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

HARMONIZED HISTORIES Georgia (10000 respondents)

Karolin Kubisch
Max Planck Institute for Demographic Research Rostock

15.2. 2013
Updated 3.6.2014
Updated 27.10.2015

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.

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

October 2015: Please note that the partnership histories were modified in October 2015. More precisely, we changed the sorting of the variable UNION_$ (Union order). Prior to that date, we had sorted the unions by start year of the union. This involved that unions which start dates were missing were always listed as last unions. In the modified version, we sorted the partnerships no longer by relying on the start year of the union, but by relying on the order of the union as they appear in the original dataset. For Georgia it affects 0 cases.

In connection with this modification, some smaller consistency changes were made to the data. In particular, we recoded the following constellations:

- Events (Union, Marriage, Separation, Divorce) before age 12 of respondent
- Event before age 12 of partner
- Negative difference between partnership date and marriage date
- Negative difference between separation date and union or marrige date and negative difference between divorce date and union or marriage date
- Sucessive partnerships mar-mar[_n-1]<=0 or par-par[_n-1]<=0
- Differences between separation date and next partnership date sep>par[_n+1]

All modifications made October 2015 are described in the updated documentation.

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

Source: UN Data: GGS_Wave1_Georgia_V.4.1.dta

Interview dates Georgia GGS: March to July 2006

1. Part Basic Information

RESPID: ID number to be assigned at merging — LEAVE BLANK

ARID: ID number from raw data (original ID number) — used: arid
10000 respondents

COUNTRY: Country and survey — used: acountry
acountry: code 13: Georgia
COUNTRY: code: 2681: Georgia GGS
no missing cases

MONTH_S: Month of survey
Not included

IMONTH_S: Month of survey, including imputed dates
Interview between March and July 2006

YEAR_S: Year of survey — used: ayear
2006

SEX: Sex of the respondent — used: ahg4_1
No missing cases
Sex structure of the Georgian respondents:
Male: 4405 and Female: 5595

BORN_Y: Year of birth of respondent — used: ahg6y_1
1926-1988

BORN_M: Month of birth of respondent — used: ahg6m_1

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

2. Part LEAVING HOME

LEAVE_1: Indicator of whether “left home”
used: GRID=1 go to a5117a
=0 go to a5116m/y
a5117a=1 go to a5117bm/y

Definition:
*Respondent did not leave home (code 0) if: a parent lives in the household (GRID=1) and respondent never lived separately from parents (a5117a=2)
*Respondent left home (code 1) if: there is no parent in household (GRID=0) or there is a parent in household (GRID=1) and respondent ever left home (a5117a=1)

LEAVE_1 0: 1750 / 1: 8249

LEAVE_Y1: Year of first time leaving home used: a5116y and a5117by

Filter: LEAVE_Y1/LEAVE_M1 to .b if LEAVE_1==0 (547)
Missing cases: 35

LEAVE_M1: Month of first time leaving home used: a5116m and a5117bm

Missing cases: .b 1750 .a 82

ILEAVE_M1: Month of first time leaving home and imputed months: used: LEAVE_M1

Harmonized: random variables according to manual

---

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

UNINUM: Total number of unions used: UNION_1 to _3

Syntax:
```
forvalues x=1/3 {
replace UNINUM=UNINUM+1 if UNION_`x'>0
            }
```
UNINUM:
0: 2192
1: 7604
2: 196
3: 8

UNION_$: UNION order

For the chapters union /marriage and divorce/ and a part of partners characteristics an reshaping program was used, which includes partnership histories and questions to the current partner

Definition (Union_1 to UNION_x):
➔an union exists if there is an answer in at least one of the questions about the current partner (a301m – a308) or in partnership histories (a334m – a349y)

UNION_1: 7808
UNION_2: 204

UNION_3: 8
No missing cases

UNION_Y$: Year of start union used: a301y and a334y

Filter: UNION_Yx=.b if UNION_x==0

UNION_Y2 missing values: 1

UNION_M$: Month of start UNION used: a301m and a334m

Filter: UNION_Mx=.b if UNION_x==0

UNION_M1 missing values: 26
UNION_M2 missing values: 2

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

SEP_$: Dissolution of UNION used: a343 (only histories)

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

Order of Union	Number of unions	number of separations	death of partner
1	7808	475	1080
2	204	22	20
3	8		

SEP_Y$: Year of end of UNION used: a344y (only histories)

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

SEP_Y1 missing values: 26
SEP_Y2 missing values: 1

SEP_M$: Month of end of UNION used: a344m (histories only)

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

SEP_M1 missing values: 51
SEP_M2 missing values: 2

ISEP_M$: Month of end of UNION and imputed months used: SEP_M$

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: a302a and a335a

Filter: MARR_x=.b if UNION_x==0

MARR_2 missing values: 1

Order of Union	Number of unions	number of marriages
1	7808	6748
2	204	107
3	8	2

MARR_Y$: Year of marriage used: a302by and a335y

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

MARR_Y2 missing values: 1

MARR_M$: Month of marriage used: a302bm and a335m

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

MARR_M1 missing values: 24
MARR_M2 missing values: 1

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

DIV_$: Indicator of whether divorce occurred used: a349a, a343 (only histories)

Filter: DIV_x=.b if UNION_x==0
DIV_x=.b if MARR_x==0
DIV_x=.d if a343_x==2

DIV_2 missing values: 1

Order of Union	Number of unions	number of marriages	number of divorces
1	7808	6748	205

2	204	107	5
3	8	2	

DIV_Y$: Year of divorce used: a349y

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_Y2 missing values: 1

DIV_M$: Month of divorce used: a349m

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: 2
DIV_M2 missing values: 1

IDIV_M$: Month of divorce used: DIV_M$
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 used: ahg4_1, ahg4_2

Filter: SEXP_x=.b if UNION_x==0

Partner	Number of unions	Number male	Number female
1	7808	4535	3273
2	204	79	125
3	8	1	7

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

Filter: YEARBIRP_x=.b if UNION_x==0

YEARBIRP_1 missing cases: 19
YEARBIRP_2 missing cases: 3

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

Filter: MONBIRP_x=.b if UNION_x==0

MONBIRP_1 missing cases: 61
MONBIRP_2 missing cases: 4

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

Filter: IMONBIRP_x=.b if UNION_x==0

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

for current partner:
a)children of partner (household members): relation of household member to respondent : code 4: stepchild: my current partners child not adopted by me➔ ahg3_2 to ahg3_8
b)non-resident stepchildren: a226==1 and a231
c)for partnership histories: a338_1 to a338_6
also: year of start of union(a301y) and year of birth of stepchild (ahg6y_x and a230_x)

Problem: The question: When you started living together, how many children did your partner have? (a338)- exists only for partnership histories
-for current partnership it had to be created with help of the number of stepchildren, year of start of union and year of birth of stepchild

Definition: in the number of children of current partner are included:
* all stepchildren of respondent living at the moment of interview in household grid and were born before the start of the union
* all nonresident stepchildren at the time of interview - partners children born before partnership (year start union-birth year>0)
* the number of partner`s children at start of a union in partnership history (a338_1 to a338_6)

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

NUMCHP_2: missing values: 1

NUMCLIV_$: Number of children of partner lived with respondent

a341_1 - a341_6 not included in survey

Union	Number of unions	NUMCHP
1	7808	1:71 2:40 3:11 4:5 6:1 7:1
2	204	1:16 2:31 3:7 4:4
3	8	1:2 2:1

Summary: The variable NUMCHP had to be created for the current partnership. The variable NUMCLIV is not included in dataset.

6. Part Birth histories (biological kids)

For the chapter "Birth histories" a reshaping program was used, which includes biological children in household and questions to the nonresident biological children

To create the number of biological children (KID_1 to KID_x) the following definition was applied:
➔a biological child exists in household if there is code 2 or 3 (biological child by current or previous partner) in the relationship to respondent (ahg3_)
➔a nonresident biological child exists if a213_==1

KID_$: Indicator of child order

used: ahg1_ and generated variable obnr (at least 1 answer in questions a212 to a224)

no missing cases

Child order	number of children
1	7501
2	5923
3	2384
4	740
5	260
6	96
7	47
8	10
9	5
10	4
11	2
12	2
13	1

KID_Y$: Year of birth of child used: ahg6y_ and a216y

Filter: KID_Yx=.b if KID_x==0

KID_M$: Month of birth of child used: ahg6m and a216m

Filter: KID_Mx=.b if KID_x==0

KID_M1 3 missing cases
KID_M2 2 missing cases
KID_M3 5 missing cases
KID_M4 2 missing cases
KID_M5 1 missing cases
KID_M6 1 missing cases

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

Filter: IKID_M_x=.b if KID_x==0

KID_S$: Sex of child used: ahg4 and a212

Filter: KID_Sx=.b if KID_x==0

KID_S1 missing cases: 2

Child order	number of children	male	female
1	7501	3816	3685
2	5923	3056	2867
3	2384	1223	1161
4	740	397	343
5	260	142	118
6	96	51	45
7	47	22	25
8	10	3	7
9	5	4	1
10	4	2	2
11	2		2
12	2	1	1
13	1		1

KID_D$: Death of child used: a211b

Filter: KID_Dx=.b if KID_x==0

2014: changes in KID_Dx

Child order	number of children	death
1	7501	222
2	5923	170
3	2384	93
4	740	27
5	260	18
6	96	7
7	47	2
8	10	2
9	5	
10	4	
11	2	
12	2	
13	1	

KID_DY$: Year of death of child used: a217y

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

KID_DY2 missing values: 1

KID_DM$: Month of death of child used: a217m

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

KID_DM1 missing values: 4
KID_DM2 missing values: 1
KID_DM3 missing values: 2

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

KID_L$: Child left home used: a220y/a220m

Child's parental home leave variable (KID_L) was not constructed perfectly as it was created in wide format instead of long. Namely the error occurred assuming that child's order would perfectly match of those living outside the household. More specifically, if child from outside household changes its order (because of preceding foster/adopted or a step child) and in household grid is reported biological child of the same order, then this particular child will be coded as "0" (did not leave home). Furthermore some children living in the household were coded as left home.

Initially both KID_LY (year of child's home leave) and KID_M (month of child's home leave) variables were constructed correctly, however due to reason that KID_L variable serves as filter for both variables then these variables eventually were changed to either ".b" (does not apply) or ".a " (unknown).

Since june 2014 KID_L is constructed in a long format. In addition 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.

Definition: Child left home if a220m_x or a220y_x!=.

Filter: KID_Lx=.b if KID_x==0

Child order	number of children	Left home
1	7501	2317
2	5923	1780
3	2384	790
4	740	300
5	260	122
6	96	46
7	47	23
8	10	3
9	5	2
10	4	1
11	2	
12	2	
13	1	

KID_LY$: Year child left home used: a220y

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

Missing cases KID_LY_1: 21

Missing cases KID_LY_2: 19
Missing cases KID_LY_3: 13
Missing cases KID_LY_4: 6
Missing cases KID_LY_5: 2
Missing cases KID_LY_6: 2
Missing cases KID_LY_7: 1

KID_LM$: Month child left home used: a220m

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

Missing cases KID_LM_1: 43
Missing cases KID_LM_2: 33
Missing cases KID_LM_3: 20
Missing cases KID_LM_4: 14
Missing cases KID_LM_5: 7
Missing cases KID_LM_6: 3
Missing cases KID_LM_7: 1

IKID_LM$: Month of death of child used: KID_LM
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

INSCHOOL: Currently studying at the time of interview used: a151

Currently studying: 86

EDU_COU: Highest level of education, country specific used: 148

These data exist in the harmonized dataset in an ISCED97 coded form.

These country specific codes include:
* a 3-digit country prefix(268)
* a 1-digit survey code (Georgia GGS=1) and
* a 2-digit country specific code for level of education (0-6 levels of education)

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

Harmonized:

ISCED	Number
1	436
2	941
3	3637

4	2232
5	2714
6	40

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

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

Level	Number
High	2754
medium	5869
low	1377

EDU_Y: Year highest level of education achieved used: a150y

Missing values: .a 3211

replace EDU_Y=.a if (EDU_Y<BORN_Y)

EDU_M: Month highest level of education achieved

Missing values: .a 3269

IEDU_Y: Year highest level education achieved and imputed year

IMPUTATION of missing years by level of Education
➔find the modal age of education with help of birth year and graduation year. Year of graduation for missing cases then is calculated by adding modal age of graduation to the birth date.

Missing values: .a 154

IEDU_M: Month highest education achieved and imputed month

Missing values: .a 154

Summary:
The EDU_COU data exist in a country specific ISCED97 form.

---

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

NATIVE: Born in country used: a105

Born in country: 9807
Born elsewhere: 193

ETHNOS: Ethnicity/nationality used: a110

NOT INCLUDED IN SURVEY

BIRTH_COU: Country of birth used: a106b

Country specific variable (268+1+code)

Filter: BIRTH_COU=.b if a105==1

MIG_Y: Year of migration used: a107y

Filter: MIG_Y=.b if a105==1

MIG_M: Month of migration used: 107m

Filter: MIG_M=.b if a105==1

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 used: a5106a_s

0 – 12 sisters

BRO_NO: Number of brothers used: a5106a_b

0 – 9 brothers

missing cases: 108

SIBS: Total number of sibs used: a5106a_s and a5106a_b

0-14 sibs

DECISION: If number of sisters is known and number of brothers is unknown or number of brothers is known and number of sisters is unknown: the number of known brothers or sisters is used

if number of brothers and number of sisters is unknown the value remains (missing .a)

SIS_DIED: Number of sisters that died
used: a5106a_s and a5106b_s
(number of sisters respondent have ever had – number of alive sisters)

Filter: SIS_DIED=.b if a5106a_s==0
Missing cases: 57

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

Filter: BRO_DIED=.b if a5106a_b==0
Missing cases: 150

ISCED_MO: Mother`s highest level of education used: a5115

ISCED	Number
1	1765
2	1667
3	2882
4	1664
5	1301
6	20
.a	701

ISCED_FA: Father`s highest level of education used: a5113

1	1552
2	1477
3	2553
4	1347
5	1550
6	32
.a	1489

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

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

Level	Number
High	1321
medium	4546
low	3432
.a	701

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

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

Level	Number
High	1582
medium	3900
low	3029
.a	1489

WORK_MO: Mother`s occupation, when respondent was 15
Country codes used: 5114

WORK_FA: Father`s occupation, when respondent was 15
Country codes used: 5112

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

High non manual: 1967
Non manual: 629
Manual: 2576

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

High non manual: 2226
Non manual: 647
Manual: 5569

NATIVE_MO: Mother born in country used: 513a

NOT INCLUDED IN SURVEY

NATIVE_FA: Father born in country used: 533a

NOT INCLUDED IN SURVEY

BIRTHCO_MO: Mother`s country of origin used: a513b

NOT INCLUDED IN SURVEY

BIRTHCO_FA: Father`s country of origin used: a533b

NOT INCLUDED IN SURVEY

PARDIVEV: Parents ever divorced/separated used: a550/a552

Missing values: 74

PARDIV_15: Parents divorced before age of 15 used: a550/a552
a551/ a511/ ahg6y_1

missing values: 79

Background variables (region, size of location)

REGION: Country region at time of interview

Country specific variable (268+1 +code) used: aregion

No missing cases

SIZE: Size of place of residence at time of interview used: atype

Country specific variable (268+1+code)

No missing cases

ISIZE: Size of place of residence at time of interview

Standardized code

SIZE_15: Size of place of residence at age 15 used: a5108

Country specific variable (268+1+code)

ISIZE_15: Size of place of residence at age 15

Standardized code

---

11. Part Other background variables

RELIGION: Religious affiliation at time of interview used: a1101

IRELIGION: Religious affiliation at time of interview

Standardized code

ADOPT: Number of adopted children of respondent
used: ahg3_2-ahg3_5, ahg3_8 (code5) and a213 (code 2)

FOSTER: Number of foster children of respondent
Used: ahg3_2-ahg3_6 (code 6) and a213 (code 3)

STEP: Number of stepchildren of respondent
Used: ahg3_2-ahg3_8 (code 4) and a226/ a229

Number of children	Adopt	Foster	Step
1	27	43	61
2	3	6	43
3		3	9
4		1	6
5			
6			
7			1

---

12. Part Weights

HHWGT: Household weight – not available in survey

PERSWGT: Personal weight - aweight

KISHWGT: Kishweight - not available in survey