

WP 3.1.2: Ethnicity, migration, and integration

Work package 3 - Task 3.1 - Deliverable D3.2

Authors:

Sandra Florian, Mathieu Ichou, Lidia Panico

(Institut National d'Etudes Démographiques, INED, Paris)

Version 1.4

Delivery date: Month 30

Submission date: April 20, 2020

Dissemination level: Confidential, only for members
of the consortium (including the Commission Services)

Table of Contents

Table of Contents	2
1. Protocol for subtask 3.1.2	3
1.1. Children’s migration generation (migen_child)	3
1.2. Parents’ region of birth (region_mo & region_fa)	4
1.3. Parents’ educational selectivity (reledu_mo & reledu_fa)	5
2. Codebook for new harmonised variables for WP 3.1.2	8
3. Descriptive statistics of participating cohorts	11
Table 3.1. Children’s migration status	11
Table 3.2. Mother’s region of birth	12
Table 3.3. Father's region of birth	13
Figure 3.1. Children’s migration status	14
Figure 3.2. Mother’s region of birth (immigrant only)	14
Figure 3.3. Father's region of birth (immigrant only)	15
Table 3.4. Mother's relative level of education	15
Table 3.5. Father's relative level of education	15
Figure 3.4. Mother's relative level of education	16
Figure 3.5. Father's relative level of education	16
4. Variable list for WP 3.1.2	17
5. Next steps	18
Annex 1: ISO3 Country codes “Countries-WP3.1.2.xls”	19
Annex 2: Integrated educational distribution dataset “EducAttain.csv” (Extract Sample)	25

1. Protocol for subtask 3.1.2

For this subtask, three sets of variables were created and harmonized: child's migration generation; mother's and father's region of birth; mother's and father's educational selectivity.

Not all cohorts have been able to produce the three sets of variables. Cohorts that have created the harmonized variables for subtask 3.1.2 are the following:

- **BiB**: Born in Bradford (UK)
- **ELFE**: Étude Longitudinale Française depuis l'Enfance (France)
- **GEN R**: Generation R (Netherlands)
- **INMA**: INfancia y Medio Ambiente Project (Spain)
- **NINFEA**: Nascita e INFanzia, gli Effetti dell'Ambiente (Italy)
- **RAINE**: Western Australian Pregnancy Cohort Study (Australia)
- **RHEA**: Mother Child Cohort in Crete (Greece)
- **PiccoliPiù** (Italy)
- **GECKO**: Groningen Expertise Centrum voor Kinderen met Overgewicht (Netherlands)
- **CHOP**: Childhood Obesity Programme study (Germany)

The following pages provide the detailed protocol used to create each set of variables. For sake of clarity and cross-country harmonization, we have suggested variable names for each variable.

1.1. Children's migration generation (**miggen_child**)

Harmonized input variables:

abroad_child

Binary variable indicating whether child is born abroad (0 = No, 1= Yes)

abroad_mo

Binary variable indicating whether mother is born abroad (0 = No, 1= Yes)

abroad_fa

Binary variable indicating whether father is born abroad (0 = No, 1= Yes)

Created variable: `miggen_child`

Child migration generation status. Categorical variable.

Categories:

- “1st generation”: cohort member is born abroad¹;
- “2nd generation”: cohort member is born in the host country, *both* parents are born abroad;
- “2.5 generation – mother”: cohort member born in the host country, father born in the host country, mother born abroad;
- “2.5 generation – father”: cohort member born in the host country, mother born in the host country, father born abroad
- “Natives”: cohort member is born in the host country; *both* parents born in the host country².

1.2. Parents’ region of birth (`region_mo` & `region_fa`)

Use the **countries-WP3.1.2.xls** spreadsheet, attached in Annex 1, with the ISO3 country codes and country names for each region, as well as conversions from ISO2 and UN M49 numeric codes. The spreadsheet can also be used to link countries to the regions as listed below. The **countries-WP3.1.2.xls** spreadsheet was distributed to all participating cohorts and is available on the [LifeCycle Intranet](#).

Harmonized input variables³:

`cntry_mo`

Country of birth of the mother, in ISO3 country codes

`cntry_fa`

Country of birth of the father, in ISO3 country codes

Created variables

`region_mo`

Mother’s region of birth, adapted from the World Bank classification (see below)

¹ Most cohorts are birth cohorts, thus the first generation category does not apply to them. However, a small number of cohorts do include first generation children, so we have kept this category, but given the small sample sizes, we will not be able to use the first generation category in analyses.

² Potentially includes grandchildren and more distant descendants of immigrants (often called “third + generations”).

³ IMMA cohort: Dutch Antilles were classified as Curacao; Palestine as West Bank and Gaza; Sahara was left as missing, as suggested by the subtask leaders.

region_fa⁴

Father's region of birth, adapted from the World Bank classification (see below)

Categories: 0 = Host country-born, 1 = Western EU/EEA, 2 = Eastern EU, 3 = Other Europe & Central Asia, 4 = East Asia & Pacific, 5 = South Asia, 6 = Middle East & North Africa, 7 = Sub-Saharan Africa, 8 = Latin America & Caribbean, 9 = North America

Documentation: The geo-political classification is based on the World Bank Country Groups classification

(<https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups>). We have adapted it to allow a further distinction within the "Europe and Central Asia" region between: "Western EU/EEA", "Eastern EU" and "Other Europe & Central Asia".

1.3. Parents' educational selectivity (reledu_mo & reledu_fa)

Reminder of the variable definition: immigrant parents' position within the educational attainment distribution of the population of the same age and sex in their country of origin. This measure is introduced and explained in greater detail in Ichou (2014).

Harmonized input variables:

cntry_mo / cntry_fa

Country of birth of the mother/father in ISO3 country codes

yob_mo / yob_fa

Year of birth of the mother/father in four digits

edu_mo / edu_fa⁵

Educational attainment of mother; educational attainment of father, in recoded ISCED 2011 codes, as follows: 0=no/pre-primary education (ISCED 0); 1=primary education (ISCED 1); 2=lower secondary (ISCED 2); 3=upper secondary (ISCED 3); 4=post-secondary and tertiary (ISCED 4 through 8).

⁴ In the NINFEA cohort (Italy), data on paternal country and year of birth is available only for a subset of the cohort (3656/7645 children). A flag variable indicates whether the observation is in the subsample.

⁵ INMA cohort considered the old EGB (basic school) as low secondary school (ISCED 2). In the GECKO cohort, due to a fault in the baseline questionnaire, the education at baseline probably has an underestimation of the categories medium (3 levels, used in core variables) / post-secondary (5 categories, used in this subtask) and an overestimation of the categories low (3 levels, used in core variables) / lower secondary (5 categories, used in this subtask).

Created variables

reledu_mo / reledu_fa⁶

Mother's /father's educational selectivity: percentile position in the country of origin's educational attainment distribution of women born in the same year (variable values form 0 – least negatively selected to 100 – most positively selected)

Protocol

Integrated educational distribution dataset (EducAttain.csv): To minimize the amount of data management and reshaping for each task participant, we provided a ready-to-use integrated dataset (based on Barro-Lee educational attainment data and Wittgenstein Centre education Data), with one line per country (in ISO3 codes) and year of birth and educational distribution for each sex (mother and father). The **EducAttain.csv** file is too long to be included in this document, however Annex 2 presents an extract sample of **EducAttain.csv**. The complete **EducAttain.csv** file was distributed to all participating cohorts and is available on the [LifeCycle Intranet](#).

Cohort dataset shaped as one line per child (i.e. household)

- To create mother's educational selectivity variable:

- Merge cohort data with the integrated education dataset based on mother's country of birth (**cntry_mo**), year of birth (**yob_mo**) and educational attainment (**edu_mo**).
- The type of merge required is a "many-to-one" merge, since there can be many cases (mothers) in your cohort data that are matched to one combination of country of birth/year of birth/educational attainment in the EducAttain database⁷.
- For the purpose of this merge, keep only the variable reledu_mo from the EducAttain dataset.

- To create father educational selectivity variable, follow an identical process using the father's variables (**cntry_fa**, **yob_fa** & **edu_fa**) and keeping only the variable reledu_fa from the EducAttain dataset.

⁶ Cause of missingness: missing data on education, or born after 1994.

⁷ For example, in Stata, the syntax used is :

```
merge m:1 cntry_mo yob_mo edu_mo using "EducAttain.dta", keepusing(reledu_mo)
```

At the end of the process, **reledu_mo** & **reledu_fa** are two new variables measuring the educational selectivity of mothers and fathers of each child in the cohort.

Documentation: The original Barro-Lee data and documentation is available here: <http://www.barrolee.com/>, the latest version of the data is described in Barro and Lee (2013). For the original Wittgenstein Centre data, see: <http://dataexplorer.wittgensteincentre.org/shiny/wic/>. A general introduction to the data is available in the Lutz et al. (2014) book. For a paper specifically describing the educational data, its underlying sources and assumptions, see Goujon et al. (2016). For ISCED 2011 classifications and country mappings, see: <http://uis.unesco.org/en/isced-mappings>.

Bibliography

- Robert J. Barro and Jong Wha Lee (2013) A new data set of educational attainment in the world, 1950–2010. *Journal of Development Economics* 104 (September 2013), 184–98. doi:10.1016/j.jdeveco.2012.10.001.
- Goujon, A., K.C., S., Springer, M., Barakat, B., Potancoková, M., Eder, J., Lutz, W. (2016). A Harmonized Dataset on Global Educational Attainment between 1970 and 2060 – An Analytical Window into Recent Trends and Future Prospects in Human Capital Development. *Journal of Demographic Economics*, 82(3), 315-363. doi:10.1017/dem.2016.10
- Ichou, Mathieu. 2014. "Who They Were There: Immigrants' Educational Selectivity and Their Children's Educational Attainment." *European sociological review* 30:750-765.
- Lutz W, Butz WP, & KC S (2014). *World Population & Human Capital in the Twenty-first Century: Executive Summary*. IIASA, Laxenburg, Austria (October 2014) <https://pure.iiasa.ac.at/11189/>.

2. Codebook for new harmonised variables for WP 3.1.2

Variable	Variable label	Type	Range	Categories
miggen_child	Child's migration generation	Categorical	1-5	1= 1st generation 2= 2nd generation 3= 2.5 generation – mother 4= 2.5 generation – father 5 = natives . = missing

Comments:

“1st generation”: cohort member is born abroad; “2nd generation”: cohort member is born in the host country, both parents are born abroad; “2.5 generation – mother”: cohort member born in the host country, father born in the host country, mother born abroad; “2.5 generation – father”: cohort member born in the host country, mother born in the host country, father born abroad “Natives”: cohort member is born in the host country; both parents born in the host country.

region_mo	Mother's region of birth	Categorical	0-9	0= Host country-born 1= Western EU/EEA 2= Eastern EU 3= Other Europe & Central Asia 4= East Asia & Pacific 5= South Asia 6= Middle East & North Africa 7= Sub-Saharan Africa 8= Latin America & Caribbean 9= North America . = Missing
------------------	--------------------------	-------------	-----	--

Variable	Variable label	Type	Range	Categories
region_fa	Father's region of birth	Categorical	0-9	0= Host country-born 1= Western EU/EEA 2= Eastern EU 3= Other Europe & Central Asia 4= East Asia & Pacific 5= South Asia 6= Middle East & North Africa 7= Sub-Saharan Africa 8= Latin America & Caribbean 9= North America . = Missing

reledu_mo	Mother's relative level of education	Continuous	0-100	. = Missing
------------------	--------------------------------------	------------	-------	-------------

Comments:

Percentile position in the country of birth's educational attainment distribution of women born in the same year (form 0 – least negatively selected to 100 – most positively selected).
Examples: 13.94, 45.28, 70, 99.999992.

reledu_fa	Father's relative level of education	Continuous	0-100	. = Missing
------------------	--------------------------------------	------------	-------	-------------

Comments:

Percentile position in the country of birth's educational attainment distribution of men born in the same year (form 0 – least negatively selected to 100 – most positively selected).
Examples: 13.94, 45.28, 70, 99.999992.

Variable	Variable label	Type	Range	Categories
abroad_child	Child's born abroad	Binary	0-1	0 = No 1 = Yes

abroad_mo	Mother's born abroad	Binary	0-1	0 = No 1 = Yes

abroad_fa	Father's born abroad	Binary	0-1	0 = No 1 = Yes

3. Descriptive statistics of participating cohorts

Table 3.1. Children's migration status

	BiB		ELFE		NINFEA*		Piccoli Piu		INMA	
	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%
1st generation	0	0.0	0	0.0	35	1.0	0	0.0	0	0.0
2nd generation	2579	18.6	1241	6.8	28	0.8	122	3.7	114	5.0
2.5 generation (mother)	1622	11.7	831	4.5	110	3.0	155	4.6	86	3.8
2.5 generation (father)	1897	13.7	1030	5.6	95	2.6	107	3.2	131	5.8
natives	5361	38.7	12107	66.1	3348	91.1	2954	88.5	1780	78.4
Missing	2399	17.3	3120	17.0	59	1.6	0	0.0	159	7.0
Total	13858	100.0	18329	100.0	3675	100.0	3338	100.0	2270	100.0

Table 3.1. Children's migration status... ctd.

	GECKO		CHOP		Raine		RHEA		Gen R	
	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%
1st generation	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
2nd generation	26	0.9	8	0.5	797	27.8	75	5.1	1913	19.3
2.5 generation (mother)	60	2.1	12	0.7	419	14.6	46	3.2	766	7.7
2.5 generation (father)	64	2.3	26	1.5	485	16.9	15	1.0	1041	10.5
natives	2445	86.0	476	28.4	1167	40.7	1188	81.5	4558	46.0
Missing	249	8.8	1156	68.9	.	.	134	9.2	1623	16.4
Total	2844	100	1678	100.0	2868	100.0	1458	100.0	9901	100

*NINFEA subsample. No information on paternal country of birth and education. 816 observations without information on birth level. Total sample 7645.

Table 3.2. Mother's region of birth

	BiB		ELFE		NINFEA*		Piccoli Piu		INMA	
	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%
Host country-born	7263	52.4	13589	74.1	3516	96.2	3061	91.7	1996	87.9
Western EU/EEA	95	0.7	278	1.5	50	1.4	40	1.2	29	1.3
Eastern EU	256	1.9	101	0.6	32	0.9	110	3.3	23	1.0
Other Europe & Crtl Asia	28	0.2	140	0.8	12	0.3	38	1.1	6	0.3
East Asia & Pacific	124	0.9	129	0.7	1	0.0	6	0.2	3	0.1
South Asia	3347	24.1	23	0.1	1	0.0	4	0.1	0	0.0
Middle East & N. Africa	83	0.6	858	4.7	3	0.1	16	0.5	5	0.2
Sub-Saharan Africa	238	1.7	604	3.3	2	0.1	4	0.1	4	0.2
Latin America & Caribb.	14	0.1	125	0.7	37	1.0	53	1.6	161	7.1
North America	15	0.1	18	0.1	2	0.1	6	0.2	0	0.0
Missing	2395	17.3	2464	13.4	0	0.0	0	0.0	43	1.9
Total	13858	100.0	18329	100.0	3656	100.0	3338	100.0	2270	100.0

Table 3.2. Mother's region of birth... ctd.

	GECKO		CHOP		Raine		RHEA		Gen R	
	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%
Host country-born	2548	89.6	506	30.2	1652	57.6	1311	89.9	5907	59.7
Western EU/EEA	16	0.6	5	0.3	643	22.4	6	0.4	446	4.5
Eastern EU	7	0.2	3	0.2	34	1.2	31	2.1	55	0.6
Other Europe & Crtl Asia	11	0.4	2	0.1	0	0.0	79	5.4	590	6.0
East Asia & Pacific	18	0.6	0	0.0	149	5.2	3	0.2	206	2.1
South Asia	7	0.2	0	0.0	211	7.4	0	0.0	31	0.3
Middle East & N. Africa	9	0.3	2	0.1	15	0.5	2	0.1	533	5.4
Sub-Saharan Africa	6	0.2	2	0.1	64	2.2	0	0.0	371	3.7
Latin America & Caribb.	16	0.6	5	0.3	16	0.6	1	0.1	897	9.1
North America	1	0.0	1	0.1	12	0.4	1	0.1	39	0.4
Missing	205	7.2	1152	68.6	72	2.5	24	1.7	826	8.3
Total	2844	100.0	1678	100.0	2868	100.0	1458	100.0	9901	100.1

*NINFEA subsample. No information on paternal country of birth and education. 816 observations without information on birth level. Total sample 7645.

Table 3.3. Father's region of birth

	BiB		ELFE		NINFEA*		Piccoli Piu		INMA	
	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%
Host country-born	6984	50.4	12947	70.6	3476	95.1	3108	93.1	1956	86.2
Western EU/EEA	76	0.6	282	1.5	51	1.4	42	1.3	48	2.1
Eastern EU	200	1.4	59	0.3	21	0.6	67	2.0	17	0.8
Other Europe & Crtl Asia	27	0.2	148	0.8	12	0.3	30	0.9	3	0.1
East Asia & Pacific	100	0.7	84	0.5	2	0.1	2	0.1	2	0.1
South Asia	3612	26.1	27	0.2	1	0.0	5	0.2	1	0.0
Middle East & N. Africa	114	0.8	1090	6.0	14	0.4	32	1.0	53	2.3
Sub-Saharan Africa	265	1.9	607	3.3	5	0.1	4	0.1	9	0.4
Latin America & Caribb.	31	0.2	87	0.5	20	0.6	40	1.2	136	5.9
North America	9	0.1	20	0.1	4	0.1	7	0.2	3	0.1
Missing	2440	17.6	2978	16.3	50	1.4	1	0.0	42	2.0
Total	13858	100.0	18329	100.0	3656	100.0	3338	100.0	2270	100.0

Table 3.3. Father's region of birth... ctd.

	GECKO		CHOP		Raine		RHEA		Gen R	
	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%
Host country-born	2511	88.3	489	29.1	1586	55.3	1240	85.1	5329	53.8
Western EU/EEA	20	0.7	10	0.6	652	22.7	4	0.3	271	2.7
Eastern EU	2	0.1	4	0.3	44	1.5	15	1.0	40	0.4
Other Europe & Crtl Asia	7	0.2	3	0.2	0	0.0	66	4.5	571	5.8
East Asia & Pacific	7	0.2	0	0.0	148	5.2	1	0.1	154	1.6
South Asia	5	0.2	0	0.0	168	5.9	0	0.0	34	0.3
Middle East & N. Africa	20	0.7	5	0.3	19	0.7	6	0.4	612	6.2
Sub-Saharan Africa	10	0.4	9	0.5	75	2.6	0	0.0	381	3.8
Latin America & Caribb.	17	0.6	3	0.2	25	0.9	0	0.0	883	8.9
North America	4	0.1	0	0.0	28	1.0	0	0.0	29	0.3
Missing	241	8.5	1155	68.8	123	4.3	126	8.6	1597	16.1
Total	2844	100.0	1678	100.0	2868	100.1	1458	100.0	9901	99.9

*NINFEA subsample. No information on paternal country of birth and education. 816 observations without information on birth level. Total sample 7645.

Fig. 3.1. Children’s migration status: Frequencies distributions by cohort

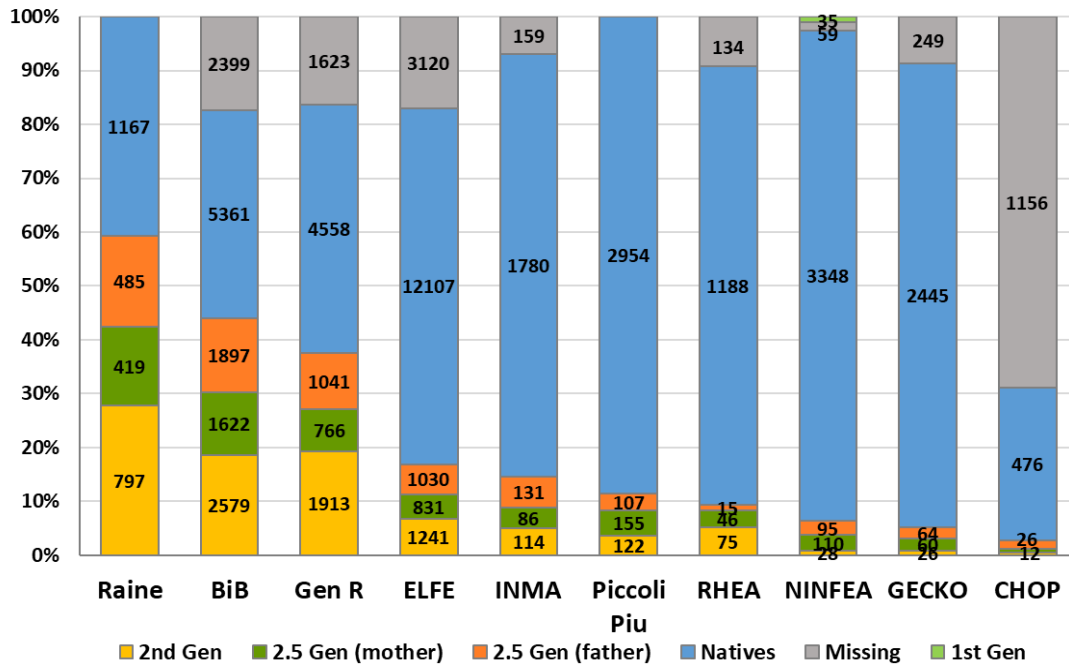
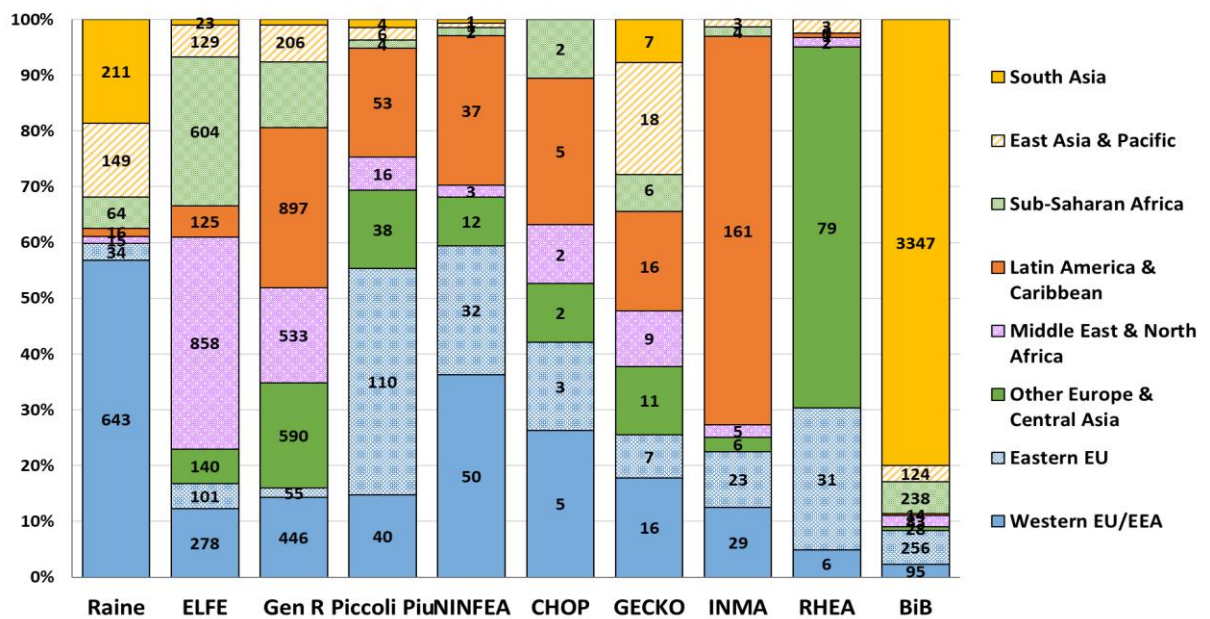
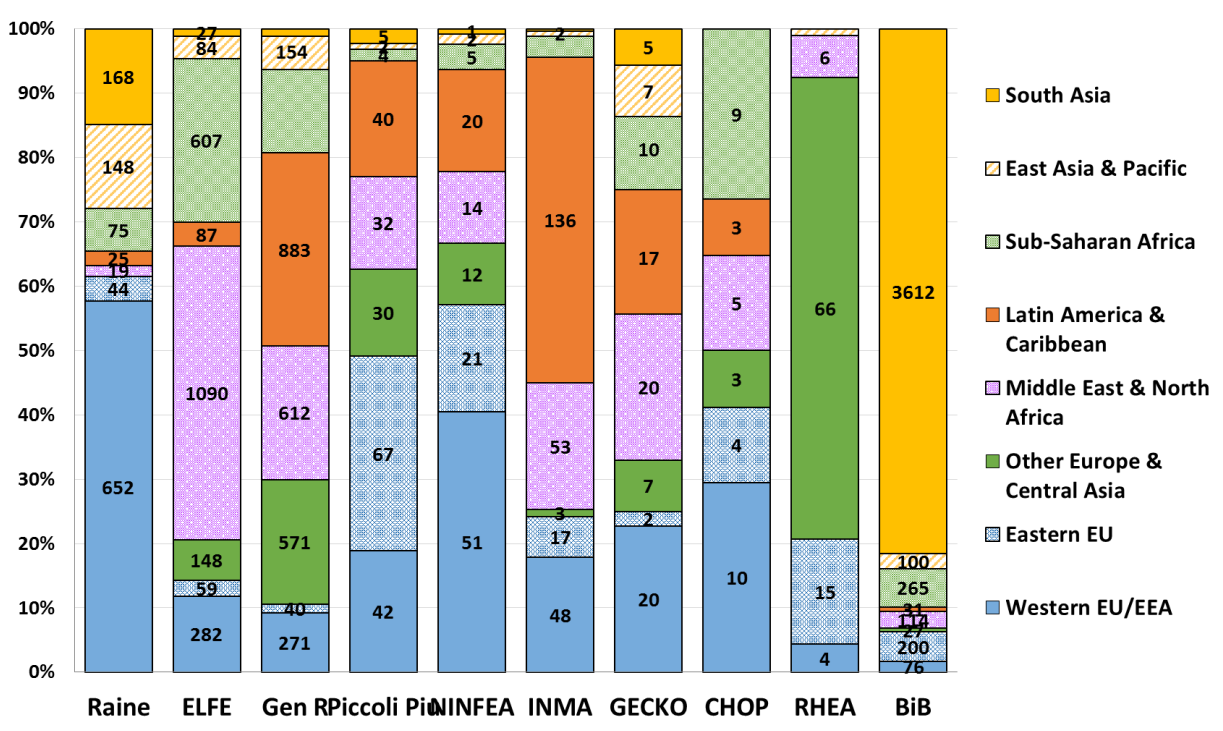


Fig. 3.2. Mother’s region of birth (immigrant only)*



* Excludes North America due to small counts

Fig. 3.3. Father's region of birth (immigrant only)*



* Excludes North America due to small counts

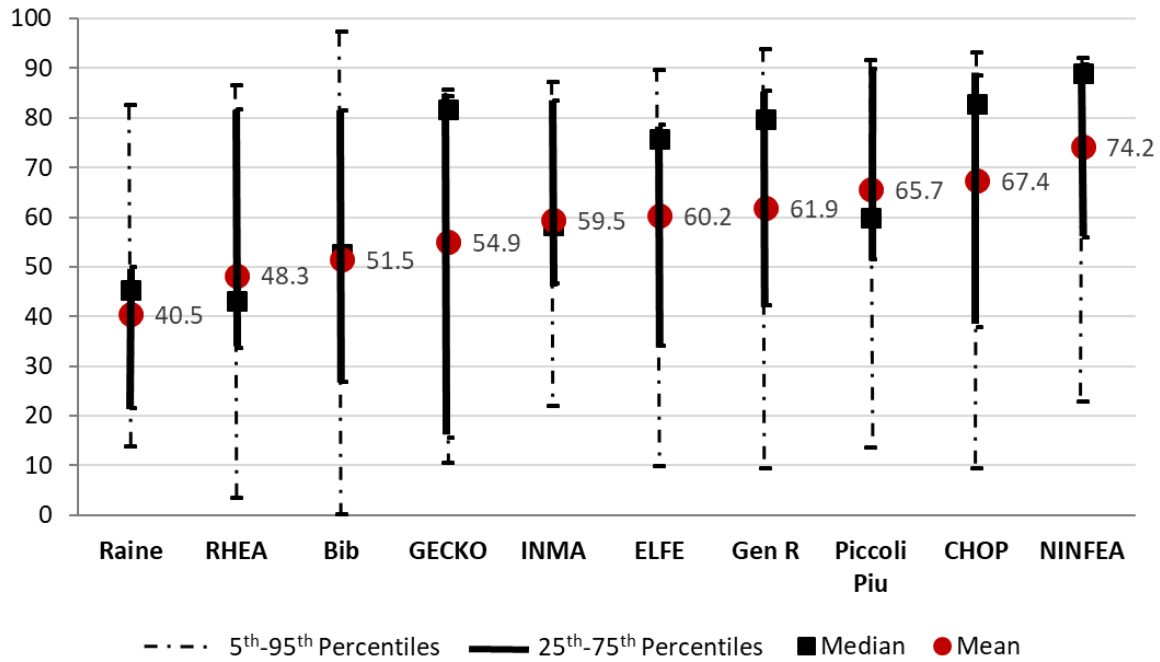
Table 3.4. Mother's relative level of education

Percentile	Raine	RHEA	Bib	GECKO	INMA	ELFE	Gen R	PiccoliPiu	CHOP	NINFEA
5%	13.9	3.5	0.2	10.6	22.0	9.9	9.4	13.5	9.4	22.9
25%	21.5	33.6	26.9	15.5	46.6	34.2	42.3	51.6	37.9	56.0
50%	45.3	43.1	52.7	81.7	58.4	75.8	79.8	60.0	82.9	89.1
75%	49.9	81.8	81.5	84.3	83.5	78.6	85.5	89.8	88.6	90.7
95%	82.6	86.6	97.4	85.8	87.3	89.7	93.8	91.6	93.2	92.0
Mean	40.5	48.3	51.5	54.9	59.5	60.2	61.9	65.7	67.4	74.2

Table 3.5. Father's relative level of education

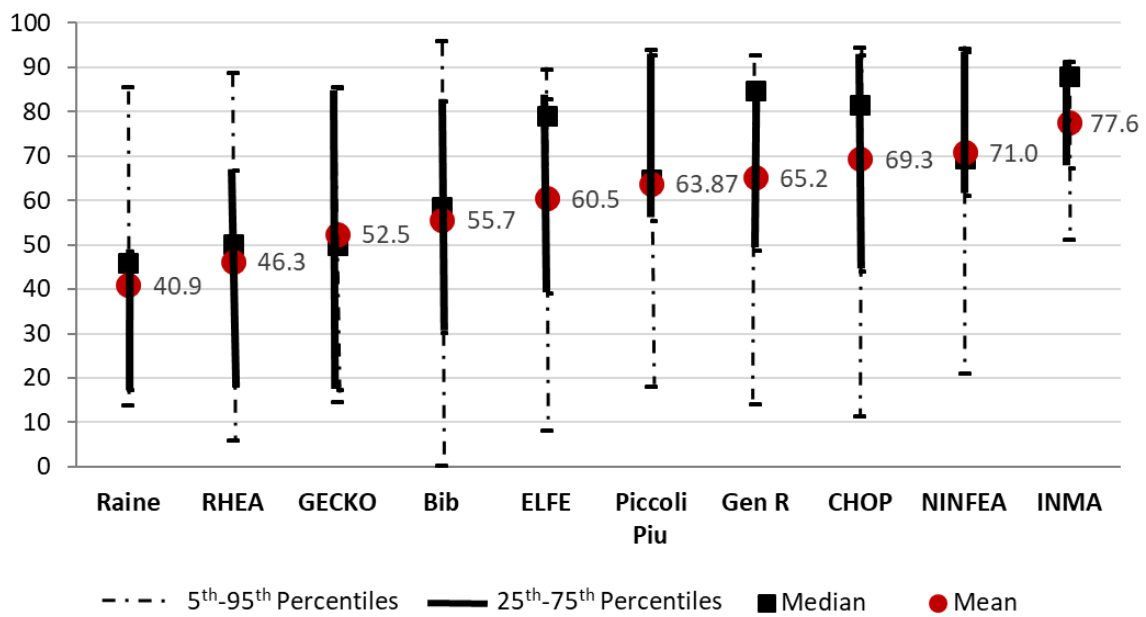
Percentile	Raine	RHEA	GECKO	Bib	ELFE	PiccoliPiu	Gen R	CHOP	NINFEA	INMA
5%	13.9	5.8	14.6	0.3	8.2	18.0	14.0	11.3	20.9	51.1
25%	17.2	18.3	17.2	30.2	38.9	55.3	48.6	43.9	61.1	67.3
50%	46.0	50.1	49.8	58.6	79.2	64.7	84.7	81.6	69.5	88.0
75%	48.5	66.7	85.2	82.4	82.8	92.7	85.4	92.6	93.4	89.7
95%	85.4	88.8	85.4	96.0	89.6	94.0	92.6	94.5	94.2	91.3
Mean	40.9	46.3	52.5	55.7	60.5	63.87	65.2	69.3	71.0	77.6

Fig 3.4. Mother's relative level of education (unweighted)*



*The numbers represent the mean of the distribution for each cohort

Fig 3.5. Father's relative level of education (unweighted)*



*The numbers represent the mean of the distribution for each cohort

4. Variable List for WP 3.1.2

Variable name	Label/description	Values	Units	Type	Comments
miggen_child	Child's migration generation	1 = 1st generation 2 = 2nd generation 3 = 2.5 generation – mother 4 = 2.5 generation – father 5 = Natives		Categorical	1= 1st generation, cohort member is born abroad 2= 2nd generation, cohort member is born in the host country, both parents are born abroad 3= 2.5 generation (mother), cohort member born in the host country, father born in the host country, mother born abroad 4= 2.5 generation (father), cohort member born in the host country, mother born in the host country, father born abroad 5 = natives, cohort member is born in the host country; both parents born in the host country
region_mo	Mother's region of birth	0 = Host country-born 1 = Western EU/EEA 2 = Eastern EU 3 = Other Europe & Central Asia 4 = East Asia & Pacific 5 = South Asia 6 = Middle East & North Africa 7 = Sub-Saharan Africa 8 = Latin America & Caribbean 9 = North America		Categorical	Mother's region of birth, adapted from the World Bank classification
region_fa	Father's region of birth	0 = Host country-born 1 = Western EU/EEA 2 = Eastern EU 3 = Other Europe & Central Asia 4 = East Asia & Pacific 5 = South Asia 6 = Middle East & North Africa 7 = Sub-Saharan Africa 8 = Latin America & Caribbean 9 = North America		Categorical	Father's region of birth, adapted from the World Bank classification
reledu_mo	Mother's relative level of education	Possible values from 0 to 100 Examples: 0, 32.599998, 29.70297, 75, 99.999	Percentile	Continuous	Mother's percentile position in the country of origin's educational attainment distribution of women born in the same year (variable values from 0 – 0th percentile (least selected or negatively selected) to 100 – 100th percentile (most positively selected)).
reledu_fa	Father's relative level of education	Possible values from 0 to 100 Examples: 0, 32.599998, 29.70297, 75, 99.999	Percentile	Continuous	Father's percentile position in the country of origin's educational attainment distribution of men born in the same year (variable values from 0 – 0th percentile (least selected or negatively selected) to 100 – 100th percentile (most positively selected)).
abroad_child	Child's born abroad	0 = No, 1 = Yes		Binary	Whether child cohort member was born abroad. Note: The ELFE sample does not include children born abroad, thus, all observations are coded as 0.
abroad_mo	Mother's born abroad	0 = No, 1 = Yes		Binary	Whether child's mother was born abroad
abroad_fa	Father's born abroad	0 = No, 1 = Yes		Binary	Whether child's father was born abroad

5. Next steps

5.1. Participating cohort status

As of June 2019, all cohorts have completed creating the new harmonized variables. We do not expect other cohorts to participate in this subtask.

5.2. Paper plan

- **Paper 1:** Children of migrants' health at birth. *A descriptive paper to kick-off analyses with relatively easy-to-harmonize outcomes (birthweight, gestational age...)*

Target journal: International Epidemiology Journal

- **Paper 2:** The effect of immigrants' educational selectivity on early child health (pregnancy outcomes).
- **Paper 3:** One or series of descriptive papers describing differences across migrant groups at different ages (including health trajectories where possible).
- **Paper 4:** Mechanisms to explain child health disparities across migrant groups: do they change across countries?

5.3. Uploading data to Opal

- We are working on files needed to upload data to Opal. We are expecting all cohorts to upload the new variables by May 2020.
- We are expecting to work on the analyses for papers 1 starting May 2020.

Annex 1

ISO3 Country codes and country names for each region

Countries-WP3.1.2.xls

<u>countryname</u>	<u>ISO2</u>	<u>ISO3</u>	ISO Numerical		<u>Region</u>	<u>Region value</u>
			<u>Code</u>	<u>UN M49</u>		
Afghanistan	AF	AFG		4	South Asia	5
Albania	AL	ALB		8	Eastern (non-EU) Europe & Central Asia	3
Algeria	DZ	DZA		12	Middle East & North Africa	6
American Samoa	AS	ASM		16	East Asia & Pacific	4
Andorra	AD	AND		20	Western EU/EEA	1
Angola	AO	AGO		24	Sub-Saharan Africa	7
Antigua and Barbuda	AG	ATG		28	Latin America & Caribbean	8
Argentina	AR	ARG		32	Latin America & Caribbean	8
Armenia	AM	ARM		51	Eastern (non-EU) Europe & Central Asia	3
Aruba	AW	ABW		533	Latin America & Caribbean	8
Australia	AU	AUS		36	East Asia & Pacific	4
Austria	AT	AUT		40	Western EU/EEA	1
Azerbaijan	AZ	AZE		31	Eastern (non-EU) Europe & Central Asia	3
Bahamas, The	BS	BHS		44	Latin America & Caribbean	8
Bahrain	BH	BHR		48	Middle East & North Africa	6
Bangladesh	BD	BGD		50	South Asia	5
Barbados	BB	BRB		52	Latin America & Caribbean	8
Belarus	BY	BLR		112	Eastern (non-EU) Europe & Central Asia	3
Belgium	BE	BEL		56	Western EU/EEA	1
Belize	BZ	BLZ		84	Latin America & Caribbean	8
Benin	BJ	BEN		204	Sub-Saharan Africa	7
Bermuda	BM	BMU		60	North America	9
Bhutan	BT	BTN		64	South Asia	5
Bolivia	BO	BOL		68	Latin America & Caribbean	8
Bosnia and Herzegovina	BA	BIH		70	Eastern (non-EU) Europe & Central Asia	3
Botswana	BW	BWA		72	Sub-Saharan Africa	7

Brazil	BR	BRA	76	Latin America & Caribbean	8
British Virgin Islands	VG	VGB	92	Latin America & Caribbean	8
Brunei Darussalam	BN	BRN	96	East Asia & Pacific	4
Bulgaria	BG	BGR	100	Eastern EU	2
Burkina Faso	BF	BFA	854	Sub-Saharan Africa	7
Burundi	BI	BDI	108	Sub-Saharan Africa	7
Cabo Verde	KH	CPV	116	Sub-Saharan Africa	7
Cambodia	CM	KHM	120	East Asia & Pacific	4
Cameroon	CA	CMR	124	Sub-Saharan Africa	7
Canada	CV	CAN	132	North America	9
Cayman Islands	KY	CYM	136	Latin America & Caribbean	8
Central African Republic	CF	CAF	140	Sub-Saharan Africa	7
Chad	TD	TCD	148	Sub-Saharan Africa	7
Channel Islands		CHI		Western EU/EEA	1
Chile	CL	CHL	152	Latin America & Caribbean	8
China	CN	CHN	156	East Asia & Pacific	4
Colombia	CO	COL	170	Latin America & Caribbean	8
Comoros	KM	COM	174	Sub-Saharan Africa	7
Congo, Dem. Rep.	CD	COD	180	Sub-Saharan Africa	7
Congo, Rep.	CG	COG	178	Sub-Saharan Africa	7
Costa Rica	CR	CRI	188	Latin America & Caribbean	8
Cote d'Ivoire	CI	CIV	384	Sub-Saharan Africa	7
Croatia	HR	HRV	191	Eastern EU	2
Cuba	CU	CUB	192	Latin America & Caribbean	8
Curacao		CUW		Latin America & Caribbean	8
Cyprus	CY	CYP	196	Western EU/EEA	1
Czech Republic	CZ	CZE	203	Eastern EU	2
Denmark	DK	DNK	208	Western EU/EEA	1
Djibouti	DJ	DJI	262	Middle East & North Africa	6
Dominica	DM	DMA	212	Latin America & Caribbean	8
Dominican Republic	DO	DOM	214	Latin America & Caribbean	8
Ecuador	EC	ECU	218	Latin America & Caribbean	8
Egypt, Arab Rep.	EG	EGY	818	Middle East & North Africa	6
El Salvador	SV	SLV	222	Latin America & Caribbean	8
Equatorial Guinea	GQ	GNQ	226	Sub-Saharan Africa	7
Eritrea	ER	ERI	232	Sub-Saharan Africa	7
Estonia	EE	EST	233	Eastern EU	2
Ethiopia	ET	ETH	231	Sub-Saharan Africa	7
Faroe Islands	FO	FRO	234	Western EU/EEA	1
Fiji	FJ	FJI	242	East Asia & Pacific	4
Finland	FI	FIN	246	Western EU/EEA	1
France	FR	FRA	250	Western EU/EEA	1
French Polynesia	PF	PYF	258	East Asia & Pacific	4

Gabon	GA	GAB	266	Sub-Saharan Africa	7
Gambia, The	GM	GMB	270	Sub-Saharan Africa	7
Georgia	GE	GEO	268	Eastern (non-EU) Europe & Central Asia	3
Germany	DE	DEU	276	Western EU/EEA	1
Ghana	GH	GHA	288	Sub-Saharan Africa	7
Gibraltar	GI	GIB	292	Western EU/EEA	1
Greece	GR	GRC	300	Western EU/EEA	1
Greenland	GL	GRL	304	Western EU/EEA	1
Grenada	GD	GRD	308	Latin America & Caribbean	8
Guam	GU	GUM	316	East Asia & Pacific	4
Guatemala	GT	GTM	320	Latin America & Caribbean	8
Guernsey	GG	GGY	831	Western EU/EEA	7
Guinea	GN	GIN	324	Sub-Saharan Africa	7
Guinea-Bissau	GW	GNB	624	Sub-Saharan Africa	7
Guyana	GY	GUY	328	Latin America & Caribbean	8
Haiti	HT	HTI	332	Latin America & Caribbean	8
Honduras	HN	HND	340	Latin America & Caribbean	8
Hong Kong SAR, China	HK	HKG	344	East Asia & Pacific	4
Hungary	HU	HUN	348	Eastern EU	2
Iceland	IS	ISL	352	Western EU/EEA	1
India	IN	IND	356	South Asia	5
Indonesia	ID	IDN	360	East Asia & Pacific	4
Iran, Islamic Rep.	IR	IRN	364	Middle East & North Africa	6
Iraq	IQ	IRQ	368	Middle East & North Africa	6
Ireland	IE	IRL	372	Western EU/EEA	1
Isle of Man	IM	IMN	833	Western EU/EEA	1
Israel	IL	ISR	376	Middle East & North Africa	6
Italy	IT	ITA	380	Western EU/EEA	1
Jamaica	JM	JAM	388	Latin America & Caribbean	8
Japan	JP	JPN	392	East Asia & Pacific	4
Jersey	JE	JEY	832	Western EU/EEA	1
Jordan	JO	JOR	400	Middle East & North Africa	6
Kazakhstan	KZ	KAZ	398	Eastern (non-EU) Europe & Central Asia	3
Kenya	KE	KEN	404	Sub-Saharan Africa	7
Kiribati	KI	KIR	296	East Asia & Pacific	4
Korea, Dem. People's Rep.	KP	PRK	408	East Asia & Pacific	4
Korea, Rep.	KR	KOR	410	East Asia & Pacific	4
Kosovo		XKX		Eastern (non-EU) Europe & Central Asia	3
Kuwait	KW	KWT	414	Middle East & North Africa	6

Kyrgyz Republic	KG	KGZ	417	Eastern (non-EU) Europe & Central Asia	3
Lao PDR	LA	LAO	418	East Asia & Pacific	4
Latvia	LV	LVA	428	Eastern EU	2
Lebanon	LB	LBN	422	Middle East & North Africa	6
Lesotho	LS	LSO	426	Sub-Saharan Africa	7
Liberia	LR	LBR	430	Sub-Saharan Africa	7
Libya	LY	LBY	434	Middle East & North Africa	6
Liechtenstein	LI	LIE	438	Western EU/EEA	1
Lithuania	LT	LTU	440	Eastern EU	2
Luxembourg	LU	LUX	442	Western EU/EEA	1
Macao SAR, China	MO	MAC	446	East Asia & Pacific	4
Macedonia, FYR	MK	MKD	807	Eastern (non-EU) Europe & Central Asia	3
Madagascar	MG	MDG	450	Sub-Saharan Africa	7
Malawi	MW	MWI	454	Sub-Saharan Africa	7
Malaysia	MY	MYS	458	East Asia & Pacific	4
Maldives	MV	MDV	462	South Asia	5
Mali	ML	MLI	466	Sub-Saharan Africa	7
Malta	MT	MLT	470	Western EU/EEA	1
Marshall Islands	MH	MHL	584	East Asia & Pacific	4
Mauritania	MR	MRT	478	Sub-Saharan Africa	7
Mauritius	MU	MUS	480	Sub-Saharan Africa	7
Mexico	MX	MEX	484	Latin America & Caribbean	8
Micronesia, Fed. Sts.	FM	FSM	583	East Asia & Pacific	4
Moldova	MD	MDA	498	Eastern (non-EU) Europe & Central Asia	3
Monaco	MC	MCO	492	Western EU/EEA	1
Mongolia	MN	MNG	496	East Asia & Pacific	4
Montenegro	ME	MNE	499	Eastern (non-EU) Europe & Central Asia	3
Morocco	MA	MAR	504	Middle East & North Africa	6
Mozambique	MZ	MOZ	508	Sub-Saharan Africa	7
Myanmar	MM	MMR	104	East Asia & Pacific	4
Namibia	NA	NAM	516	Sub-Saharan Africa	7
Nauru	NR	NRU	520	East Asia & Pacific	4
Nepal	NP	NPL	524	South Asia	5
Netherlands	NL	NLD	528	Western EU/EEA	1
New Caledonia	NC	NCL	540	East Asia & Pacific	4
New Zealand	NZ	NZL	554	East Asia & Pacific	4
Nicaragua	NI	NIC	558	Latin America & Caribbean	8
Niger	NE	NER	562	Sub-Saharan Africa	7
Nigeria	NG	NGA	566	Sub-Saharan Africa	7

Northern Mariana Islands	MP	MNP	580	East Asia & Pacific	4
Norway	NO	NOR	578	Western EU/EEA	1
Oman	OM	OMN	512	Middle East & North Africa	6
Pakistan	PK	PAK	586	South Asia	5
Palau	PW	PLW	585	East Asia & Pacific	4
Palestinian Territory	PS	PSE	275	Middle East & North Africa	6
Panama	PA	PAN	591	Latin America & Caribbean	8
Papua New Guinea	PG	PNG	598	East Asia & Pacific	4
Paraguay	PY	PRY	600	Latin America & Caribbean	8
Peru	PE	PER	604	Latin America & Caribbean	8
Philippines	PH	PHL	608	East Asia & Pacific	4
Poland	PL	POL	616	Eastern EU	2
Portugal	PT	PRT	620	Western EU/EEA	1
Puerto Rico	PR	PRI	630	Latin America & Caribbean	8
Qatar	QA	QAT	634	Middle East & North Africa	6
Romania	RO	ROU	642	Eastern EU	2
Russian Federation	RU	RUS	643	Eastern (non-EU) Europe & Central Asia	3
Rwanda	RW	RWA	646	Sub-Saharan Africa	7
Samoa	WS	WSM	882	East Asia & Pacific	4
San Marino	SM	SMR	674	Western EU/EEA	1
Sao Tome and Principe	ST	STP	678	Sub-Saharan Africa	7
Saudi Arabia	SA	SAU	682	Middle East & North Africa	6
Senegal	SN	SEN	686	Sub-Saharan Africa	7
Serbia	RS	SRB	688	Eastern (non-EU) Europe & Central Asia	3
Seychelles	SC	SYC	690	Sub-Saharan Africa	7
Sierra Leone	SL	SLE	694	Sub-Saharan Africa	7
Singapore	SG	SGP	702	East Asia & Pacific	4
Sint Maarten (Dutch part)		SXM		Latin America & Caribbean	8
Slovak Republic	SK	SVK	703	Eastern EU	2
Slovenia	SI	SVN	705	Eastern EU	2
Solomon Islands	SB	SLB	90	East Asia & Pacific	4
Somalia	SO	SOM	706	Sub-Saharan Africa	7
South Africa	ZA	ZAF	710	Sub-Saharan Africa	7
South Sudan	SS	SSD	728	Sub-Saharan Africa	7
Spain	ES	ESP	724	Western EU/EEA	1
Sri Lanka	LK	LKA	144	South Asia	5
St. Kitts and Nevis	KN	KNA	659	Latin America & Caribbean	8
St. Lucia	LC	LCA	662	Latin America & Caribbean	8
St. Martin (French part)	MF	MAF	663	Latin America & Caribbean	8
St. Vincent and the Grenadines	VC	VCT	670	Latin America & Caribbean	8

Sudan	SD	SDN	736	Sub-Saharan Africa	7
Suriname	SR	SUR	740	Latin America & Caribbean	8
Swaziland	SZ	SWZ	748	Sub-Saharan Africa	7
Sweden	SE	SWE	752	Western EU/EEA	1
Switzerland	CH	CHE	756	Western EU/EEA	1
Syrian Arab Republic	SY	SYR	760	Middle East & North Africa	6
Taiwan, China	TW	TWN	158	East Asia & Pacific	4
				Eastern (non-EU) Europe & Central	
Tajikistan	TJ	TJK	762	Asia	3
Tanzania	TZ	TZA	834	Sub-Saharan Africa	7
Thailand	TH	THA	764	East Asia & Pacific	4
Timor-Leste	TL	TLS	626	East Asia & Pacific	4
Togo	TG	TGO	768	Sub-Saharan Africa	7
Tonga	TO	TON	776	East Asia & Pacific	4
Trinidad and Tobago	TT	TTO	780	Latin America & Caribbean	8
Tunisia	TN	TUN	788	Middle East & North Africa	6
				Eastern (non-EU) Europe & Central	
Turkey	TR	TUR	792	Asia	3
				Eastern (non-EU) Europe & Central	
Turkmenistan	TM	TKM	795	Asia	3
Turks and Caicos Islands	TC	TCA	796	Latin America & Caribbean	8
Tuvalu	TV	TUV	798	East Asia & Pacific	4
Uganda	UG	UGA	800	Sub-Saharan Africa	7
				Eastern (non-EU) Europe & Central	
Ukraine	UA	UKR	804	Asia	3
United Arab Emirates	AE	ARE	784	Middle East & North Africa	6
United Kingdom	GB	GBR	826	Western EU/EEA	1
United States	US	USA	840	North America	9
Uruguay	UY	URY	858	Latin America & Caribbean	8
				Eastern (non-EU) Europe & Central	
Uzbekistan	UZ	UZB	860	Asia	3
Vanuatu	VU	VUT	548	East Asia & Pacific	4
Venezuela, RB	VE	VEN	862	Latin America & Caribbean	8
Vietnam	VN	VNM	704	East Asia & Pacific	4
Virgin Islands (U.S.)	VI	VIR	850	Latin America & Caribbean	8
West Bank and Gaza	PS	PSE	275	Middle East & North Africa	6
Yemen, Rep.	YE	YEM	887	Middle East & North Africa	6
Zambia	ZM	ZMB	894	Sub-Saharan Africa	7
Zimbabwe	ZW	ZWE	716	Sub-Saharan Africa	7

Annex 2

ISO3 Integrated educational distribution dataset

EducAttain.csv – EXTRACT SAMPLE*

*The entire EducAttain.csv file is available on the [LifeCycle Intranet](#)

cntry_mo	cntry_fa	yob_mo	yob_fa	edu_mo	edu_fa	reledu_mo	reledu_fa
ALB	ALB	1970	1970	Pre-primary	Pre-primary	0.50505048	0.5
ALB	ALB	1970	1970	Primary	Primary	1.5151515	2.5
				Lower	Lower		
ALB	ALB	1970	1970	secondary	secondary	25.252525	25
				Upper	Upper		
ALB	ALB	1970	1970	secondary	secondary	70.202019	68
ALB	ALB	1970	1970	Tertiary	Tertiary	95.959595	95
ALB	ALB	1971	1971	Pre-primary	Pre-primary	0.4040404	0.49900991
ALB	ALB	1971	1971	Primary	Primary	1.4121212	2.5940595
				Lower	Lower		
ALB	ALB	1971	1971	secondary	secondary	24.50202	24.752476
				Upper	Upper		
ALB	ALB	1971	1971	secondary	secondary	69.261612	67.667328
ALB	ALB	1971	1971	Tertiary	Tertiary	95.767677	95.009903
...							
CHE	CHE	1980	1980	Pre-primary	Pre-primary	0	0
CHE	CHE	1980	1980	Primary	Primary	0.50505048	0.50505048
				Lower	Lower		
CHE	CHE	1980	1980	secondary	secondary	7.575757	7.0707068
				Upper	Upper		
CHE	CHE	1980	1980	secondary	secondary	43.939392	38.383835
CHE	CHE	1980	1980	Tertiary	Tertiary	86.868683	81.818176
...							
CHN	CHN	1991	1991	Pre-primary	Pre-primary	0.49504951	0
CHN	CHN	1991	1991	Primary	Primary	6.4356437	4.5454545
				Lower	Lower		
CHN	CHN	1991	1991	secondary	secondary	36.633663	34.848484
				Upper	Upper		
CHN	CHN	1991	1991	secondary	secondary	71.782181	71.717171
CHN	CHN	1991	1991	Tertiary	Tertiary	91.089111	91.414139

...							
FRA	FRA	1982	1982	Pre-primary	Pre-primary	0.5	0.49504951
FRA	FRA	1982	1982	Primary	Primary	3.9000001	4.0594063
				Lower	Lower		
FRA	FRA	1982	1982	secondary	secondary	10.6	11.386139
				Upper	Upper		
FRA	FRA	1982	1982	secondary	secondary	33.400002	38.217823
FRA	FRA	1982	1982	Tertiary	Tertiary	76.199997	80.396042
...							
HND	HND	1984	1984	Pre-primary	Pre-primary	12.1	14
HND	HND	1984	1984	Primary	Primary	43.5	48.299999
				Lower	Lower		
HND	HND	1984	1984	secondary	secondary	67.699997	73.099998
				Upper	Upper		
HND	HND	1984	1984	secondary	secondary	81.900002	84.400002
HND	HND	1984	1984	Tertiary	Tertiary	95.599998	95.599998
...							
POL	POL	1993	1993	Pre-primary	Pre-primary	0	0
POL	POL	1993	1993	Primary	Primary	0	0
				Lower	Lower		
POL	POL	1993	1993	secondary	secondary	0.99009901	1.5
				Upper	Upper		
POL	POL	1993	1993	secondary	secondary	31.683168	39
POL	POL	1993	1993	Tertiary	Tertiary	80.693069	87.5
...							
SYR	SYR	1992	1992	Pre-primary	Pre-primary	14.851485	14.356436
SYR	SYR	1992	1992	Primary	Primary	42.574257	43.069309
				Lower	Lower		
SYR	SYR	1992	1992	secondary	secondary	60.89109	62.871288
				Upper	Upper		
SYR	SYR	1992	1992	secondary	secondary	73.762375	75.742577
SYR	SYR	1992	1992	Tertiary	Tertiary	90.594063	91.58416
...							
YEM	YEM	1981	1981	Pre-primary	Pre-primary	38.127537	20.956982
YEM	YEM	1981	1981	Primary	Primary	80.791367	51.986019



Instructions for Quality Control
of Harmonized WP 3.1.2 Variables

WP 3.1.2: Ethnicity, migration, and integration

Table of Contents

Rationale..... 3

Step 1: Verify list of variables and formats 4

Step 2: Check univariate distributions 5

Step 3: Check internal validation..... 7

Step 4: Complete the Online Catalogue 7

Rationale

As the final part of harmonizing the data in Work Package 3.1.2 of LifeCycle, the list of harmonized variables needs a local validation. In this document, instructions for such local quality control are described.

Accordingly, each cohort in LifeCycle needs to perform quality checks to validate and evaluate their construction of the harmonized variables, and align this information with that documented in the *Online Catalogue*: <https://molgenis88.gcc.rug.nl/>. The Quality Control must be in accordance with the *WP 3.1.2 Variable List*, which was updated in April 2020.

To this end, we ask each cohort to carefully read these instructions for a local validation, and follow the syntax from the Stata code to replicate the quality control on their own cohort data.

The aim of this Quality Control is firstly to check that each cohort's harmonized WP 3.1.2 variables match those described in the LifeCycle Catalogue, and secondly to ensure the quality of harmonization. If any errors are found these should be amended accordingly. Each cohort should carry out the quality checks and any necessary corrections themselves.

If you have any queries about the quality control of the harmonized variables please make contact to Sandra Florian (sandra.florian@ined.fr), Mathieu Ichou (mathieu.ichou@ined.fr), or Lidia Panico (lidia.panico@ined.fr) from WP 3.1.2.

Step 1: Verify list of variables and formats

Please, verify that your cohort-specific harmonized WP 3.1.2 variables completely match with the information provided in the WP3.1.2 Variable List.

Each cohort needs to check that the *name* and *data type* of each of the variables correspond exactly to the WP3.1.2. Variable List (see the Excel file attached). Please, see the following examples of Step 1 in the Quality Control:

Example: miggen_child

Variable name	Label/description	Values	Units	Type	Comments
miggen_child	Child's migration generation	1 = 1st generation 2 = 2nd generation 3 = 2.5 generation – mother 4 = 2.5 generation – father 5 = Natives		Categorical	1= 1st generation, cohort member is born abroad 2= 2nd generation, cohort member is born in the host country, both parents are born abroad 3= 2.5 generation (mother), cohort member born in the host country, father born in the host country, mother born abroad 4= 2.5 generation (father), cohort member born in the host country, mother born in the host country, father born abroad 5 = natives, cohort member is born in the host country; both parents born in the host country
region_mo	Mother's region of birth	0 = Host country-born 1 = Western EU/EEA 2 = Eastern EU 3 = Other Europe & Central Asia		Categorical	Mother's region of birth, adapted from the World Bank classification

1. Check the variable name ('miggen_child')
2. Check that the number of categories match (five). Note: the first category (1 = 1st generation) will be empty for most cohorts.

Example: reledu_mo

Variable name	Label/description	Values	Units	Type	Comments
reledu_mo	Mother's relative level of education	Possible values from 0 to 100 Examples: 0, 32.599998, 29.70297, 75, 99.999	Percentile	Continuous	Mother's percentile position in the country of origin's educational attainment distribution of women born in the same year (variable values from 0 – least negatively selected to 100 – most positively selected)
reledu_fa	Father's relative level of education	Possible values from 0 to 100 Examples: 0, 32.599998, 29.70297, 75, 99.999	Percentile	Continuous	Father's percentile position in the country of origin's educational attainment distribution of men born in the same year (variable values from 0 – least negatively selected to 100 – most positively selected)

1. Check the variable name ('reledu_mo')
2. Check that values are continuous

If any mismatch is observed, please correct the errors accordingly.

Step 2: Check univariates distributions

Please, generate univariate distributions for all variables, and check for outliers and improbable values. Also, for variables that have been reported in papers/publications, verify that distributions or summary statistics of the harmonized LifeCycle variable match those of the reported/published variables.

For continuous variables, check that there are no outliers, i.e. values out of the minimum and maximum range based on the variable list. Please, use your scientific knowledge and practical sense when making this quality check so as not to drop interesting outliers. Errors are probably caused by an error in your harmonization script, so please check this carefully and correct where required.

Example: reledu_mo & reledu_fa

Check whether values of **reledu_mo** and **reledu_fa** are in the expected range.

The variables measure the mother's (or father's) percentile position in the country of origin's educational attainment distribution of women (men) born in the same year. The possible values range from 0 – 0th percentile (least selected or negatively selected) to 100 – 100th percentile (most positively selected).

```
. sum reledu_mo reledu_fa
```

Variable	Obs	Mean	Std. Dev.	Min	Max
reledu_mo	15,774	60.24596	25.85315	0	99.99999
reledu_fa	14,319	60.54097	26.49996	0	99.6

Output from STATA

For categorical variables, check that there are no improbable values, i.e. values not corresponding to the categories defined in the Variable List. Please, correct errors where relevant.

Example: region_mo

Variable name	Label/description	Values	Units	Type	Comments
region_mo	Mother's region of birth	0 = Host country-born 1 = Western EU/EEA 2 = Eastern EU 3 = Other Europe & Central Asia 4 = East Asia & Pacific 5 = South Asia 6 = Middle East & North Africa 7 = Sub-Saharan Africa 8 = Latin America & Caribbean 9 = North America		Categorical	Mother's region of birth, adapted from the World Bank classification
region_fa	Father's region of birth	0 = Host country-born 1 = Western EU/EEA		Categorical	Father's region of birth, adapted from the World Bank classification

Check that data is actually coded into the same categories as the Variable List (see output):

```
. tab region_mo
```

Mother's region of birth	Freq.	Percent	Cum.
0. Host country-born	14,839	86.70	86.70
1. Western EU/EEA	278	1.62	88.33
2. Eastern EU	101	0.59	88.92
3. Other Europe & Central Asia	140	0.82	89.73
4. East Asia & Pacific	129	0.75	90.49
5. South Asia	23	0.13	90.62
6. Middle East & North Africa	858	5.01	95.64
7. Sub-Saharan Africa	604	3.53	99.16
8. Latin America & Caribbean	125	0.73	99.89
9. North America	18	0.11	100.00
Total	17,115	100.00	

Output from STATA

Example: abroad_mo

This is a binary variable with two possible values

Variable name	Label/description	Values	Units	Type	Comments
abroad_mo	Mother's born abroad	0 = No, 1 = Yes		Binary	
abroad_fa	Father's born abroad	0 = No, 1 = Yes		Binary	

Check that data actually take those two possible values according to the Variable List:

```
. tab abroad_mo
```

Mother's born abroad	Freq.	Percent	Cum.
0. No	14,865	86.16	86.16
1. Yes	2,387	13.84	100.00
Total	17,252	100.00	

Output from STATA

Step 3: Check internal validation

Internal validation is an important part of the local quality control. Thus, within reason cross-tabulate the variables against other variables to check for consistency.

For example, run a cross-tab between region_mo and miggen_child. The first row for region_mo (0) Host country) should be empty (0) for the columns of miggen_child: 2) 2nd generation & 3) 2.5 generation-mother. A cross-tab between region_mo and abroad_mo will be also informative.

Step 4: Complete the Online Catalogue

As part of the Quality Control, each cohort must ensure that the Online Catalogue is aligned accordingly. This means that in the Online Catalogue description of harmonization is complete and information in all three tabs ('description', 'variables used' and 'script syntax') are completed in full for the 8 variables in WP 3.1.2.

Please follow Step 5 of the instructions for quality control of WP1 to update the Catalogue.