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

HARMONIZED HISTORIES Germany Pairfam Data (13891 respondents)

Karolin Kubisch
Max Planck Institute for Demographic Research Rostock

Michaela Kreyenfeld
Max Planck Institute for Demographic Research Rostock

September 2015
Updated October 2015

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.
All problem cases as well as the treatment of these cases are described in detail. At the end of each module a summary of the main findings is displayed (in red).

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

Source: Pairfam („Panel Analysis of Intimate Relationships and Family Dynamics"), Release 6.0

Interview dates: Panel data 2008-2014

October 2015: Leaving home histories included

For Germany, we include data from the German Family Panel (pairfam) into the Harmonized Histories. We include data from the German Family Panel instead of data from the Generations and Gender Study (GGS), because data evaluations have shown that the partnership and fertility histories are rather unreliable in the German GGS (Kreyenfeld, Hornung, & Kubisch, 2012).The drawback of including data from the German Family Panel is that it follows a cohort approach and includes only respondents of the birth cohorts 1971-73, 1981-83, 1991-93. Also note that German Family Panel includes an oversample of East Germans, so that descriptive analysis need to be weighted.

Kreyenfeld, M., Hornung, A., & Kubisch, K. (2012). The German Generations and Gender Survey: Some critical reflections on the validity of fertility histories. *Comparative Population Studies*, *38*(1), 3-28.

PAIRFAM - The German Family Panel

What is pairfam?
The German Family Panel pairfam ("Panel Analysis of Intimate Relationships and Family Dynamics") is a multi-disciplinary, longitudinal study for researching partnership and family dynamics in Germany. Essential elements of the project are to provide user-friendly prepared data together with a comprehensive documentation and to release these data as scientific use file for the (inter)national and interdisciplinary scientific community.

Design:
The survey data is collected annually on a national level. The initial sample comprises more than 12,000 randomly selected individuals from the three birth cohorts 1971-73, 1981-83 and 1991-93. The multi-actor survey additionally includes the current partners of these young adults, up to three (step)parents and children at the age of 8 to 15 years. The resulting data base opens up new potentials for the analysis of partnership and family relations as they develop over time.

Study Themes:
The pairfam study focuses on the formation and the development of partnerships, processes of starting and expanding families, parenting and child development, and intergenerational relationships. The questions also cover various aspects from other areas of life and provide extensive socio-demographic information. Most of the instruments are implemented in every wave of the survey (core modules), some instruments are either used alternating in every second or third wave (in-depth modules) or only once (special modules). TNS Infratest Social Research in Munich is the responsible survey agency that organizes the fieldwork and data collection.

Organizational structure:
The German Family Panel pairfam, launched in 2008, is a cooperation between Chemnitz University of Technology, the University of Bremen and the Ludwig Maximilian University of Munich. It is funded by the German Research Foundation (DFG), since 2010 it has been part of the DFG long-term program. Pairfam is accredited as a research data center by the German Data Forum (RatSWD).

(downloaded from http://www.gesis.org/en/services/data-analysis/survey-data/pairfam/ on 02.07.2015)

Additional Notes

In 2009, an additional subsample of about 1,400 respondents was drawn. This sample (called "DemoDiff") includes East German respondents of the cohorts 1971-73 and 1981-83. The German data that was generated for the Harmonized Histories includes, in addition to the original pairfam data, also this East German subsample. Thus, weights need to be used

for descriptive analyses. The weight provided in the Harmonized Histories is a combined design and post-stratification weight.

The union, marriage and fertility histories for the Harmonized Histories were taken from the files “biochild.dta” and “biopart.dta” that is generated by the pairfam-group. Missing starting and ending states of partnerships were already imputed in this data (by the random number operator of STATA). (for more information, **see http://www.pairfam.de/en)**

For every variable the information is given from which datafile the variable was built. ($SOURCE)

Part Basic Information

RESPID: ID number to be assigned at merging LEAVE BLANK

ARID: ID number from raw data (original ID number) used: id
13891 respondents

"$SOURCE\anchor1.dta" and "$SOURCE\anchor1_DD.dta"

COUNTRY: Country and survey
Harmonized: code: 2762: Germany Pairfam
no missing cases

"$SOURCE\anchor1.dta" and "$SOURCE\anchor1_DD.dta"

MONTH_S: Month of survey used:intm

"$SOURCE\anchor1_DD"+"$SOURCE\anchor2.dta"+"$SOURCE\anchor3.dta"+"$SOURCE\anchor4.dta"+"$SOURCE\anchor5.dta"+"$SOURCE\anchor6.dta"

IMONTH_S: Month of survey, including imputed dates
According to manual page 4: random variables

"$SOURCE\anchor1_DD"+"$SOURCE\anchor2.dta"+"$SOURCE\anchor3.dta"+"$SOURCE\anchor4.dta"+"$SOURCE\anchor5.dta"+"$SOURCE\anchor6.dta"

YEAR_S: Year of survey used:inty
2008-2014
No missing cases

"$SOURCE\anchor1_DD"+"$SOURCE\anchor2.dta"+"$SOURCE\anchor3.dta"+"$SOURCE\anchor4.dta"+"$SOURCE\anchor5.dta"+"$SOURCE\anchor6.dta"

SEX: Sex of the respondent used:sex_gen
2 missing cases
Sex structure of the respondents:
Male: 6760 and Female: 7129

"$SOURCE\anchor1.dta" and "$SOURCE\anchor1_DD.dta"

BORN_Y: Year of birth of respondent used: doby
1970-1974, 1980-1983, 1991-1993
no missing cases

"$SOURCE\anchor1.dta" and "$SOURCE\anchor1_DD.dta"

BORN_M: Month of birth of respondent used: dobm_gen

"$SOURCE\anchor1.dta" and "$SOURCE\anchor1_DD.dta"

IBORN_M: Month of birth of respondent
including imputed months
Harmonized: random variable between 1-12

"$SOURCE\anchor1.dta" and "$SOURCE\anchor1_DD.dta"

---

2. Part LEAVING HOME

The question for the time of the first leaving of the parental household was asked twice in pairfam – in wave 1 and wave 3. Due to methodical reasons the information gathered in wave 3 has been mainly used. Merely missing values have been supplemented by information of wave 1. If the year was known, but only ambiguous seasonal information or no information on the month was available, the date has been randomly imputed for the respective time span. To include the moves happening after wave 3 we used the data on the household type and the data on the last change of residence. If they were not living anymore in a parental household and one change of residence occurred, the time of the change of residence has been defined as time of leaving home. Again, exact dates have been imputed if no exact information on the time was available.

LEAVE_1: Indicator of whether left home

LEAVE_1 0: 1778 / 1: 9219 .a 2894

LEAVE_Y1: Year of first time leaving home used: igr9y,
rtr27y, ehc21, ehc29, hc4h1, resbeg
"$SOURCE\anchor1.dta"+"$SOURCE\anchor1_DD.dta"+"$SOURCE\anchor3.dta"+"$SOURCE\anchor4.dta"+"$SOURCE\anchor5.dta"+"$SOURCE\anchor6.dta"+"$SOURCE\biomob_ehc.dta"

.a 3280

LEAVE_M1: Month of first time leaving home used: igr9m,
rtr27m, ehc21, ehc29, hc4h1, resbeg

"$SOURCE\anchor1.dta"+"$SOURCE\anchor1_DD.dta"+"$SOURCE\anchor3.dta"+"$SOURCE\anchor4.dta"+"$SOURCE\anchor5.dta"+"$SOURCE\anchor6.dta"+"$SOURCE\biomob_ehc.dta"

.a 3971

ILEAVE_M1: Month of first time leaving home and imputed months
used: LEAVE_M1, inty (wave 1-6), intm (wave 1-6), ehc21

"$SOURCE\anchor1.dta"+"$SOURCE\anchor1_DD.dta"+"$SOURCE\anchor3.dta"+"$SOURCE\anchor4.dta"+"$SOURCE\anchor5.dta"+"$SOURCE\anchor6.dta"+"$SOURCE\biomob_ehc.dta"

.a 2894

---

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

UNINUM: Total number of unions used: UNION_1 to _7

Syntax:
```
forvalues x=1/7 {
replace UNINUM=UNINUM+1 if UNION_`x'>0
            }
```
0: 5398
1: 6506
2: 1518
3: 337
4: 54
5: 12
6: 3
7: 2
.a:61

UNION_$: UNION order used: cohbeg

UNION_1: 8172
UNION_2: 1893
UNION_3: 402
UNION_4: 68
UNION_5: 14
UNION_6: 5
UNION_7: 2

missing cases:
UNION_1: 321
UNION_2: 94
UNION_3: 67
UNION_4: 64
UNION_5: 64
UNION_6: 61
UNION_7: 61

use "$SOURCE\biopart.dta"

UNION_Y$: Year of start union used: cohbeg

Filter: UNION_Yx=.b if UNION_x==0

UNION_Y1 missing values: 229
UNION_Y2 missing values: 94
UNION_Y3 missing values: 67
UNION_Y4 missing values: 64
UNION_Y5 missing values: 64
UNION_Y6 missing values: 61
UNION_Y7 missing values: 61

"$SOURCE\biopart.dta"

UNION_M$: Month of start UNION used: cohbeg

Filter: UNION_Mx=.b if UNION_x==0

UNION_M1 missing values: 229
UNION_M2 missing values: 94
UNION_M3 missing values: 67
UNION_M4 missing values: 64
UNION_M5 missing values: 64
UNION_M6 missing values: 61
UNION_M7 missing values: 61

"$SOURCE\biopart.dta"

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

Filter: IUNION_Mx=.b if UNION_x==0

(because of error corrections)

IUNION_M1 missing values: 61
IUNION_M2 missing values: 61
IUNION_M3 missing values: 61
IUNION_M4 missing values: 61
IUNION_M5 missing values: 61
IUNION_M6 missing values: 61
IUNION_M7 missing values: 61

"$SOURCE\biopart.dta"

SEP_$: Dissolution of UNION used: cohend

Filter: SEP_x=.b if UNION_x==0
* in case of current partner: no separation

SEP_1 missing cases: 192
SEP_2 missing cases: 95
SEP_3 missing cases: 69
SEP_4 missing cases: 64
SEP_5 missing cases: 64

SEP_6 missing cases: 61
SEP_7 missing cases: 61

Order of Union	Number of unions	number of separations	death of partner
1	8172	2943	33
2	1893	662	11
3	402	142	2
4	68	33	
5	14	5	
6	5	3	
7	2		

"$SOURCE\biopart.dta"

SEP_Y$: Year of end of UNION used: cohend

Filter: SEP_Yx=.b if UNION_x==0
SEP_Yx=.b if SEP_x==0

SEP_Y1 missing cases: 192
SEP_Y2 missing cases: 95
SEP_Y3 missing cases: 69
SEP_Y4 missing cases: 64
SEP_Y5 missing cases: 64
SEP_Y6 missing cases: 61
SEP_Y7 missing cases: 61

"$SOURCE\biopart.dta"

SEP_M$: Month of end of UNION used: cohend

Filter: SEP_Mx=.b if UNION_x==0
SEP_Mx=.b if SEP_x==0

SEP_M1 missing cases: 192
SEP_M2 missing cases: 95
SEP_M3 missing cases: 69
SEP_M4 missing cases: 64
SEP_M5 missing cases: 64
SEP_M6 missing cases: 61
SEP_M7 missing cases: 61

"$SOURCE\biopart.dta"

ISEP_M$: Month of end of UNION and imputed months according to manual page 4 (random) used: SEP_M$

Filter: ISEP_Mx=.b if UNION_x==0
ISEP_Mx=.b if SEP_x==0

(because of error corrections)

```
ISEP_M1  missing values: 61
ISEP_M2  missing values: 61
ISEP_M3  missing values: 61
ISEP_M4  missing values: 61
ISEP_M5  missing values: 61
ISEP_M6  missing values: 61
ISEP_M7  missing values: 61

"$SOURCE\biopart.dta"
```

---

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

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

Filter: MARR_x=.b if UNION_x==0

```
MARR_1 missing values: 84
MARR_2 missing values: 62
MARR_3 missing values: 61
```

➔Only available untilup to marriage 3

Order of Union	Number of unions	number of marriages
1	8172	5288
2	1893	252
3	402	8
4	68	
5	14	
6	5	
7	2	

"$SOURCE\biopart.dta"

MARR_Y$: Year of marriage used: marbeg

Filter: MARR_Yx=.b if UNION_x==0
MARR_Yx=.b if MARR_x==0

```
MARR_Y1 missing values: 84
MARR_Y2 missing values: 62
MARR_Y3 missing values: 61

"$SOURCE\biopart.dta"
```

MARR_M$: Month of marriage used: marbeg

Filter: MARR_Mx=.b if UNION_x==0
MARR_Mx=.b if MARR_x==0

```
MARR_M1 missing values: 84
```

MARR_M2 missing values: 62
MARR_M3 missing values: 61

"$SOURCE\biopart.dta"

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

Filter: IMARR_Mx=.b if UNION_x==0
IMARR_Mx=.b if MARR_x==0

(because of error corrections)
IMARR_M1 missing values: 61
IMARR_M2 missing values: 61
IMARR_M3 missing values: 61

"$SOURCE\biopart.dta"

DIV_$: Indicator of whether divorce occurred — used: marend (only histories)

Filter: DIV_x=.b if UNION_x==0
DIV_x=.b if MARR_x==0
DIV_x=.d if marend==-66

DIV_1 missing values: 99
DIV_2 missing values: 65
DIV_3 missing values: 61

➔Only available up to divorce 3

Order of Union	Number of unions	number of marriages	number of divorces
1	8172	5288	820
2	1893	252	27
3	402	8	2
4	68		
5	14		
6	5		
7	2		

"$SOURCE\biopart.dta"

DIV_Y$: Year of divorce — used: marend

Filter: DIV_Yx=.b if UNION_x==0
DIV_Yx=.b if MARR_x==0
DIV_Yx=.b if DIV_X==0 or .d

DIV_Y1 missing values: 99
DIV_Y2 missing values: 65
DIV_Y3 missing values: 61

"$SOURCE\biopart.dta"

DIV_M$: Month of divorce used: marend

Filter: DIV_Mx=.b if UNION_x==0
DIV_Mx=.b if MARR_x==0
DIV_Mx=.b if DIV_x==0 or .d

DIV_M1 missing values: 99
DIV_M2 missing values: 65
DIV_M3 missing values: 61

"$SOURCE\biopart.dta"

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

(because of error corrections)
IDIV_M1 missing values: 61
IDIV_M2 missing values: 61
IDIV_M3 missing values: 61

Filter: IDIV_Mx=.b if UNION_x==0
IDIV_Mx=.b if MARR_x==0
IDIV_Mx=.b if DIV_x==0 or .d

"$SOURCE\biopart.dta"

---

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

SEXP_$: Partner`s sex used: ahg4_2, ahg4_1, a352a

Filter: SEXP_x=.b if UNION_x==0

Unavailable in survey

YEARBIRP_$: Year of birth of partner Used: ahg6y_2 and a336y

Filter: YEARBIRP_x=.b if UNION_x==0

Unavailable in survey

MONBIRP_$: Month of birth of partner used: ahg6m_2 and a336m

Filter: MONBIRP_x=.b if UNION_x==0

Unavailable in survey

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

Filter: IMONBIRP_x=.b if UNION_x==0

Unavailable in survey

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

Filter: NUMCHP_$=.b if UNION_X==0

Unavailable in survey

NUMCLIV_$: Number of children of partner lived with respondent

Unavailable in survey

---

6. Part Birth histories (biological kids)

KID_$: Indicator of child order (stausk==1)

Child order	number of children	missings
1	5685	49
2	3532	24
3	1085	9
4	273	6
5	89	4
6	25	3
7	10	2
8	6	2
9	3	2
10	2	2

"$SOURCE\biochild.dta"

KID_Y$: Year of birth of child used: dobk

Filter: KID_Yx=.b if KID_x==0

KID_Y1 missing values: 46
KID_Y2 missing values: 8
KID_Y3 missing values: 5
KID_Y4 missing values: 3
KID_Y5 missing values: 3
KID_Y6 missing values: 3
KID_Y7 missing values: 2
KID_Y8 missing values: 2
KID_Y9 missing values: 2
KID_Y10 missing values: 2

"$SOURCE\biochild.dta"

KID_M$: Month of birth of child used: dobk

Filter: KID_Mx=.b if KID_x==0

KID_M1 missing values: 46
KID_M2 missing values: 8
KID_M3 missing values: 5
KID_M4 missing values: 3
KID_M5 missing values: 3
KID_M6 missing values: 3
KID_M7 missing values: 2
KID_M8 missing values: 2
KID_M9 missing values: 2
KID_M10 missing values: 2

"$SOURCE\biochild.dta"

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

Filter: IKID_M_x=.b if KID_x==0

(because of error corrections)
IKID_M1 missing values: 2
IKID_M2 missing values: 2
IKID_M3 missing values: 2
IKID_M4 missing values: 2
IKID_M5 missing values: 2
IKID_M6 missing values: 2
IKID_M7 missing values: 2
IKID_M8 missing values: 2
IKID_M9 missing values: 2
IKID_M10 missing values: 2

"$SOURCE\biochild.dta"

KID_S$: Sex of child used: sexk

Filter: KID_Sx=.b if KID_x==0

KID_S1 missing cases: 3
KID_S2 missing cases: 2
KID_S3 missing cases: 2
KID_S4 missing cases: 2
KID_S5 missing cases: 2
KID_S6 missing cases: 2
KID_S7 missing cases: 2
KID_S8 missing cases: 2
KID_S9 missing cases: 2
KID_S10 missing cases: 2

Child order	number of children	male	female
1	5685	2921	2807
2	3532	1803	1735
3	1085	587	501
4	273	140	134
5	89	48	42
6	25	13	13
7	10	6	4

8	6	4	2
9	3	3	0
10	2	1	1

"$SOURCE\biochild.dta"

KID_D$: Death of child used: dodk

Filter: KID_Dx=.b if KID_x==0

KID_D1 missing cases: 5
KID_D2 missing cases: 3
KID_D3 missing cases: 2
KID_D4 missing cases: 2
KID_D5 missing cases: 2
KID_D6 missing cases: 2
KID_D7 missing cases: 2
KID_D8 missing cases: 2
KID_D9 missing cases: 2
KID_D10 missing cases: 2

Child order	number of children	death
1	5685	10
2	3532	10
3	1085	1
4	273	1
5	89	
6	25	
7	10	
8	6	
9	3	
10	2	

"$SOURCE\biochild.dta"

KID_DY$: Year of death of child used: dodk

Filter: KID_DYx=.b if KID_x==0
KID_DYx=.b if KID_Dx==0

KID_DY1 missing cases: 5
KID_DY2 missing cases: 3
KID_DY3 missing cases: 2
KID_DY4 missing cases: 2
KID_DY5 missing cases: 2
KID_DY6 missing cases: 2
KID_DY7 missing cases: 2
KID_DY8 missing cases: 2
KID_DY9 missing cases: 2
KID_DY10 missing cases: 2

"$SOURCE\biochild.dta"

KID_DM$: Month of death of child used: dodk

Filter: KID_DMx=.b if KID_x==0
KID_DMx=.b if KID_Dx==0

KID_DM1 missing cases: 5
KID_DM2 missing cases: 3
KID_DM3 missing cases: 2
KID_DM4 missing cases: 2
KID_DM5 missing cases: 2
KID_DM6 missing cases: 2
KID_DM7 missing cases: 2
KID_DM8 missing cases: 2
KID_DM9 missing cases: 2
KID_DM10 missing cases: 2

"$SOURCE\biochild.dta"

IKID_DM$: Month of death of child and imputed months — used: KID_DM

according to manual page 4 (random)

Filter: IKID_DMx=.b if KID_x==0
IKID_DMx=.b if KID_Dx==0

(because of error corrections)
IKID_DM1 missing cases: 2
IKID_DM2 missing cases: 2
IKID_DM3 missing cases: 2
IKID_DM4 missing cases: 2
IKID_DM5 missing cases: 2
IKID_DM6 missing cases: 2
IKID_DM7 missing cases: 2
IKID_DM8 missing cases: 2
IKID_DM9 missing cases: 2
IKID_DM10 missing cases: 2

"$SOURCE\biochild.dta"

KID_L$: Child left home

Filter: KID_Lx=.b if KID_x==0

Unavailable in survey

KID_LY$: Year child left home

Filter: KID_LYx=.b if KID_x==0
KID_LYx=.b if KID_Lx==0

Unavailable in survey

KID_LM$: Month child left home

Filter: KID_LMx=.b if KID_x==0
KID_LMx=.b if KID_Lx==0

Unavailable in survey

IKID_LM$: Month of death of child and imputed months

according to manual page 4 (random variable)

Filter: IKID_LMx=.b if KID_x==0
IKID_LMx=.b if KID_Lx==0

Unavailable in survey

---

7. Part Education

INSCHOOL: Currently studying at the time of interview used: isced

Currently studying: 2958

Note that only respondents who were in school during all years if survey are coded here with 0.

```
"$SOURCE\anchor1_DD"+"$SOURCE\anchor2.dta"+"$SOURCE\anchor3.dta"+"$SOUR
CE\anchor4.dta"+"$SOURCE\anchor5.dta"+"$SOURCE\anchor6.dta"
```

EDU_COU: Highest level of education, country specific used: isced

Highest level of education across all survey years

Missing values: 17

Harmonized: these country specific codes include:
* a 3-digit country prefix(276)
* a 1-digit survey code (Pairfam Germany=2) and
* a 2-digit country specific code for level of education

```
"$SOURCE\anchor1_DD"+"$SOURCE\anchor2.dta"+"$SOURCE\anchor3.dta"+"$SOUR
CE\anchor4.dta"+"$SOURCE\anchor5.dta"+"$SOURCE\anchor6.dta"

by id: egen EDU_COU=max(isced)
```

ISCED_7: Highest level of education
Achieved according to ISCED 1997 used: EDU_COU

Definition:

```
replace ISCED_7=.b if EDU_COU==276200
replace ISCED_7=1 if EDU_COU==276201
replace ISCED_7=2 if EDU_COU==276202
replace ISCED_7=2 if EDU_COU==276203
replace ISCED_7=3 if EDU_COU==276204
replace ISCED_7=3 if EDU_COU==276205
replace ISCED_7=4 if EDU_COU==276206
replace ISCED_7=5 if EDU_COU==276207
replace ISCED_7=6 if EDU_COU==276208
replace ISCED_7=.a if ISCED_7==.
```

Missing cases: 17

Harmonized:

ISCED	Number
0+1	301
2	1158
3	5404
4	1110
5	2782
6	161

"$SOURCE\anchor1_DD"+"$SOURCE\anchor2.dta"+"$SOURCE\anchor3.dta"+"$SOURCE\anchor4.dta"+"$SOURCE\anchor5.dta"+"$SOURCE\anchor6.dta"

EDU_3: Highest level of education ISCED used: ISCED_7
Collapsed into 3 categories

Definition: High: ISCED_7=5 or 6
Medium: ISCED_7=3 or 4
Low: ISCED_7=1 or 2

Level	Number
High	2943
medium	6514
low	1459
missing cases	17

"$SOURCE\anchor1_DD"+"$SOURCE\anchor2.dta"+"$SOURCE\anchor3.dta"+"$SOURCE\anchor4.dta"+"$SOURCE\anchor5.dta"+"$SOURCE\anchor6.dta"

EDU_Y: Year highest level of education achieved used: a150y

Unavailable in survey

EDU_M: Month highest level of education achieved used: a150m

Unavailable in survey

IEDU_Y: Year highest level education achieved and imputed year

Unavailable in survey

IEDU_M: Month highest education achieved and imputed month

Unavailable in survey

---

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

NATIVE: Born in country used: ethni

Born in country: 10329, 339 missing cases
Born elsewhere: 3223

"$SOURCE\anchor1.dta" and "$SOURCE\anchor1_DD.dta"

ETHNOS: Ethnicity/nationality used: ethni

Country specific variable (276+2+code)

339 missing cases

"$SOURCE\anchor1.dta" and "$SOURCE\anchor1_DD.dta"

BIRTH_COU: Country of birth used: mig4

Country specific variable (276+2+code)

7 missing cases

"$SOURCE\anchor1.dta" and "$SOURCE\anchor1_DD.dta"

MIG_Y: Year of migration used: mig5y

Missing cases: 35

Filter: MIG_Y=.b if a105==1

"$SOURCE\anchor1.dta" and "$SOURCE\anchor1_DD.dta"

MIG_M: Month of migration used: mig5m

Missing cases: 170+seasonal codes

Filter: MIG_M=.b if a105==1

"$SOURCE\anchor1.dta" and "$SOURCE\anchor1_DD.dta"

IMIG_M: Month of migration and imputed months used: MIG_M
according to manual page 4 (random)

"$SOURCE\anchor1.dta" and "$SOURCE\anchor1_DD.dta"

---

9. Part Background variables (parental background)

SIS_NO: Number of sisters

Unavailable in survey

BRO_NO: Number of brothers

Unavailable in survey

SIBS: Total number of sibs used: igr19

82 missing cases

"$SOURCE\anchor1.dta" and "$SOURCE\anchor1_DD.dta"

SIS_DIED: Number of sisters that died

Unavailable in survey

BRO_DIED: Number of brothers that died used: a5106a_b and a5106b_b

Unavailable in survey

ISCED_MO: Mother`s highest level of education used: misced

ISCED	Number
0+1	381
2	883
3	5396
4	520
5	1790
6	72
missing	4849

"$SOURCE\anchor1.dta" and "$SOURCE\anchor1_DD.dta"

ISCED_FA: Father`s highest level of education used: fisced

ISCED	Number
0+1	214
2	432
3	4928
4	390
5	2407
6	139
missing	5381

"$SOURCE\anchor1.dta" and "$SOURCE\anchor1_DD.dta"

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

Definition: 1 (high) if ISCED_MO=5 or 6
2 (medium) if ISCED_MO=3 or 4
3 (low) if ISCED_MO=1 or 2

Level	Number
High	1862
medium	5916
low	1264

missing cases	4849

"$SOURCE\anchor1.dta" and "$SOURCE\anchor1_DD.dta"

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

Definition: 1 (high) if ISCED_FA=5 or 6
2 (medium) if ISCED_FA=3 or 4
3 (low) if ISCED_FA=1 or 2

Level	Number
High	2546
medium	5318
low	646
missing cases	5381

"$SOURCE\anchor1.dta" and "$SOURCE\anchor1_DD.dta"

WORK_MO: Mother`s occupation, when respondent was 15
Country codes

Unavailable in survey

WORK_FA: Father`s occupation, when respondent was 15
Country codes

Unavailable in survey

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

Unavailable in survey

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

Unavailable in survey

NATIVE_MO: Mother born in country

No: 2817 used: mcob

"$SOURCE\anchor1.dta" and "$SOURCE\anchor1_DD.dta"

NATIVE_FA: Father born in country

No: 3157 used: fcob

"$SOURCE\anchor1.dta" and "$SOURCE\anchor1_DD.dta"

BIRTHCO_MO: Mother`s country of origin, country specific
Used: mcob

BIRTHCO_MO missing cases: 118

"$SOURCE\anchor1.dta" and "$SOURCE\anchor1_DD.dta"

BIRTHCO_FA: Father`s country of origin, country specific
used: fcob

BIRTHCO_FA missing cases: 312

"$SOURCE\anchor1.dta" and "$SOURCE\anchor1_DD.dta"

PARDIVEV: Parents ever divorced/separated

Unavailable in survey

PARDIV_15: Parents divorced before age of 15

Unavailable in survey

---

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

REGION: Country region at time of interview

Country specific variable (276+2 +code) used: hc1p1i2

37 missing cases

"$SOURCE\anchor1.dta" and "$SOURCE\anchor1_DD.dta"

SIZE: Size of place of residence at time of interview

Country specific variable (276+2 +code) used: gkpol

No missing cases

"$SOURCE\anchor1.dta" and "$SOURCE\anchor1_DD.dta"

ISIZE: Size of place of residence at time of interview

Standardized code

SIZE_15: Size of place of residence at age 15

Unavailable in survey

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 (276+1 +code) used: sd30

57 missing values

"$SOURCE\anchor1.dta" and "$SOURCE\anchor1_DD.dta"

IRELIGION: Religious affiliation at time of interview

Standardized code

ADOPT: Number of adopted children of respondent

"$SOURCE\biochild.dta"

FOSTER: Number of foster children of respondent

"$SOURCE\biochild.dta"

STEP: Number of stepchildren of respondent

Number of children	Adopt	Foster	Step
1	6	5	29
2	5	5	33
3	1	4	33
4	1	2	24
5	2	2	35
6	5	1	37
7		1	1
8			4
9			
10		1	4
11			
12	1	1	5
13			1
15			1
18			2
24			1

"$SOURCE\biochild.dta"

---

12. Part Weights

HHWGT: Household weight Unavailable in survey

PERSWGT: Personal weight – used: d1ca1weight

"$SOURCE\anchor1.dta" and "$SOURCE\anchor1_DD.dta"

KISHWGT: Kishweight Unavailable in survey

Corrections made in data file

1)replace all partnership variables with .a if differences between UNION_1=0 and Part_max>0 or UNION_1==0 and Part_MAX==-7

Part_max= maximal number of partners over all waves

```
g PARTNER=.
     replace PARTNER=0 if relstat==1 | relstat==2
     tab PARTNER
     recode PARTNER .=1
     replace PARTNER=-7 if relstat==-7
     by id: egen PART_max=max(PARTNER)

replace  UNION_1=.a if UNION_1==0 & PART_max>0
(67 real changes made, 67 to missing)
```

2)replace all partnership variables with .a if there is a flag between flag_6 and flag_11,

flags were created by pairfam team:
flag6: Inconsistency beginning current and end previous cohabitation (current partner)
flag7: Inconsistency beginning current and end previous cohabition (different partners)
flag8: Inconsistency beginning current and end previous marriage (different partners)
flag9: Inconsistency divorced/widowed and no partner before current relationship
flag10: Inconsistency divorce from a partner to whom never married
flag11: Inconsistency separation through death/divorce current spouse

```
replace  UNION_`i'=.a if FLAG==1

(53 real changes made, 53 to missing)
(61 real changes made, 61 to missing)
(61 real changes made, 61 to missing)
(61 real changes made, 61 to missing)
(61 real changes made, 61 to missing)
```

```
(61 real changes made, 61 to missing)
(61 real changes made, 61 to missing)

replace  UNION_Y`i'=.a if FLAG==1

(56 real changes made, 56 to missing)
(61 real changes made, 61 to missing)
(61 real changes made, 61 to missing)
(61 real changes made, 61 to missing)
(61 real changes made, 61 to missing)
(61 real changes made, 61 to missing)
(61 real changes made, 61 to missing)
…

3)replace all fertility variables with .a if prepared
variable KIDS==-7

by id: egen KIDS=max(nkidsbio)

replace  KID_`i'=.a if KIDS==-7

 (2 real changes made, 2 to missing)
(2 real changes made, 2 to missing)
(2 real changes made, 2 to missing)
(2 real changes made, 2 to missing)
(2 real changes made, 2 to missing)
(2 real changes made, 2 to missing)
(2 real changes made, 2 to missing)
(2 real changes made, 2 to missing)
(2 real changes made, 2 to missing)

4)replace fertility variables with .a if differences
between order of kid and prepared variable KIDS

replace  KID_`i'=.a if KIDS>=`i' & KID_`i'==0

(3 real changes made, 3 to missing)
(16 real changes made, 16 to missing)
(4 real changes made, 4 to missing)
(3 real changes made, 3 to missing)
(1 real change made, 1 to missing)
(0 real changes made)
(0 real changes made)
(0 real changes made)
(0 real changes made)
(0 real changes made)
```