harmonised BHPS: Waves 1-18, 1991-2009. [data collection]. 12th Edition. UK Data Service. SN: 6614, http://doi.org/10.5255/UKDA-SN-6614-13.

University of Essex. Institute for Social and Economic Research. (2018). *British Household Panel Survey: Waves 1-18, 1991-2009*. [data collection]. *8th Edition.* UK Data Service. SN: 5151, http://doi.org/10.5255/UKDA-SN-5151-2

7. Acknowledgements

Understanding Society is an initiative funded by the Economic and Social Research Council and various Government Departments, with scientific leadership by the Institute for Social and Economic Research, University of Essex, and survey delivery by NatCen Social Research and Kantar Public. The research data are distributed by the UK Data Service.

We would like to thank Stephanie Auty, Yanchun Bao, Mike Brewer, Paulino Fiorini, Nikita Jacob and Victoria Nolan for their help and feedback and Graham Jolliffe and the UKDS team for help with the data deposit.

8. Appendix

Table A1: Description of the variables from the source data files that were used to create the variables in this file

Variable name	Waves	Data File[1]	Survey
pidp, hhorig, pid	-	xwavedat	SN 6614
w_ppid, a_mastat_dv, a_nmar, a_cohab_dv, a_lcmcby4, a_lcmcbm, a_lcmcoh, a_lcmspy4, a_lcmspm a_lcsby4, a_lcsbm a_istrtdaty, a_istrtdatm, a_mrgendt,	1	a_indresp	
a_cohabno, a_lcsby4, a_lcsbm, a_lcsey4, a_lcsem	1	a_cohab	
a_marno, a_lcmarm, a_lcmary4, a_lmarm, a_lmary4, a_lmarm, a_lcmarm	1	a_marriage	
w_ppid, w_mastat_dv, w_cohab_dv, w_statcy4*, w_statcm*, w_lmcby4, w_lmcbm, w_coh1by, w_coh1bm, w_lmar1y, w_lmar1m, w_istrtdaty, w_istrtdatm	2-6	w_indresp (w = a…i)	
pp, partner, separation, start_date, stop_date, marital, cohabitation	-	Family	SN 5629
pid, wmarstat	1-18	windresp (w = a,…r)	SN 5151