Documentation for the Standardization of the UK Harmonized Histories Data File for birth, partnership histories, leaving home questions and background variables

HARMONIZED HISTORIES UK
(14539 respondents)

Karolin Kubisch
Max Planck Institute for Demographic Research Rostock

Wendy Sigle-Rushton
London School of Economics

2009

The following documentation gives a description of all input variables and the consequent preparation of the output variables according to the manual for the preparation of comparative fertility and union histories.

Missing values are coded:
.a unknown
.b does not apply
.c unavailable in survey

Source: DatabaseAugust2009.zip

(Source prepared by Wendy Sigle-Rushton (LSE))

Original BHPS datasets used: ohhsamp.dta, oindresp.dta, family.dta, bindall.dta, cindall.dta, dindall.dta, eindall.dta, findall.dta, gindall.dta, hindall.dta, iindall.dta, jindall.dta, kindall.dta, lindall.dta, mindall.dta, nindall.dta, oindall.dta, bchildnt.dta, kchildnt.dta, lchildnt.dta, dindsamp.dta, eindsamp.dta, findsamp.dta, gindsamp.dta, hindsamp.dta, iindsamp.dta, jindsamp.dta, kindsamp.dta, lindsamp.dta, mindsamp.dta, nindsamp.dta, oindsamp.dta, xwavedat.dta, mindresp, nindresp.dta, oindresp.dta

With Notes on the UK Data (August 2009) from W. Sigle-Rushton:

1.)I have dropped those people who had already been married when entering the panel and for whom I do not have a history.

Interview dates UK BHPS: 2005/2006

1. Part Basic Information

RESPID: ID number to be assigned at merging LEAVE BLANK

ARID: ID number from raw data (original ID number)
14539 respondent

COUNTRY: Country and survey
Original: string
8261 UK BHPS

MONTH_S: Month of survey
Missing cases: 8

IMONTH_S: Month of survey, including imputed dates

2.) *When not available, I imputed the month of interview as Sept because more than half the sample gets interviewed in September.*

YEAR_S: Year of survey
2005:13708 /2006:823
8 missing cases➔ imputed 2005

SEX: Sex of the respondent
No missing cases
Sex structure of the UK respondents:
Male: 6683 and Female: 7856

BORN_Y: Year of birth of respondent
0 missing cases
1925-1989

BORN_M: Month of birth of respondent
0 missing cases

IBORN_M: Month of birth of respondent
including imputed months

2.Part LEAVING HOME

LEAVING HOME QUESTIONS ARE NOT AVAILABLE IN THE SURVEY

LEAVE_1: Indicator of whether left home

LEAVE_Y1: Year of first time leaving home

LEAVE_M1: Month of first time leaving home

ILEAVE_M1: Month of first time leaving home
and imputed months:

---

3. Part UNIONS AND DISSOLUTION ($=order of union)

UNINUM: Total number of unions

0: 2824
1: 9446
2: 1769
3: 382
4: 85
5: 24
6: 6
7: 3

UNION_$: UNION order

UNION_1: 11715
UNION_2: 2269
UNION_3: 500
UNION_4: 118
UNION_5: 33
UNION_6: 9
UNION_7: 3

No missing cases

UNION_Y$: Year of start union

UNION_Y1 missing values: 3333
UNION_Y2 missing values: 94
UNION_Y3 missing values: 13
UNION_Y4 missing values: 5

LTUNION_Y$: Left truncated start year of union (country specific variable for UK)

LTUNION_Y1: 3289 cases
LTUNION_Y2: 71 cases
LTUNION_Y3: 3 cases
LTUNION_Y4: 4 cases

UNION_M$: Month of start UNION

Harmonized:
UNION_Mx=.b if UNION_x==0
UNION_M1 missing values: 4017
UNION_M2 missing values: 432
UNION_M3 missing values: 108
UNION_M4 missing values: 32
UNION_M5 missing values: 10
UNION_M6 missing values: 4
UNION_M7 missing values: 3

LTUNION_M$: Left truncated start month of union
(country specific variable for UK)

LTUNION_M1: 3270 cases 19 unknown
LTUNION_M2: 71 cases
LTUNION_M3: 3 cases
LTUNION_M4: 4 cases

IUNION_M$: Month of start UNION
and imputed months
according to manual page 4 (random)

➔ Imputation if months is unknown and year is known
Leave unknown cases (.a):
IUNION_M1: 3333
IUNION_M2: 90
IUNION_M3: 13
IUNION_M4: 5

SEP_$: Dissolution of UNION

SEP_1 missing cases: 2
SEP_2 missing cases: 1
SEP_3 missing cases: 3

Order of Union	Number of unions	number of separations	death of partner
1	11715	3164	733
2	2269	830	69
3	500	186	7
4	118	56	
5	33	18	
6	9	6	
7	3	2	

SEP_Y$: Year of end of UNION

SEP_Y1 missing values: 580
SEP_Y2 missing values: 100
SEP_Y3 missing values: 26
SEP_Y4 missing values: 5
SEP_Y5 missing values: 1

SEP_M$: Month of end of UNION

SEP_M1 missing values: 1521
SEP_M2 missing values: 481
SEP_M3 missing values: 116
SEP_M4 missing values: 44
SEP_M5 missing values: 16
SEP_M6 missing values: 4
SEP_M7 missing values: 2

ISEP_M$: Month of end of UNION
and imputed months

➔ Imputation if months is unknown and year is known
Leave unknown cases (.a):

ISEP_M1 missing values: 580
ISEP_M2 missing values: 100
ISEP_M3 missing values: 26
ISEP_M4 missing values: 5
ISEP_M5 missing values: 1

---

4. Part MARRIAGE AND DIVORCE ($=order of union)

MARR_$: Indicator of whether marriage took place
and type of marriage

Order of Union	Number of unions	number of marriages
1	11715	9587
2	2269	1444
3	500	272
4	118	66
5	33	11
6	9	1
7	3	

MARR_Y$: Year of marriage

MARR_Y1 missing values: 3242
MARR_Y2 missing values: 68
MARR_Y3 missing values: 11
MARR_Y4 missing values: 4

LTMARR_Y$: left truncated year of marriage
(country specific variable for UK)

LTMARR_Y1 3230 cases
LTMARR_Y2 63 cases
LTMARR_Y3 7 cases
LTMARR_Y4 4 cases

MARR_M$: Month of marriage

MARR_M1 missing values: 3472
MARR_M2 missing values: 173
MARR_M3 missing values: 45
MARR_M4 missing values: 14
MARR_M5 missing values: 2

LTMARR_M$: Left truncated month of marriage
(country specific variable for UK)

LTMARR_M1 3230 missing cases
LTMARR_M2 63 cases

LTMARR_Y3 7 missing cases
LTMARR_Y4 4 cases

IMARR_M$: Month of marriage
and imputed months
according to manual page 4 (random)

IMARR_M1: 3230
IMARR_M2: 66
IMARR_M3: 9
IMARR_M4: 4

DIV_$: Indicator of whether divorce occurred

DIV_1 missing values: 2
DIV_2 missing values: 1
DIV_3 missing values: 3

Order of Union	Number of unions	number of marriages	number of divorces
1	11715	9587	1860
2	2269	1444	306
3	500	272	57
4	118	66	16
5	33	11	2
6	9	1	
7	3		

DIV_Y$: Year of divorce

DIV_Y1 missing values: 48
DIV_Y2 missing values: 8
DIV_Y3 missing values: 5

DIV_M$: Month of divorce

DIV_M1 missing values: 509
DIV_M2 missing values: 94
DIV_M3 missing values: 24
DIV_M4 missing values: 8

IDIV_M$: Month of divorce
and imputed months
according to manual page 4 (random)

➔ Imputation if months is unknown and year is known

IDIV_M1 missing values: 509
IDIV_M2 missing values: 48
IDIV_M3 missing values: 8
IDIV_M4 missing values: 5

5. Part PARTNER`S CHARACTERISTICS ($=order of union)

SEXP_$: Partner`s sex

SEXP_1: missing cases: 3062
SEXP_2: missing cases: 649
SEXP_3: missing cases: 146
SEXP_4: missing cases: 38
SEXP_5: missing cases: 12
SEXP_6: missing cases: 5
SEXP_7: missing cases: 2

Partner	Number of unions	Number male	Number female
1	11715	4502	4151
2	2269	920	700
3	500	195	159
4	118	49	31
5	33	13	8
6	9	1	3
7	3		1

YEARBIRP_$: Year of birth of partner

YEARBIRP_1 missing cases: 3062
YEARBIRP_2 missing cases: 649
YEARBIRP_3 missing cases: 146
YEARBIRP_4 missing cases: 38
YEARBIRP_5 missing cases: 12
YEARBIRP_6 missing cases: 5
YEARBIRP_7 missing cases: 2

MONBIRP_$: Month of birth of partner

MONBIRP_1 missing cases: 3062
MONBIRP_2 missing cases: 649
MONBIRP_3 missing cases: 146
MONBIRP_4 missing cases: 38
MONBIRP_5 missing cases: 12
MONBIRP_6 missing cases: 5
MONBIRP_7 missing cases: 2

IMONBIRP_$: Month of birth of partner and imputed months according to manual page 4 (random)

➔ Imputation if months is unknown and year is known

IMONBIRP_1 missing cases: 3062
IMONBIRP_2 missing cases: 649
IMONBIRP_3 missing cases: 146
IMONBIRP_4 missing cases: 38
IMONBIRP_5 missing cases: 12

IMONBIRP_6 missing cases: 5
IMONBIRP_7 missing cases: 2

NUMCHP_$: Number of children of partner at start of union$

NUMCHP_1: missing values: 6241
NUMCHP_2: missing values: 725
NUMCHP_3: missing values: 154
NUMCHP_4: missing values: 41
NUMCHP_5: missing values: 12
NUMCHP_6: missing values: 5
NUMCHP_7: missing values: 2

NUMCLIV_$: Number of children of partner lived with respondent

NOT INCLUDED IN SURVEY

Union	Number of unions	NUMCHP
1	11715	1: 429 2: 278 3: 103 4: 47 5: 12 6: 4 7: 1 9: 1 11: 1
2	2269	1: 223 2: 198 3: 94 4: 39 5: 13 6: 5 7: 1
3	500	1: 50 2: 55 3: 16 4: 7 5: 2
4	118	1: 11 2: 8 3: 5
5	33	1: 3 2: 2 3: 1 5: 1
6	9	1: 1
7	3	

6. Part Birth histories (biological kids)

KID_$: Indicator of child order

used:

Child order	number of children
1	9472
2	7312
3	3306
4	1219
5	448
6	177
7	76
8	39
9	17
10	9
11	7
12	4
13	3
14	2
15	2

KID_Y$: Year of birth of child

KID_Y1 missing values: 18
KID_Y2 missing values: 9
KID_Y3 missing values: 7
KID_Y4 missing values: 5
KID_Y5 missing values: 4
KID_Y6 missing values: 2
KID_Y7 missing values: 1
KID_Y8 missing values: 1

KID_M$: Month of birth of child

KID_M1 missing values: 53
KID_M2 missing values: 62
KID_M3 missing values: 46
KID_M4 missing values: 28
KID_M5 missing values: 16
KID_M6 missing values: 7
KID_M7 missing values: 6
KID_M8 missing values: 5
KID_M9 missing values: 1
KID_M10 missing values: 1
KID_M11 missing values: 1
KID_M12 missing values: 2
KID_M13 missing values: 1

IKID_M$: Month of birth of child
and imputed months
according to manual page 4 (random)

➔ Imputation if months is unknown and year is known

IKID_M1 missing values: 18
IKID_M2 missing values: 9
IKID_M3 missing values: 7
IKID_M4 missing values: 5
IKID_M5 missing values: 4
IKID_M6 missing values: 2
IKID_M7 missing values: 1
IKID_M8 missing values: 1

KID_S$: Sex of child

KID_S1 missing cases: 377
KID_S2 missing cases: 203
KID_S3 missing cases: 115
KID_S4 missing cases: 57
KID_S5 missing cases: 23
KID_S6 missing cases: 15
KID_S7 missing cases: 7
KID_S8 missing cases: 5
KID_S9 missing cases: 4
KID_S10 missing cases: 2
KID_S11 missing cases: 2
KID_S12 missing cases: 2
KID_S13 missing cases: 1

Child order	number of children	male	female
1	9472	4735	4360
2	7312	3534	3575
3	3306	1645	1546
4	1219	597	565
5	448	222	203
6	177	79	83
7	76	38	31
8	39	19	15
9	17	6	7
10	9	3	4
11	7	2	3
12	4	2	2
13	3	2	1
14	2	2	1
15	2	1	1

KID_D$: Death of child

QUESTIONS ABOUT THE DEATH OF A CHILD ARE NOT AVAILABLE IN THE SURVEY

2)Whether children have died is very incomplete. Parents were asked when providing histories whether children they have given birth to had died, but they weren't asked in subsequent waves about those children. So I know if a birth reported in 1992 died before 1992. I don't know if a child reported as still alive and

who was not part of the sample died after that. There was so much missing information I decided not to include this variable.

UNAVAILABLE IN SURVEY

KID_DY$: Year of death of child

UNAVAILABLE IN SURVEY

KID_DM$: Month of death of child

UNAVAILABLE IN SURVEY

IKID_DM$: Month of death of child
and imputed months

UNAVAILABLE IN SURVEY

KID_L$: Child left home

3.)Whether children have left home is difficult to measure. Parents were asked when they stopped living with children - not when children left home. For men, this is often the date the partnership ended, not the date the child left home.

LEAVING HOME QUESTIONS ARE NOT AVAILABLE IN THE SURVEY

KID_LY$: Year child left home

UNAVAILABLE IN SURVEY

KID_LM$: Month child left home

UNAVAILABLE IN SURVEY

IKID_LM$: Month of death of child
and imputed months

UNAVAILABLE IN SURVEY

---

7. Part Education

INSCHOOL: Currently studying at the time of interview

studying: 967
Missing cases: 16

EDU_COU: Highest level of education, country specific

TRANSFORMATION:
Recode 8260xx➔8261xx (Code 1 for UK BHPS)
826099➔.a

Missing values: 294
Proxy respondent: 902

The country specific codes include:
* a 3-digit country prefix(826)
* a 1-digit survey code (UK BHPS=1) and
* a 2-digit country specific code for level of education (1-13 levels of education)

ISCED_7: Highest level of education
Achieved according to ISCED 1997

Definition: ISCED_7=1 (ISCED 0+1) code 12+13
ISCED_7=2 code 11
ISCED_7=3 code 6-10
ISCED_7=5 code 2-5
ISCED_7=6 code 1

Missing cases: 1196

Harmonized:

ISCED	Number
0+1	2475
2	93
3	5045
4	0
5	5317
6	413

EDU_3: Highest level of education ISCED
Collapsed into 3 categories

High: ISCED_7=5+6
Medium: ISCED_7=3
Low: ISCED_7=1+2

Level	Number
High	5730
medium	5045
low	2568
missing cases	1196

NOT AVAILABLE IN THE SURVEY ARE:

EDU_Y: Year highest level of education achieved

EDU_M: Month highest level of education achieved

IMPUTATION RULES:

```
gen birthcm=(12*BORN_Y)+IBORN_M;
** Assume start school at age 5 but based on age 31 August;
** so those who are 4 on 21 August but 5 on 1 Sept should wait a year, I
think;
gen begschoolcm=.;
local i=1;
local q=68;
while `i'<9 {;
        replace begschoolcm=birthcm+ `q' if IBORN_M== `i';
local i= `i'+1;
local q= `q'-1;
};

local i=9;
local q=72;
while `i'<=12 {;
        replace begschoolcm=birthcm+ `q' if IBORN_M== `i';
local i= `i'+1;
local q= `q'-1;
};
** This assumes leave school at 15 -- 10 years after entry age:5;
gen educmonths = (10*12)-2 if EDU_COU==826012 | EDU_COU==826013;

** Leave school at 16;
replace educmonths = (11*12)-2 if EDU_COU<=826011 & EDU_COU>=826007;

** Leaves school at 18;
replace educmonths = (13*12)-2 if EDU_COU==826006;

** Leaves school at 21;
replace educmonths = (16*12)-2 if EDU_COU<=826005 & EDU_COU>=826002;

** Leaves school at 23;
replace educmonths = (18*12)-2 if EDU_COU==826001;

replace educmonths =. if ISCED_7==.a;

gen lvschoolcm=begschoolcm+educmonths;

gen EDU_Y=.c;
gen EDU_M=.c;

gen int IEDU_Y=1900+(lvschoolcm/12);
gen int IEDU_M=(lvschoolcm-(12*IEDU_Y));
```

replace IEDU_Y=.b if INSCHOOL==1;
replace IEDU_Y=.a if ISCED_7==.a;

replace IEDU_M=.b if INSCHOOL==1;
replace IEDU_M=.a if ISCED_7==.a;

drop educmonths lvschoolcm birthcm begschoolcm;

IEDU_Y: Year highest level education achieved and imputed year
➔.a if imputed year >2006
Missing (.a) 1246
Filter: .b 892

IEDU_M: Month highest education achieved and imputed month

IMPUTATION: JULY for all
Missing (.a) 1196
Filter: .b 892

---

8. Part Background variables (ethnicity, nationality etc.)

NATIVE: Born in country

TRANSFORMATION: Code 0➔ code 2 (born elsewhere)

Born in country: 13453, 464 missing cases
Born elsewhere: 622

ETHNOS: Ethnicity/nationality

Country specific variable (826+1+code)

Missing values: 72

BIRTH_COU: Country of birth

Country specific variable (826+1+code)

Missing values: 122

4.)Country of birth was only coded for people who were not born in the UK. I have kept this convention and coded native born individuals as not applicable.
5.)Note that for BIRTH_COU there is a value that is not labeled (8260007); this is in the original data and is probably entered incorrectly. ➔.a

MIG_Y: Year of migration

Missing values: 145

MIG_M: Month of migration

6.)There is no information on month of migration for anyone so I have not imputed it but made IMIG_M missing for everyone.

NOT INCLUDED IN SURVEY

IMIG_M: Month of migration and imputed months

NOT INCLUDED IN SURVEY

9. Part Background variables (parental background)

SIS_NO: Number of sisters
NOT INCLUDED IN SURVEY

BRO_NO: Number of brothers
NOT INCLUDED IN SURVEY

SIBS: Total number of sibs
Missing values: 2016
0-17 sibs

SIS_DIED: Number of sisters that died
NOT INCLUDED IN SURVEY

BRO_DIED: Number of brothers that died
NOT INCLUDED IN SURVEY

ISCED_MO: Mother`s highest level of education

7)Information on parents' education is very crude and difficult to code into ISCED. I have kept the original codes and put the UK code in front (8261) so it is clear that is was coded differently from the other countries.

Country specific variable (826+1+code)

	Number
1 Never went to school	120
2 Left school no qualifications	6320
3 Left school some qualifications	2843
4 Further educational qualifications	1815
5 University/higher degree	502
.a	2939

ISCED_FA: Father`s highest level of education used:

	Number
1 Never went to school	115
2 Left school no qualifications	5562
3 Left school some qualifications	1919
4 Further educational qualifications	2880
5 University/higher degree	763
.a	3300

ISCED_MO/ ISCED_FA are different from the other countries with the ISCED 1-7 codes!!!

EDU3_MO: Highest level of education of mother
ISCED 1997, collapsed into 3 categories

Definition: 1 (high) if code 5
2 (medium) if code 3+4
3 (low) if code 1+2

Level	Number
High	502
medium	4658
low	6440
missing cases	2939

EDU3_FA: Highest level of education of father
ISCED 1997, collapsed into 3 categories

Definition: 1 (high) if code 5
2 (medium) if code 3+4
3 (low) if code 1+2

Level	Number
High	763
medium	4799
low	5677
missing cases	3300

WORK_MO: Mother`s occupation, when respondent was 15

Missing cases: 318

WORK_FA: Father`s occupation, when respondent was 15

Missing cases: 260

ISCO3_MO: Mother`s occupation, when respondent was 15
3 categories

Definition: according to manual page 7
* Group 1: High non manual: 1, 2, 3
* Group 2: Non manual: 4, 5, 0
* Group 3: Manual: 6,7,8,9

Level	Number
1	1182
2	2230
3	1565
.b	9244
.a	318

ISCO3_FA: Father`s occupation, when respondent was 15
3 categories

Definition: according to manual page 7
* Group 1: High non manual: 1, 2, 3
* Group 2: Non manual: 4, 5, 0
* Group 3: Manual: 6,7,8,9

Level	
1	2759
2	943
3	4851
.b	5726
.a	260

NATIVE_MO: Mother born in country
NOT INCLUDED IN SURVEY

NATIVE_FA: Father born in country
NOT INCLUDED IN SURVEY

BIRTHCO_MO: Mother`s country of origin

Country specific variable (826)

NOT INCLUDED IN SURVEY

BIRTHCO_FA: Father`s country of origin

Country specific variable (826)

NOT INCLUDED IN SURVEY

PARDIVEV: Parents ever divorced/separated
NOT INCLUDED IN SURVEY

PARDIV_15: Parents divorced before age of 15

8)The variable PARDIV_15 is based on a question about whether the respondent lived with both natural parents at age 16.

Missing cases: 2047

Yes: 2414

10. Part Background variables (region, size of location)

REGION: Country region at time of interview

Country specific variable (826 +1 +code)

TRANSFORMATION: Recode 8260xx➔8261xx (Code 1 for UK BHPS)

missing cases: 20

SIZE: Size of place of residence at time of interview
NOT INCLUDED IN SURVEY

ISIZE: Size of place of residence at time of interview

Standardized code

SIZE_15: Size of place of residence at age 15

TRANSFORMATION:
Recode 8260xx➔8261xx (Code 1 for UK BHPS)

9)SIZE_15 is based on a question about your childhood up to age 15, not at exact age 15 as in the other surveys.

Missing values: 2014

ISIZE_15: Size of place of residence at age 15

Standardized code

11. Part Other background variables

RELIGION: Religious affiliation at time of interview

Country specific variable (826+1 +code)

Missing cases: 1673

IRELIGION: Religious affiliation at time of interview

Standardized code

ADOPT: Number of adopted children of respondent NOT INCLUDED IN SURVEY

FOSTER: Number of foster children of respondent
NOT INCLUDED IN SURVEY

STEP: Number of stepchildren of respondent
NOT INCLUDED IN SURVEY

---

12. Part Weights

HHWGT: Household weight

PERSWGT: Personal weight

KISHWGT: Kishweight – NOT INCLUDED IN SURVEY