



*Understanding Society: Marital and Cohabitation Histories,
1991-2015 User Guide*

Alita Nandi

Seetha Menon

Institute for Social and Economic Research

University of Essex

CITATION FOR THE DATA:

University of Essex. Institute for Social and Economic Research. (2019).
Understanding Society: Marital and Cohabitation Histories, 1991-2015. [data
collection]. UK Data Service. SN: 8473, <http://doi.org/10.5255/UKDA-SN-8473-1>

CITATION FOR THE USER GUIDE:

Nandi, Alita and Menon, Seetha. (2019). Understanding Society: Marital and
Cohabitation Histories, 1991-2015 User Guide. Colchester: University of Essex

Table of Contents

1. Introduction	1
2. File structure and variable descriptions.....	2
3. Variable creation, missing data and imputations	3
4. Data Description	5
5. Data Quality and reporting errors.....	8
6. References	9
7. Acknowledgements.....	9
8. Appendix	10

1. Introduction

The *Understanding Society* partnership history file, “*Understanding Society: Marital and Cohabitation Histories, 1991-2015*” (SN 8473) contains information about partnership spells reported by adult respondents in all *Understanding Society* (UKHLS) and British Household Panel Survey (BHPS) samples except the Immigrant and Ethnic Minority Boost sample up until Wave 6. An updated version of this file which includes information for the IEMB sample and information collected since Wave 6 will be available in January 2020.

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, 1991-2015* dataset is based on the following datasets which are available from the UK Data Service:

SN 6614: *Understanding Society: Waves 1-7, 2009-2016 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 included¹. 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 we describe the long format file and its content. 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. It only includes persons who have provided information about at least one union 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_mastst_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.
- **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.
- **t1l_married, t1l_civil_partnership, t1l_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 are asked about changes to their marital status and information about these changes. They are also asked about any new cohabitation spells since the last interview.

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

For ongoing spells, the end date is 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, we imputed the month 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 dates that could not be imputed

Missing values:

We used the same missing value codes as in the BHPS and Understanding Society datafiles, that is,

- 8: valid skip
- 9: missing

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

4. Data Description

Here is a brief description of the dataset.

```

obs:      107,962      Substantive data for responding adults (16+), incl. proxies
vars:      32          23 Jul 2018 21:40
size:     14,682,832  (_dta has notes)

```

variable name	storage type	display format	value label	variable label
pidp	long	%12.0g	pidp	cross-wave person identifier (public release)
pid	long	%12.0g	pid	person number identification
hhorig	float	%27.0g	hhorig	Sample origin, household
spellno	float	%9.0g		spell no, earliest first
spellnoR	float	%9.0g		spell no, recent first
status	float	%41.0g	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
t1l_spells	float	%9.0g		total number of any partnership spells
t1l_married	float	%9.0g		total number of marriage spells
t1l_civil_pa~p	float	%9.0g		total number of civil partnership spells
t1l_cohabit	float	%9.0g		total number of cohabitation spells
ever_married	float	%9.0g		ever married
ever_civil_pa~p	float	%9.0g		ever in civil partnership
ever_cohabit	float	%9.0g		ever cohabit
lastintdate	float	%tm		last interview date
lastinty	float	%9.0g		last interview year
lastintm	float	%9.0g		last interview month

Here are the summary statistics of all the variables included in the dataset.

Variable	Obs	Mean	Std. Dev.	Min	Max
pidp	107,962	6.57e+08	4.90e+08	2727	1.64e+09
pid	107,962	1.80e+07	3.79e+07	-8	1.89e+08
hhorig	107,962	2.402299	2.285031	1	16
spellno	107,962	1.649571	1.023837	1	14
spellnoR	107,962	1.649571	1.023837	1	14
status	107,962	5.376355	3.949675	2	10
partner	107,962	3.94e+08	4.89e+08	-9	1.64e+09
startdate	107,962	329.43	265.6775	-720	671
starty	107,962	1934.87	326.1163	-9	2015
startm	107,962	6.193392	3.992536	-9	12
enddate	107,962	476.5384	259.0474	-720	676
endy	107,962	1939.993	349.0926	-9	2016
endm	107,962	6.058604	4.196927	-9	12
divorcedate	107,962	-619.7413	323.8873	-721	675
divorcey	107,962	173.3548	574.4025	-9	2016
divorcem	107,962	-6.700543	4.207655	-9	12
mrgend	107,962	-2.885589	4.461171	-8	4
cohend	107,962	-4.107241	4.583547	-8	2
ongoing	107,962	.4425909	.4966956	0	1
start_if	107,962	.0877716	.3357782	0	2
end_if	107,962	.1251829	.4119035	0	2
divorce_if	107,962	-7.257878	2.353003	-8	2
t1l_spells	107,962	2.299142	1.439862	1	14
t1l_married	107,962	1.140577	.7359082	0	5
t1l_civil~p	107,962	.00427	.0675778	0	2
t1l_cohabit	107,962	1.154295	1.211981	0	12
ever_married	107,962	.8466034	.360371	0	1
ever_civil~p	107,962	.0041126	.0639975	0	1
ever_cohabit	107,962	.676877	.4676714	0	1
lastintdate	107,962	621.3809	58.55981	377	676
lastinty	107,962	2011.309	4.941021	1991	2016
lastintm	107,962	6.668569	3.508475	1	12

The data file only includes adult respondents who have reported at least one partnership spell. It includes 64,906 adult respondents and 107,962 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	Marriage	Civil partnership	Cohabitation
0		11,736	64,736	32,535
1	38,673	45,191	166	23,058
2	15,707	7,000	4	6,724
3	6,547	868		1,806
4	2,571	105		522
5	880	6		160
6	319			46
7	124			26
8	40			9
9	23			7
10	12			7
11	4			4
12	3			2
13	1			
14	2			
Total number of adult respondents	64,906	64,906	64,906	64,906

Table 2 shows the distribution of respondents with at least one marital/civil partnership or (non-marital) cohabitation spell across the different samples.

Table 2: Distribution of respondents with at least one marital/civil partnership or (non-marital) cohabitation spell across the different samples

Sample	Number of adult respondents
UKHLS GP-GB sample	34,875
UKHLS GP-NI sample	1,695
UKHLS EMB sample	5,018
BHPS GB sample	13,503
BHPS Scottish Boost sample	2,739
BHPS Welsh Boost sample	2,870
BHPS Northern Irish Boost sample	2,996
ECHP – SCPR	550
ECHP – ONS	508
ECHP - NI	152
Total number of adult respondents	64,906

5. Data Quality and reporting errors

There are 92 spells where the end date is earlier than the start date, 5 spells where the end date is later than the divorce date and 1427 spells where the start date is earlier than the end date of the previous spell.

	Spell end date missing?	
Spell start date missing?	No	Yes
No	103,294	1,720
Yes	1,316	1,632

	Spell end date imputed?	
Spell start date imputed?	No	Yes
No	93,488	6,762
Yes	4,206	3,506

Comparing this data with the partnership status and partnership PIDP as reported in the individual files, **w_indresp**, we found no mismatch in partner PIDP, but some mismatch in marital status: 872 cases in Wave 1, 521 in Wave 2, 329 in Wave 3, 336 in Wave 4, 231 in Wave 5 and 53 in Wave 6.

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.

Knies, Gundi (ed.) (2018). *Understanding Society: Waves 1-8, 2009-2017 and harmonised British Household Panel Survey: Waves 1-18, 1991-2009, 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. (2018). *Understanding Society: Waves 1-8, 2009-2017 and Harmonised BHPS: Waves 1-18, 1991-2009*. [data collection]. 11th Edition. UK Data Service. SN: 6614, <http://doi.org/10.5255/UKDA-SN-6614-12>

University of Essex. Institute for Social and Economic Research. (2010). *British Household Panel Survey: Waves 1-18, 1991-2009*. [data collection]. 7th Edition. UK Data Service. SN: 5151, <http://doi.org/10.5255/UKDA-SN-5151-1>

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.

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_lcmcb4, a_lcmcbm, a_lcmcoh, a_lcmspy4, a_lcmspm a_lcsby4, a_lcsbm a_istrtdaty, a_istrtdatm, a_mrgendt, a_cohabno, a_lcsby4, a_lcsbm, a_lcsey4, a_lcsem	1	a_indresp	
a_marno, a_lcmarm, a_lcmarmy4, a_lmarm, a_lmarmy4, a_lmarm, a_lcmarm	1	a_cohab	
w_ppid, w_mastat_dv, w_cohab_dv, w_statcy4*, w_statcm*, w_lmcb4, w_lmcbm, w_coh1by, w_coh1bm, w_lmar1y, w_lmar1m, w_istrtdaty, w_istrtdatm	1	a_marriage	
pp, partner, separation, start_date, stop_date, marital, cohabitation	2-6	w_indresp (w = a...f)	
pid, wmarstat	-	family	SN 5629
	1-18	windresp (w = a,...r)	SN 5151