

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

**HARMONIZED HISTORIES USA NSFG 1995 (10847 female
respondents)**

Karolin Kubisch
Max Planck Institute for Demographic Research Rostock

Katherine Michelmore
Cornell University

Oktober 2010
Updated 2014

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.

Missing values are coded:

.a unknown
.b does not apply
.c unavailable in survey

Source: NSFG 1995

Interview dates NSFG 1995: From January to October 1995

In the 1995 file only females were interviewed.

June 2014: Corrections in the variables to leaving home histories of children (KID_L, KID_LY, KID_LM)

1. Part Basic Information

RESPID: ID number to be assigned at merging LEAVE BLANK

ARID: ID number from raw data (original ID number) used: CASEID
10847 respondents

COUNTRY: Country and survey
COUNTRY: code: 8401: USA NSFG 1995
no missing cases

MONTH_S: Month of survey used: A_DOI

codes: 1-10
no missing cases

IMONTH_S: Month of survey, including imputed dates

YEAR_S: Year of survey used: A_DOI
YEAR_S: 1995
no missing cases

SEX: Sex of the respondent
No missing cases
Sex structure of the respondents:
Female: 10847

BORN_Y: Year of birth of respondent used: BDAYCENM
1950-1980: no missing cases

BORN_M: Month of birth of respondent used: BDAYCENM
no missing cases

IBORN_M: Month of birth of respondent

2. Part LEAVING HOME

LEAVE_1: Indicator of whether "left home" used: FSTONOWN

LEAVE_1: 0: 1725 / 1: 9097
25 missing cases

LEAVE_Y1: Year of first time leaving home used: FSTONOWN

1953-1995, 25 missing cases
.b 1725

LEAVE_M1: Month of first time leaving home used: FSTONOWN
Missing cases: 25
.b 1725

ILEAVE_M1: Month of first time leaving home
and imputed months:
including imputed months randomly variable between 1-12

Filter: .b 1725

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

Marriages/Unions:

There were a couple of issues with the marriage/union histories. The NSFG computes variables such as marriage start and end dates, but the end dates include separation and divorce dates. So for couples that separate before they divorce, I only know the separation date, not the divorce date. There are raw variables that code for the actual divorce dates, but some of the values are spurious--divorce dates occur before the actual marriage occurs, so I thought it best to use the cleaned variable. This can be changed as desired. I have indicated which dates have been imputed with the imputation variables described in the manual (e.g. IDIV_M). Individuals with imputed dates for marital dissolution do not have values for DIV_M, the manual indicates that this variable only contain respondent answers, but they are included in the IDIV_M variable, which contains information on divorce dates regardless of whether the answer was imputed.

Calculation:

The biggest challenge for creating the union start and end dates was synthesizing the separate information about cohabiting partners and marital partners. The NSFG has three sets of variables for start and end dates of unions for respondents. There are variables for previous cohabiting partners who they did not marry, marriage start and end dates for their first five marriages, and pre-marital cohabitation start dates for those marriages. I calculate the marriage and union histories by taking the earliest start date mentioned in any of the three sets of variables-- cohabitation start date, marriage start date, and per-marital cohabitation start date. Once I find the earliest date, I map on the rest of the information about the partner- the birth dates and the number of children they have had with previous mates. Once I gather all of the information from the first union/marriage, I set the raw start and end date variables equal to missing so they are no longer considered in future unions. I continue in this manner until all of the cohabitations and marriages have

been enumerated. In 1995 the NSFG asks about 5 marriages and 7 cohabiting partners, and the 2007 NSFG asks about 5 marriages and 5 cohabiting partners for women, and 4 marriages and 3 previous partners for men.

Finally, the NSFG does ask all respondents how many times they have been married and how many cohabiting partners they have ever had, which is how I ascertained the total number of unions, UNINUM. It's possible, however, that the NSFG would not have collected information about all of these unions, so that the value for UNINUM may exceed the number of unions for which we have information on start and end dates, and birth dates. This is especially true for the 2007 males, who are not asked about any of their previous cohabitation partners aside from their first partner and any partners in the last 12 months. I thought it best to utilize all of the data we have though, so I let the UNINUM value include all previous unions, not just those with explicit information on start and end dates in the NSFG.

UNINUM: Total number of unions used: TIMESMAR, HMOTHMEN

Syntax:
Egen UNINUM=rsum(TIMESMAT HMOTHMEN)

UNINUM:
0: 2784
1: 5558
2: 1765
3: 534
4: 137
5: 45
6: 13
7: 8
8: 1
9: 2

UNION_\$: UNION order used: UNINUM

Definition (UNION 1 to UNION x)

A union exists if the respondent reports at least x unions in the UNINUM variable.

UNION_1: 8063
UNION_2: 2505
UNION_3: 740
UNION_4: 206
UNION_5: 69
UNION_6: 24
UNION_7: 11
UNION_8: 3

UNION_9: 2

No missing cases

UNION_Y\$: Year of start union used: STRTOTHx MARDATx STRTOGHx

Filter: UNION_Yx=.b if UNION_x==0

Missing cases: UNION_Y1:5
Missing cases: UNION_Y2:11
Missing cases: UNION_Y3:21
Missing cases: UNION_Y4:11
Missing cases: UNION_Y5:7
Missing cases: UNION_Y6:3
Missing cases: UNION_Y7:3
Missing cases: UNION_Y8:1
Missing cases: UNION_Y9:1

UNION_M\$: Month of start UNION

Filter: UNION_Mx=.b if UNION_x==0

No missing values

IUNION_M\$: Month of start UNION
and imputed months

Filter: IUNION_Mx=.b if UNION_x==0

SEP_\$: Dissolution of UNION used: STOPOTHx MARDIS0x

Filter: SEP_x=.b if UNION_x==0

No missing cases

Order of Union	Number of unions	number of separations	death of partner
1	UNION 1: 8063	3771	
2	UNION 2: 2505	1206	
3	UNION 3: 740	364	
4	UNION 4: 206	106	
5	UNION 5: 69	35	
6	UNION 6: 24	12	
7	UNION 7: 11	4	
8	UNION 8: 3	2	
9	UNION 9: 2	0	

SEP_Y\$: Year of end of UNION

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

No missing cases

SEP_M\$: Month of end of UNION

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

No missing cases

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

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

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

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

used: MARDATx STRTOGHx

Filter: MARR_x=.b if UNION_x==0

No missing cases

Order of Union	Number of unions	number of marriages
1	UNION 1: 8063	6349
2	UNION 2: 2505	1540
3	UNION 3: 740	394
4	UNION 4: 206	94
5	UNION 5: 69	29
6	UNION 6: 24	13
7	UNION 7: 11	4
8	UNION 8: 3	1
9	UNION 9: 2	2

MARR_Y\$: Year of marriage

used: MARDATx

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

MARR_Y1 missing values: 4
MARR_Y2 missing values: 10
MARR_Y3 missing values: 22
MARR_Y4 missing values: 11
MARR_Y5 missing values: 7
MARR_Y6 missing values: 3
MARR_Y7 missing values: 3
MARR_Y8 missing values: 1
MARR_Y9 missing values: 1

MARR_M\$: Month of marriage

used: MARDATx

Filter: MARR_Mx=.b if UNION_x==0

MARR_Mx=.b if MARR_x==0

MARR_M1 missing values: 4
MARR_M2 missing values: 10
MARR_M3 missing values: 22
MARR_M4 missing values: 11
MARR_M5 missing values: 7
MARR_M6 missing values: 3
MARR_M7 missing values: 3
MARR_M8 missing values: 1
MARR_M9 missing values: 1

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

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

DIV_\$: Indicator of whether divorce occurred used: MARENDx

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

No missing cases

Order of Union	Number of unions	number of marriages	number of divorces
1	UNION 1: 8063	6349	1963
2	UNION 2: 2505	1540	399
3	UNION 3: 740	394	91
4	UNION 4: 206	94	27
5	UNION 5: 69	29	3
6	UNION 6: 24	13	3
7	UNION 7: 11	4	0
8	UNION 8: 3	1	0
9	UNION 9: 2	2	0

DIV_Y\$: Year of divorce used: MARDISx

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

No missing cases

DIV_M\$: Month of divorce used: MARDISx

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

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

Filter: IDIV_Mx=.b if UNION_x==0

IDIV_Mx=.b if MARR_x==0
IDIV_Mx=.b if DIV_x==0 or .d

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

SEXP_ \$: Partner`s sex

The gender of partner was not asked, they asked only about male cohabitators.

Filter: SEXP_x=.b if UNION_x==0

No missing cases

YEARBIRP_ \$: Year of birth of partner Used: H1BORN, WNCBPRN

Filter: YEARBIRP_x=.b if UNION_x==0

YEARBIRP_1 missing cases: 592
YEARBIRP_2 missing cases: 575
YEARBIRP_3 missing cases: 347
YEARBIRP_4 missing cases: 266
YEARBIRP_5 missing cases: 250
YEARBIRP_6 missing cases: 244
YEARBIRP_7 missing cases: 244
YEARBIRP_8 missing cases: 244
YEARBIRP_9 missing cases: 243

MONBIRP_ \$: Month of birth of partner used: H1BORN, WNCBPRN

Filter: MONBIRP_x=.b if UNION_x==0

MONBIRP_1 missing cases: 592
MONBIRP_2 missing cases: 575
MONBIRP_3 missing cases: 347
MONBIRP_4 missing cases: 266
MONBIRP_5 missing cases: 250
MONBIRP_6 missing cases: 244
MONBIRP_7 missing cases: 244
MONBIRP_8 missing cases: 244
MONBIRP_9 missing cases: 243

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

Filter: IMONBIRP_x=.b if UNION_x==0

KID_Š: Indicator of child order used: BABDOB

no missing cases

Child order	number of children
1	6911
2	4839
3	2183
4	766
5	249
6	93
7	41
8	19
9	6
10	4
11	1

KID_YŠ: Year of birth of child used: BABDOB

Filter: KID_Yx=.b if KID_x==0

KID_Y1 missing values: 10
KID_Y2 missing values: 5
KID_Y3 missing values: 8
KID_Y4 missing values: 1
KID_Y7 missing values: 1

KID_MŠ: Month of birth of child used: BABDOB

Filter: KID_Mx=.b if KID_x==0

KID_M1 missing values: 10
KID_M2 missing values: 5
KID_M3 missing values: 8
KID_M4 missing values: 1
KID_M7 missing values: 1

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

Filter: IKID_M_x=.b if KID_x==0

KID_SŠ: Sex of child used: SEXAx

Filter: KID_Sx=.b if KID_x==0

No missing cases

Child order	number of children	male	female
1	6911	3692	3219
2	4839	2524	2315
3	2183	1131	1052
4	766	390	376
5	249	126	123
6	93	49	44
7	41	19	22
8	19	14	5
9	6	4	2
10	4	4	0
11	1	1	0

KID_D\$: Death of child used: ALIVENWA

Filter: KID_Dx=.b if KID_x==0

missing cases:

KID_D1: 7

Child order	number of children	death
1	6911	102
2	4839	49
3	2183	26
4	766	10
5	249	4
6	93	3
7	41	2
8	19	0
9	6	0
10	4	0
11	1	0

KID_DY\$: Year of death of child used: WHENDIEA

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

No missing values

KID_DM\$: Month of death of child used: WHENDIEA

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

No missing cases

IKID_DM\$: Month of death of child
and imputed months

according to manual page 4 (random)

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

KID_L\$: Child left home used: WHENLEFA

Filter: KID_Lx=.b if KID_x==0

KID_L1 missing cases: 8

KID_L2 missing cases: 2

2014: children which died were excluded from KID_L=1 and are now coded with special missing code .d and KID_LY and KID_LM for dead children is coded as .b.

Child order	number of children	Left home
1	6911	1023
2	4839	496
3	2183	160
4	766	58
5	249	16
6	93	4
7	41	3
8	19	1
9	6	1
10	4	0
11	1	0

KID_LY\$: Year child left home used: WHENLEFA

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

KID_LY1 missing cases: 8

KID_LY2 missing cases: 2

KID_LM\$: Month child left home used: WHENLEFA

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

KID_LM1 missing cases: 8

KID_LM2 missing cases: 2

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

7. Part Education

Education:

Education histories were created using the highest degree attained variables in the NSFG. I classified individuals into the highest ISCED category (6) if the individual reported having a graduate degree, they received a 5 if they obtained a Bachelors' degree, a 4 if some college, but no degree, 3 if high school, 2 for lower-secondary, and 1 for just primary school. I then collapsed these categories by including all post-high school individuals in the top category, high school graduates in the middle, and less than high school in the bottom category. These can also be switched if a better model is proposed. I used a similar method for classifying the education of the respondent's parents. For the education degree dates, I only included dates for individuals who are no longer in school. To me, that made the most sense, and I recoded the EDU_Y to .b, not applicable. I can easily switch this though, if an actual date is preferred. On the 1995 file, there are completion years even if the respondent is still in school and in the 2007 file, we only know the completion dates for high school and Bachelors' degree.

There were some cases in the 2007 file of missing values for the year of completion dates-- respondents refused or didn't remember when they completed their highest level of education. All of these individuals had a high school degree or less, and I knew the actual grade completed. I used the following method for assigning completion dates for these individuals:

if the individual reported completing the 12th grade, I assigned their education completion year (IEDU_Y) to their birth year plus 18 years, the average age at completing 12th grade. If the respondent reported completing 11th grade, I added 17 years to their birth year for the education completion year. I continued in this way down to the 9th grade, which was the lowest grade reported of the individuals with missing education years. I assigned all of the education months (IEDU_M) to 6, for June, as most schooling is finished in June in the United States.

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

Currently studying: 2105 respondents
No missing cases

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

No missing cases

Definition:

The country specific codes include:

- * a 3-digit country prefix(840)
- * a 1-digit survey code (NSFG 1995=1) and
- * a 2-digit country specific code for level of education (0-7 levels of education)

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

Definition: ISCED_7=1 HIEDUC<=3
ISCED_7=2 HIEDUC >3 & <=8
ISCED_7=3 HIEDUC ==9
ISCED_7=4 HIEDUC==10,11
ISCED_7=5 HIEDUC==12
ISCED_7=6 HIEDUCY>=13

Harmonized:

ISCED	Number
0+1	182
2	2111
3	3865
4	2704
5	1553
6	432

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

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

Level	Number
High	4689
medium	3865
low	2293
missing cases	0

EDU_Y: Year highest level of education achieved used: CMHIDEG

Missing cases: 20

EDU_M: Month highest level of education achieved used: CMHIDEG

Missing cases: 21

IEDU_Y: Year highest level education achieved and imputed year

Definition for imputation:

Imputation for year of education is described above. If the completion grade is known, the completion year is defined as the birth year plus the average age at which an individuals completes that grade.

IEDU_M: Month highest education achieved and imputed month

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

NATIVE: Born in country used: BRNOUT

Born in country: 8933
Born elsewhere: 1117
797 missing cases

ETHNOS: Ethnicity/nationality used: HISPRACE

Country specific variable (840+1+code)

No missing cases

BIRTH_COU: Country of birth used: BRNCNTCD

Country specific variable (840+1+code)

Specialty: The country codes have 3 digits

missing cases: 6

MIG_Y: Year of migration used: STRUS

missing cases: 15

MIG_M: Month of migration used: STRUS

15 missing cases

IMIG_M: Month of migration and imputed months used: MIG_M

→according to manual page 4 (random)

9. Part Background variables (parental background)

SIS_NO: Number of sisters

not available in survey

BRO_NO: Number of brothers

not available in survey

SIBS: Total number of sibs

not available in survey

SIS_DIED: Number of sisters that died

not available in survey

BRO_DIED: Number of brothers that died

not available in survey

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

ISCED	Number
0+1	701
2	1344
3	5415
4	1462
5	967
6	417
.b	72
missing	469

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

ISCED	Number
0+1	778
2	1534
3	4214
4	1036
5	1123
6	787
.b	650
missing	725

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

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

Level	Number
High	2846
medium	5415
low	2045
.b	
missing cases	541

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

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

Level	Number
High	2946
medium	4214
low	2312
.b	
missing cases	1375

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

not available in survey

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

not available in survey

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

not available in survey

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

not available in survey

NATIVE_MO: Mother born in country

not available in survey

NATIVE_FA: Father born in country

not available in survey

BIRTHCO_MO: Mother`s country of origin

not available in survey

BIRTHCO_FA: Father`s country of origin

not available in survey

PARDIVEV: Parents ever divorced/separated used: PARDIVQ

Definition:

1 if PARDIVQ=1
 2 if PARDIVQ=2

5 if PARDIVQ=.

Missing cases: 28

PARDIV_15: Parents divorced before age of 15
used: WNPARDIV, WNMOMDIE, WNDADDIE

27 missing cases

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

REGION: Country region at time of interview

Country specific variable (840+1+code) used: REGION

No missing cases

SIZE: Size of place of residence at time of interview,

Country specific variable (840+1+code)

not available in survey

ISIZE: Size of place of residence at time of interview

Standardized code

SIZE_15: Size of place of residence at age 15

not available 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 (840+1+code) used: RELIGION

No missing cases

IRELIGION: Religious affiliation at time of interview

Standardized code

ADOPT: Number of adopted children of respondent used: ADPTOT

FOSTER: Number of foster children of respondent used: OTHKDF

STEP: Number of stepchildren of respondent used: RELOTH

Number of children	Adopt	Foster	Step
1	66	50	191
2	20	21	104
3	3	14	29
4	1	5	2
5	1	5	1
6		1	2
7		1	
8		1	
9		2	
11		2	

12. Part Weights

HHWGT: Household weight - not available in survey

PERSWGT: Personal weight - used: POST_WT

KISHWGT: Kishweight - not available in survey