DATA MANUAL
Harmonized Histories Estonia GGS II – V1.0

Almut Schumann
Federal Institute for Population Research (BiB)

01/2023

AHEAD:

About
The data manual gives an overview and description of all items included in the Harmonized Histories, the items used to construct these histories from the original dataset and information about cleaning and imputing cases according to the Working Paper for the preparation of comparative fertility and union histories. All details of the procedure to harmonize the data can be found in the respective Working Paper by Perelli-Harris, Kreyenfeld and Kubisch 2015 (see: https://www.demogr.mpg.de/papers/working/wp-2010-011.pdf). Further information on the Harmonized Histories can be found here: https://www.ggp-i.org/data/harmonized-histories/. The published material is provided for purposes of transparency and comprehensibility. We do not guarantee complete correctness and reserve the right to make possible corrections in the future.

Structure
The data manual starts with a list of all items constructed for the Harmonized Histories as well as their description grouped according to the thematic sections (highlighted in yellow). The items used to construct these histories from the original dataset are mentioned for the purpose of comprehensibility of the operationalization. The extra information provides general notes on the construction and information on adjustments and imputations according to the Working Paper.
After each section a description of the standardized and country-specific coding for particular items is provided (highlighted in red). In addition, respondents with illogical information in their fertility or partnership biography according to the Working Paper are indicated after the respective section (highlighted in blue). The data user can decide on their own whether to delete these cases on the basis of the provided respondent ID "ARID" or to keep them in the analyses.

Data
Source: GGS II, Estonia, First wave
Dataset: GGSII_Wave1_EE_V_1_0
Sampling: Random sampling from person registers.
Target population: Residents in Estonia aged 18 to 59 years.
Time and mode of conduction: Data was collected from October 2021 to March 2022; only CAWI-interviews.
Note: Respondents dropped out during the online interview, and answered the questionnaire only halfway. These observations were marked accordingly in the original data.
Net case numbers: n=8,992
Country Code for Estonia (used for country-specific items): 2332

Missing values coded in Harmonized Histories
.a = unknown
.b = does not apply
.c = unavailable in survey

Contact
If you have further questions, please contact: ggp@nidi.nl

BASIC INFORMATION			
Item name	Item label	Item used	Extra information
RESPID	ID number to be assigned at merging	respid	
ARID	ID number from raw data (Original ID number)	respid	
COUNTRY	Country and survey	country	*Country code for GGS II Estonia: 2332.
YEAR_S	Year of survey	intdatey	
MONTH_S	Month of survey	intdatem	*Set to .a missing if no information on year of survey (YEAR_S) is available *(no changes).*
IMONTH_S	Month of survey, including imputed dates	MONTH_S	*No imputation needed, because no missing value in MONTH_S.
SEX	Sex of the respondent	dem01	
BORN_Y	Year of birth of respondent	dem02y	*Set to .a-missing if age lies out of age range from target population: BORN_Y < 1960 or BORN_Y >2003 *(no changes).*
BORN_M	Month of birth of respondent	dem02m	*Set to .a-missing if no information on year of birth (BORN_Y) is available *(no changes).*
IBORN_M	Month of birth of respondent, including imputed dates	BORN_M	*Random imputation of 1-12 for .a-missing if information on year of birth (BORN_Y) is available *(changes=1).*

LEAVING HOME			
Item name	Item label	Item used	Extra information
LEAVE_1	Indicator of whether left home	gen52 hhd04_$ hhd05_$	*Case is coded as 0 (did not leave home) if a parent lives in the same household (hhd04_$/ hhd05_$ =7) at time of interview. All other cases, (including missing values) are coded as 1 (left home).
LEAVE_Y1	Year of leaving home	gen52ay	*Set to .a-missing if LEAVE_Y1 < 1900 or LEAVE_Y1 >2022 *(no changes)* *Recoded if date can be identified as a typo *(no changes).*
LEAVE_M1	Month of leaving home	gen52am	*Set to .a missing if no information on year of leaving home (LEAVE_Y1) is available *(no changes).*
ILEAVE_M1	Month of leaving home and imputed values	LEAVE_M1	*Random imputation of 1-12 for .a-missing *(changes=118).*

PARTNERSHIP HISTORIES			
Item name	Item label	Item used	Extra information
UNINUM	Total number of unions	UNION_$	
UNION_$	Indicator of union order	lhi04_m$ dem28a dem30a	*Only partnerships where partners are cohabiting or married count as a union and are coded as 1. Unmarried living-apart-together-partnerships do not count as unions and are coded as 0. *Unions with missing information on year of start of cohabitation and year of marriage, are coded as 0 (no union).

			*Unions are ordered chronologically, starting with the first/oldest union.
UNION_Y$	Year of start of union	lhi04_y$ dem30by	*Set to .a-missing if UNION_Y$ < 1900 or UNION_Y$ >2022 *(no changes).* *Recoded if date can be identified as a typo *(no changes).*
UNION_M$	Month of start of union	lhi04_m$ dem30bm	*Set to .a-missing if no information on year of start of union (UNION_Y$) is available *(no changes).*
IUNION_M$	Month of start of union, including imputed month	UNION_M$	*Random imputation of 1-12 for .a-missing if information on year of start of union (UNION_Y$) is available *(changes=111).*
SEP_$	Dissolution of union	lhi13_$	
SEP_Y$	Year of end of union	lhi14_y$	*Set to .a-missing if SEP_Y$ < 1900 or SEP_Y$ >2022 *(no changes).* *Recoded if date can be identified as a typo *(no changes).*
SEP_M$	Month end of union	lhi14_m$	*Set to .a-missing if no information on year of separation (SEP_Y$) is available *(no changes).*
ISEP_M$	Month end of union, and imputed month	SEP_M$	*Random imputation of 1-12 for .a-missing if information on year of separation (SEP_Y$) is available *(changes=71).*
MARR_$	Indicator of whether marriage took place and type of marriage	lhi05a_$ dem28a	
MARR_Y$	Year of marriage	lhi05b_y$ dem28by	*Set to .a-missing if MARR_Y$ < 1900 or MARR_Y$ >2022 *(no changes).* *Recoded if date can be identified as a typo *(no changes).*
MARR_M$	Month of marriage	lhi05b_m$ dem28bm	*Set to .a-missing if no information on year of marriage (MARR_Y$) is available *(no changes).*
IMARR_M$	Month of marriage and imputed marriage month	MARR_M$	*Random imputation of 1-12 for .a-missing if information on year of marriage (MARR_Y$) is available *(changes=30).*
DIV_$	Indicator of whether divorce occurred	lhi15a_$	
DIV_Y$	Year of divorce	lhi15b_y$	*Set to .a-missing if DIV_Y$ < 1900 or DIV_Y$ >2022 *(no changes).* *Recoded if date can be identified as a typo *(no changes).*
DIV_M$	Month of divorce	lhi15b_m$	*Set to .a-missing if no information on year of divorce (DIV_Y$) is available *(no changes).*
IDIV_M$	Month of divorce and imputed months of divorce	DIV_M$	*Random imputation of 1-12 for .a-missing if information on year of divorce (DIV_Y$) is available *(changes=28).*

CHECK: ILLOGICAL CASES IN PARTNERSHIP HISTORIES	
Check	ARID
Union histories, where the date of divorce occurs before the date of marriage for the same union and where marriages and unions occur before the birth of the respondent or before the respondent turned age 12. These cases can be excluded from the analysis on the basis of the ARID.	EST233FHW | EST2BZ2MP | EST3RLFLE | EST7QQDE8 | ESTCF8LS9 | ESTDAM7EP | ESTDJ8S2P | ESTELFJRQ | ESTFML2F7 | ESTGHCQ9X | ESTJ82BW5 | ESTNFZNYF | ESTQEGC8S | ESTS4BCS1 | ESTX8ZADY | ESTZPRPDK | ESTZYLX46

PARTNER´S CHARACTERISTICS			
Item name	Item label	Item used for construction	Extra information
SEXP_$	partner's sex	lhi17_$ dem23	
YEARBIRP_$	partner's year of birth	lhi06_y$ dem22y	*Set to .a-missing if YEARBIRP_$ < 1900 or YEARBIRP_$ >2022 *(no changes).* *Recoded if date can be identified as a typo*(no changes).*
MONBIRP_$	partner's month of birth	lhi06_m$ dem22m	*Set to .a-missing if no information on year of partner´s birth (YEARBIRP_$) is available *(no changes).*
IMONBIRP_$	Partner's month of birth and imputed months of birth	MONBIRP_$	*Random imputation of 1-12 for .a-missing if information on year of partner´s birth (YEARBIRP_$) is available *(changes=32).*
NUMCHP_$	Number of children of partner at start of union	lhi12_$ dem46	
NUMCLIV_$	Number of children of partner lived with respondent	not available in survey	

BIRTH HISTORIES			
Item name	Item label	Item used	Extra information
KID_$	Indicator of child order	lhi26_$	*Only biological children count as children and are coded as 1. *Children with missing information on year of birth are coded as 0 (no child). *Births of children are ordered chronologically, starting with the first/oldest birth.
KID_Y$	Year of birth of child	lhi29_y$	*Set to .a-missing if KID_Y$ < 1900 or KID_Y$ >2022 *(no changes).* *Recoded if date can be identified as a typo *(no changes).*
KID_M$	Month of birth of child	lhi29_m$	*Set to .a-missing if no information on year of birth of child (KID_Y$) is available *(no changes).*
IKID_M$	Month of birth of child and imputed months"	KID_M$	*Random imputation of 1-12 for .a-missing if information on year of birth of child (KID_Y$) is available *(changes=42).*

KID_S$	Sex of child	lhi28_$	
KID_D$	Death of child	lhi25_$	
KID_DY$	Year of death of child	lhi30_y$	*Set to .a-missing if KID_Y$ < 1900 or KID_Y$ >2022 *(no changes).*
KID_DM$	Month of death of child	lhi30_m$	*Set to .a-missing if no information on year of death of child (KID_DY$) is available *(no changes).*
IKID_DM$	Month of death of child and imputed months	KID_DM$	*Random imputation of 1-12 for .a-missing if information on year of death of child (KID_DY$) is available *(changes=3).*
KID_L$	Child left home	lhi31_$	*Only information if a child is living in the same household as respondent is available. *Case is coded as 0 (child not left home) if child is always (lhi31_$=1), most of the time (lhi31_$=2) and some of the time (lhi31_$ =3) living in the same household as the respondent. Case is coded as 1 (child left home) if child never lives in the same household as respondent (lhi31_$ =4).
KID_LY$	Year child left home	not available in survey	
KID_LM$	Month child left home	not available in survey	
IKID_LM$	Month child left home and imputed months	not available in survey	

CHECK: ILLOGICAL CASES IN BIRTH HISTORIES	
Check	ARID
Birth dates of children occur before respondent´s birth date or before respondent turned age 12. These cases can be excluded from the analysis on the basis of the ARID.	EST233FHW | EST33BCMQ | EST38F1X4 | EST42D7MT | EST5LDXL7 | EST5M3HW5 | EST6PYQXW | EST727BT5 | EST78HWCT | EST7PSTMK | EST7QQDE8 | EST8JMT7C | ESTBG8YLT | ESTCN5WBZ | ESTE386CP | ESTGPYAH6 | ESTGXY4QR | ESTHBCHPF | ESTJLS18C | ESTLRSBFE | ESTMLMAK2 | ESTNFZNYF | ESTNKNE92 | ESTNNG36Q | ESTNZEC8K | ESTTMWHEH | ESTWPBKL7 | ESTZYLX46

EDUCATION			
Item name	Item label	Item used	Extra information
INSCHOOL	Currently studying at the time of the interview	dem06	*Case is coded as 1 (in school), when respondent is in education or training (dem06 =1) or answers “does not apply”. For all other cases (except “unknown”-missings) it is assumed that they were not in school anymore and coded as 2 (not in school).
EDU_COU	Highest level of education achieved; country specific	dem07	
ISCED_7	Highest level of education achieved according to ISCED 1997	EDU_COU	
EDU_3	Highest level of education, ISCED collapsed into 3 categories	ISCED_7	

EDU_Y	Year highest level of education achieved	dem08y	*Set to .a-missing if EDU_Y < 1900 or EDU_Y >2022 *(no changes).*
IEDU_Y	Year highest level education achieved and imputed year	EDU_Y	*The median of years between year of birth and graduation is calculated for all observations by ISCED-level: *If the ISCED-level of the observation with .a-missing in the year of education is known, the respective median for the years between birth and graduation is imputed *(changes=554).* *If the ISCED level or year of birth is unknown, the .a-missing remains. *If the imputed value based on the median (this is especially the case for younger respondents with .a-missing) exceeds the time of interview (here:2022), the case remains .a-missing.
EDU_M	Month highest level of education achieved	dem08m	*Is <u>not</u> set to .a-missing if no information on year of education is available because you can use the imputed year of education variable.
IEDU_M	Month highest education achieved and imputed month	EDU_M	*Imputation of 6 (June) for .a-missing if information on imputed year of education (IEDU_Y) is available *(changes=610).*

COUNTRY-SPECIFIC AND STANDARDIZED CODINGS: EDUCATION			
dem07	EDU_COU	ISCED_7	EDU_3
2200. Alghariduseta	233200.	1. ISCED 0+1	3. Low
2201. Algharidus (3-, 4- või 6-klassiline)	233201.	1. ISCED 0+1	3. Low
2202. Kutseharidus ilma põhihariduseta	233202.	1. ISCED 0+1	3. Low
2203. Põhiharidus (7-, 8- või 9-klassiline)	233203.	3. ISCED 2	3. Low
2204. Põhiharidusel põhinev kutseharidus	233204.	4. ISCED 2	3. Low
2205. Üldkeskharidus (10-, 11- või 12-klassiline)	233205.	5. ISCED 3	2. Medium
2206. Keskharidusel põhinev kutseõpe, keskeriharidus	233206.	5. ISCED 4	2. Medium
2207. Rakenduskõrgharidus, diplomiõpe, tehnikumiharidus pärast keskharidustt	233207.	5. ISCED 5	1. High
2208. Bakalaureus või sellega võrdsustatud	233208.	6. ISCED 5	1. High
2209. Magister või sellega võrdsustatud	233209.	5. ISCED 5	1. High
2210. Doktorikraad või sellega võrdsustatud	233210.	6. ISCED 6	1. High

BACKGROUND VARIABLES (NATIONALITY, ETHNICITY ETC.)			
Item name	Item label	Item used	Extra information
NATIVE	Born in country	dem03	
ETHNOS	Ethnicity/nationality; country specific	not available in survey	
BIRTH_COU	Country of birth; country specific	dem04biso dem04b	
MIG_Y	Year of migration	dem05y	*Set to .a-missing if MIG_Y < 1900 or MIG_Y >2022 *(no changes).*
MIG_M	Month of migration	dem05m	*Set to .a-missing if no information on year of migration (MIG_Y) is available *(changes=5).*
IMIG_M	Month of migration and	MIG_M	*Random imputation of 1-12 for

	imputed month"		.a-missing if information on year of migration (MIG_Y) is available *(changes=6).*

BACKGROUND VARIABLES (PARENTAL BACKGROUND)			
Item name	Item label	Item used	Extra information
SIS_NO	Number of sisters	gen39b	
BRO_NO	Number of brothers	gen39a	
SIBS	Total number of siblings	gen39a gen39b	*Set to .a-missing if number of sisters is known and number of brothers is unknown or number of brothers is known and number of sisters is unknown. Only if number of brothers and sisters is known, SIBS gets a valid value.
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	gen51	
ISCED_FA	Father's highest level of education	gen49	
EDU3_MO	Highest level of education of mother, ISCED 1997 (3 categories	ISCED_MO	
EDU3_FA	Highest level of education of father, ISCED 1997 (3 categories)	ISCED_FA	
WORK_MO	Mother's occupation when respondent was 15	gen50isco gen50	*Set to .b-missing ("does not apply") if mother or father did not work or were homemaker (.g-missing in GGS). *Original variable from GGS has a high number of missing values.
WORK_FA	Father's occupation when respondent was 15	gen48isco gen48	*Set to .b-missing ("does not apply") if mother or father did not work or were homemaker (.g-missing in GGS). *Original variable from GGS has a high number of missing.
ISCO3_MO	Mother's occupation when respondent was 15 (3 categories)	WORK_MO	*Set to .b-missing ("does not apply") if mother or father did not work or were homemaker (.g-missing in GGS). *Original variable from GGS has a high number of missing values.
ISCO3_FA	Father's occupation when respondent was 15 (3 categories)	WORK_FA	*Set to .b-missing ("does not apply") if mother or father did not work or were homemaker (.g-missing in GGS). *Original variable from GGS has a high number of missing values.
NATIVE_MO	Mother born in country	gen11 gen12iso gen12	
NATIVE_FA	Father born in country	gen25 gen26iso gen26	
BIRTHCO_MO	Mother's country of origin	gen12iso	

		gen12	
BIRTHCO_FA	Father's country of origin	gen26iso gen26	
PARDIVEV	Parents ever divorced/separated	gen01 gen02 gen03 gen38a	*Information whether parents are separated/divorced or died at time of interview is used. If there is no information available about current status of parents, information if parents ever broke up is used.
PARDIV_15	Parents divorced/separated before age of 15/16	not available in survey	

COUNTRY-SPECIFIC AND STANDARDIZED CODINGS: OCCUPATION	
ISCO3_	WORK_
1 High non manual	0-3999
2 Non manual	4000-5999
3 Manual	6000-9999

BACKGROUND VARIABLES (REGION, SIZE OF LOCATION)			
Item name	Item label	Item used	Extra information
REGION	Country region at time of interview	not available in survey	
SIZE	Size of place of residence at time of interview	not available in survey	
ISIZE	Standardized size of place of residence at time of interview	not available in survey	
SIZE_15	Size of place of residence at age 15	not available in survey	
ISIZE_15	Standardized size of place of residence at age 15	not available in survey	

OTHER BACKGROUND VARIABLES			
Item name	Item label	Item used	Extra information
RELIGION	Religious affiliation at time of interview	att08	
IRELIGION	Standardized religious affiliation at time of interview	RELIGION	
ADOPT	Number of adopted children of respondent	dem44 lhi10_$ lhi09_$ lhi23	*ADOPT is coded as 0 in the Harmonized Histories when it is .-missing in the GGS.
FOSTER	Number of foster children of respondent	hhd04_$	*FOSTER is coded as 0 in the Harmonized Histories when it is .-missing in the GGS.
STEP	Number of stepchildren of respondent	dem46 lhi11_$ lhi12_$ lhi22	*STEP is coded as 0 in the Harmonized Histories when it is .-missing in the GGS.

COUNTRY-SPECIFIC AND STANDARDIZED CODINGS: RELIGION		
att08	RELIGION	IRELIGION
1. Protestant	233201.	1. Christian
2. Roman Catholic	233202.	1. Christian
3. Buddhist	233203.	3. Other religion
4. Hindu	233204.	3. Other religion
5. Muslim	233205.	2. Muslim
6. Jewish	233206.	3. Other religion
7. Sikh	233207.	3. Other religion
8. Orthodox (e.g. Greek or Russian)	233208.	1. Christian
9. Other Christian	233209.	1. Christian
10. Other Religion	233210.	3. Other religion
11. None	233211.	4. No religion

WEIGHTS			
Item name	Item label	Item used	Extra information
HHWGT	Household weight	not available in survey	
PERSWGT	Personal weight	weight	
KISHWGT	Aweight	not available in survey	