ÖAW AUSTRIAN ACADEMY OF SCIENCES

VIENNA INSTITUTE OF DEMOGRAPHY

WORKING PAPERS

02/2019

GLOBAL RECONSTRUCTION OF EDUCATIONAL ATTAINMENT, 1950 TO 2015: METHODOLOGY AND ASSESSMENT

MARKUS SPERINGER, ANNE GOUJON, SAMIR K.C., MICHAELA POTANČOKOVÁ, CLAUDIA REITER, SANDRA JURASSZOVICH AND JAKOB EDER

Vienna Institute of Demography
Austrian Academy of Sciences
Welthandelsplatz 2, Level 2 | 1020 Wien, Österreich
vid@oeaw.ac.at | www.oeaw.ac.at/vid

Vienna Institute of Demography

Abstract

This paper documents the rationale, the data and the methodology for reconstructing the population of 185 countries by levels of educational attainment for the period 1950–2015, by age and sex. The reconstruction uses four main input types for each country: **(1)** The most recent and reliable education structure by age and sex, **(2)** any reliable historical education data by age and sex to use as marker points in the reconstruction to increase output accuracy, **(3)** a set of age- and sex-specific mortality differentials and education transition by education and **(4)** population estimates by age and sex. The methodology relies on the fact that education is acquired at young ages and does not change much over the life course. In the first part we present the reconstruction principle. In the second one, we document the methodology and the data. The third section compares the reconstructed estimates to other existing estimates including the past reconstruction effort of the Wittgenstein Centre for Demography and Human Capital. The data are available at: www.wittgensteincentre.org/dataexplorer (version 2.0). Supplementary to this Working Paper a detailed data documentation Excel file can be downloaded via: https://www.oeaw.ac.at/vid/publications/serial-publications/vid-working-papers/.

Keywords

Educational attainment, human capital, reconstruction, back-projection, modelling.

Authors

Markus Speringer (corresponding author), Wittgenstein Centre for Demography and Global Human Capital (IIASA, VID/ÖAW, WU), Vienna Institute of Demography, Austrian Academy of Sciences and Department for Geography and Regional Research, University of Vienna. Email: markus.speringer@oeaw.ac.at | markus.speringer@univie.ac.at

Anne Goujon, Wittgenstein Centre for Demography and Global Human Capital (IIASA, VID/ÖAW, WU), Vienna Institute of Demography, Austrian Academy of Sciences and World Population Program, International Institute for Applied Systems Analysis. Email: anne.goujon@oeaw.ac.at

Samir K.C., Wittgenstein Centre for Demography and Global Human Capital (IIASA, VID/ÖAW, WU), International Institute for Applied Systems Analysis and Asian Demographic Research Institute, University of Shanghai (ADRI). Email: kc@iiasa.ac.at

Michaela Potančoková, Wittgenstein Centre for Demography and Global Human Capital (IIASA, VID/ÖAW, WU), Vienna Institute of Demography, Austrian Academy of Sciences and World Population Program, International Institute for Applied Systems Analysis.

Claudia Reiter, Wittgenstein Centre for Demography and Global Human Capital (IIASA, VID/ÖAW, WU), International Institute for Applied Systems Analysis. Email: reiter@iiasa.ac.at

Sandra Jurasszovich, Wittgenstein Centre for Demography and Global Human Capital (IIASA, VID/ÖAW, WU), Vienna Institute of Demography, Austrian Academy of Sciences.

Jakob Eder, Wittgenstein Centre for Demography and Global Human Capital (IIASA, VID/ÖAW, WU), Vienna Institute of Demography, Austrian Academy of Sciences.

Acknowledgements

Many institutions, colleagues and individuals have helped us in the collection of the base year and historical data. We are particularly thankful to Robert McCaa and his team at the Minnesota Population Center (IPUMS), to Patrick Gerland and his colleagues at the United Nations Population Division (UNPD), to Ariel Lebowitz at the Dag Hammarskjöld Library (United Nations Library), to Lucy McCann and her colleagues at the Bodleian Library (University of Oxford), to David W. Waters and others at the Library of the U.S. Congress, to Dominique Diguet and Karin Sohler at the library of the French Institute for Demographic Studies (INED), to Marie-France Scansaroli and André Lebrun and others at the library of the National Institute for Statistics and Economic Studies (INSEE). We would also like to thank many anonymous employees who have answered our data requests at National Statistics Offices (NSO) and Archives. Special thanks also to Siegfried Gruber (University of Graz), Gilles Pison (INED), Richard Gisser, Wolfgang Lutz (both Wittgenstein Centre) and Ramon Bauer (MA23) for valuable feedback and leads how to pursue this work, and to Guy Abel, Dilek Yildiz (both Wittgenstein Centre), Christian Wegner-Siegmundt and Marcus Wurzer for tips and tricks when it came to modelling in R.

Table of Contents

Global Reconstruction of Educational Attainment, 1950 to 2015: Methodology and Assessment

Markus Speringer, Anne Goujon, Samir K.C., Michaela Potančoková, Claudia Reiter, Sandra Jurasszovich, Jakob Eder

1 Introduction

The research presented here is part of the several major ongoing efforts to reconstruct past levels of educational attainment (Lutz et al. 2007a, 2007b; Cohen and Soto 2007; Fuente and Doménech 2013; Cohen and Leker 2014; Barro and Lee 2015; Goujon et al. 2016; Speringer, Goujon and Jurasszovich 2018). As mentioned in Goujon et al. (2016), the need for a reconstruction of time series on educational attainment is justified on two main grounds related to demand and supply. The demand side are the global modelling exercises that usually require educational attainment as an input or control variable to assess the impact of educational attainment in some of the major past and present changes whether they are of socio-economic, technological or environmental nature in the medium to long run. On the supply side, we cannot help noticing that data on educational attainment suffer from several flaws that prevent comparison across years or countries.

The several reconstruction exercises above vary in the methodology and in the empirical data that are used for the reconstruction. The main problem is usually with the latter as one observes jumps if the data are not checked carefully. This is the major strength of the methodology proposed here: all data points, whether base-year or historical data, have been checked thoroughly and harmonised when needed to fit the selection criteria for inclusion into the reconstruction.

This paper documents the work to update, extend and improve the previous reconstruction dataset, which will be further referred to as *"WIC 2014"*, on educational attainment (Goujon et al. 2016). The updated reconstruction dataset will be referred to as *"WIC 2018"* dataset.

What is new in the present WIC 2018 reconstruction? – First of all, the period covered by the WIC 2014 dataset was extended from a previous coverage from 1970 to 2010 (Goujon et al. 2016) to a coverage from 1950 to 2015. This became possible due to the increasing availability of more recent base-year (2010-round census) data and the use of historical data in the reconstruction process. ***Second***, the geographical coverage has been increased from 171 (WIC 2014 dataset) to 185 countries (WIC 2018) through a thorough search of available historical data sources.

After a short report on the reconstruction principles in Section 2, we present the process behind these novelties in Section 3: assembling the base-year and historical data and documenting the reconstruction methodology. The latter is based on the fact that people rarely change their levels of educational attainment once they have acquired them during childhood and at young ages. As a result, information about the education of a 50-year old in 2015 can be back-projected or reconstructed to 1995 when this person was 30-year old. What is possible at the individual level can be done at the aggregate level as well, after adjusting for mortality and migration.

In Section 4, we document the assessment of the reconstructed WIC 2018 dataset by comparing it with two alternative reconstruction exercises. The first one is the earlier version of this dataset, namely the WIC 2014 reconstruction (Goujon et al. 2016), and the second refers to the latest reconstruction exercise by Barro and Lee (2015). Not surprisingly, the most visible differences can be observed in more recent years due to different base-year datasets. The data are available online (version 2.0) in the Wittgenstein Centre Data Explorer (Wittgenstein Centre 2018).[1]

2 Reconstruction Principles

The need for a global reconstruction of educational attainment for the period 1950–2015 emerges because of the lack of empirical time series on educational attainment. We developed a distinct research design to create such a globally comprehensive dataset. It makes use of recent and historical educational attainment data by age and sex as input data for the reconstruction model called ***Iterative Multi-dimensional Cohort-component Reconstruction*** (IMCR) model. The reconstruction exercise consists of multiple steps which are listed below before contextualising the necessity of each step and referencing to the detailed description of the workflow in Section 3:

(A) Collection, adjustment and harmonisation of **base-year data** (see Section 3.1);

(B) Collection, adjustment and harmonisation of **historical data** (see Section 3.2), including the preparation of data on population and mortality (see Section 3.2.1);

(C) **Reconstruction** of past educational composition (see Section 0), including education-specific mortality (see Section 3.3.1), education transitions (see Section 3.3.2) and projection to unified baseline year 2015 (see Section 3.3.3);

(D) **Assessment** and comparison of reconstructed dataset **with alternative datasets** (see Section 4), including the WIC 2014 (see Section 4.1) and Barro & Lee (see Section 4.2) datasets;

[1] Available at: www.wittgensteincentre.org/dataexplorer

The reconstruction follows the principle of back projections. Like in a forward projection, back projections require the availability of **(ad A)** a **base-year dataset**, in this case the most recent valid data on country-specific educational attainment (see Section 3.1). We assembled data from multiple sources and years. The most accurate dataset was chosen based on several criteria detailed in Section 3.1.1. The dataset was further adjusted and harmonised to fit standard education categories (see Section 3.1.2). A feasible base-year dataset as primary input data determines to an important extent the quality of the reconstruction exercise.

The base year is used **(ad C)** in the **reconstruction model** to go back in time using the general principles that **(1)** education is predominantly acquired at young ages, and **(2)** that education is acquired in a unidirectional mode. Individuals can only add skills and educational levels until reaching their personal highest educational level, which becomes a fixed attribute for the remaining life. Already achieved educational levels cannot be reversed. This allows to follow the educational progress of an individual back in time along cohort lines (Goujon et al. 2016; Speringer, Goujon and Jurasszovich 2018).

Therefore, a person *i* should have the same education at time *t* and at time *t-5* in the period after leaving school/university until death. This is also valid at the aggregate level: the share of the population by level of educational attainment can be back-projected in time along cohort lines. In country *j*, the share *s* of population with education *k* should be similar at time *t* and at time *t-5*. However, two factors could upset this equivalence and affect the educational distribution:

- **Mortality:** if the probability of surviving between time *t-5* and time *t* differs by level of education *k;*
- **Migration:** if the probability of (in- or out-)migration between time *t-5* and time *t* differs by level of education *k;*

To account for education-specific differentials in mortality, we used information on mortality rates from the United Nations Population Division (2017) and applied standard mortality differentials by levels of education as developed by Lutz, Butz and KC (2014). The differentials vary between genders and across the reconstruction period (see Sections 3.2.2.1 and 3.3.1). It is not possible to account for education-specific differentials in migration since these cannot be standardised in the same way as for mortality. However, by using historical data points as mentioned below, we expect that the effect of migration on the education structure will be taken care of.

While the above reconstruction procedure is valid for the out-of-school/university population, this is not necessarily the case for the schooling/studying age groups of 15 to 34 years.[2] Therefore we developed a procedure to calculate over the reconstruction period

[2] Disaggregation by education starts at age 15. Based on evidence, we consider that most education transitions happen by the age of 35.

country-specific education transition rates that are applied to the population below the age of 35 years (see Section 3.3.2).

It is worth remembering that the reconstruction is done for the education distribution (in %) by age and sex, from 1950 to 2015. These shares are applied to the population by age and sex for the same period as estimated for each country of the world by the United Nations Population Division (2017). The reconstruction principles are to a large extent the same as those used in the WIC 2014 dataset, though the periodical adaptation of education-specific mortality rates is a novelty in this version, and the estimation of education transitions is carried out in a different way in WIC 2018. Another innovation in WIC 2018 is the use of information contained **(ad B)** in historical datasets on education (and literacy) when available for the period 1950 to 2015 (see Section 3.2). The collected historical datasets were checked for accuracy and usability (see Section 3.2.1). They are used for two main purposes. First, they replaced missing education data when age groups are depleted through the reconstruction. For instance, the share of the population by levels of education for the age group 100+ in 2010 will be used to calculate the share of the population in age group 95–99 in 2005. However, the share of the 100+ population in 2005 will most likely not be available and has to be estimated. In this way we incorporate historical cohort information on the educational composition in the reconstruction process (see Section 3.2.2.3). The second purpose of historical data is to check the accuracy of the reconstruction output at each step.[3] In a final step, **(ad D)** the quality of the reconstructed WIC 2018 dataset is assessed and compared with alternative reconstruction exercises.

3 Data and Methodology

3.1 Assembling the Base-year Data

3.1.1 Base-year Data Collection: Coverage & Data Sources

The base-year data on population by age, sex and educational attainment described below serves as a starting point for both the multistate population projections (Lutz et al. 2018) and the update of the reconstruction that is documented in this report. The previous reconstruction exercise, WIC 2014, was based on the collected and harmonised base-year data for 171 countries (Bauer et al. 2012). However, this dataset relied mostly on the information from the 2000 census round, as the more recent data from the 2010 census round were not yet accessible at the time of the data collection. We also took the opportunity to fill data gaps and improve the quality of the dataset.[4]

[3] In the WIC 2014 we were also using the historical data, but only to assess reconstruction results and not as marker points in the reconstruction model itself.

[4] Furthermore, we have implemented the new ISCED 2011 classification (UNESCO 2012) to disaggregate the post-secondary education category into three education categories of higher education for countries with available data (mostly OECD countries). We document this work in

The starting point for the update of the educational attainment by age and sex was the previous dataset that included information for 171 countries (Bauer et al. 2012). As previously, the aim was to collect as recent as possible information on population by age, sex and educational attainment for 201 countries listed in the 2017 Revision of the UN World Population Prospects. In terms of geographical coverage, it was possible to collect and harmonise data for 185 countries (92% of all countries), covering 99% of the world's population. This makes this dataset the most comprehensive internationally.[5]

In comparison to the WIC 2014 dataset, the country coverage has improved from 171 countries covering 88% of all countries and 97.4% of the world's population (Goujon et al. 2016). Table 1 summarises the data availability and lists newly added countries and those with missing education data. The countries with missing education data are not reconstructed as they are lacking the base-year information.[6]

Some of the 16 new countries that were added to the dataset came to existence due to political changes—e.g. Sudan split into Sudan and South Sudan, while several others were added to the data collection through population growth as they have exceeded or came close to the 100,000 population size threshold, e.g. Kiribati. To highlight some modifications in comparison to the WIC 2014 dataset: in the WIC 2018 dataset the Netherlands Antilles were replaced by Curaçao, the only one of the now independent entities with a population exceeding the 100,000 threshold. In addition, Taiwan, which had been included under China in WIC 2014, was added as a separate entity.

Looking at the continents, Table 1 shows that the data coverage increased most for Oceania (up from covering 76% to 80% of the region's population), followed by Africa, (from 96% to 99% coverage) and Asia (97% to 99%) and with small improvements for Latin America. The coverage did not change for the remaining regions. In Africa, some sizeable countries such as Angola and Botswana were added. New datasets also became available for Afghanistan, Oman, North Korea, Sri Lanka and Yemen.

Annex B as it is mostly relevant for the projections (to 2100—see Lutz et al. 2018) and not for the back-projections.

[5] For comparison, the Barro and Lee dataset covers in its latest version 146 counties (Barro and Lee 2015).

[6] However, they are used in the projections, and their educational compositions is then imputed using proxy countries from the region (see Annex A).

Table 1. Country coverage of the updated WIC 2018 dataset grouped by UN regions

UN region	All countries	Countries with education data	Countries covered (in %)	Population covered (in %)	Countries with **missing** education data	**New** countries covered in WIC 2018*
Europe	40	39	97.5	99.9	Channel Islands	-
Asia	51	49	96.1	99.3	Brunei, Uzbekistan	Afghanistan, North Korea, Oman, Sri Lanka, Taiwan, Yemen
Africa	57	50	87.7	98.6	Djibouti, Eritrea, Libya, Mauritania, Mayotte, Seychelles Western Sahara	Angola, Botswana, South Sudan, Sudan, Togo
Northern America	2	2	100.0	100.0	-	-
Latin America & Caribbean	38	34	89.5	99.9	Antigua and Barbuda, Barbados, Grenada, U.S. Virgin Islands	Curaçao
Oceania	13	11	84.6	79.6	Papua New Guinea, Guam	Fiji, Kiribati, Micronesia, Solomon Islands
World	***201***	***185***	***92.0***	***99.2***	-	-

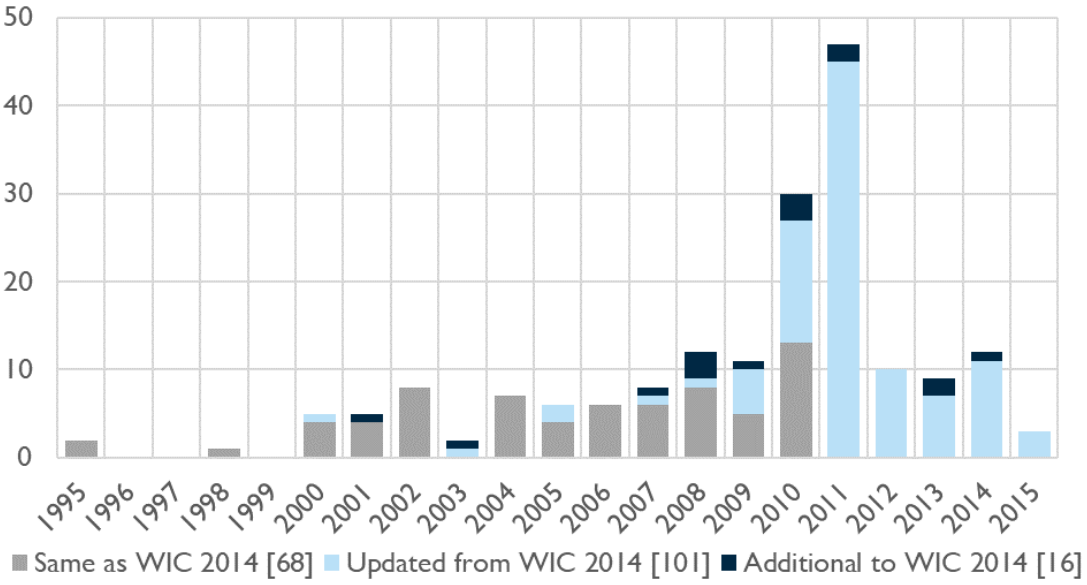
* compared to WIC 2014

Due to lack of data, insufficient level of detail of the published educational data or due to other data quality issues, information for 16 countries out of 201 countries is missing (down from 24 in WIC 2014). Among the populous countries with missing education data are Uzbekistan (the last full census was implemented in 1989 and the only data available comes from a Demographic and Health Survey (DHS) in 1996) and Papua New Guinea where our request for data was not answered.

Besides data for 16 new countries (blue in Figure 1), information for 101 countries was updated (orange in Figure 1), either in terms of more recent census or survey, or in terms of data quality. For 68 countries, the same data source as in the previous baseline (WIC 2014) is used (grey in Figure 1). The WIC 2018 dataset has information for 112 countries with data pertaining to 2010–2015 (up from 12 in WIC 2014 dataset). Good quality data are still hard to get or for many countries in the Middle East, Africa and Central Asia, in many

cases due to political unrest and violent conflicts. For example, it is still necessary to rely on data sources from before the year 2000 for the Central African Republic (DHS 1995), Pakistan (1998 census as the 2017 census in not yet available) and Turkmenistan (last census in 1995).

Figure 1. Update of the WIC 2018 educational data *(Source: authors' own calculation)*



An increase in data quality was achieved by avoiding pre-compiled data sources provided by UNESCO or other agencies as these data compilations often suffer from biases. Thus, for instance, census data were mostly collected from *National Statistical Institutes* (NSO) and in the greatest possible level of detail when it comes to age and education categories, avoiding pre-defined aggregate education categories that often do not correspond to the ISCED 2011 definitions, and are thus of limited use or require additional adjustments.

In general, register and census data tend to provide more complete (for example capture all age groups not only persons in reproductive or economically active ages) and reliable information than most surveys.[7] For countries without census data, we used the new waves of surveys such as DHS or *Multiple Indicator Cluster Survey* (MICS), when available.

Figure 2 and Figure **3** show the change in the type of primary data source for the country-specific base-year data between the WIC 2018 and the WIC 2014 datasets. The latter is composed of more register or census data for more countries than the former. The WIC

[7] This may not be the case in all countries as it has been documented that some census results were flawed, e.g. Nigeria 2006 census or Chile 2010 census. In such cases, we have either used an older census or alternative data, such as DHS and national households surveys. The selection of the final data source is subject to careful time-series analysis and cohort checks of the educational attainment data.

2018 dataset includes census data for 126 countries from the 2010 census round, and for 20 countries from the 2000 census round (Figure 3). For five countries (Austria, Switzerland, Estonia, Norway and Sweden), the most up-to-date register data at the time of data collection was used. We replaced the previous WIC 2014 survey data with newer information (mostly from DHS) for 25 countries. If register or census data were not available, reliable or of sufficient quality, representative sample surveys (6 countries), DHS (19 countries) or other household surveys (9 countries) have been used.

Figure 2. Data types used in WIC 2018 dataset for 185 countries *(Source: authors' own calculation)*

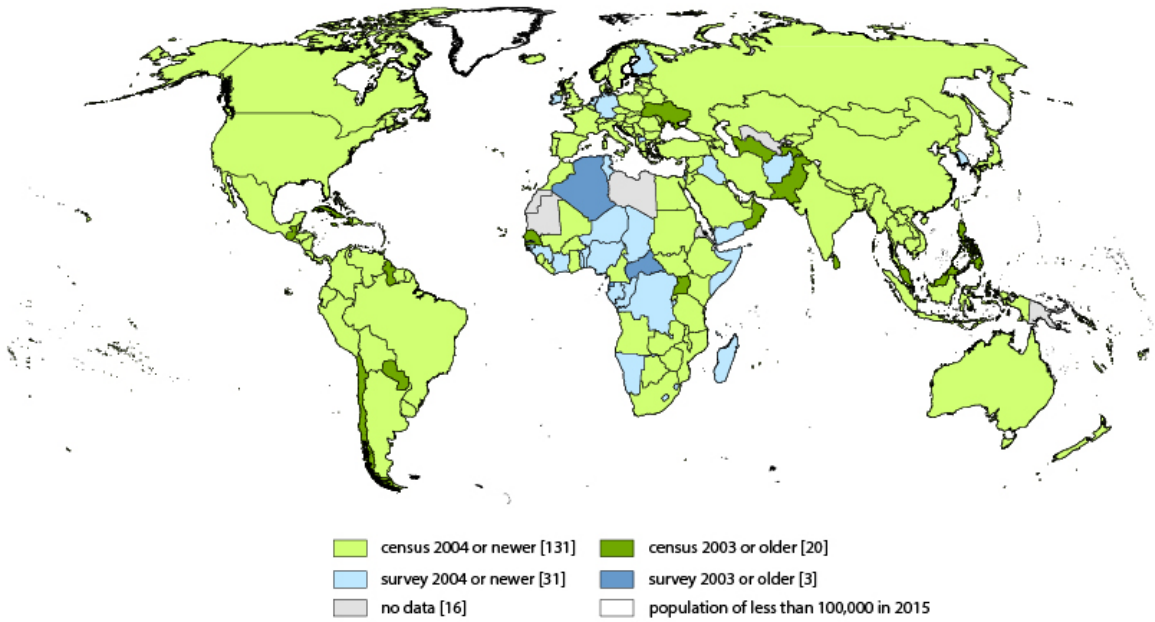
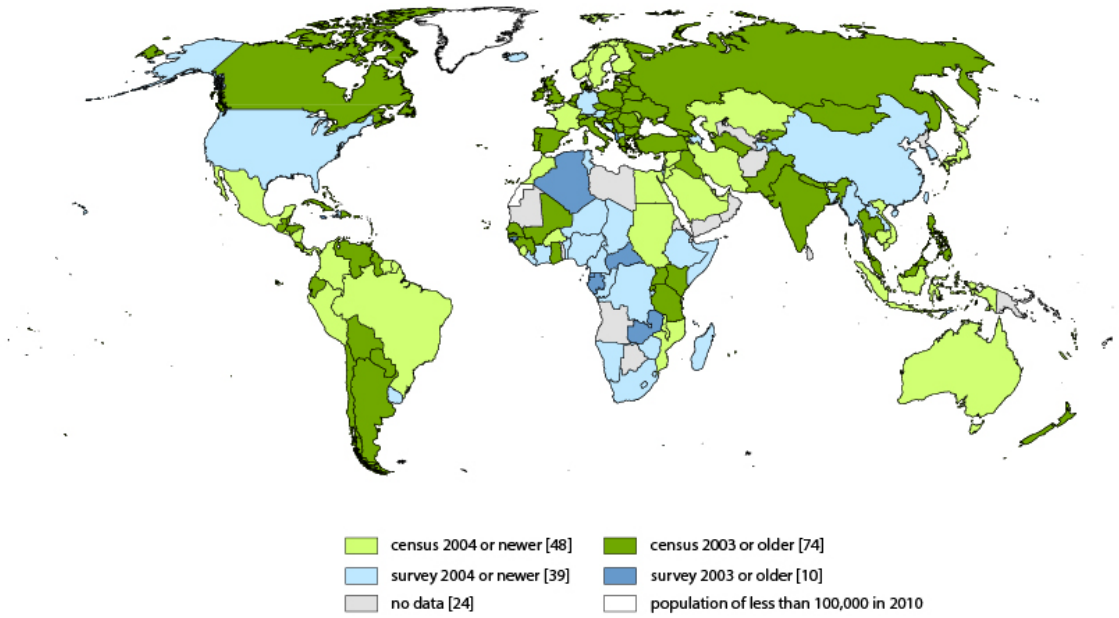


Figure 3. Data types used in WIC 2014 dataset for 171 countries (Source: author's own calculation)



Countries collect and publish education data according to their definitions and needs. *Low and Middle-income Countries* (LMICs) that strive mostly to increase enrolment in primary and secondary, along the recommendations of international goals (such as the Millennium Development Goals or the Sustainable Development Goals) often publish detailed statistics by level and grade completed. It was possible to collect all main six educational categories for 150 countries (Case 1 countries in Table 2).

In contrast, *High-income Countries* (HICs) often focus the data collection on the higher end of the educational spectrum as virtually the entire working-age population has achieved lower secondary education.[8] However, this hides the progress in educational attainment that many societies have made during the past century and that should still be visible among older age groups. Such information is important when assessing the educational compositions of whole populations, including older age groups who have studied under very different educational systems and in times when compulsory education was set at lower levels[9] and not strictly implemented. To fill those data gaps, alternative or

[8] As foreign-born populations increase in many economically advanced countries there is nonetheless a need for educational data among the immigrants who may have lower educational attainments compared to the native-born population. This can be the case because compulsory schooling requirements in their countries of origin differ but also because some migrants, especially in school-age may have been out of school if they are coming from countries that do not enforce compulsory education or they had disrupted educational trajectory due to war, refugee situation or as a result of migration process.

[9] In most HICs, compulsory education is set at completing lower secondary level (i.e. min 8 or 9 years of schooling), while in the past it was often set at the level of primary.

older datasets have been collected and estimates have been made. This was not possible for many OECD countries, see for example the United Kingdom (Case 6) as an extreme case in point (see Table 2).

Table 2. Overview of available[10] education categories by number of countries

Case	No education	Incompl. primary	Complete primary	Lower secondary	Upper secondary	Post-secondary	No. of countries
1	✓	✓	✓	✓	✓	✓	153
2	✓	►	✓	✓	✓	✓	14
3	►	►	✓	✓	✓	✓	8
4	✓	►	✓	►	✓	✓	5
5	►	✓	✓	✓	✓	✓	3
6	✓	►	►	✓	✓	✓	1
7	►	►	►	✓	✓	✓	1
N	173	156	183	180	185	185	185

Several adjustments were made to the data, similarly to what was implemented for WIC 2014 (Bauer et al. 2012). These are documented in the following sub-sections.

3.1.2 Base-year Data Adjustments

3.1.2.1 Applying ISCED 2011

Educational categories surveyed in the censuses and surveys tend to be based on national educational programs. Due to the variety of nationally distinct educational systems, UNESCO designed the *International Standard Classification of Education* (ISCED). In WIC 2014 the six educational categories (no formal education, incomplete primary, completed primary, completed lower secondary, completed upper secondary and post-secondary) were based on ISCED 1997 (UNESCO Institute for Statistics 1997) as ISCED mappings were not available for the 2011 revision (UNESCO 2012). WIC 2018 is following ISCED 2011, whereby the changes concern the post-secondary category. As mentioned before, we disaggregated the post-secondary education category for a limited number of countries (60). The allocation procedure is detailed in Annex B. The population of these countries divided in eight categories was projected to 2100, but not back projected for lack of historical data. Therefore, in the reconstruction, we us the same six education categories as in the WIC 2014 dataset (seeTable **3**).

[10] Available categories include cases where data were readily available from the primary source as well as cases where they were estimated using additional/alternative data sources. A country-specific overview about available education categories can be found in Annex D as well as in the supplementary data documentation files.

Table 3. The WIC 2014 educational attainment categories according to ISCED 1997 and 2011 classification

ISCED 2011		ISCED 1997		WIC 2014	
01	Early childhood development	-	-	e1	No education
02	Pre-primary	0	Pre-primary		
1	Primary	1	Primary	e2	Incomplete primary
				e3	Complete primary
2	Lower secondary	2	Lower secondary	e4	Lower secondary
3	Upper secondary	3	Upper secondary	e5	Upper secondary
4	Post-secondary (non-tertiary)	4	Post-secondary (non-tertiary)	e6	Post-secondary
5	Short-cycle tertiary	5B	First stage tertiary		
6	Bachelor's or equivalent	5A			
7	Master's or equivalent	5A			
8	Doctoral or equivalent	6	Second stage tertiary		

3.1.2.2 Estimating Unavailable Education Categories

As already mentioned, it was not always possible to collect as detailed education categories as required. While information on (some) post-secondary education has been available for all countries, the most frequently unavailable category was incomplete primary education that was missing in 34.6% of countries (see Table 4), merged with completed primary education. In many cases, primary completed and lower secondary were aggregated together in one category, which sometimes also included all education categories below lower secondary education (including or excluding the no education category).

Table 4. Frequency of unavailable and estimated educational categories

Cases	No education	Incompl. primary	Complete primary	Lower secondary	Upper secondary	Post-secondary
N estimated	4	35	3	9	2	0
N missing	12	29	2	5	0	0
Unavailable	**16**	**64**	**5**	**14**	**2**	**0**
	8.6%	**34.6%**	**2.7%**	**7.6%**	**1.1%**	**0.0%**

We attempted to estimate unavailable categories by **a)** using alternative data sources (triangulation) or **b)** in well-grounded cases by analogy using information from a very

similar country. To give an example, a combination of census and Labour Force survey[11] (LFS) data was used to fill in the gaps (on post-secondary and lower secondary education and less) for some European countries—given that **a)** LFS provided more detailed categories, and **b)** there was a good correspondence between the LFS and census data. Other surveys or older censuses were also used in the data triangulation. The use of older data was often limited to get hold of information for the lower education categories as the share of population with such low levels is very small in HICs and completion of compulsory education is assumed (among the working-age population).

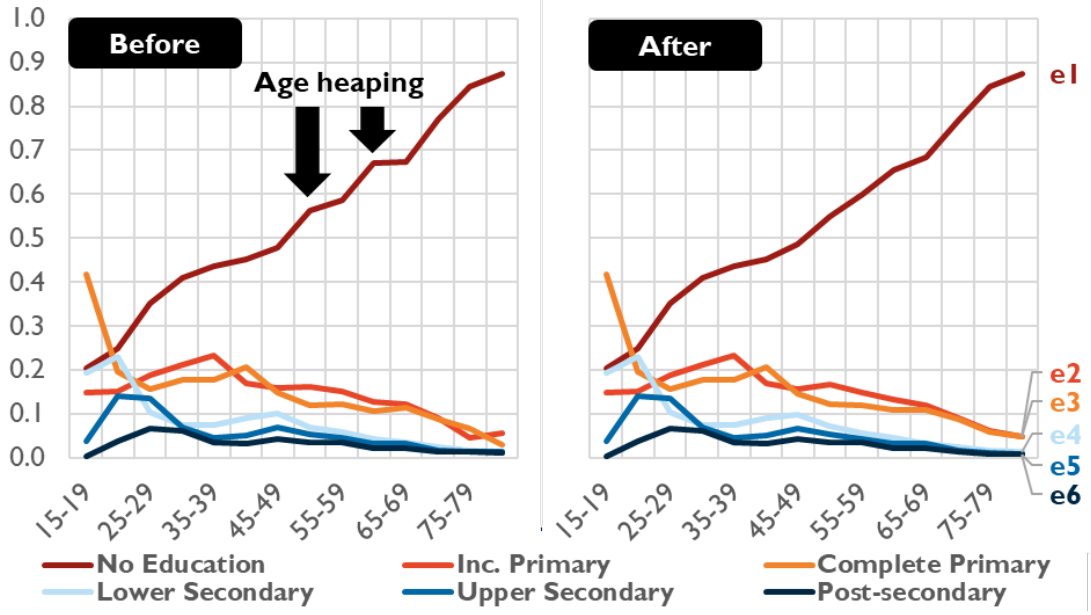
When no detailed base data were available for adjustments, we refrained from *guesstimating* the missing data. While this may be inconvenient for some users we are convinced that it increases the quality of the dataset.

3.1.2.3 Adjustments of age heaping in developing countries

When updating the WIC dataset with the aim to improve the quality, we decided to look at the issue of age heaping in census and survey data in some LMICs, mostly in sub-Saharan Africa. The misreporting of age leads to peaks in the age distribution with people rounding their age to a number ending in 0 and to a lesser extent in 5. This was especially present at age 50 or 60, but in fewer cases also at age 40 and other ages. It was most common among persons with no or low education and less among those with a higher education. Figure 4 shows an example of age heaping in the Benin DHS 2011. Interestingly, more DHS datasets showed age heaping for men compared to women, perhaps because the male subsample is smaller and therefore the problem becomes more apparent.

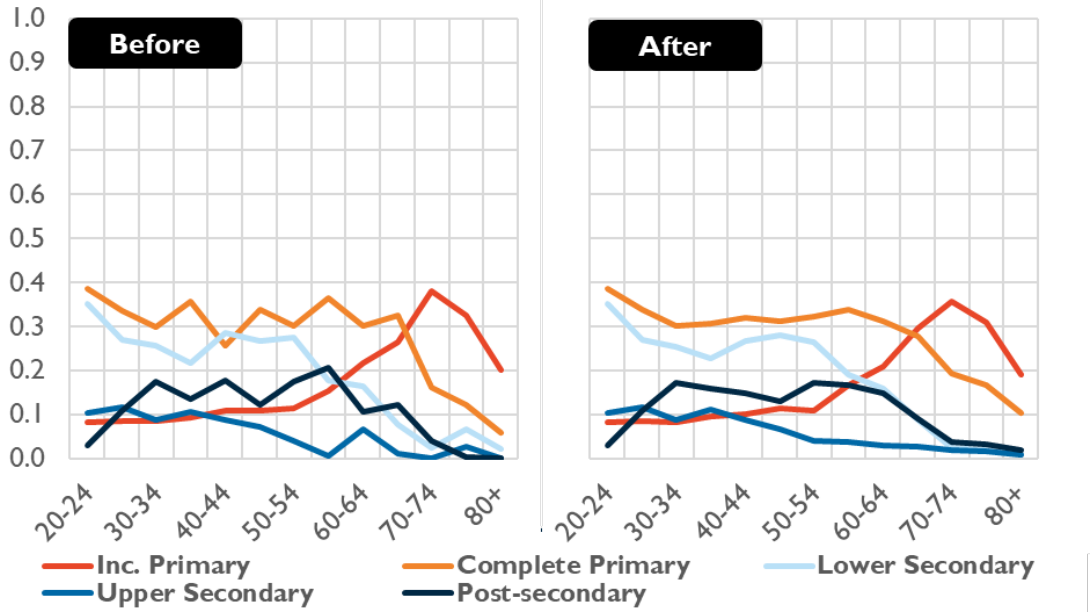
[11] ISCED 2011 categories have been implemented in LFS only from 2014 onwards, thus are the time of the data collection we had data for two years.

Figure 4. Example of age heaping in DHS 2012 data for Benin, educational attainment shares before (left panel) and after smoothing (right panel) (author's illustration)



We applied smoothing techniques (functions or moving averages if rounding was only visible in one age group) to correct age heaping in the relevant education category and adjusted iteratively the shares in the other education categories. The same smoothing techniques were used to adjust other irregularities in the data, such as uneven trends in education shares by age. The DHS for Gabon is one of the few extreme cases. Figure 5 shows the data before and after smoothing.

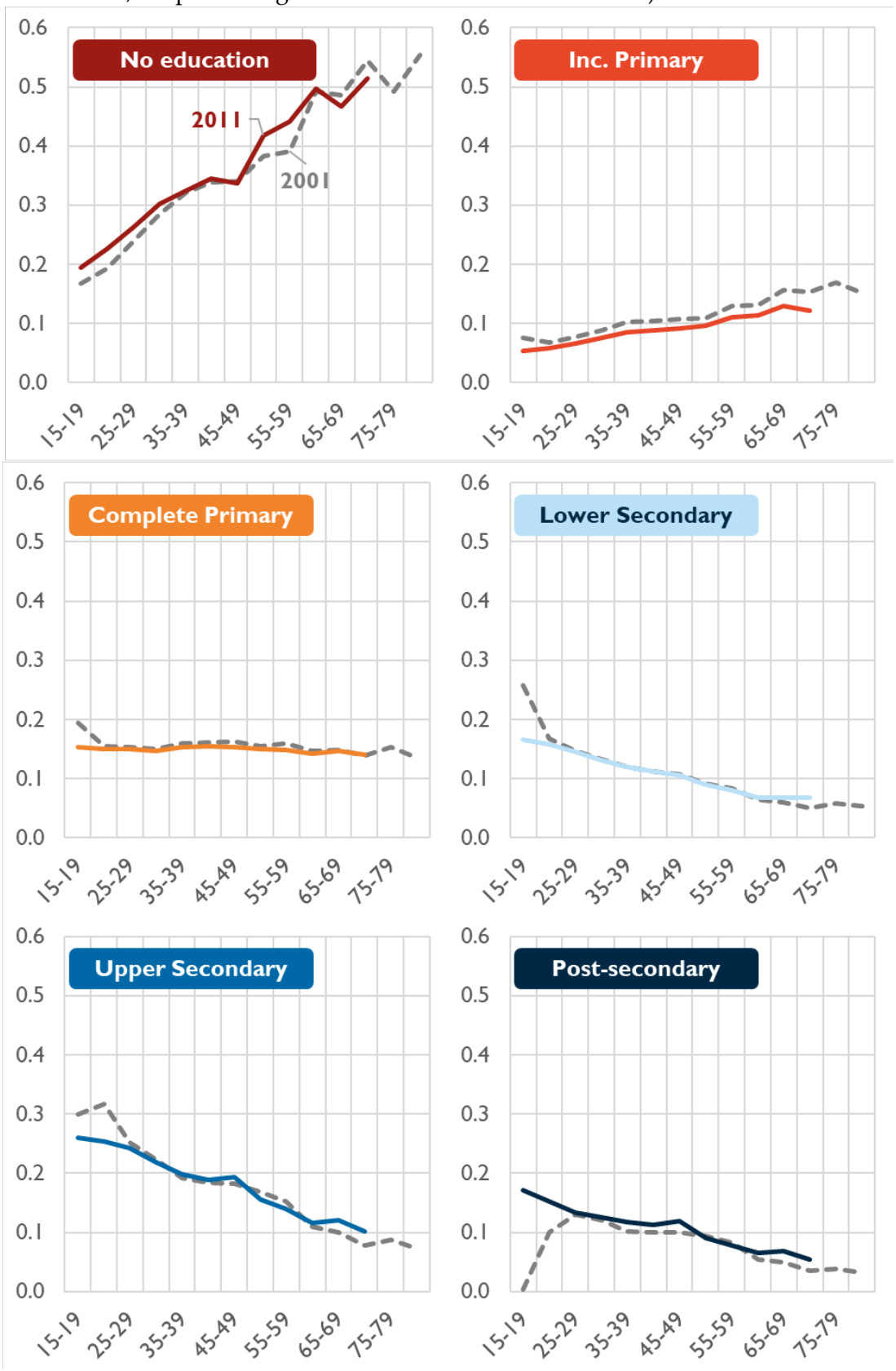
Figure 5. Example of age heaping in Gabon DHS 2012 data, educational attainment shares before (left panel) and after smoothing (right panel) (author's illustration)



3.1.2.4 Base-year Data Validation

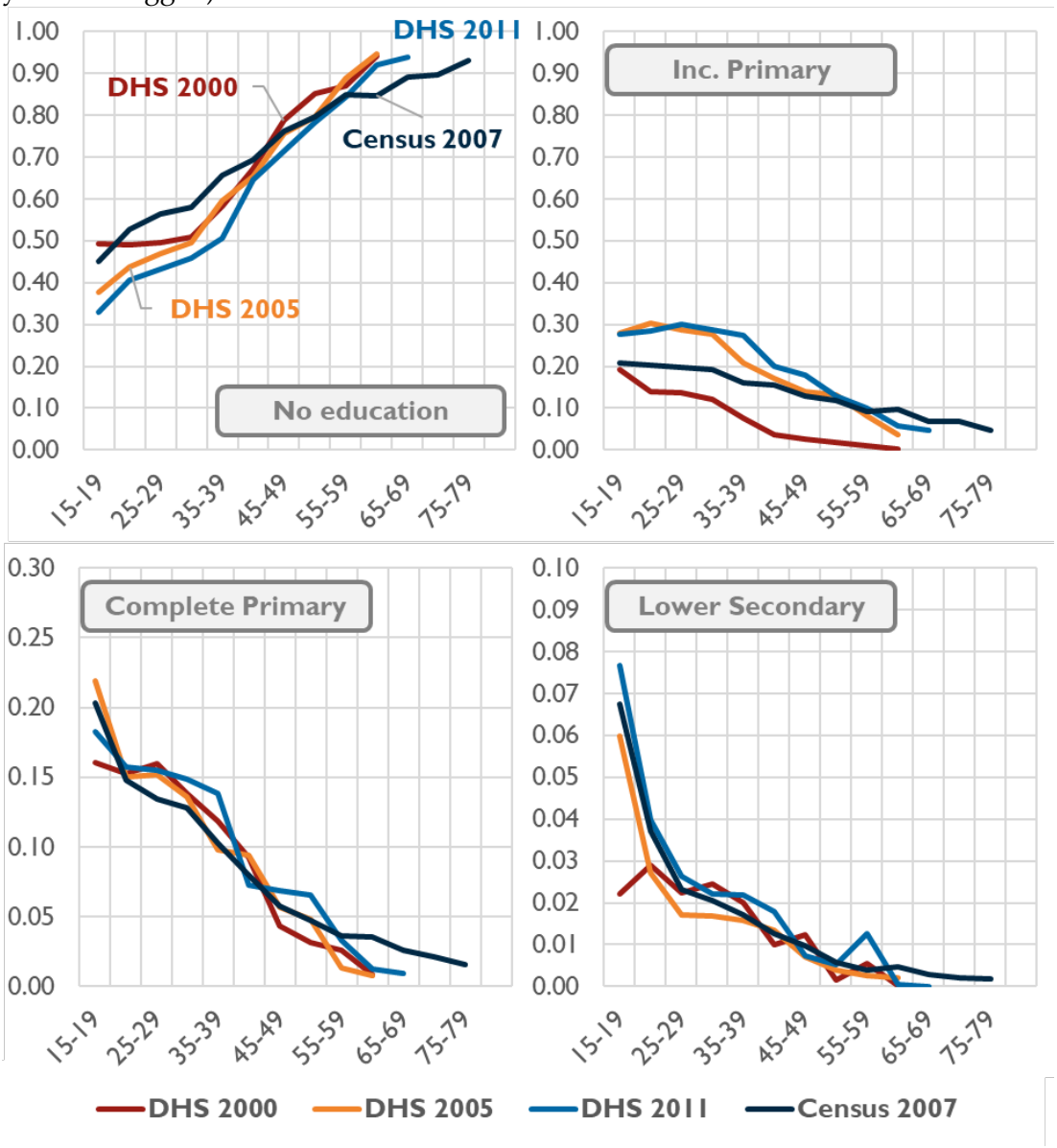
To inspect the data, identify possible flaws or to identify the best data source among those collected, we performed a validation exercise across and within data sources, mostly looking and comparing the educational attainment of cohorts born during the same period. Cohort validation is a good technique for checking the consistency of educational data. In an ideal case, the education shares should be similar in populations with low migration and at ages with relatively low mortality and for ages after the end of schooling. The cohort validation was routinely implemented, separately by sex and each education category, to check the newly collected census and surveys against the data collected for the WIC 2014 dataset. Allocation of the national education categories to ISCED is often not straightforward and cohort validation enabled us to identify unusual data patterns and inconsistencies, which were corrected. Figure 6 shows an example of good correspondence of the data for India in most categories across the two latest censuses. The slightly higher share of population with no education in the 2011 census may actually be a sign of better data quality—completeness of the data (including remote areas with less educated rural populations) and less misreporting. Also age heaping at higher ages is less pronounced in 2011 compared to 2001.

Figure 6. Cohort validation of educational shares in Indian censuses (x axis plots age at census 2001, i.e. persons age 15–19 were 25–29 in census 2011)



Cohort validation was used to routinely assess DHS rounds IV-VI and census data when available. While in some countries, such as Kenya, the data are consistent across censuses, in other countries such as Ethiopia (Figure 7) it is rather difficult to say which data source is the most accurate. In the end it was decided to use the census data as the education patterns resembled more that of other countries and due to better coverage of the population in the census data, especially for persons above 50 years of age.

Figure 7. Ethiopia DHS and census 2007 cohort validation (x axis shows age in 2000, other years are lagged)



One common pattern was found in comparing census and DHS data for the same country: censuses tend to report a higher share of uneducated population than DHS does, and a lower share of those with completed primary level.

3.2 Assembling the Historical Data

The model for the reconstruction of educational attainment (IMCR model) requires on top of the base-year data two obligatory and one optional historical dataset:

- Population by age (5-year age groups) and sex for 185 countries from 1950 to 2015 in 5-year steps **(obligatory)**
- Life table by sex for all 185 countries from 1950 to 2015 in 5-year steps **(obligatory)**
- Population by age (5-year age groups), sex, literacy and/or highest educational attainment for up to 185 countries from 1950 to 2015 in 5-year steps **(optional)**

3.2.1 Historical Data Collection: Coverage & Data Sources

Historical data points are needed for the reconstruction exercise to increase its fitting and accuracy. As depicted earlier the availability and quality of global data on educational attainment is substantial today. However, the awareness of the importance of data on educational attainment, especially in an internationally consistent and comparable matter, just arose quite recently. Most countries in the world were not consistently collecting education before the 1950s. Nevertheless, the available historical education data can be retrieved from the NSOs and international organisations such as the European Union (Eurostat 2018), OECD (OECD 2018), World Bank (World Bank 2018), CELADE (CELADE/CEPAL 2014) and IPUMS (Minnesota Population Center 2018), to assemble and to harmonise demographic and socio-economic data for cross-country comparisons (Speringer et al. 2015; Speringer, Goujon and Jurasszovich 2018).

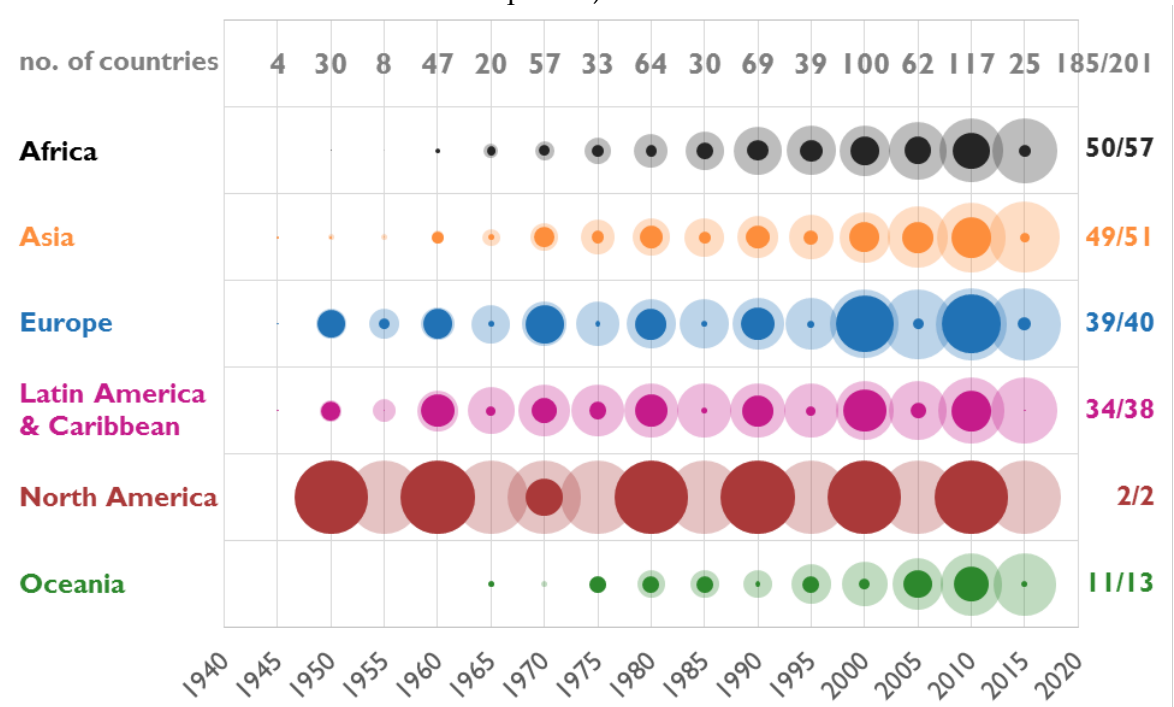
Out of the 185 countries for which we have base-year educational data, it was possible to gather historical data on educational attainment for 141 countries, (see Section 3.1) for at least one other data point in the period 1950 to country-specific base year. In total, it was possible to collect data on educational attainment for 705 data points (incl. historical and base-year data) in the period 1950 to 2015, covering 32% of the data needed (2220 data points: 185 countries x 12 data points). The historical data points show different degrees of data quality. As the availability and quality of data on educational attainment is quite limited, this reconstruction exercise gains importance to create such a comprehensive, consistent and comparable dataset.

Despite the limited availability of historical data on educational attainment, the structure and quality of the available data are quite different. The biggest challenges with historical data on educational attainment are **(A)** the cross-country comparability of educational categories, especially before the ISCED classification became widely implemented, **(B)** the quality of the data, **(C)** the available age-structure, which might contain aggregated age groups and **(D)** that the available data point years not necessarily match with the reconstruction intervals (e.g. 1971 vs 1970).

Figure 8 illustrates the availability of data points for the 185 countries in 5-year intervals (1950–2015) by continent, whereby the full coloured bubbles refer to empirically available data points as a share of all countries in the continent and the shaded bubbles highlight the cumulative availability of data points as a share of all countries. The most data points are available for the years 2010 (117 data points) and 2000 (100 data points).

Despite the limited availability of historical data on educational attainment, the structure and quality of the available data are quite different. The biggest challenges with historical data on educational attainment are **(A)** the cross-country comparability of educational categories, especially before the ISCED classification became widely implemented, **(B)** the quality of the data, **(C)** the available age-structure, which might contain aggregated age groups and **(D)** that the available data point years not necessarily match with the reconstruction intervals (e.g. 1971 vs 1970).

Figure 8. Overview of collected and used historical data on educational attainment by UN regions (filled bubbles refer to available data in the period and transparent bubbles to the cumulative number of countries in the period)



The sources and years for the historical datasets on educational attainment are available in the supplementary file. The file also provides information on the challenges that were faced in recoding the historical data in the WIC 2018 education categories. Furthermore, we also highlight the validation processes that were undertaken. Hence, the file also lists the available data that were not used when the age-cohort analysis suggested data reliability issues or other issues.

Other essential historical datasets for the reconstruction, are data on the population age and sex structure as well as mortality information. For that, we use the latest assessment of the UN (UN 2017). The Population Division of the UN *Department of Economic and Social Affairs* (DESA) thereby collects, processes and provides with 201 countries in the world the most data. The DESA publishes in regular intervals the World Population Prospects (WPP), with the WPP 2017 as latest version, which include historical estimates and projections of demographic data. The WPP 2017 is an essential data source for this reconstruction exercise (United Nations 2015, 2017) when it comes to population and mortality in the reconstruction model (Speringer, Goujon and Jurasszovich 2018).

3.2.2 Historical Data Adjustments

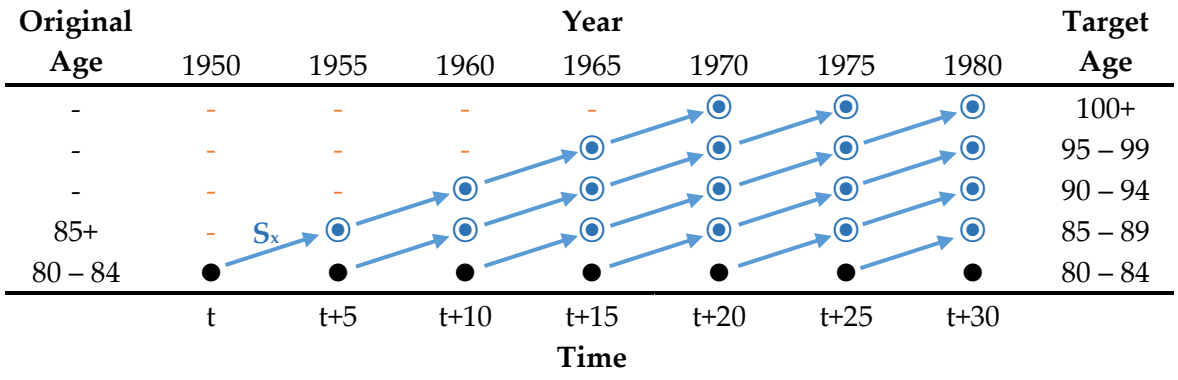
3.2.2.1 Adjustments to the Mortality Data

Some adjustments to the life tables provided by the UN (2017) are needed in order to cover the 1950–2015 period for all ages. The life table information provided by the UN has an open-ended age group of 100+ years for the 1985–2015 period, and of 85+ years for the 1950–1980 period (United Nations 2017). Since the reconstruction covers all age groups to 100+, the ***life tables for 1950 to 1980 were extended*** by applying a logistic extrapolation on the values for nq_x, L_x and a_x for the missing age groups. We calculated the age-specific *Survivorship Ratios* (S_x) and *Life Expectancy at age 15 years* (e15) from the extended life tables and use them in the reconstruction model as explained later.

3.2.2.2 Adjustments to the Age Structure Data

A similar adjustment had to be implemented for the population age-structure by sex, since the open-ended age group for the period 1950 to 1980 provided by the UN is 85+ years (United Nations 2017). Therefore, the ***age-structure by sex had to be extended***. We applied the estimated S_x to the available age-structure (●) to estimate the age-structure up to 100+ years (◉) as shown in Figure 9.

Figure 9. Schematic illustration of the age-structure extension



The remaining empty data points (-) are filled iteratively by reversely proportionally splitting the open-ended age group based on data points with the full age-structure. For instance, after forecasting the age group 80 to 84 years in 1950 by applying S_x in 5-year steps, the age-structure in 1970 is filled up to the open-ended age group of 100+ years. This information can be used as proxy to calculate the age-specific shares for the age groups 95 to 99 years and 100+ years in 1970 and reversely applying these shares on the age group 95+ years in 1965 to estimate the age distribution for this year. This approach is iteratively applied until the whole time series is filled.

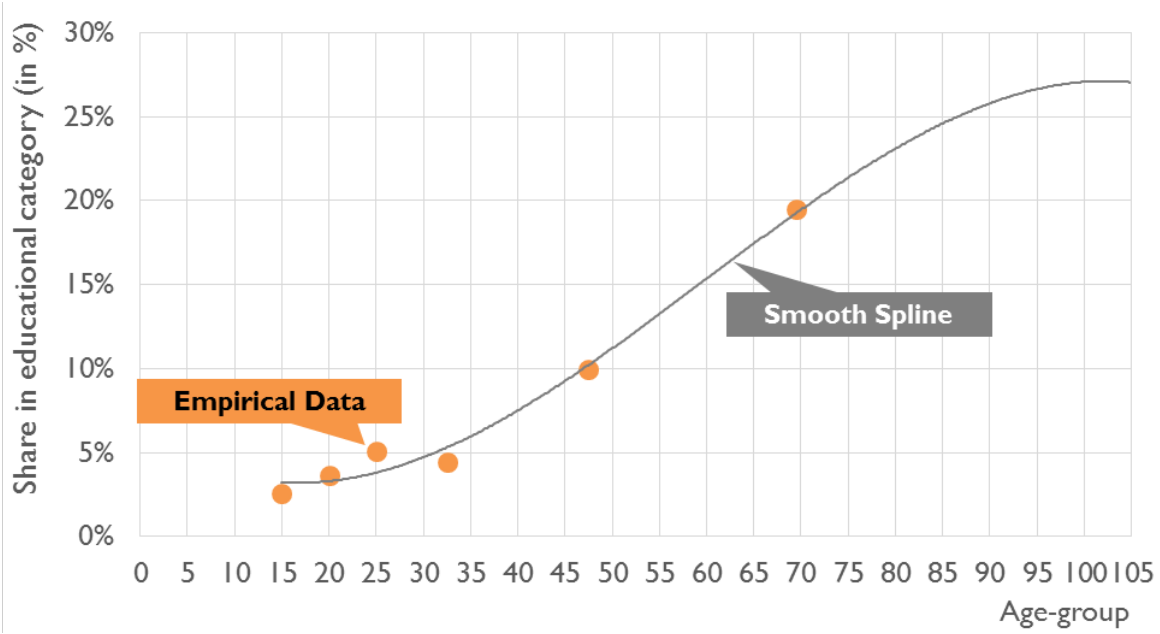
Both, the life table and age-structure extensions are new methods applied in WIC 2018 in comparison to WIC 2014 to create comprehensive and consistent mortality and population time series that can be used in the IMCR model (see Section 0).

3.2.2.3 Adjustments to the Education Data

New estimation procedures were also applied to the (historical and base-year) education information, compared to WIC 2014. To process the available historical data into the required structure of population by age (5-year age groups from 0 to 100 years and higher), sex, literacy and/or highest educational attainment (6 educational categories) (see Table **3**) estimation and harmonisation procedures were applied. The estimation procedures included new methodological approaches compared with the WIC 2014 reconstruction exercise, like **(A)** the smoothing and extension of the age-specific education distribution, **(B)** the shift of data points to 0/5 round years and **(C)** the usage of supplementary historical cohort data in the education distribution extension.

After harmonising the educational categories in the historical data according to ISCED, the educational distribution by age and sex was further processed **(ad A)** to ***smooth and extend the age-specific educational distribution*** using cubic spline function Therefore, the age group mid-points had to be calculated to get markers for applying an R cubic *"smooth.spline"*-function (see Figure 10). This was the case when the age pattern contained irregular or aggregated age groups as well as when the open-ended age group was too low (e.g. 75+ instead of 100+).

Figure 10. Schematic illustration of education-specific age-disaggregation and smoothing



In this way it was possible to create a smooth education-specific 5-year age distribution up to the available open-ended age group. However, the methodology requires an open-ended age group of 105+ years. Although only showing the age-specific education distribution up to the open-ended age group 100+ years, the one extra education group is essential in the reconstruction exercise: At each reconstruction step the age-specific educational distribution by sex is shifted by 5 years to the past (*t-5*) creating an empty open-ended age group of 100+ years (equals 95–99 years at *t-5*). Since the 105+ age group is not available, the ***age-specific education distribution (shares) had to be extended*** by applying a logarithmic extension to the five available age groups beneath the open-ended age group (e.g. if open-end age group is 85+, the logarithmic extension was applied to the 60–64 to 80–84 age groups). The result is an age-specific education distribution by sex for each historical data point and base year up to the age of 105+ years. The education distribution extension is necessary to have an estimated educational distribution for the age group 105+ years at time *t* to fill the age group 100+ year at *t-5*. This filling up of the required open-ended age group 100+ years is conducted in each reconstruction iteration up to 1950.

Obviously, data are not always available for 0/5 round years, but normally the documented data points are deviating 1—2 years in one or another direction. Therefore **(ad B)** the available data for the ***age-specific educational distribution had to be shifted to 0/5 round*** years for the base-year and the later used historical data. For instance, the data for the years 2008 to 2012 would be shifted to 2010. In earlier reconstruction exercises, conducted in the Wittgenstein Centre (Lutz et al. 2007a, [b] 2007; Goujon et al. 2016), the available data were simply recoded to be equal in the next 0/5 round years under the assumption that the education distribution might not differ significantly. However, in this

reconstruction exercise it was decided to shift and adjust the educational distributions to the nearer 0/5 round years.

To achieve this shift a linear trend interpolation was used to follow the age-specific education distribution pattern. For instance, the age group 40–49 in 2012 would be shifted back by 2 years or 40% of the 5-year reconstruction interval, and similarly for the whole age-specific education distribution. The education distribution by age is proportionally fitted to 100%.

To increase the accuracy of age-specific educational attainment extension, in this version we made use of the historical cohort-specific educational distributions as markers. The implementation of **(ad C)** ***historical data in the age extension*** of the base-year education distribution is straightforward. For instance, if we know that in country *i* at time t-25 (e.g. 1990) about 36 per cent of the female population aged 75 to 79 years (birth cohort: 1915) had lower secondary education, this can be used as marker information to estimate the education distribution for the age group 100 to 104 years at time t (e.g. 2015). If multiple historical data points for the same birth cohort are available the median of the birth-cohort-specific education distribution for each education group can be calculated. Same principle applies for historical data points as well, if older data are available, when extending the educational distribution.

The usage of historical data precedes the condition that the historical data are reliable, comparable and consistent over time. Therefore, it is necessary to identify the same birth-cohorts in multiple data points and validate if they are consistent. This is done for each country and educational group to identify potential data flaws. Inconsistent data points are excluded from the dataset. However, cohort information is normally not comprehensively available and often shows inconsistent educational categorisations that might differ over time. Therefore, it is necessary to filter out empty data points and group historical information by birth cohorts that reflect a certain target age group to be estimated (see Figure 11).

Figure 11. Schematic illustration of cohort identification process

Original		**Education Category**[12]						**Target**	
Period	Age	e1	e2	e3	e4	e5	e6	Age	Period
1925	20 – 24	-	-	-	-	-	-	105 – 109	2010
1930	25 – 29	●	-	-	-	-	-	105 – 109	2010
1935	30 – 34	-	-	-	-	-	-	105 – 109	2010
1940	35 – 39	●	-	-	-	-	-	105 – 109	2010
1945	40 – 44	●	-	-	-	-	-	105 – 109	2010
	(…)				(…)			(…)	
1980	75 – 79	●	-	●	-	●	●	105 – 109	2010
1985	80 – 84	-	-	-	-	-	-	105 – 109	2010
1990	85 – 89	●	-	●	●	●	●	105 – 109	2010
1995	90 – 94	-	-	-	-	-	-	105 – 109	2010
2000	95 – 99	●	●	●	●	●	●	105 – 109	2010
2005	100+	-	-	-	-	-	-	105 – 109	2010
Median		**?**	**?**	**?**	**?**	**?**	**?**	**105 – 109**	**2010**

● ... empirically / estimated population by age and educational attainment for the base year

To cope with the inconsistent availability of historical education categories, the reverse cumulative sum for each age-cohort is calculated, which sums to 100 per cent in the highest education category (post-secondary). The reverse cumulative sum was chosen as some countries do not provide information on the no-education category. On the contrary, the post-secondary education is always available if multiple education categories exist. Based on this reverse education progression ratios are calculated between the neighbouring education categories (e.g. from upper secondary to lower secondary). In the case education categories were missing a progression to or from a missing category is not possible and results in a 'not available' (NA).

Figure 12. Schematic illustration of logit regression extrapolation

Original		**Reverse Progression Ratio**					**Target**	
Period	Age	e1 ‹ e2	e2 ‹ e3	e3 ‹ e4	e4 ‹ e5	e5 ‹ e6	Age	Period
1980	75 – 79	NA	●	NA	●	●	105 – 109	2010
1990	85 – 89	NA	●	●	●	●	105 – 109	2010
2000	95 – 99	●	●	●	●	●	105 – 109	2010

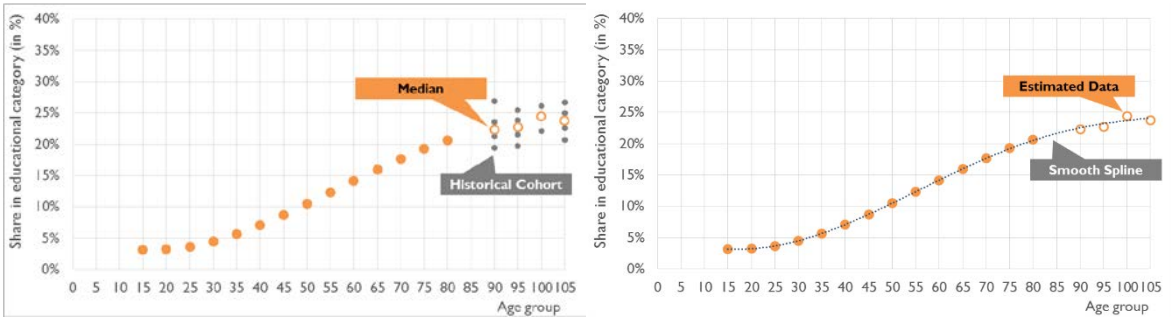
The existing reverse progression values can be used to calculate an intercept and slope value as basis for a logit regression extrapolation to estimate the missing progression ratio, if at least two data points are available. In the case of only one usable data point, we estimated the same progression ratio for the other data points. After filling the progression

[12] Education categories: (e1) no education, (e2) incomplete primary, (e3) completed primary, (e4) lower secondary, (e5) upper secondary, and (e6) post-secondary education;

ratios we translate the progression ratios and cumulative sums into education distribution values for each data point (see Figure 12).

Based on this estimated distribution for historical data points it is possible to calculate the median over all existing data points to derive the education distribution in the target age group 105–109 years in 2010. This procedure is iteratively repeated for all target age groups in the base year so as to obtain multiple median values as proxy information for the education extension to which we apply a smooth spline function (see Figure 13). This distribution is then logistically extrapolated to generate the full age-structure by educational attainment.

Figure 13. Schematic illustration of education-specific age structure extension by using historical data



This methodological approach makes it possible to implement historical cohort information in the extension of the age-specific education distribution for the base year, which marks the basis for the reconstruction model. As the aim of the reconstruction is to estimate most accurately the historical educational development for 185 countries, the newly implemented estimation procedures increase the estimation accuracy, especially due to the consideration of historical information. This is a major asset of the WIC 2018 reconstruction exercise compared to earlier WIC reconstruction exercises.

3.3 Reconstructing the Past Educational Composition

The reconstruction model applied in this research was originally developed by researchers at the Wittgenstein Centre (Lutz et al. 2007a, [b] 2007). It has since then been further developed in the so-called ***Iterative Multi-dimensional Cohort-component Reconstruction*** (IMCR). The model follows the reconstruction principles as described in Section 2 and comprises ***six model elements*** that are iteratively applied for each reconstruction iteration, namely:

- The **projection/reconstruction of the age-specific educational distribution** (share) by sex from base year t to t-5, following cohort lines;
- The **open-ended age group 100+ years** at time t-5 gets empty due to the cohort projection and needs to be retrieved from the estimated education distribution extension;

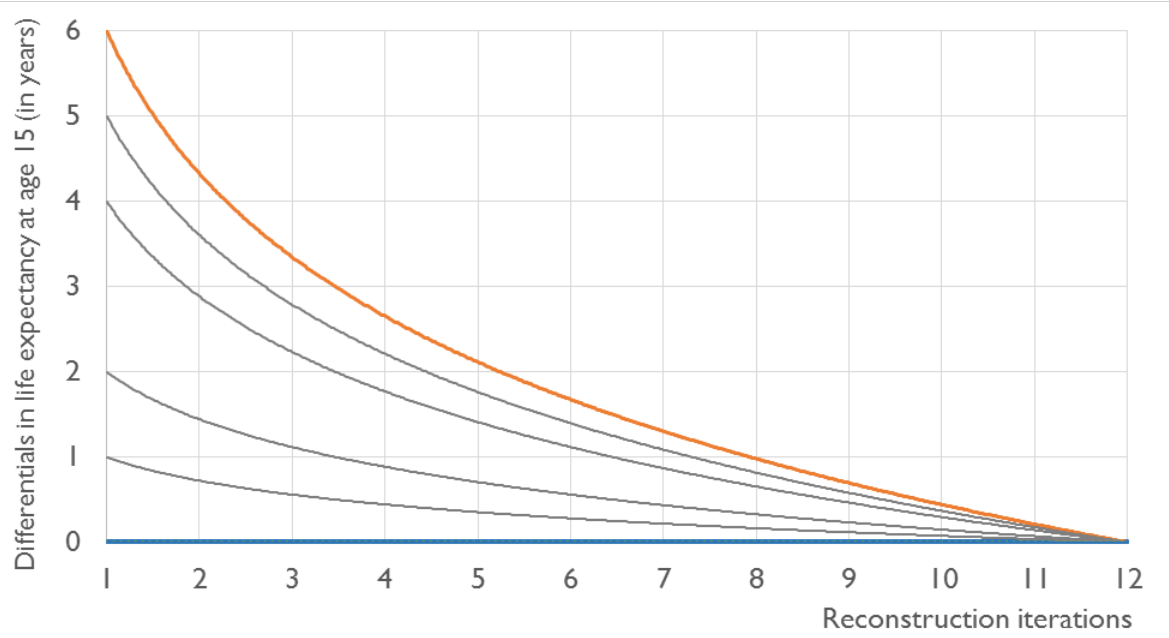
- The **education-specific mortality differentials** standard schedules by gender are applied at t-5;
- The **education transition** in the age groups 15 to 34 years is applied at t-5;
- The ***adaptation*** of the reconstructed educational distribution ***to UN population totals*** by age and sex;
- The ***population projection*** by age, sex and educational attainment from the country-specific base years to 2015 to generate a consistent dataset;

3.3.1 Estimation of Education-Specific Mortality Differentials

The ***education-specific mortality differentials*** are derived from gender-specific standard schedules of *life expectancy education differentials at age 15* (e15) defined in previous reconstruction exercises (Lutz, Butz and KC 2014). The maximum differentials in life expectancy at age 15 between the no education and the post-secondary education categories is set at 6 years for men and 4 years for women. Furthermore we assumed the education differentials in e15 for men to have a 1-1-2-1-1 year pattern between the no education, some primary, completed primary, lower, upper and post-secondary education levels, respectively—and the same proportionally for women (Lutz et al. 2007a; Lutz, Butz and KC 2014; Goujon et al. 2016). The initial mortality differentials in terms of education-specific differentials in life expectancy at age 15 (e15) are then converging in the model to no differentials in life expectancy at age 15 (e15) by 1950 along a logarithmic trend extrapolation (see Figure 14). This is different to WIC 2014 that held the mortality differentials constant over the entire reconstruction period (Goujon et al. 2016). The convergence to zero differentials seems to be more fitting the epidemiological transition theory whereby in the 1950s infectious diseases might have been more prevalent and less discriminatory by levels of education compared to chronic diseases that are more dominant nowadays.

From the education-specific life expectancy differentials at the age of 15 (e15), new education-specific life tables are created and used in the reconstruction. Hereby, the education-specific *Survivorship Ratios* (S_x) are applied in each iteration to the reconstructed population by age, sex and education.

Figure 14. Schematic illustration of the logarithmic convergence of education-specific life expectancy differentials at age 15 years (e15) in the reconstruction model (authors illustration)



3.3.2 Estimation of Education Transitions in the Reconstruction

Education transitions, meaning the mobility of population of schooling age between education categories, from here onwards referred to as *Transition Ratios* (TR), are allowed in the model between the age groups of 15 and 34 years.[13] Thereby, the maximum transition age group ("⊗") is gradually increasing. For instance, the transition to primary education is only allowed until a cohort reaches the age group 15–19 years as in most societies primary education usually gets achieved in the age group 10–14 years. Further transitions from completed primary (e3) to lower secondary education (e4) are allowed up to 20–24 years, the transitions to upper secondary (e5) and post-secondary education (e6) are possible until 25–29 and 30–34 years respectively (see Figure 15).

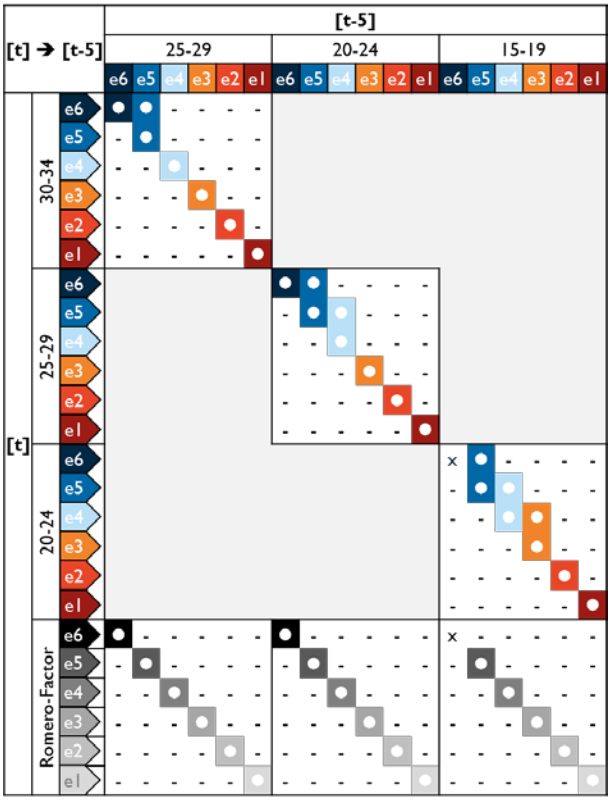
[13] We do not distinguish the population by education between the age of 0 and 14 years and consider that most transitions would occur before the age of 35.

Figure 15. Schematic illustration of the maximum education transition age groups and required transitions

age	TR_{e12}	TR_{e23}	TR_{e34}	TR_{e45}	TR_{e56}
15--19	⊗	⊗	?	?	?
20--24	○	○	⊗	?	?
25--29	○	○	○	⊗	?
30--34	○	○	○	○	⊗

Legend: ⊗ Maximum age group for transition between dedicated education categories

Figure 16. Schematic illustration of education transition scheme in the reconstruction model from time t to t-5 (author's illustration)



In the reconstruction process, as illustrated in Figure 16 the education transition schedules allow people either to remain in their level of educational attainment or in certain age groups to back-transit to the next lower education category. For instance, when shifting education-specific population aged 30–34 years at time *t* to time *t–5*, to the age group 25–29 years, some people with post-secondary education (e6) will eventually back-transit to upper secondary education (e5) while others remain in e6. This is because not all back-transitions from e6 to e5 have to occur in these age groups. Other transitions, e.g. from e4 to e3, are not possible yet at this ages, as shifts from primary (e3) to lower secondary education (e4) are considered to happen in younger ages below 15 years of age.

To calculate TR, we would ideally need the age-specific educational distribution by sex for two consecutive time points (five years apart) to follow the changing educational composition in a cohort (see Equation 1).

Equation 1.
$$TR_{e56}^{25\to30,t\to t+5,} = \frac{P_{e6}^{30,t+5} - P_{e6}^{25,t}}{P_{e5}^{25,t}} = \frac{20\% - 5\%}{25\%} = 60\%$$

, with:
TR ... Transition Ratio
$P_{e6}^{30,t+5}$ ...Share of population aged 30–34 years at time t+5 with upper secondary education (e5)

While, TR gives us forward transition ratios from an education attainment level to the next, in the reconstruction exercise, we need a *backward Transition Ratios* (bTR) and here we define it as *"what proportion remained in an education category"* (see Equation 2).

Equation 2. $$\boldsymbol{bTR_{e66}^{30\rightarrow 25,t+5\rightarrow t}} = \frac{\boldsymbol{P_{e6}^{25,t}}}{\boldsymbol{P_{e6}^{30,t+5}}} = \frac{\mathbf{5\%}}{\mathbf{20\%}} = \mathbf{25\%}$$

Given the data availability for a time point, say t, the denominator is not available. To estimate the values for time t+5, we extrapolated the available cross-sectional data by creating a time series based on age specific education attainment distribution.
The cross-sectional age-specific educational distribution at time t is converted into reverse cumulative proportions over education categories by age groups, which refers to "at least" education attainment levels. This results in from the highest (e6 – post-secondary education) to the lowest education category (e1+, at least no education) aggregated proportions, which leads to overall age groups consistent 100 per cent in the category e1, as everyone has at least no education (see Equation 3).

Equation 3. $$\boldsymbol{P_{i+}^{j}} = \sum_{i}^{6} \boldsymbol{P_{i}^{j}}$$

, with:
$\boldsymbol{P_{i+}^{j}}$ … Share of population in age group *j* with at least *i* as educational level (*e*)
i … Educational level[14]

Based on this table *Education Attainment Progression Ratio* (EAPR) are calculated by dividing the consecutive education cumulative values, namely e2+/e1+, e3+/e2+, e4/e3, e5/e4, e6/e5, for each age group. This means for six educational categories, five EAPRs (eapr12, eapr23 and so on) by age *j* are calculated (see Equation 4).

Equation 4. $$\boldsymbol{EAPR_{i\rightarrow i+1}^{j}} = \frac{\boldsymbol{P_{i+1}^{j}}}{\boldsymbol{P_{i}^{j}}}$$

, with:
$\boldsymbol{EAPR_{i\rightarrow i+1}^{j}}$ … EAPR of population in age group *j* between educational level ***i*** and ***i+1***

Equation 5. $$\boldsymbol{logitEAPR_{i\rightarrow i+1}} = \mathbf{log}\left(\frac{\boldsymbol{EAPR_{i\rightarrow i+1}}}{\mathbf{1} - \boldsymbol{EAPR_{i\rightarrow i+1}}}\right)$$

These EAPRs are converted into *logits of the EAPR* (logitEAPR) (see Equation 5) that are extrapolated by applying linear regression equations. Those projected l*ogitEAPR's* are converted to EAPRs and in a further step to a so-called *extrapolated cumulative educational distribution* for a population in age j (eP_{i+}^{j}) at time t+5, t+10 and so on (see Figure 17). The P_{i+}^{j} and eP_{i+}^{j} are the basis for calculating the bTRs between t and t-5 (see Figure **18**).

[14] Educational Levels: e1 – No education, e2 – Incomplete primary, e3 – Completed primary, e4 – Lower secondary, e5 – Upper secondary, and e6 – Post-secondary;

Figure 17. Schematic illustration of education transition calculation with the cumulative education distribution by age at time t ($\boldsymbol{P}^{\boldsymbol{j}}_{\boldsymbol{i+}}$, left panel) and projected cumulative population distribution by education ($\boldsymbol{eP}^{\boldsymbol{j}}_{\boldsymbol{i+}}$, right panel) [colours illustrate what cells are used together to calculate TR at time t-5 in Figure 18]

age	e1+	e2+	e3+	e4+	e5+	e6
15--19	⊗	⊗	⊗	●	●	○
20--24	○	○	○	⊗	●	●
25--29	○	○	○	○	⊗	●
30--34	○	○	○	○	○	⊗

t+10

t+5

e1+	e2+	e3+	e4+	e5+	e6	time
○	○	○	○	○	○	t
○	○	○	●	●	●	t+5
○	○	○	○	●	●	t+10
○	○	○	○	○	○	t+15

Legend:		
	●	Age groups by education where education transitions happen
	⊗	Maximum age group for transition between dedicated education categories
	○	No education transitions are happening OR not required for EAPR calculation

The colours indicate which cells are connected to calculate the bTRs. The bTRs are calculated in each reconstruction iteration and applied on the age groups 15 to 34 years to proportionally shift population from higher to lower education categories. For instance, when calculating the bTR at time t-5 for an age group the empirical cumulative age-specific education distribution at time *t* (P^{j}_{i+}) is divided by the projected/estimated cumulative age-specific education distribution at time *t-n* (eP^{j}_{i+}). The *j* is determined by the education-specific maximum age group, where education transitions are supposed to end ("⊗") within the transition matrix.

The calculation of the bTRs starts with the highest education category (e.g. e6) and highest age group that undergoes education transitions (e.g. 30–34 years) at time t to estimate the proportion of people shifting in the reconstruction iteration to time t–5 to age group 25–29 years to estimate how many people remain in e6 and shift to e5 (see Figure 18). To calculate the $bTR^{25\rightarrow 20}_{e66}$ for the age group 20–24 years at time t–5 that remain in e6 (see Equation 7), the projected cumulative age-specific education distribution has to be extracted for time t+10 as the maximum age group for this transition is 30–34 years (see Figure **18**). For the $bTR^{30\rightarrow 25}_{e66}$ for the age group 25–29 years, the eP^{j}_{i+}-value for t+5 has to be taken (see Figure 17 & Figure 18).

Figure 18. Schematic illustration of the backward Transition Ratio (bTR) calculation between t and t-5 (the numbers in the brackets refer to the related formula)

age	bTR_{e22}	bTR_{e33}	bTR_{e44}	bTR_{e55}	bTR_{e66}
15--19	⊗	⊗	(10) $bTR_{e44}^{20\rightarrow15}$	(9)$bTR_{e55}^{20\rightarrow15}$	○
20--24	○	○	⊗	(8) $bTR_{e55}^{25\rightarrow20}$	(7) $bTR_{e66}^{25\rightarrow20}$
25--29	○	○	○	⊗	(6) $bTR_{e66}^{30\rightarrow25}$
30--34	○	○	○	○	⊗

, with:

Equation 6.
$$bTR_{e66}^{30\rightarrow25,t+5\rightarrow t} = \frac{P_{e6}^{25,t}}{eP_{e6}^{30,t+5}}$$

Equation 7.
$$bTR_{e66}^{25\rightarrow20,t+5\rightarrow t} = \frac{P_{e6}^{20,t}}{\left(eP_{e6}^{30,t+10} * bTR_{e66}^{30\rightarrow25,t+5\rightarrow t}\right)}$$

Equation 8.
$$bTR_{e55}^{25\rightarrow20,t+5\rightarrow t} = \frac{P_{e5+}^{20,t}}{eP_{e5+}^{25,t+5}}$$

Equation 9.
$$bTR_{e55}^{20\rightarrow15,t+5\rightarrow t} = \frac{P_{e5+}^{15,t}}{\left(eP_{e5+}^{25,t+10} * bTR_{e55}^{25\rightarrow20,t+5\rightarrow t}\right)}$$

Equation 10.
$$bTR_{e44}^{20\rightarrow15,t+5\rightarrow t} = \frac{P_{e4+}^{15,t}}{eP_{e4+}^{20,t+5}}$$

For the calculation of the $bTR_{e66}^{25\rightarrow20}$at time t-5 the age group is not next to the maximum transition age of 30–34 years, therefore the estimated bTR ($bTR_{e66}^{30\rightarrow25,t+5\rightarrow t}$) of the mid-age group have to be multiplied with the $eP_{e6}^{30,t+10}$-value to withdraw population that already made the transition in the mid-age group (see Figure 18 & Equation 7).

The calculated *bTR*s are applied on the for time t-5 reconstructed cumulative age-specific education structure (absolute) to determine how many people remain in the age group and how many to shift in the age groups 15 to 34 years from higher to lower education categories and to create a *"final"* reconstructed educational distribution at time t-5. This procedure gets iteratively repeated in 5-year steps from the available base year until 1950 resulting in a comprehensive, consistent and comparable time series of educational attainment by age and sex.

Box 3.3.2 Example of Computation of Education Transitions in the Reconstruction for Italy (Men), 2010

Basis: Educational distribution by age at time t (in %) [P_i]

Time	Age	P_{e1}	P_{e2}	P_{e3}	P_{e4}	P_{e5}	P_{e6}
t	15-19	0.1%	0.3%	2.0%	77.3%	20.4%	0.0%
t	20-24	0.1%	0.5%	1.3%	25.6%	64.1%	8.5%
t	25-29	0.1%	0.7%	2.0%	27.5%	53.1%	16.7%
t	30-34	0.4%	0.7%	2.7%	31.5%	45.5%	19.2%

Step1: Applying Equation 3 to get cumulative educational distribution (in %) [P_{i+}]

Time	Age	P_{e1+}	P_{e2+}	P_{e3+}	P_{e4+}	P_{e5+}	P_{e6}
t	15-19	100.0%	99.9%	99.6%	97.6%	20.4%	0.0%
t	20-24	100.0%	99.9%	99.4%	98.1%	72.6%	8.5%
t	25-29	100.0%	99.9%	99.1%	97.2%	69.7%	16.7%
t	30-34	100.0%	99.6%	98.9%	96.2%	64.7%	19.2%

Step 2: Applying Equation 4 to get Educational Attainment Progression Ratios [*EAPR*]

Time	Age	*EAPR12*	*EAPR23*	*EAPR34*	*EAPR45*	*EAPR56*
t	15-19	0.999	0.997	0.980	0.209	0.000
t	20-24	0.999	0.995	0.987	0.740	0.116
t	25-29	0.999	0.993	0.980	0.717	0.239
t	30-34	0.996	0.993	0.973	0.673	0.297

Step 3: Applying Equation 5 to get logit of the EAPR [*logitEAPR*]

Time	Age	*logitEAPR12*	*logitEAPR23*	*logitEAPR34*	*logitEAPR45*	*logitEAPR56*
t	15-19	7.344	5.676	3.911	-1.333	-9.924
t	20-24	7.014	5.344	4.328	1.043	-2.026
t	25-29	6.587	4.915	3.908	0.931	-1.159
t	30-34	5.406	5.005	3.569	0.720	-0.861

Step 4: The logitEAPR's get extrapolated by applying linear regression equations, which get converted to EAPRs and in a further step into a so-called extrapolated cumulative educational distribution [eP_{i+}]

Time	Age	eP_{e1+}	eP_{e2+}	eP_{e3+}	eP_{e4+}	eP_{e5+}	eP_{e6}
t	15-19	100.0%	99.9%	99.5%	98.1%	69.0%	20.5%
t+5	20-24	100.0%	100.0%	99.6%	98.5%	72.5%	22.1%
t+10	25-29	100.0%	100.0%	99.7%	98.9%	75.7%	23.7%
t+15	30-34	100.0%	100.0%	99.7%	99.1%	78.6%	25.2%

Step 5: Applying Equation 6 to 10 to get estimated backward Transition Ratio's [*bTR*]

Time	Age	bTR_{e22}	bTR_{e33}	bTR_{e44}	bTR_{e55}	bTR_{e66}
t→t-5	15-19	⊗	⊗	99.1%	26.9%	○
t→t-5	20-24	○	○	⊗	100.0%	47.4%
t→t-5	25-29	○	○	○	⊗	75.3%
t→t-5	30-34	○	○	○	○	⊗

Note: The calculation of the *bTR* uses an integrated rule that its values cannot surpass 100%, although the calculation might show another result, here the case for $bTR_{e55}^{25\rightarrow 20}$.
Data source: Adapted Italy 2011 Census Data after smoothing and data adjustments described in Section 3.2.2.

3.3.3 Projecting to the Baseline Year 2015

While the reconstruction (until 1950) starts from the latest available year (ending with 0 or 5), population projections start from 2015. In order to complete the gap until 2015, the educational attainment from the latest available year (ending with 0 or 5) had to be projected to 2015 by first applying a similar algorithm as for the reconstruction by applying education-specific mortality differential (see section 0). Following the Wittgenstein Centre's Global Education Trend Scenario developed for all countries starting from the latest available year (Lutz et al. 2018; KC et al. 2018), medium education scenarios were applied to project the education distribution among the 15–34 years old. Finally, the resulting age-sex distribution was adjusted to exactly match the UN's estimates of age-sex distribution (United Nations 2017), while not changing the education distribution within each age–sex group.

4 Assessment and Comparison

4.1 Comparison with the Wittgenstein Centre 2014 Dataset

As compared to WIC 2014, the new WIC 2018 dataset provides updated estimates on past levels of education, expanding both the reconstruction period to 1950 (as compared to previously 1970) and the coverage to 185 countries. While the Netherlands Antilles are no longer considered as a separate country within the new dataset, 16 additional countries have been added: Afghanistan, Angola, Botswana, Curaçao, the Federated States of Micronesia, Fiji, Israel, Kiribati, North Korea, Oman, Solomon Islands, South Sudan, Sri Lanka, Taiwan, Togo and Yemen.

Differences in estimates between the two datasets result both from different base years and different data sources, as well as from changes in methodology as described in Section 3. Figure 19 gives the example for Egypt, a country where both the baseline (2006 IPUMS) and input data did not change in the revised version. Consequently, differences in the reconstruction of educational attainment depicted below solely result from the updated methodology. Even though the overall trend remains the same in both datasets, new estimates project on average lower shares for 'No education' and higher shares for 'Incomplete primary education' for 25–39 year-old Egyptians as compared to the previous version.

In total, 169 comparable countries and 9 comparable years make for 1521 data points which are available to compare the WIC 2014 dataset to the WIC 2018 dataset. Figure 20 summarises the extent of deviation over time for both 25–39 year-olds and the population aged over 25. When looking at the total population aged 25–39, more than half of these data points (793 data points or 52.1 per cent) deviate by less than 5 percentage points (pp) in all education categories, whereof 442 data points (29.1 per cent) have an absolute difference in

all levels of education even lower than 2.5 pp. About 21.5 per cent or 327 data points deviate by more than 10 pp in one or more education category; with only 76 data points (5.0 per cent) revealing differences higher than 20 pp. Estimates between the two data revisions are most similar for the years 1990 to 2005. Deviations hardly vary between sexes and reveal very similar results for the total population aged over 25. In this open-ended age group, differences between the datasets are even slightly smaller than when compared to only 25–39 year-olds: 57.9 per cent (881 data points) of all data points deviate by less than 5pp, whereof 493 data points (32.4 per cent) demonstrate a maximum difference of 2.5 pp in all education levels. Again, estimates are most similar around the turn of the millennium.

With regard to differences by educational categories, the largest deviation, on average, applies to the higher secondary education level. Figure 21, depicting the mean discrepancy from the WIC 2018 to the WIC 2014 dataset by education category, shows that the new estimates, on average, project larger shares of higher secondary education, while lower education categories are estimated smaller compared to the previous version.

Figure 19. Comparison between the WIC 2014 and the WIC 2018 estimates on past levels of education, Egypt, total population aged 25–39, 1970–2010

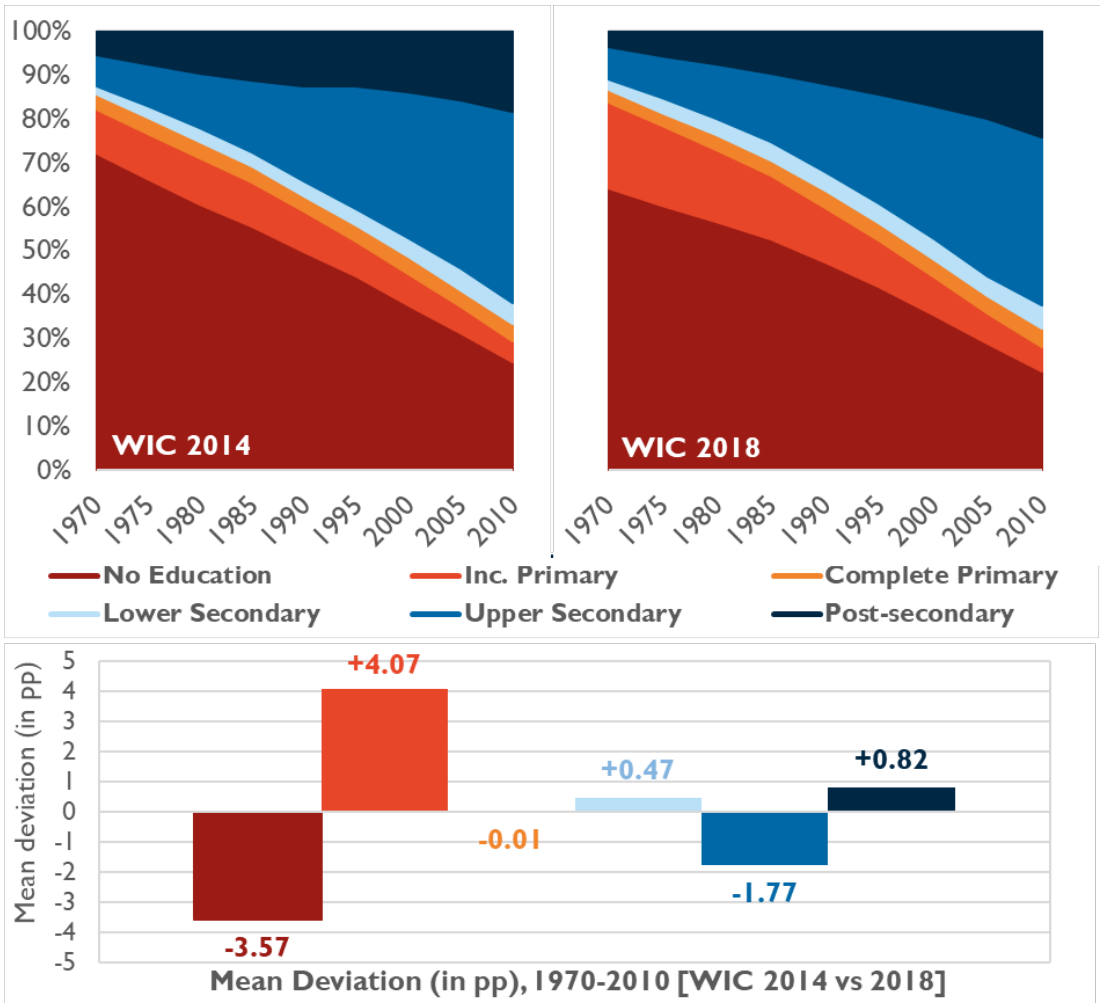


Figure 20. Extent of deviation of the WIC 2018 to the WIC 2014 dataset in number of data points, total population aged 25–39 [left panel] and 25+ [right panel], all countries, 1970–2010

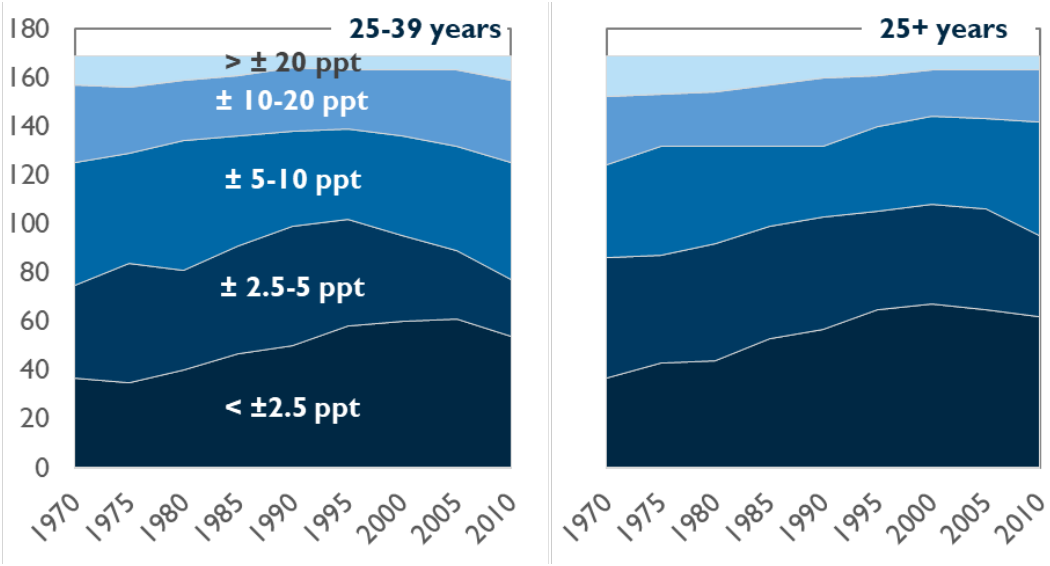
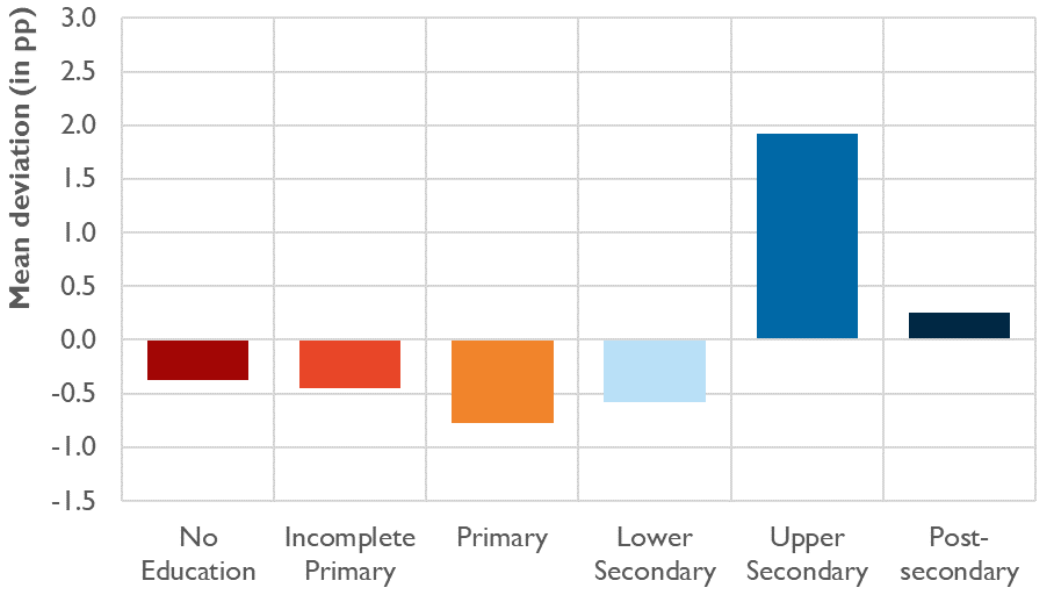


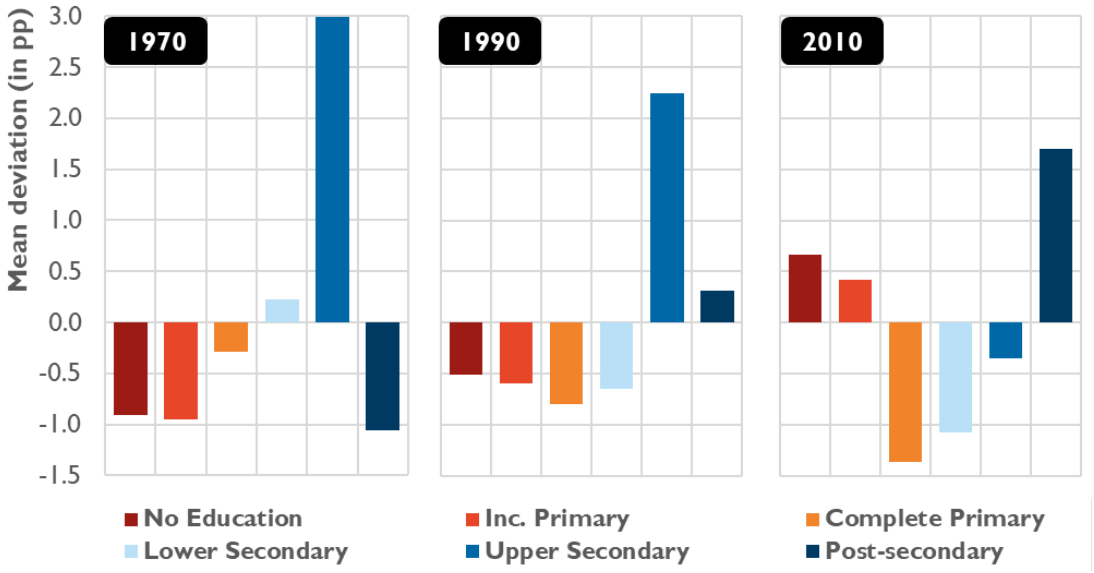
Figure 21. Mean deviation of the WIC 2018 to the WIC 2014 dataset by education category in percentage points, total population aged 25–39, all countries, 1950–2010



However, as shown in Figure 22, the average discrepancy by education category changes remarkably over time. While for the year 1970 the share of 'Higher secondary education' is estimated much higher in WIC 2017 than in WIC 2014, the opposite is true for the year 2010. Instead, more people are projected to have attained a tertiary degree. Generally, estimates of the share of the least educated ('No education') in the revised WIC

2017 dataset tend to be more optimistic for early years and more pessimistic for recent years, as compared to the WIC 2014 dataset. By contrast, post-secondary education estimates in the revised dataset are on average more pessimistic for 1970, but more optimistic for 2010.

Figure 22. Mean deviation of the WIC 2018 to the WIC 2014 dataset by education category in percentage points, total population aged 25–39, all countries, selected years



However, deviations discussed above represent only average discrepancies for all countries available for comparison and can thus differ significantly on the country level. Countries demonstrating the largest discrepancies between the two datasets are often countries with limited empirical data points, including countries with no usable historical data on educational attainment, such as French Guiana, Iceland, Laos, Latvia, Palestine, Moldova, Tajikistan, Ukraine or the United Kingdom.

Nevertheless, discrepancies between the two datasets are for the most part minor and can be explained by changes in base years, data sources and an updated methodology. Moreover, the general trend remains the same for each country, resulting in a high overall matching accuracy. Consequently, updated estimates of past levels of education within the new WIC 2018 dataset can be considered as stringent and coherent.

4.2 Comparison with the Barro & Lee 2015 Dataset

The Barro & Lee datasets are the most widely used reconstructed datasets on past levels of education (Barro and Lee 1993, 2010, 2013, 2015). While previously applying the *Perpetual Inventory Method* to estimate past educational attainment, the authors have updated their methodology in their latest revision (Barro and Lee 2015). Based on the collection of empirical data points, mainly from the *UNESCO Institute for Statistics* (UIS), Barro and Lee now use observations in 5-year age intervals for the previous or subsequent 5-year periods.

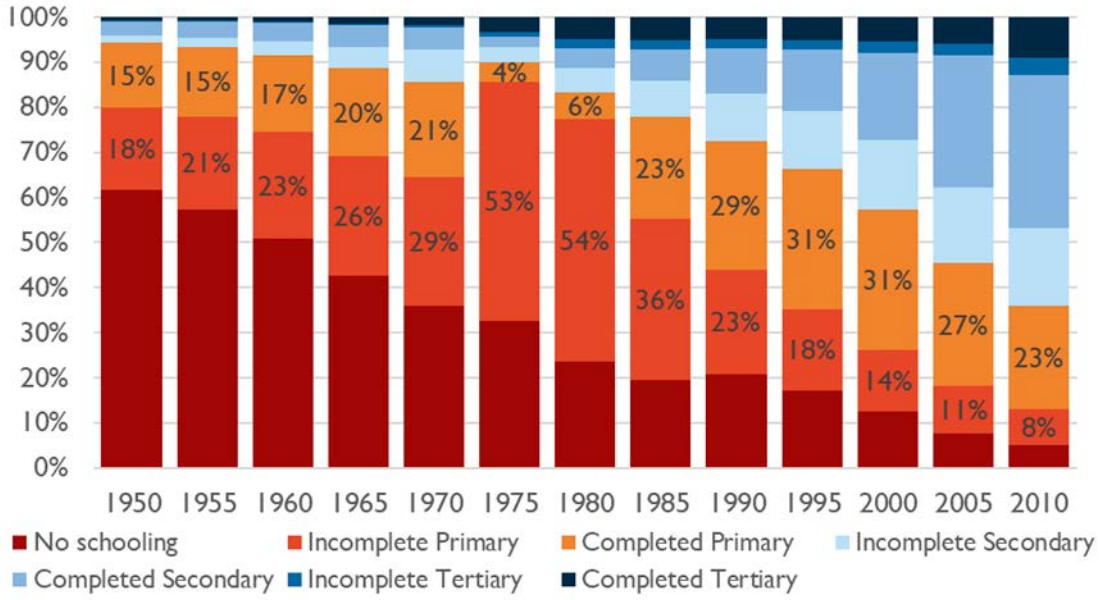
Similar to the WIC approach, they assume differential mortality by education for the population aged 65+, broadly distinguishing between groups of OECD and non-OECD countries as well as between two broad educational groups: a less educated population, having attained primary education at most, and a more educated population, with at least secondary schooling. Based on enrolment data, aggregated and overlapping education categories from censuses are split up into four education classes (no formal education, primary, secondary and tertiary education) which are further disaggregated into seven categories by means of age- and sex-specific completion ratios (Barro and Lee 2013).

A main reason for the low matching accuracy of Barro & Lee with the WIC 2017 dataset is their heavy reliance on unharmonised UIS data. As highlighted by Speringer and colleagues (2015), inconsistencies in coding as well as overlapping of different levels of education occur frequently in different variations and intensities in the UIS dataset (Speringer et al. 2015). Educational categories, such as incomplete and complete primary education, are often aggregated into one single category, resulting in issues related to disentangling. The decomposition method used by Barro and Lee to separate incomplete from completed education levels causes several oddities in the time series. For instance, in the case of Brazil, a country that provides detailed time series via National Statistical Offices and IPUMS, the Barro & Lee dataset shows an 86 per cent growth rate of incomplete primary education from 1970 to 1975 for 25–39 year-olds and in the same period a decrease of completed primary education from 21 per cent to 4 per cent. From 1980 to 1985 the share of people between 25 and 39 with completed primary education suddenly almost quadruples again from 6 per cent to 23 per cent (see Figure 23).

Barro and Lee's strong confidence in UIS data becomes particularly problematic for the reconstruction of countries that are based on just one or two data points, which is the case for more than half of the countries in their dataset (Barro and Lee 2013). However, Barro and Lee recently showed awareness of these problems and managed to considerably reduce inconsistencies in time series in their latest data revisions. Nevertheless, remaining oddities in addition to differences in methodology, data sources and base years contribute to the high deviations between the WIC 2018 and the Barro & Lee 2013 dataset.

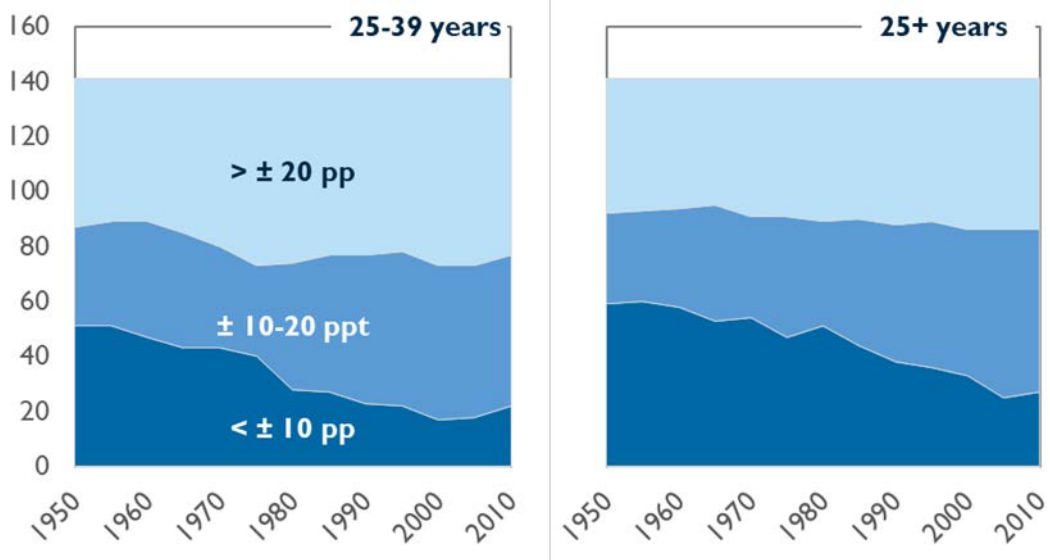
The latest Barro and Lee dataset (2013 v. 2.2, June 2018) provides educational attainment data, disaggregated by age and sex, from 1950 to 2010 in 5-year time intervals. It contains estimates for 146 countries, including five countries that are not listed in the WIC 2018 dataset: Barbados, Brunei, Libya, Mauritania and Papua New Guinea. As Barro and Lee do not differentiate between lower and higher secondary education, the comparison is limited to the following five categories: no formal education, incomplete primary, primary, secondary and post-secondary education (Barro and Lee 2013, 2015).

Figure 23. Total population aged 25-–39 by education, Brazil, 1950–2010 (Barro & Lee 2013) [authors' illustration]



In the case of 25–39 year-olds, out of the 141 countries and 1833 data points available for comparison, only 432 data points or 23.6 per cent show an absolute difference of less than 10 pp in all education categories. A vast majority of comparable data points (76.4 per cent) deviates by more than 10 pp in one or more education categories, with 43.7 per cent even deviating by more than 20 pp. As can be seen on Figure 24, differences become greater over time. No significant differences between genders are observable; however, deviations are slightly smaller for females than for males.

Figure 24. Extent of deviation of the WIC 2018 to the Barro & Lee 2013 dataset in number of data points, total population aged 25–39 and 25+, all countries, 1950-2010



Estimates of the two datasets are slightly closer when comparing the total population aged over 25 years. In this case, 585 out of the 1833 data points or 31.9 per cent deviate by less than 10 pp in all education categories. Another 31.9 per cent show discrepancies between 10 and 20 pp, while a substantial share of 36.2 per cent of all comparable data points deviates by even more than 20 pp in one or more of the five education categories. In terms of time, the general pattern remains the same: deviations are higher in recent years and lower in early years.

As regards differences by education category, the estimated shares of primary and secondary education show the greatest deviations. When looking at the mean discrepancy from the WIC 2018 to the Barro & Lee 2013 dataset by education for all countries and all years (see Figure 25), primary and secondary education categories deviate on average by more than 10 pp, whereas other categories show considerably less deviation.

While our dataset estimates on average significantly shares for lower primary educational attainment than Barro & Lee, the opposite is true for secondary education. This does not change over the years; however, the extent of the deviation varies with time (Figure 26). As regards the share of people attaining post-secondary education, the positive difference between WIC 2018 and Barro & Lee 2013 considerably increases with time, suggesting that estimates of tertiary education by Barro & Lee are, on average, much more pessimistic—particularly for recent years. Again, these figures only represent average deviations and can differ widely between specific countries.

Figure 25. Mean deviation of the WIC 2018 dataset to the Barro & Lee 2013 dataset by education category in percentage points, total population aged 25–39, all countries, 1950–2010

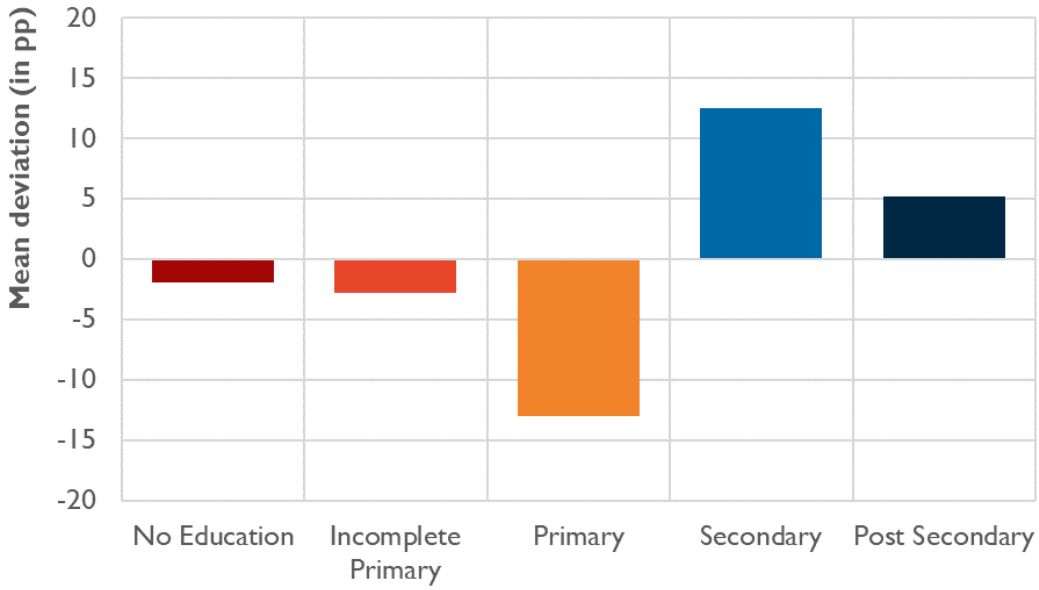
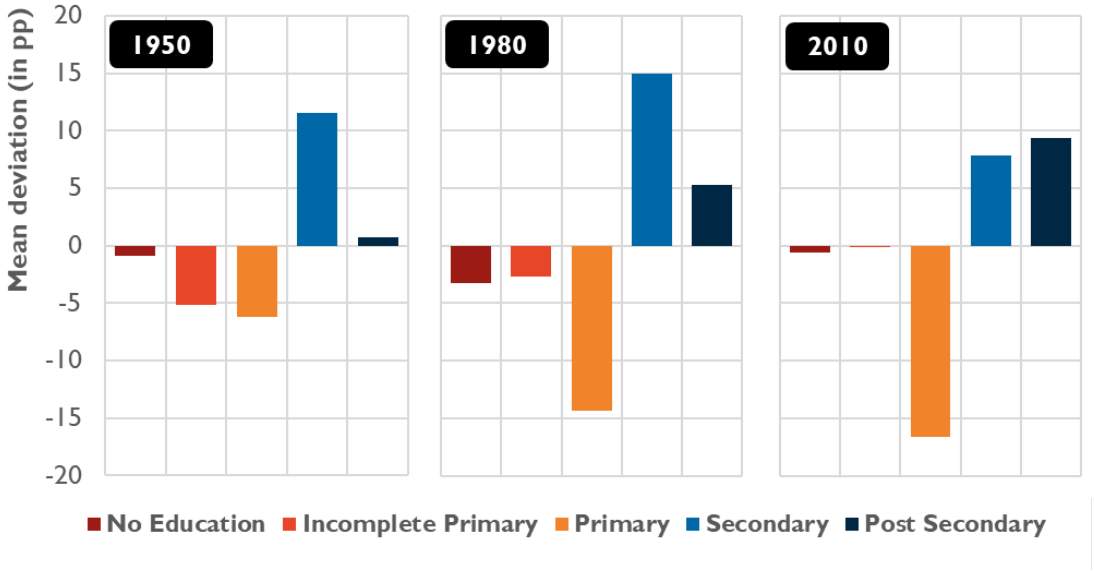


Figure 26. Mean deviation of the WIC 2018 dataset to the Barro & Lee 2013 dataset by education category in percentage points, total population aged 25–39, all countries, selected years



Although the overall matching accuracy between the WIC 2018 and the Barro & Lee 2013 dataset is undeniably low, discrepancies seem to be traceable and plausible. Due to considerable differences in methodologies and data sources used as well as potential shortcomings identified in Barro and Lee's work related to unharmonised UIS data and issues of disentangling, high deviations of their estimates on past levels of education from our reconstruction of educational attainment do not necessarily contradict with the coherency and validity of our data.

5 Conclusion

The WIC 2018 dataset has increased the time and geographic coverage compared to the WIC 2014 version. It provides data on population by age, sex and level of educational attainment for the period 1950 to 2015 for 185 countries. Not only were the data time series extended, but also validated, which will be important for future global comparative research on patterns of change in educational attainment and on examining the role of educational attainment in different context. Beside the methodological improvements in assembling the base-year data and the reconstruction model itself, the main improvement lies in the incorporation of historical data in the reconstruction. This yields a notable increase in the accuracy of the reconstruction output as the historical data points serve as marker points.

This dataset can be further improved, for instance by further increasing the coverage of historical and base year data points, e.g. additional data mining. An improvement that we

have undertaken for a limited number of countries[15] is to extend the reconstruction timeframe beyond 1950 to 1900 to cover the entire 20th century in order to analyse the interplay between the education expansion and other socio-economic and technological developments.

[15] See www.edu20C.org (accessed on 23/11/2018)

References

Barro, Robert J., and Jong Wha Lee. 1993. "International Comparison of Educational Attainment." *Journal of Monetary Economics* 32 (3): 363–394.

———. 2010. "A New Data Set of Educational Attainment in the World, 1950-2010." NBER Working Paper No.15902. Cambridge, Massachusetts: National Bureau of Economic Research. http://www.nber.org/papers/w15902.

———. 2013. "A New Data Set of Educational Attainment in the World, 1950–2010." *Journal of Development Economics* 104 (September): 184–98. https://doi.org/10.1016/j.jdeveco.2012.10.001.

———. 2015. *Education Matters: Global Schooling Gains from the 19th to the 21st Century*. Oxford: Oxford University Press.

Bauer, Ramon, Michaela Potančoková, Anne Goujon, and Samir KC. 2012. "Populations for 171 Countries by Age, Sex, and Level of Education around 2010: Harmonized Estimates of the Baseline Data for the Wittgenstein Centre Projections." Interim Report IR-12-016. Laxenburg, Austria: International Institute for Applied Systems Analysis (IIASA). http://pure.iiasa.ac.at/10259/.

CELADE/CEPAL. 2014. "Redata Informa. Software para procesar y mapear datos de censos y encuestas para análisis local y regional." Redatam Informa. 2014. http://www.cepal.org/cgi-bin/getprod.asp?xml=/redatam/noticias/paginas/8/14188/P14188.xml&xsl=/redatam/tpl/p18f.xsl&base=/redatam/tpl-i/top-bottom.xsl.

Cohen, Daniel, and Laura Leker. 2014. "Health and Education: Another Look with the Proper Data." Paris, France. http://www.parisschoolofeconomics.eu/docs/cohen-daniel/cohen-leker-health-and-education-2014.pdf.

Cohen, Daniel, and Marcelo Soto. 2007. "Growth and Human Capital: Good Data, Good Results." *Journal of Economic Growth* 12 (1): 51–76. https://doi.org/10.1007/s10887-007-9011-5.

Eurostat. 2018. "Eurostat. Your Key to European Statistics." EUROSTAT Database. 2018. http://ec.europa.eu/eurostat/data/database?p_p_id=NavTreeportletprod_WAR_NavTreeportletprod_INSTANCE_nPqeVbPXRmWQ&p_p_lifecycle=0&p_p_state=normal&p_p_mode=view&p_p_col_id=column-2&p_p_col_pos=1&p_p_col_count=2.

Fuente, Angel de la, and Rafael Doménech. 2013. "Cross-Country Data on the Quantity of Schooling: A Selective Survey and Some Quality Measures." Working Paper 720. Barcelona, Spain: Barcelona Graduate School of Economics. http://ideas.repec.org/p/bge/wpaper/720.html.

Goujon, Anne, Samir K.c, Markus Speringer, Bilal Barakat, Michaela Potančoková, Jakob Eder, Erich Striessnig, Ramon Bauer, and Wolfgang Lutz. 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): 315–63. https://doi.org/10.1017/dem.2016.10.

KC, Samir, Wolfgang Lutz, Michaela Potančoková, Guy J. Abel, Bilal Barakat, Jakob Eder, Anne Goujon, et al. 2018. "Approach, Methods and Assumptions." In *Demographic and Human Capital Scenarios for the 21st Century. 2018 Assessment for 201 Countries.*, 19–28. Luxembourg: Publications Office of the European Union.

Lutz, Wolfgang, William P. Butz, and Samir KC, eds. 2014. *World Population and Human Capital in the Twenty-First Century*. Oxford, UK: Oxford University Press. http://ukcatalogue.oup.com/product/9780198703167.do.

Lutz, Wolfgang, Anne Goujon, Samir KC, and Warren C. Sanderson. 2007a. "Reconstruction of Populations by Age, Sex and Level of Educational Attainment for 120 Countries for 1970-2000." *Vienna Yearbook of Population Research* 2007: 193–235.

———. 2007b. "Reconstruction of Populations by Age, Sex and Level of Educational Attainment for 120 Countries for 1970-2000." Interim Report IR-07-002. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Lutz, Wolfgang, Anne Valia Goujon, Samir KC, Marcin Stonawski, and Nikolaos Stilianakis. 2018. "Demographic and Human Capital Scenarios for the 21st Century: 2018 Assessment for 201 Countries." European Union.

Minnesota Population Center. 2018. "Integrated Public Use Microdata Series, International: Version 7.0 [Dataset]." IPUMS International. 2018. https://doi.org/10.18128/D020.V7.0.

OECD. 2018. "Data Warehouse." Doi:https://doi.org/10.1787/data-00900-en. 2018. https://www.oecd-ilibrary.org/content/data/data-00900-en.

Potančoková Michaela, Samir KC, and Anne Goujon. 2014. "Global Estimates of Mean Years of Schooling: A New Methodology". *IIASA Interim Report*. IIASA, Laxenburg, Austria: IR-14-005

Speringer, Markus, Anne Goujon, Jakob Eder, Samir KC, Ramon Bauer, and Michaela Potančoková. 2015. "Validation of the Wittgenstein Centre Back-Projections for Populations by Age, Sex, and Level of Education from 1970 to 2010." Interim Report IR-15-008. Laxenburg, Austria: International Institute for Applied Systems Analysis (IIASA).

Speringer, Markus, Anne Goujon, and Sandra Jurasszovich. 2018. "Inequality in Educational Development from 1900 to 2015." In *Classes - From National to Global Class Formation*, edited by Hardy Hanappi. London, UK: InTechOpen.

UNESCO. 2006. "International Standard Classification of Education: ISCED 1997 (Reprint)." Montreal, Canada: UNESCO Institute for Statistics. http://www.uis.unesco.org/Library/Documents/isced97-en.pdf.

———. 2012. “International Standard Classification of Education, ISCED 2011.” Montreal, Canada: UNESCO Institute for Statistics. http://unesdoc.unesco.org/images/0021/002191/219109e.pdf.

UNESCO Institute for Statistics. 1997. “ISCED 1997 Mappings.” 1997. http://www.uis.unesco.org/Education/ISCEDMappings/Pages/default.aspx.

United Nations. 2015. *World Population Prospects: The 2015 Revision, Key Findings and Advance Tables*. Working Paper, ESA/P/WP.241. New York: United Nations Population Division | Department of Economic and Social Affairs. https://www.popline.org/node/639412.

———. 2017. “World Population Prospects: The 2017 Revision.” New York, NY: Department of Economic and Social Affairs, Population Division. http://esa.un.org/unpd/wpp/.

Wittgenstein Centre. 2018. “Wittgenstein Centre Data Explorer Version 2.0.” Database. Wittgenstein Centre Data Explorer. 2018. www.wittgensteincentre.org/dataexplorer.

World Bank. 2018. “World Development Indicators.” 2018. https://datacatalog.worldbank.org/dataset/world-development-indicators.

Acronyms

CELADE	Latin American and Caribbean Demographic Centre
DHS	Demographic and Health Survey
DESA	United Nations Department of Economic and Social Affairs
EAPR	Education Attainment Progression Ratio
HCDL	Human Capital Data Lab
HIC	High-income Countries
IIASA	International Institute for Applied Systems Analysis
IMCR	Iterative Multi-dimensional Cohort-component Reconstruction Model
INED	Institut national d'études démographiques (engl. French Institute for Demographic Studies)
INSEE	Institut national de la statistique et des études économiques (engl. National Institute of Statistics and Economic Studies)
IPUMS	Integrated Public Use Microdata Series
ISCED	International Standard Classification of Education
ISO	International Organization for Standardization
JRC	Joint Research Centre, European Commission
LFS	Labour Force Survey
LMIC	Low and Middle-income Countries
MICS	Multiple Indicator Cluster Survey
NSO	National Statistical Office
ÖAW	Österreichische Akademie der Wissenschaften (engl. Austrian Academy of Sciences)
OECD	Organisation for Economic Co-operation and Development
S_x	Survivorship Ratios
UIS	UNESCO Institute for Statistics
UN	United Nations
VID	Vienna Institute of Demography
WIC	Wittgenstein Centre for Demography and Global Human Capital
WPP	World Population Prospects

Annex: Methodological Notes

A. Filling the data gaps—educational compositions for 16 countries with missing educational data

For the 16 countries with missing educational data (listed in Table 1) data on the age and sex compositions for 2015 from the UN WPP 2017 revision were available. The educational compositions for these countries were imputed using information from the most similar country in the region or regional average. The approximation countries and procedures are the same as in the previous round of projections (Lutz, Butz and KC 2014). This imputation was necessary for the global population projections to have information for all 201 countries (Lutz et al. 2018; KC et al. 2018). In the reconstruction exercise, those countries were not used, which limits the number of countries in the reconstruction dataset to 185. For analyses of prospective educational trends, however, it is advisable to be cautious when using the data for the 16 countries listed in Table A1.

Table A1. Overview of imputed and the source countries for the imputation:

Case	Country	Approximated with
1	Antigua and Barbuda	Saint Lucia
2	Barbados	Saint Lucia
3	Brunei Darussalam	Indonesia
4	Mayotte	Réunion
5	Eritrea	Average of Somalia, Ethiopia, Sudan
6	Djibouti	Average of Somalia, Ethiopia, Sudan
7	Grenada	Saint Lucia
8	Guam	Fiji
9	Libya	Average of Morocco, Tunisia, Egypt
10	Mauritania	Senegal
11	Papua New Guinea	Solomon Islands
12	Seychelles	Réunion
13	Western Sahara	Senegal
14	Channel Islands	United Kingdom
15	US Virgin Islands	Saint Lucia
16	Uzbekistan	Tajikistan

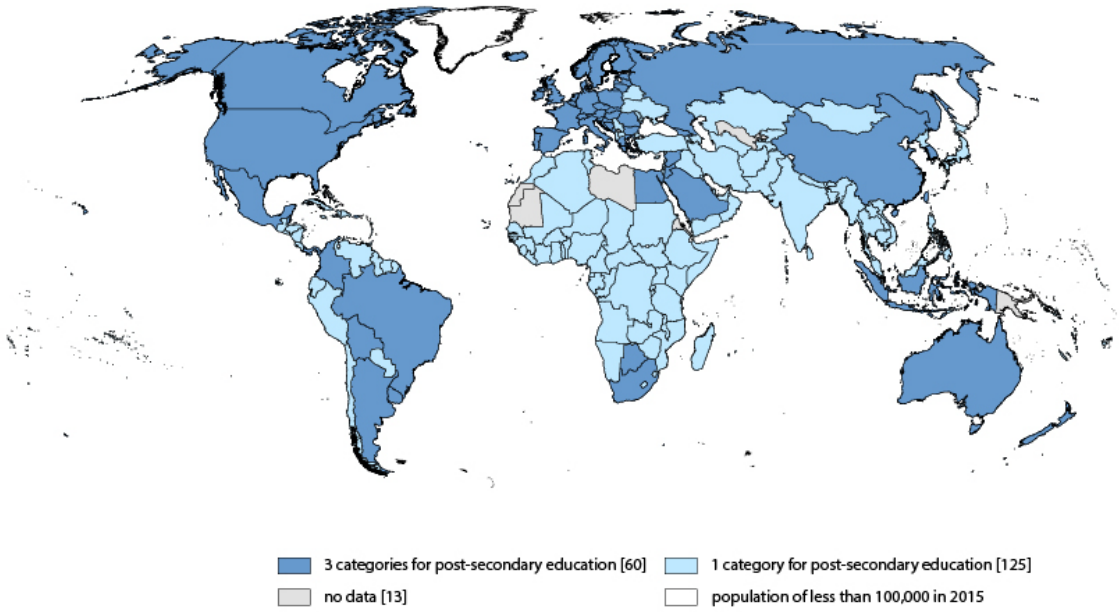
B. Post-secondary Subset

The 2011 ISCED revision clearly separates between the tertiary degrees and allows us to create more meaningful subcategories at the post-secondary level. We now differentiate between short-cycle post-secondary, bachelor's or equivalent degrees, and master's degree or its equivalent and higher degrees. This was not possible previously as in ISCED 1997, level 5A included both bachelor and master's degrees that were both classified as first

stages of tertiary education and therefore it was not possible to clearly disentangle between bachelor's degrees that take on average 3–4 years to accomplish, and master's degrees that usually take an additional 2 years.

We were able to collect and harmonise data into more detailed post-secondary categories for the 60 countries where a) data were available, and b) the share of persons with post-secondary education reached at least 10% among those aged 25 to 34 years. We restrict the number of countries with information on detailed post-secondary levels for two reasons: data availability and relevance. The subset of the countries with more detailed post-secondary categories covers 60 countries shown in Figure B1. All three subcategories are available for 58 countries. In Brazil and Russia the organisation of the census data covers only ISCED 6–8 categories, and the short-track post-secondary category is missing.[16]

Figure B1. Post-secondary education categories in the WIC 2018 dataset



In most data archives the data are still organised into ISCED 1997 levels (UNSD, EUROSTAT census hub), thus we had to request the detailed data directly from the statistical offices, or triangulate between different data sources (mostly LFS) to obtain full range of categories. All the data were meticulously checked and compared for correspondence and in case the correspondence was good enough we used the alternative data source to split the census data into more detailed categories. This was in fact done not only to split the post-secondary category into the three more detailed ones, but also in cases

[16] In Russia Persons with ISCED 4 or 5 diplomas are included in completed secondary education category (ISCED 3-4-5). In Brazil ISCED 4 level programs do not exist and the marginal ISCED 5 programs (Diploma de curso superior sequencial) are not surveyd as a separate category but are also included with the regular completed secondary education (ISCED 3 and 5).

where the census data collected only information for the aggregate *"low education"* or *"no diploma or certificate"* or other categories comprising several ISCED categories at the lower educational spectrum.

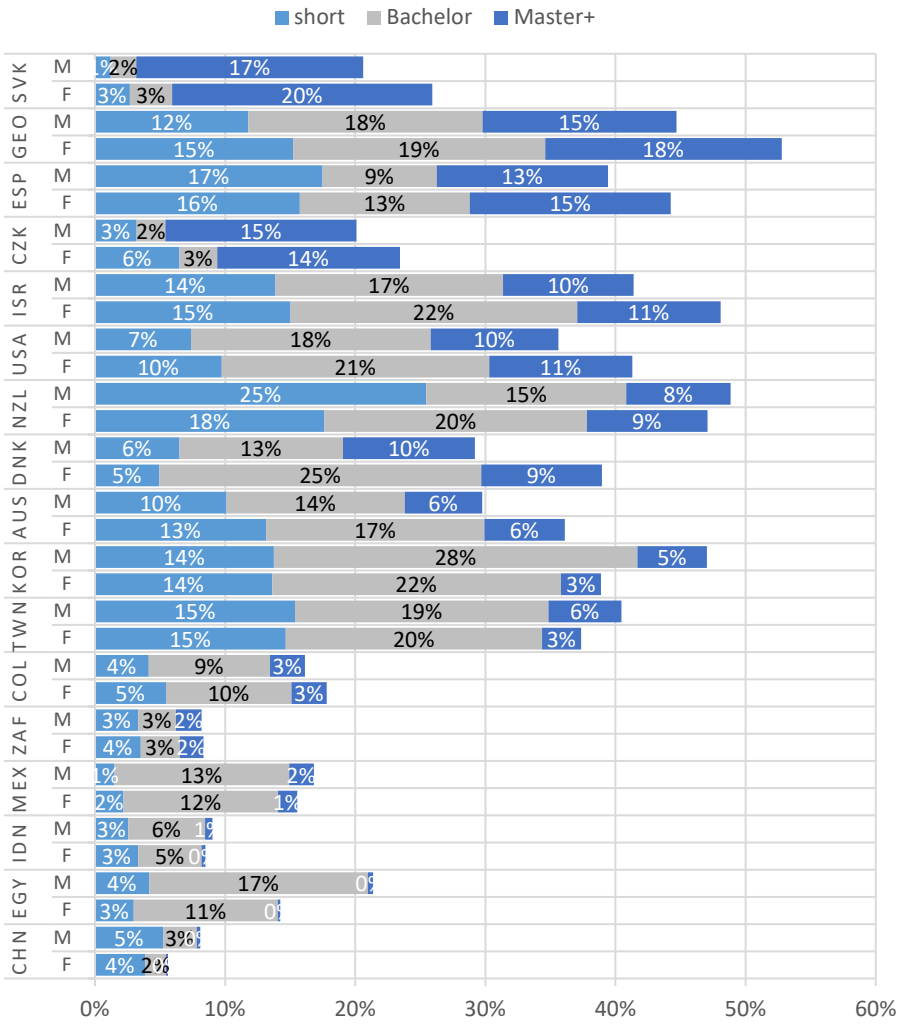
Originally we were aiming to obtain detailed post-secondary education categories for all OECD countries. However, for some countries the census data were not organised by completed degrees (Japan). On the other hand, we found sufficiently detailed data for many non-OECD countries.

The main obstacle in organising the data into eight categories for all 185 countries in the dataset were data limitations. In many developing countries we had to rely on DHS and similar surveys; these surveys only ask if respondents have some higher education but not the actual degrees. Similarly, many censuses only differentiate between post-secondary non-university and university education, without specification of degrees or diplomas awarded, which complicates any effort at categorisation.

The distinction between kinds of post-secondary education brings new perspective into cross-country comparisons. It becomes clear that in the countries with overall high shares of post-secondary educated persons (Georgia, Israel, New Zealand, South Korea, USA) a substantial fraction holds short-track non-university degrees or only first-cycle university degrees (bachelor or equivalent) while only a smaller share has completed a master's degree. This share is surprisingly similar at about 10%–15% among the 25–59 year olds in many countries (Figure B2). In most countries about 25–35% among post-secondary educated persons would achieve master's degree. Furthermore, in countries with a high share of post-secondary education it is often the case that women have higher education than men as the share of women with at least a master's degree is in many countries higher or at least on par with men (Slovakia, Georgia, Spain, Israel, USA). Thus it is not universally true that where women have an advantage in education over men they would rather hold lower degrees, although such patterns exist in some countries.

In some post-socialist countries, such as Slovakia or the Czech Republic, we find exceptionally high shares of population with at least a master's degree. Bachelor-level degrees and two-cycle university studies did not traditionally exist and were introduced only in the late 1990s with the Bologna process. Thus a majority of those with a post-secondary degree (60–70%) completed full university education and hold at least a master's degree. Only the late introduction of bachelor-level studies, however, cannot explain the exceptionally high share of those competing master in some countries. What may play a role is preferences of the population, the higher prestige of full university education compared to "just" the bachelor title among those studying and among employers as well, but also the educational system and practices at the universities that are often financed based on the number of students.

Figure B2. Different patterns and distributions of post-secondary levels among the selected countries (countries are ranked by the share of Master's+)



In contrast, the share of the working-age population (25–59) with at least a master's degree is very low in middle-income countries where the overall share of population with post-secondary level is even smaller (Indonesia, China, South Africa). These countries are currently experiencing an often rapid educational expansion, and the population with a university degree is increasing among the younger cohorts. Thus, the differences in educational composition are particularly pronounced between the age groups. The share of Chinese women with a post-secondary education quadrupled comparing women aged 45–49 and roughly the generation of their daughters aged 25–29 at 2010 census.

Table B1. Educational expansion in female population in China and South Korea

	CHINA – women				South Korea – women			
	Short	Bachelor	Master+	Post-secondary	Short	Bachelor	Master+	Post-secondary
25 to 29	11%	8%	1%	20%	26%	42%	3%	71%
30 to 34	9%	5%	1%	15%	25%	34%	5%	64%
35 to 39	6%	3%	0%	9%	18%	26%	4%	48%
40 to 44	4%	2%	0%	6%	12%	21%	3%	36%
45 to 49	4%	2%	0%	5%	7%	15%	3%	25%
50 to 54	3%	1%	0%	3%	4%	9%	2%	15%
55 to 59	2%	1%	0%	3%	3%	7%	1%	10%
60 to 64	1%	1%	0%	2%	2%	5%	1%	7%
65 to 69	1%	1%	0%	2%	1%	3%	0%	5%
70 to 74	1%	1%	0%	2%	1%	2%	0%	3%
75 to 79	1%	1%	0%	1%	0%	1%	0%	2%
80 to 84	0%	0%	0%	1%	0%	1%	0%	1%
85+	0%	0%	0%	1%	0%	1%	0%	1%

South Korea has been going through what may have been the fastest educational expansion in history and the inter-generational inequality in education is particularly high there. Although 71% of women aged 25–29 had a post-secondary education, a very small share them completed a master's degree (the shares are double among men in almost every age group).

C. Mean Years of Schooling

Mean years of schooling (MYS) is a commonly used indicator that expresses a population's educational characteristics in a single number. The methodology for the MYS computations in WIC 2014 dataset is described in Potančoková et al. (2014). The method is the same in the WIC 2018 dataset, with the following modifications:

1. For the sub-set of the 60 countries with the detailed post-secondary education categories we use the following standard durations of schooling that are added to the standard duration of completed secondary level (usually 12 or 13 years of schooling): short-track non-university education +2 years; bachelor +3 years; master or higher 5.5 years. We assume that persons entered the completed level upon completion of upper secondary (ISCED 3) level. For countries with a broad post-secondary category we use a standard duration of +4 years as described in Potančoková et al. (2014).
2. We updated the simple models for estimating durations of incomplete primary education by more up-to-date DHS data, thus the parameters have slightly changed.

3. We updated the standard durations of schooling using the UIS database (last accessed November 2017). For the projections we are using the values of the last available year.

D. Additional Tables

Table D1. Availability of educational attainment categories by data source and type by country and base year

ISO[17]	Country	CC[18]	Base year	Type	Source	No education	incompl. primary	Complete primary	Lower secondary	Upper secondary	Post-secondary	Status [19]
Africa												
DZA	Algeria	12	2002	survey	SURVEY	✓	✓	✓	✓	✓	✓	[A]
AGO	Angola	24	2014	census	CENSUS	✓	✓	✓	✓	✓	✓	[C]
BEN	Benin	204	2011	survey	DHS	✓	✓	✓	✓	✓	✓	[B]
BWA	Botswana	72	2011	census	CENSUS	✓	✓	✓	✓	✓	✓	[C]
BFA	Burkina Faso	854	2006	census	CENSUS	✓	✓	✓	✓	✓	✓	[A]
BDI	Burundi	108	2010	survey	DHS	✓	✓	✓	✓	✓	✓	[A]
CMR	Cameroon	120	2005	census	CENSUS	✓	✓	✓	✓	✓	✓	[B]
CPV	Cape Verde	132	2000	census	CENSUS	✓	✓	✓	✓	✓	✓	[A]
CAF	Central African Republic	140	1995	survey	DHS	✓	✓	✓	✓	✓	✓	[A]
TCD	Chad	148	2014	survey	DHS	✓	✓	✓	✓	✓	✓	[B]
COM	Comoros	174	2012	survey	DHS	✓	✓	✓	✓	✓	✓	[B]
CIV	Côte d'Ivoire	384	2011	survey	DHS	✓	✓	✓	✓	✓	✓	[B]
COD	Congo DR	180	2013	survey	DHS	✓	✓	✓	✓	✓	✓	[B]
DJI	Djibouti	262	No Data									
EGY	Egypt	818	2006	census	CENSUS	✓	✓	✓	✓	✓	✓	[A]

[17] ISO refers to the ISO 3166 alpha-3 codes, which are three-letter country codes published by the International Organization for Standardization (ISO)

[18] CC refers to the M49 codes. In 1970, the UN Statistical Office (as the United Nations Statistics Division was then known) published a document on "United Nations Standard Country Code", this was catalogued as Series M, No. 49. As a result, subsequent revisions of this document became known as M49. (see https://unstats.un.org/unsd/methodology/m49/)

[19] [A] Same Data as in WIC 2014, [B] Updated Data from WIC 2014, and [C] Additional Data compared to WIC 2014;

ISO[17]	Country	CC[18]	Base year	Type	Source	No education	incompl. primary	Complete primary	Lower secondary	Upper secondary	Post-secondary	Status [19]
GNQ	Equatorial Guinea	226	2000	survey	MICS	✓	[est.]	✓	[est.]	✓	✓	[A]
ERI	Eritrea	232	No Data									
ETH	Ethiopia	231	2007	census	CENSUS	✓	✓	✓	✓	✓	✓	[B]
GAB	Gabon	266	2012	survey	DHS	✓	✓	✓	✓	✓	✓	[B]
GMB	Gambia	270	2013	survey	DHS	✓	✓	✓	✓	✓	✓	[B]
GHA	Ghana	288	2010	census	CENSUS	✓	✓	✓	✓	✓	✓	[B]
GIN	Guinea	324	2012	survey	DHS	✓	✓	✓	✓	✓	✓	[B]
GNB	Guinea-Bissau	624	2014	survey	MICS	✓	✓	✓	✓	✓	✓	[B]
KEN	Kenya	404	2009	census	CENSUS	✓	✓	✓	✓	✓	✓	[B]
LSO	Lesotho	426	2014	survey	DHS	✓	✓	✓	✓	✓	✓	[B]
LBR	Liberia	430	2008	census	CENSUS	✓	✓	✓	✓	✓	✓	[B]
LBY	Libya	434	No Data									
MDG	Madagascar	450	2008	survey	DHS	✓	✓	✓	✓	✓	✓	[A]
MWI	Malawi	454	2008	census	CENSUS	✓	✓	✓	✓	✓	✓	[A]
MLI	Mali	466	2009	census	CENSUS	✓	✓	✓	✓	✓	✓	[B]
MRT	Mauritania	478	No Data									
MUS	Mauritius	480	2011	census	CENSUS	✓	✓	✓	✓	✓	✓	[B]
MYT	Mayotte	175	No Data									
MAR	Morocco	504	2004	census	CENSUS	✓	✓	✓	✓	✓	✓	[A]
MOZ	Mozambique	508	2007	census	CENSUS	✓	✓	✓	✓	✓	✓	[A]
NAM	Namibia	516	2013	survey	DHS	✓	✓	✓	✓	✓	✓	[B]
NER	Niger	562	2012	survey	DHS	✓	✓	✓	✓	✓	✓	[B]
NGA	Nigeria	566	2013	survey	DHS	✓	✓	✓	✓	✓	✓	[B]
COG	Republic of Congo	178	2011	survey	DHS	✓	✓	✓	✓	✓	✓	[B]
REU	Réunion	638	2008	census	CENSUS	✓	[est.]	✓	✓	✓	✓	[A]
RWA	Rwanda	646	2012	census	CENSUS	✓	►	✓	►	✓	✓	[B]

ISO[17]	Country	CC[18]	Base year	Type	Source	No education	incompl. primary	Complete primary	Lower secondary	Upper secondary	Post-secondary	Status [19]
STP	São Tomé and Príncipe	678	2009	survey	DHS	✓	✓	✓	✓	✓	✓	[A]
SEN	Senegal	686	2002	census	CENSUS	✓	✓	✓	✓	✓	✓	[A]
SYC	Seychelles	690	No Data									
SLE	Sierra Leone	694	2004	census	CENSUS	✓	✓	✓	✓	✓	✓	[A]
SOM	Somalia	706	2006	survey	MICS	✓	✓	✓	✓	✓	✓	[A]
ZAF	South Africa	710	2011	census	CENSUS	✓	✓	✓	✓	✓	✓	[B]
SSD	South Sudan	728	2008	census	CENSUS	✓	►	✓	►	✓	✓	[C]
SDN	Sudan	729	2008	census	CENSUS	✓	✓	✓	✓	✓	✓	[C]
SWZ	Swaziland	748	2006	survey	DHS	✓	✓	✓	✓	✓	✓	[A]
TZA	Tanzania	834	2012	census	CENSUS	✓	✓	✓	✓	✓	✓	[B]
TGO	Togo	768	2013	survey	DHS	✓	✓	✓	✓	✓	✓	[C]
TUN	Tunisia	788	2010	survey	SURVEY	✓	✓	✓	✓	[est.]	✓	[A]
UGA	Uganda	800	2002	census	CENSUS	✓	✓	✓	✓	✓	✓	[A]
ESH	Western Sahara	732	No Data									
ZMB	Zambia	894	2010	census	CENSUS	✓	✓	✓	✓	✓	✓	[B]
ZWE	Zimbabwe	716	2012	census	CENSUS	✓	[est.]	✓	✓	✓	✓	[B]
Asia												
AFG	Afghanistan	4	2011	survey	NRVA	✓	►	✓	►	✓	✓	[C]
ARM	Armenia	51	2011	census	CENSUS	✓	►	✓	✓	✓	✓	[B]
AZE	Azerbaijan	31	2009	census	CENSUS	✓	[est.]	✓	✓	✓	✓	[B]
BHR	Bahrain	48	2010	census	CENSUS	✓	✓	✓	✓	✓	✓	[B]
BGD	Bangladesh	50	2011	census	CENSUS	✓	✓	✓	✓	✓	✓	[B]
BTN	Bhutan	64	2005	census	CENSUS	✓	✓	✓	✓	✓	✓	[A]
BRN	Brunei	96	No Data									
KHM	Cambodia	116	2008	census	CENSUS	✓	✓	✓	✓	✓	✓	[A]
CHN	China	156	2010	census	CENSUS	✓	[est.]	✓	✓	✓	✓	[B]

ISO[17]	Country	CC[18]	Base year	Type	Source	No education	incompl. primary	Complete primary	Lower secondary	Upper secondary	Post-secondary	Status [19]
CYP	Cyprus	196	2011	census	CENSUS	✓	►	✓	✓	✓	✓	[B]
TLS	East Timor	626	2010	census	CENSUS	✓	[est.]	✓	✓	✓	✓	[B]
GEO	Georgia	268	2014	census	CENSUS	✓	✓	✓	✓	✓	✓	[B]
HKG	Hong Kong	344	2011	census	CENSUS	✓	[est.]	✓	✓	✓	✓	[B]
IND	India	356	2011	census	CENSUS	✓	✓	✓	✓	✓	✓	[B]
IDN	Indonesia	360	2010	census	CENSUS	✓	✓	✓	✓	✓	✓	[A]
IRN	Iran	364	2011	census	CENSUS	✓	[est.]	✓	✓	✓	✓	[B]
IRQ	Iraq	368	2011	survey	MICS	✓	✓	✓	✓	✓	✓	[B]
ISR	Israel	376	2004	survey	LFS	✓	✓	✓	✓	✓	✓	[A]
JPN	Japan	392	2010	census	CENSUS	✓	►	[est.]	✓	✓	✓	[A]
JOR	Jordan	400	2015	census	CENSUS	✓	✓	✓	[est.]	✓	✓	[B]
KAZ	Kazakhstan	398	2009	census	CENSUS	✓	[est.]	✓	✓	✓	✓	[A]
KWT	Kuwait	414	2011	census	CENSUS	✓	✓	✓	✓	✓	✓	[B]
KGZ	Kyrgyzstan	417	2009	census	CENSUS	✓	✓	✓	✓	✓	✓	[B]
LAO	Laos	418	2005	census	CENSUS	✓	✓	✓	✓	✓	✓	[B]
LBN	Lebanon	422	2007	census	CENSUS	✓	✓	✓	✓	✓	✓	[A]
MAC	Macao	446	2011	census	CENSUS	✓	✓	✓	✓	✓	✓	[B]
MYS	Malaysia	458	2000	census	CENSUS	✓	[est.]	✓	✓	✓	✓	[A]
MDV	Maldives	462	2006	census	CENSUS	✓	✓	✓	✓	✓	✓	[A]
MNG	Mongolia	496	2010	census	CENSUS	✓	[est.]	✓	✓	✓	✓	[B]
MMR	Myanmar	104	2014	census	CENSUS	✓	✓	✓	✓	✓	✓	[B]
NPL	Nepal	524	2011	census	CENSUS	✓	✓	✓	✓	✓	✓	[B]
PRK	North Korea	408	2008	census	CENSUS	✓	►	✓	►	✓	✓	[C]
OMN	Oman	512	2003	census	CENSUS	✓	[est.]	✓	✓	✓	✓	[C]
PAK	Pakistan	586	1998	census	CENSUS	✓	✓	✓	✓	✓	✓	[A]
PSE	Palestine	275	2007	census	CENSUS	✓	✓	✓	✓	✓	✓	[A]

ISO[17]	Country	CC[18]	Base year	Type	Source	No education	incompl. primary	Complete primary	Lower secondary	Upper secondary	Post-secondary	Status [19]
PHL	Philippines	608	2000	census	CENSUS	✓	✓	✓	✓	✓	✓	[A]
QAT	Qatar	634	2010	census	CENSUS	✓	✓	✓	✓	✓	✓	[A]
SAU	Saudi Arabia	682	2004	census	CENSUS	✓	✓	✓	✓	✓	✓	[A]
SGP	Singapore	702	2010	census	CENSUS	✓	[est.]	✓	✓	✓	✓	[A]
KOR	South Korea	410	2010	survey	SURVEY	✓	✓	✓	✓	✓	✓	[A]
LKA	Sri Lanka	144	2001	census	CENSUS	✓	✓	✓	✓	✓	✓	[C]
SYR	Syria	760	2004	census	CENSUS	✓	✓	✓	✓	✓	✓	[A]
TWN	Taiwan	158	2010	census	CENSUS	✓	✓	✓	✓	✓	✓	[C]
TJK	Tajikistan	762	2000	census	CENSUS	✓	[est.]	✓	✓	✓	✓	[B]
THA	Thailand	764	2010	census	CENSUS	✓	✓	✓	✓	✓	✓	[B]
TUR	Turkey	792	2015	census	REGISTER	✓	✓	✓	✓	✓	✓	[B]
TKM	Turkmenistan	795	1995	census	CENSUS	✓	✓	✓	✓	✓	✓	[A]
ARE	United Arab Emirates	784	2005	census	CENSUS	✓	✓	✓	✓	✓	✓	[A]
UZB	Uzbekistan	860	No Data									
VNM	Vietnam	704	2009	census	CENSUS	✓	✓	✓	✓	✓	✓	[A]
YEM	Yemen	887	2013	survey	NSPMS	✓	►	✓	►	✓	✓	[C]
Europe												
ALB	Albania	8	2011	census	CENSUS	✓	✓	✓	✓	✓	✓	[B]
AUT	Austria	40	2013	register	REGISTER	►	►	✓	✓	✓	✓	[B]
BLR	Belarus	112	2009	census	CENSUS	✓	✓	✓	✓	✓	✓	[B]
BEL	Belgium	56	2011	census	CENSUS	✓	►	✓	✓	✓	✓	[B]
BIH	Bosnia and Herzegovina	70	2013	census	LFS	✓	✓	✓	✓	✓	✓	[B]
BGR	Bulgaria	100	2011	census	CENSUS	✓	✓	✓	✓	✓	✓	[B]
-	Channel Islands	-	No Data									
HRV	Croatia	191	2011	census	CENSUS	✓	✓	✓	✓	✓	✓	[B]

ISO[17]	Country	CC[18]	Base year	Type	Source	No education	incompl. primary	Complete primary	Lower secondary	Upper secondary	Post-secondary	Status [19]
CZE	Czech Republic	203	2011	census	CENSUS	✓	►	✓	✓	✓	✓	[B]
DNK	Denmark	208	2011	census	CENSUS	✓	►	✓	✓	✓	✓	[B]
EST	Estonia	233	2012	register	REGISTER	✓	►	✓	✓	✓	✓	[B]
FIN	Finland	246	2012	survey	LFS	►	►	✓	✓	✓	✓	[B]
FRA	France	250	2011	census	CENSUS	✓	►	✓	✓	✓	✓	[B]
DEU	Germany	276	2014	survey	MICROCENSUS	►	►	✓	✓	✓	✓	[B]
GRC	Greece	300	2011	census	CENSUS	✓	✓	✓	✓	✓	✓	[B]
HUN	Hungary	348	2011	census	CENSUS	✓	►	✓	✓	✓	✓	[B]
ISL	Iceland	352	2011	census	CENSUS	►	►	✓	✓	✓	✓	[B]
IRL	Ireland	372	2014	survey	LFS	►	✓	✓	✓	✓	✓	[B]
ITA	Italy	380	2011	census	CENSUS	✓	[est.]	✓	✓	✓	✓	[B]
LVA	Latvia	428	2011	census	CENSUS	✓	✓	✓	✓	✓	✓	[B]
LTU	Lithuania	440	2011	census	CENSUS	✓	✓	✓	✓	✓	✓	[B]
LUX	Luxembourg	442	2011	census	CENSUS	►	►	✓	✓	✓	✓	[B]
MKD	Macedonia	807	2008	survey	LFS	✓	✓	✓	[est.]	✓	✓	[A]
MLT	Malta	470	2011	census	CENSUS	►	►	✓	✓	✓	✓	[B]
MDA	Moldova	498	2004	census	CENSUS	✓	✓	✓	✓	✓	✓	[A]
MNE	Montenegro	499	2011	census	CENSUS	✓	✓	✓	✓	✓	✓	[B]
NLD	Netherlands	528	2015	survey	CENSUS	►	✓	✓	✓	✓	✓	[B]
NOR	Norway	578	2014	register	REGISTER	[est.]	►	✓	✓	✓	✓	[B]
POL	Poland	616	2011	census	CENSUS	►	►	✓	✓	✓	✓	[B]
PRT	Portugal	620	2011	census	CENSUS	✓	[est.]	✓	✓	✓	✓	[B]
ROU	Romania	642	2011	census	CENSUS	[est.]	✓	✓	✓	✓	✓	[B]
RUS	Russia	643	2010	census	CENSUS	✓	✓	✓	✓	✓	✓	[B]
SRB	Serbia	688	2011	census	CENSUS	✓	[est.]	✓	✓	✓	✓	[B]
SVK	Slovakia	703	2011	census	CENSUS	✓	►	►	✓	✓	✓	[B]

ISO[17]	Country	CC[18]	Base year	Type	Source	No education	incompl. primary	Complete primary	Lower secondary	Upper secondary	Post-secondary	Status [19]
SVN	Slovenia	705	2014	census	REGISTER	✓	►	✓	✓	✓	✓	[B]
ESP	Spain	724	2011	census	CENSUS	✓	✓	✓	✓	✓	✓	[B]
SWE	Sweden	752	2014	register	REGISTER	►	►	✓	✓	✓	✓	[B]
CHE	Switzerland	756	2014	register	REGISTER	✓	►	✓	✓	✓	✓	[B]
UKR	Ukraine	804	2001	census	CENSUS	✓	✓	✓	✓	✓	✓	[A]
GBR	United Kingdom	826	2011	census	CENSUS	►	►	►	✓	✓	✓	[B]
Latin America and the Caribbean												
ATG	Antigua and Barbuda	28	No Data									
ARG	Argentina	32	2011	census	CENSUS	✓	✓	✓	✓	✓	✓	[B]
ABW	Aruba	533	2010	census	CENSUS	✓	[est.]	✓	✓	✓	✓	[A]
BHS	Bahamas	44	2010	census	CENSUS	✓	[est.]	✓	✓	✓	✓	[B]
BRB	Barbados	52	No Data									
BLZ	Belize	84	2010	census	NSO	✓	✓	✓	✓	✓	✓	[B]
BOL	Bolivia	68	2012	census	CENSUS	✓	[est.]	✓	[est.]	✓	✓	[B]
BRA	Brazil	76	2010	census	CENSUS	✓	✓	✓	✓	✓	✓	[A]
CHL	Chile	152	2002	census	CENSUS	✓	✓	✓	✓	✓	✓	[A]
COL	Colombia	170	2005	census	CENSUS	✓	✓	✓	✓	✓	✓	[A]
CRI	Costa Rica	188	2011	census	CENSUS	✓	✓	✓	✓	✓	✓	[B]
CUB	Cuba	192	2002	census	CENSUS	✓	✓	✓	✓	✓	✓	[A]
CUW	Curaçao	531	2011	census	CENSUS	[est.]	[est.]	✓	✓	✓	✓	[C]
DOM	Dominican Republic	214	2010	census	CENSUS	✓	✓	✓	✓	✓	✓	[B]
ECU	Ecuador	218	2010	census	CENSUS	✓	✓	✓	✓	✓	✓	[B]
SLV	El Salvador	222	2007	census	CENSUS	✓	✓	✓	✓	✓	✓	[A]
GUF	French Guiana	254	2008	census	CENSUS	✓	[est.]	✓	✓	✓	✓	[A]
GRD	Grenada	308	No Data									
GLP	Guadeloupe	312	2008	census	CENSUS	✓	[est.]	✓	✓	✓	✓	[A]

ISO[17]	Country	CC[18]	Base year	Type	Source	No education	incompl. primary	Complete primary	Lower secondary	Upper secondary	Post-secondary	Status [19]
GTM	Guatemala	320	2002	census	CENSUS	✓	✓	✓	✓	✓	✓	[A]
GUY	Guyana	328	2002	census	CENSUS	✓	[est.]	✓	[est.]	✓	✓	[A]
HTI	Haiti	332	2003	census	CENSUS	✓	✓	✓	✓	✓	✓	[B]
HND	Honduras	340	2013	census	CENSUS	✓	✓	✓	✓	✓	✓	[B]
JAM	Jamaica	388	2001	census	CENSUS	✓	[est.]	✓	[est.]	✓	✓	[A]
MTQ	Martinique	474	2008	census	CENSUS	✓	[est.]	✓	✓	✓	✓	[A]
MEX	Mexico	484	2010	census	CENSUS	✓	✓	✓	✓	✓	✓	[A]
NIC	Nicaragua	558	2005	census	CENSUS	✓	✓	✓	✓	✓	✓	[A]
PAN	Panama	591	2010	census	CENSUS	✓	✓	✓	✓	✓	✓	[A]
PRY	Paraguay	600	2002	census	CENSUS	✓	✓	✓	✓	✓	✓	[A]
PER	Peru	604	2007	census	CENSUS	✓	✓	✓	✓	✓	✓	[A]
PRI	Puerto Rico	630	2010	census	CENSUS	✓	✓	✓	✓	✓	✓	[B]
LCA	Saint Lucia	662	2010	census	CENSUS	✓	[est.]	✓	✓	✓	✓	[B]
VCT	Saint Vincent and the Grenadines	670	2001	census	CENSUS	✓	✓	✓	[est.]	✓	✓	[A]
SUR	Suriname	740	2004	census	CENSUS	✓	[est.]	✓	✓	✓	✓	[A]
TTO	Trinidad and Tobago	780	2011	census	CENSUS	✓	[est.]	✓	[est.]	✓	✓	[B]
URY	Uruguay	858	2011	census	CENSUS	✓	✓	✓	✓	✓	✓	[B]
VEN	Venezuela	862	2011	census	CENSUS	✓	✓	✓	[est.]	[est.]	✓	[B]
VIR	Virgin Islands	850	No Data									
Northern America												
CAN	Canada	124	2011	census	CENSUS	►	[est.]	[est.]	✓	✓	✓	[B]
USA	United States	840	2010	census	CENSUS	✓	✓	✓	✓	✓	✓	[A]
Oceania												
AUS	Australia	36	2011	census	CENSUS	✓	►	✓	✓	✓	✓	[B]

ISO[17]	Country	CC[18]	Base year	Type	Source	No education	incompl. primary	Complete primary	Lower secondary	Upper secondary	Post-secondary	Status [19]
FJI	Fiji	242	2007	census	CENSUS	✓	✓	✓	✓	✓	✓	[C]
PYF	French Polynesia	258	2007	census	CENSUS	✓	[est.]	✓	✓	✓	✓	[A]
GUM	Guam	316	No Data									
KIR	Kiribati	296	2010	census	CENSUS	✓	[est.]	✓	✓	✓	✓	[C]
FSM	Micronesia	583	2010	census	CENSUS	✓	[est.]	✓	✓	✓	✓	[C]
NCL	New Caledonia	540	2009	census	CENSUS	✓	[est.]	✓	✓	✓	✓	[A]
NZL	New Zealand	554	2013	census	CENSUS	[est.]	►	[est.]	✓	✓	✓	[B]
PNG	Papua New Guinea	598	No Data									
WSM	Samoa	882	2001	census	CENSUS	✓	✓	✓	✓	✓	✓	[A]
SLB	Solomon Islands	90	2009	census	CENSUS	✓	✓	✓	✓	✓	✓	[C]
TON	Tonga	776	2006	census	CENSUS	✓	[est.]	✓	✓	✓	✓	[A]
VUT	Vanuatu	548	2009	census	CENSUS	✓	✓	✓	✓	✓	✓	[A]

Notes:

✓ Data for respective education category is available

[est.] Data for respective education category was estimated (see supplementary material)

► Data for this education category is not available in the source data and we abstained from the possibility to estimate the data. The population in this education category is part of the next higher education category

VIENNA INSTITUTE OF DEMOGRAPHY

Working Papers

Testa, Maria Rita and Danilo Bolano, *Intentions and Childbearing in a Cross-Domain Life Course Approach: The Case of Australia*, VID Working Paper 01/2019.

Goujon, Anne, Claudia Reiter and Michaela Potančoková, *Religious Affiliations in Austria at the Provincial Level: Estimates for Vorarlberg, 2001-2018*, VID Working Paper 13/2018.

Spitzer, Sonja, Angela Greulich and Bernhard Hammer, *The Subjective Cost of Young Children: A European Comparison*, VID Working Paper 12/2018.

Jatrana, Santosh, Ken Richardson and Samba Siva Rao Pasupuleti, *The Effect of Nativity, Duration of Residence, and Age at Arrival on Obesity: Evidence from an Australian Longitudinal Study*, VID Working Paper 11/2018.

Nitsche, Natalie and Sarah Hayford, *Preferences, Partners, and Parenthood: Linking Early Fertility Desires, Union Formation Timing, and Achieved Fertility*, VID Working Paper 10/2018.

Riederer, Bernhard, *Experts' Expectations of Future Vulnerability at the Peak of the "Refugee Crisis"*, VID Working Paper 9/2018.

Cukrowska-Torzewska, Ewa and Anna Matysiak, *The Motherhood Wage Penalty: A Meta-Analysis*, VID Working Paper 8/2018.

Nitsche, Natalie and Hannah Brückner, *High and Higher: Fertility of Black and White Women with College and Postgraduate Education in the United States*, VID Working Paper 7/2018.

Beaujouan, Éva, *Late Fertility Intentions and Fertility in Austria*, VID Working Paper 6/2018.

Brzozowska, Zuzanna, Isabella Buber-Ennser, Bernhard Riederer and Michaela Potančoková, *Didn't plan one but got one: unintended and sooner-than-intended births among men and women in six European countries*, VID Working Paper 5/2018.

Berghammer, Caroline and Bernhard Riederer, *The Part-Time Revolution: Changes in the Parenthood Effect on Women's Employment in Austria*, VID Working Paper 4/2018.

Bora, Jayanta Kumar, Rajesh Raushan and Wolfgang Lutz, *Contribution of Education to Infant and Under-Five Mortality Disparities among Caste Groups in India*, VID Working Paper 3/2018.
