DEMOGRAPHIC RESEARCH
A peer-reviewed, open-access journal of population sciences

DEMOGRAPHIC RESEARCH

VOLUME 49, ARTICLE 31, PAGES 809–848
PUBLISHED 10 NOVEMBER 2023
https://www.demographic-research.org/Volumes/Vol49/31/
DOI: 10.4054/DemRes.2023.49.31

Research Article

A Bayesian model for the reconstruction of education- and age-specific fertility rates: An application to African and Latin American countries

Afua Durowaa-Boateng

Dilek Yildiz

Anne Goujon

Contents

Demographic Research: Volume 49, Article 31
Research Article

A Bayesian model for the reconstruction of education- and age-specific fertility rates: An application to African and Latin American countries

Afua Durowaa-Boateng[1]

Dilek Yildiz[1,2]

Anne Goujon[2]

Abstract

BACKGROUND
Consistent and reliable time series of education- and age-specific fertility rates for the past are difficult to obtain in developing countries, although they are needed to evaluate the impact of women's education on fertility across periods and cohorts.

OBJECTIVE
We aim to fill the existing gap by reconstructing age-specific fertility rates by level of education for a large sample of African and Latin American countries from 1970 to 2020 in 5-year steps.

METHOD
We develop a Bayesian framework to reconstruct age-specific fertility rates by level of education using prior information from the birth history module of the Demographic and Health Surveys (DHS).

RESULTS
We find that the Bayesian approach allows for estimating reliable education- and age-specific fertility rates using multiple rounds of the DHS surveys. The time series obtained confirm the main findings of the literature on fertility trends and age- and education-specific differentials.

CONCLUSION
From a methodological point of view, we show that the Bayesian reconstruction model allows for estimating missing data on fertility by level of educational attainment. This

[1] Wittgenstein Centre for Demography and Global Human Capital (VID/ÖAW, IIASA, University of Vienna), Vienna Institute of Demography, Austrian Academy of Sciences.
[2] Wittgenstein Centre for Demography and Global Human Capital (VID/ÖAW, IIASA, University of Vienna), International Institute for Applied Systems Analysis.

information is key when we account for the role of education in fertility rates and assess the impacts of education policies in countries in Africa and Latin America.

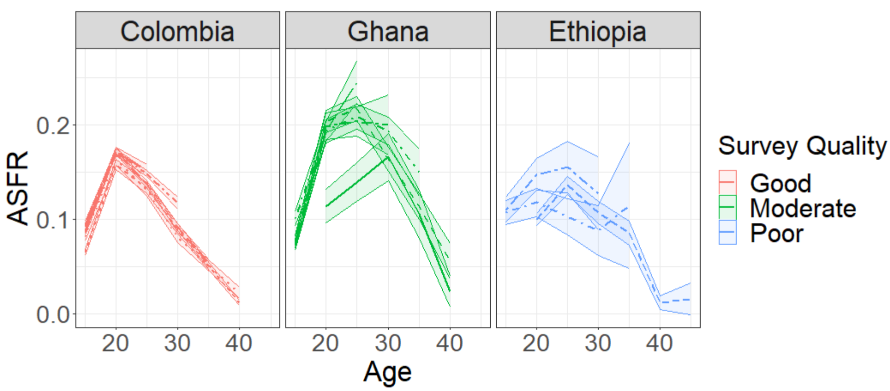
CONTRIBUTION
We propose an advanced statistical model which fills gaps in time series when data are missing, and provide complete and UN WPP-consistent age-specific fertility rates for 50 countries.

1. Introduction

The existence of a mostly negative relationship between women's education and fertility has been shown in many settings (e.g., Ahuja and Filmer 1995; Basu 2002; Bongaarts 2010; Brand and Davis 2011; Lutz and KC 2011; Martin and Juarez 1995; Weinberger, Lloyd, and Blanc 1989). The potential mechanisms at play are many, such as the fact that women with higher levels of education spend more years in school reduces their exposure to marriage and pregnancy at younger ages. They also tend to have better knowledge of and access to contraceptives to control their own fertility (e.g., Bongaarts 2010; Gebreselassie and Shapiro 2016). For example, in Latin America, Weinberger, Lloyd, and Blanc (1989) confirm that women with higher levels of education have fewer children and that 40%–67% (overall in Peru, Colombia, the Dominican Republic, and Ecuador) of the observed decline in fertility rates between the 1970s and 1980s was due to improvements in education. Bongaarts (2010) also explains the negative effects of education on fertility rates in sub-Saharan Africa by the greater demand for and use of modern contraceptives by women with higher levels of education. However, many unknowns remain about the precise pathways by which education affects fertility levels in different contexts and at different times. One of the main challenges in studying the relationship between education and fertility is the lack of comparable, consistent, and unbiased data on education- and age-specific fertility rates (EASFR), particularly in low-income countries.

Due to the limited availability of comprehensive and reliable registration systems, the main sources of data on fertility rates in high-fertility countries are various surveys (e.g., Demographic and Health Survey (DHS), Service Provision Assessment (SPA) survey, HIV/AIDS Indicator Survey (AIS), Malaria Indicator Survey (MIS)) and indirect estimates from the United Nations World Population Prospects (UN WPP 2022a) that use multiple data sources. The total fertility rates (TFR) and age-specific fertility rates (ASFR) published by the UN WPP on a regular basis are widely used. These estimates are consistent and allow for comparisons of fertility over time and across countries; as such they are considered trustworthy. However, these rates are not disaggregated by level of educational attainment. The Wittgenstein Center for Demography and Global Human Capital (WIC) publishes population projections and estimates by level of education based on data collected from multiple sources including the UN WPPs (WIC 2018), and assumptions about the future. However, the education- and age-specific fertility rates for past years are not estimated. Demographic surveys provide fertility rates for population sub-groups, including level of educational attainment. Among them, Demographic and Health Surveys (DHS) are the main and only source of fertility data in many low-income countries. However, like any survey, they can be subject to different levels of sampling error that often lead to inconsistencies, limiting comparison across countries and for the same country over time. Furthermore, for the most part they are not conducted at regular intervals, such as every five years. The most common inconsistencies have been attributed to birth omissions and displacements (e.g., Al Zalak and Goujon 2017; Pullum and Becker 2014; Schoumaker 2011, 2014). In an assessment of DHS data, Schoumaker (2014) categorizes most Latin American countries as having "good" or "moderate" data quality when available, while many African countries are categorized as having "poor" data quality, reporting different rates for the same periods across consecutive surveys (e.g., Ethiopia). Figure 1 presents the age-specific fertility rates of women with secondary education in the 1990–1995 period estimated by different rounds of the DHS surveys using the DHS birth history module. In countries with good data quality (e.g., Colombia), age-specific schedules are more complete and estimates are relatively consistent for the same years, even for different surveys (see Figure 1).

Figure 1: Quality of 1990–1995 data for secondary education ASFR by different DHS surveys[3]



[3] The 1990–1995 period and secondary education were selected due to the availability of multiple surveys and countries representing all three levels of data quality mentioned in Schoumaker (2014).

Regarding the lack of time series in demographic surveys, Schoumaker (2013) proposes a Poisson regression model to reconstruct both ASFR and TFR from different rounds of DHS surveys. The method estimates past fertility rates using the birth history module of the survey. While it fills the gaps, the estimates are not consistent with the UN WPP, which are widely used (and considered more reliable).

In this research we combine four datasets (UN WPP, WIC, estimates from Yildiz et al. 2023, and DHS) to produce reliable and consistent EASFRs over time and across countries that are compatible with the UN WPP ASFR, using a Bayesian model.

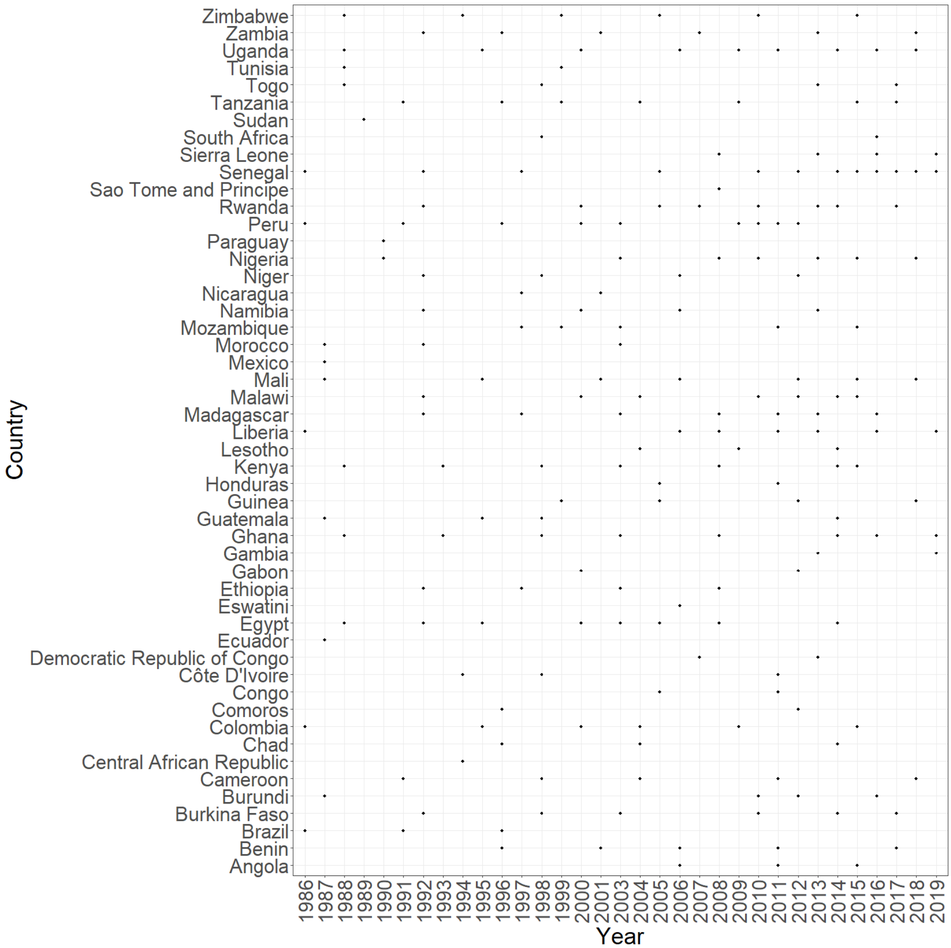
We follow Bijak and Bryant's (2016) recommendation to employ Bayesian methods in situations where sample sizes are small and the data quality is limited or poor. The Bayesian approach has been used in recent years in demographic estimates and population projections (e.g., Alkema et al. 2011; Bryant and Graham 2013; Ellison, Dodd, and Forster 2020; Hilton et al. 2019), including by the United Nations (UN) in recent rounds of population projections (Ševčíková, Raftery, and Gerland 2018). The modeling carried out in this work aims at developing a consistent and reliable dataset that will allow for a better understanding of the impact of education on fertility.

We estimate age-specific fertility rates for 41 DHS African and 9 Latin American countries by 4 levels of education between 1970–1975 and 2015–2020. Our paper contributes to the literature by proposing an advanced statistical model which fills the gap in the time series when data are missing, and by providing complete and UN WPP-consistent EASFRs for all 50 countries. We focus on Latin American and African countries in this analysis since to varying degrees they lack detailed, regular, and consistent data on EASFR for past years. Moreover, these two regions are interesting because the timing and pace of their demographic transitions are different. We plan to extend the research to more regions in the Global South in the future.

2. Data sources

We use four different data sources for this study: DHS, UN WPP (2022b), UN-consistent education-specific TFR from Yildiz et al. (2023), and WIC (2018). A total of 217 DHS conducted in Africa and Latin America are pooled for the analysis (see the full list of countries and surveys in Figure 2 and in Appendix A). Fertility rates are obtained from individual recode datasets in the DHS database (ICF 2004–2017). We consider 4 levels of education: No Education, Primary Education, Secondary Education, and Higher Education. These educational categories are collected in DHS surveys for each country and survey wave and are based on the highest level of education reported by the woman (The DHS program, n.d).

Figure 2: Countries and surveys collected, Africa and Latin America



Our analysis uses women's retrospective birth histories as collected in DHS rounds, focusing on reproductive ages 15 to 49 years. Since our goal is to reconstruct past fertility rates by level of education, for Latin American countries we focus on birth histories as far back as 30 years before the survey. For African countries the birth histories are collected for a shorter period of 15 years, given the lower quality of estimates due to long recall periods. The retrospective fertility rates by 5-year age groups and 5-year periods are obtained using the Stata 'tfr2' module, which provides fertility rates close to those published in the DHS reports (Schoumaker 2013). We utilize all available surveys with the exception of those conducted after 2020.

The second source of fertility data in our research is the 2022 edition of the UN WPP (2022b), which provides ASFRs for 5-year periods. The UN WPP collects fertility data, preferably on live births by age of mother, from civil registration systems, as well as all available data, including from the DHS. The UN WPP is updated every couple of years to achieve consistency over time and across different demographic statistics and includes new data sources which were previously omitted. This process makes the UN WPP a reliable source of global demographic data.

Further, Yildiz et al. (2023) estimate education-specific total fertility rates for sub-Saharan African countries from 1980 to 2015 by 5-year period using multiple data sources including WIC and UN WPP. The authors use a flexible hierarchical Bayesian model that allows education-specific estimates to vary with regard to degree of consistency with the UN data. The estimates are provided for sub-Saharan African countries only and focus on total fertility rates by education but not by age.

In addition to fertility rates, the size of the female population by level of education for the countries under consideration is obtained from the WIC data explorer (WIC 2018) through the 'epop' function in the 'wcde' R package (Abel 2021). These estimates are constructed using an iterative multi-dimensional cohort-component reconstruction model (IMCR) based on historical data on education and mortality (see Speringer et al. (2019) for more details). All data in our analysis concern women aged 15–49 years between 1970 and 2020.

3. Methodology

Various approaches within the Bayesian framework have been used to estimate fertility rates. For example, Ellison, Dodd, and Forster (2020), inspired by the Lexis diagram, developed a Bayesian technique to estimate cohort fertility where births were modeled to follow a Poisson distribution. Earlier, Bryant and Graham (2013) had modeled births as part of a sub-national population estimation model, in which births follow a Poisson distribution centered around expected births at the end of the period, similar to Ellison, Dodd, and Forster (2020). As mentioned above, Yildiz et al. (2023) used a flexible hierarchical Bayesian technique to estimate education-specific total fertility rates for sub-Saharan countries for the period 1980 to 2015 using existing sources. Finally, Alkema et al. (2011) estimated total fertility rates for 196 countries by considering their fertility transition phases. Each transition phase was modeled separately, taking into account the rate of decline. During the transition period, the TFR was modeled as the previous fertility rate minus the expected 5-year decrement, plus an error term. The 5-year decrement term follows a defined function while the error term follows a set of normal distributions when a set of conditions are met. For post-transition countries, the fertility rates were modeled as a normal distribution of a first-order autoregressive time series model.

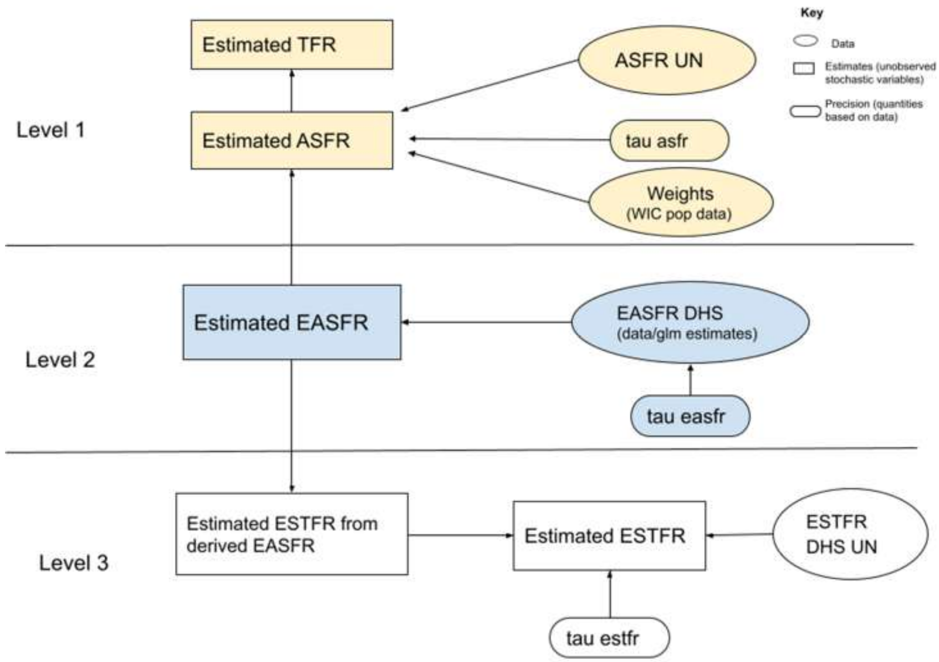
Our modeling framework consists of two main steps applied separately to African and Latin American countries. We improve and complement previous Bayesian approaches which did not integrate education and its impact on fertility rates, and did not estimate long time series of age-specific fertility rates. The first step in our approach is to estimate EASFRs for all DHS countries which enter the Bayesian model as initial values. To achieve this, we employ a generalized linear model (glm) with a Poisson link. The data for the glm model are the EASFRs obtained by the STATA tfr2 module using DHS birth histories. We adopt this approach to account for gaps in DHS estimates. We treat the glm estimates as the 'input' dataset for our Bayesian model.

The predicted estimates use information from the variables that influence fertility rates in our dataset, including the country itself, the number of women under consideration by 5-year period and age group, and their education level. In the event that a whole period schedule is missing (e.g., the 2015–2020 period for Latin American countries), the estimates use the aggregate effects of the period in question, including age group and country estimates (e.g., 2015–2020 period effect from all countries, educational effect, age group effects, country effects, and the specific interaction effects). To a large extent, the model learns from the other countries in the region since it uses all available information in the dataset to make estimates for the missing values. The glm model involves interaction terms between variables based on the assumption that the effects of these variables are not constant. The regression model for EASFR is defined as:

$$EASFR^{DHS} \sim Education + Age\ Group + Country + \text{Period} + Age\ Group * Country \\ + Country * \text{Period} + Education * Age\ Group + \text{Period} * Age\ Group. \qquad (1)$$

Using the glm model, we estimate initial values of the EASFR between 1970–1975 and 2015–2020 in 5-year intervals. The estimates produced by the model logically do not match with the UN WPP ASFRs. Also, in the event of missing or poor-quality data, the model estimates higher values than expected by the trend; for example, creating an abrupt halt or reversal of the fertility decline during the missing period. To address these issues and ensure consistency with the UN WPP ASFR, we employ a Bayesian framework in the second step, as explained in the next paragraphs. We show a graphical representation of our model in Figure 3 and then provide mathematical notations. In Figure 3, all squares are estimated values, while ovals serve as input data and half-rounded rectangles are precisions. We present level 1 in light yellow, level 2 in light blue, and level 3 in white.

Figure 3: Graphical representation of the Bayesian model



The first level (Level 1) starts with benchmarking our estimated ASFRs against UN WPP ASFRs. In Equation 2, the estimated $ASFR^{estimates}$ are adjusted to the UN WPP ASFRs for each country for all 5-year periods making use of the precision parameter τ^{asfr}, the inverse of the variance. The variance parameter σ^{asfr^2} is fixed at the variance of the UN WPP ASFRs estimates. In other words, it is used to define the degree of consistency with the UN WPP ASFRs. An exercise to investigate the sensitivity of the estimates to this parameter is presented in Appendix B. Since the UN WPP ASFRs are themselves dependent on DHS data, to avoid repeating the same information in our Bayesian model we only include ASFRs from UN WPP. We then calculate $ASFR^{estimates}$ as a weighted total of the predicted/estimated $EASFR^{estimates}$ in Equation 3. The weights, w $_{cyae}$, are the ratio of the population of women in age group *a* by level of education *e* to the total population of women aged 15–49 years for each 5-year period *y* in country *c* according to WIC.

Level 1:

$$ASFR_{cya}^{UN} \sim N_{+}(ASFR_{cya}^{estimates}, \tau^{asfr}) \tag{2}$$

$$ASFR_{cya}^{estimates} = \sum_{e=1}^{4} (EASFR_{cyae}^{estimates} \times w_{cyae}) \tag{3}$$

$$\tau^{asfr} = 1/\sigma^{asfr^2} \tag{4}$$

In the second level (Level 2, Equation 5), the 'true EASFRs,' unobserved and reconstructed $EASFR_{cyae}^{estimates}$, are sampled from a half-normal distribution centered at the $EASFR_{cyae}^{DHS}$ estimates from the regression model in Step 1 (Equation 1). The standard deviation parameter $\sigma_{e=No\ education}^{easfr}$ is the standard deviation of the standard errors from the glm estimates and to a large extent captures the variations from the 'true EASFRs.' We capture standard errors by education level. We allow τ_{cyae}^{easfr} to follow a gamma distribution and produce estimates for each country, year, and education level. We use the same 'initial values,' obtained from the 'no education' category, $\sigma_{e=No\ education}^{easfr}$, for all education levels, because this category is the education level with the highest standard deviation estimate.

Level 2:

$$EASFR_{cyae}^{estimates} \sim N_{+}(EASFR_{cyae}^{DHS}, \tau_{cyae}^{easfr}) \tag{5}$$

$$\tau_{cyae}^{easfr} \sim G(1/\sigma_{e=No\ education}^{easfr}, 2\sigma_{e=No\ education}^{easfr^2}) \tag{6}$$

The third level (Level 3, Equation 7) estimates $ESTFR_{cye}^{estimates}$ (education-specific TFR) by sampling from a half-normal distribution centered on the $ESTFR_{cye}^{derived}$ estimates calculated in Equation 8 by summing the EASFRs estimated in Equation (5) over age groups and multiplying them by 5. The ESTFRs estimates provide a perspective on the evolution of the fertility differentials by level of education at the aggregate level. The parameter σ^{estfr} follows a half-normal distribution that centers around parameter η, which is the minimum ESTFR in the region from Equation (1), and has a standard deviation h, which is fixed at 50 to allow variations in our estimates.

Level 3:

$$ESTFR_{cye}^{estimates} \sim N_{+}(ESTFR_{cye}^{derived}, \tau^{estfr}) \tag{7}$$

$$ESTFR_{cye}^{derived} = (\sum_{a=15-19}^{45-49} EASFR_{cyae}^{estimates}) \times 5 \tag{8}$$

$$\tau^{estfr} = 1/\sigma^{estfr^2} \tag{9}$$

$$\sigma^{estfr} \sim N(\eta, h) \tag{10}$$

For African countries, $ESTFR_{cye}^{derived}$ are benchmarked to "UN-fully consistent" ESTFR estimates by Yildiz et. al (2023), which are almost identical to the UN TFRs, and thus level 3 is specified as:

$$ESTFR_{cye}^{un-estfr} \sim N_{+(0,10)}(ESTFR_{cye}^{derived}, \mu_{cye}^{estfr} \tag{11}$$

$$ESTFR\,_{cye}^{estimates} \sim N_{+(0,10)}(ESTFR\,_{cye}^{derived}, \mu_{cye}^{estfr}) \tag{12}$$

$$\mu_{cye}^{estfr} \sim G(\sigma^{estfr}, \sigma^{estfr^2}) \tag{13}$$

The parameter μ follows a gamma distribution of the standard deviation, σ, of the ESTFR UN-consistent estimates for 'Higher Education.'

The full specification of the prior distributions and sensitivity analysis are presented in Appendix B. Equations 7 to 10 pertain to Latin American countries and Level 3 does not benchmark against estimates by Yildiz et al. (2023), unlike Equations 11 to 13 for African countries. Since the estimates of EASFRs are new, the model was validated using a cross-validation approach by leaving out, at random, 5%, 10%, and 15% of the data, and testing the model estimates on the omitted data. We compared the different estimates from the cross-validation exercise and found that after leaving out varying degrees of data, the estimated values from the cross-validation fell within acceptable credible intervals. We provide details of this exercise in the supplementary material (a brief description of the supplementary material is provided in Appendix D and are available at www.populationafrica.org).

4. Results

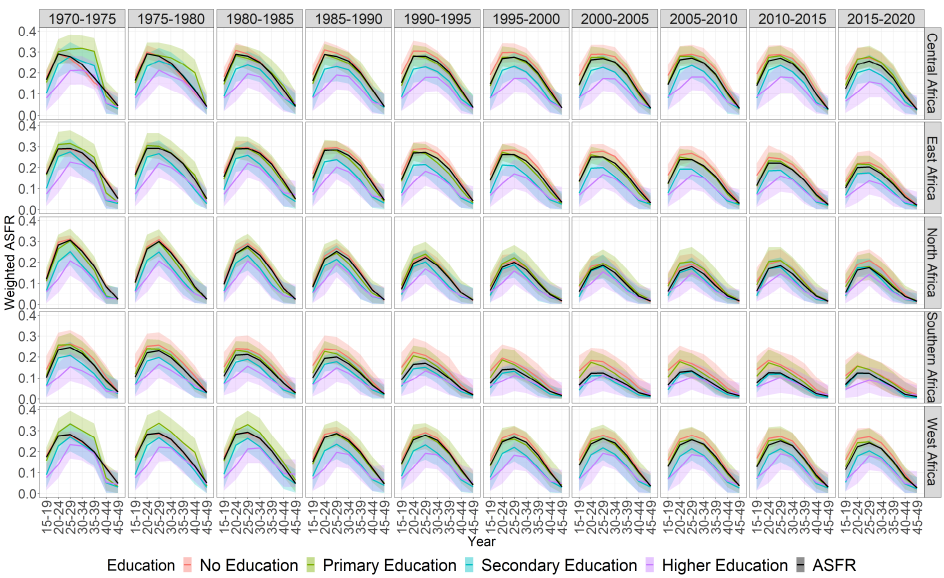
In this section we present estimates of the EASFRs for 1970–1975 to 2015–2020 for all countries under consideration, separately for Africa and Latin America, with a particular emphasis on the starting and end periods. Detailed estimates for all 5-year periods are presented in the supplementary material.

Weighted EASFRs, $EASFR_{ryae}$, are calculated for each region *r* in Africa, education level *e*, age group *a*, and period *y* as:

$$EASFR_{ryae} = \frac{\sum_{c\ in\ r}\ (EASFR_{cyae} \times Population_{cyae})}{\sum_{c\ in\ r}\ Population_{cyae}}$$

Figure 4 shows the weighted EASFRs for the African countries under consideration, grouped by UN regions. Although the lines between the subsequent levels of education sometimes cross, particularly at older ages – meaning that a higher level of education does not necessarily mean fewer children – each level of education (as categorized in this work) generally leads to lower fertility. In all regions and for all estimated years, women with secondary and/or higher education have the lowest weighted EASFR in their respective age groups, compared to women with primary education or less. Another general observation is that with increased levels of education the peak of the fertility curve occurs at older ages: this is particularly true for women with secondary or higher education in comparison with other education groups, throughout the reconstructed period. The low education levels of women in Africa translate into the average ASFR curve being very similar in terms of level and pattern to that of the ASFR of women without education in the earlier period, and with primary education in the later period. There are notable differences between regions, particularly between North and Southern Africa and the other African regions. In the former regions, the ASFR has already converged with the fertility pattern of secondary-educated women in the 1990s, indicating greater progress in education in these regions.

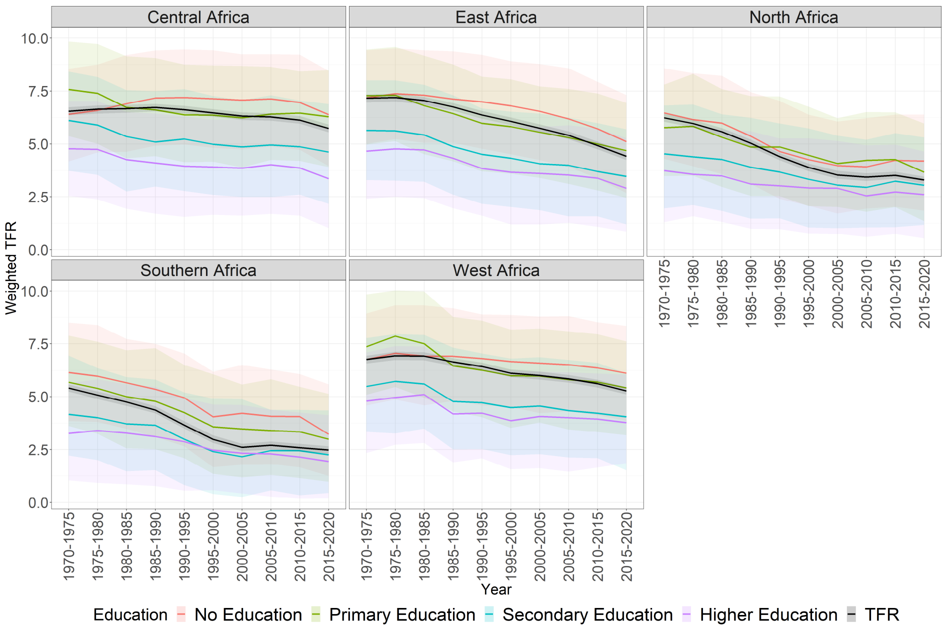
Figure 4: Weighted EASFRs and 95% CI by region, Africa, 1970–1975 to 2015–2020



Next, in Figure 5 we present weighted ESTFRs, $ESTFR_{ryae}$, estimated for each region in Africa *r*, education level *e*, age group *a*, and period *y* as:

$$ESTFR_{ryae} = \frac{\sum_{c\ in\ r} \left(ESTFR_{cyae} \times Population_{cyae}\right)}{\sum_{c\ in\ r} Population_{cyae}}$$

Figure 5: Weighted ESTFRs and 95% CI by regions, Africa, 1970–1975 to 2015–2020



The ESTFRs in Central Africa and West Africa have been declining very slowly, with some periods of stagnation and even increase; e.g., for women with no education until 2010–2015 in Central Africa. The stalls in fertility decline have occurred among women in all education categories; e.g., for women with primary or higher education in the 1985–1995 period in West Africa. In these two regions, the difference in fertility between women with the highest level of education and the lowest level of education did not decline over time and was between 3.6 and 4.9 children per woman between 1970 and 2020. In East Africa, in most education categories the fertility decline is steadier than in the two other regions, especially concerning women with secondary and higher education. In this region since 2010–2015 there appears to have been an acceleration of the decreasing trend in the previous period, particularly visible for women without education. This trend is also visible in West Africa.

Figure 6 and Figure 7 show the EASFR estimates in African countries in the 1970–1975 and 2015–2020 periods. Most of the observed patterns in the regions are visible across countries, with the overall ASFRs closely following that of lower education groups.

Figure 6: EASFR and ASFR, Africa, 1970–1975

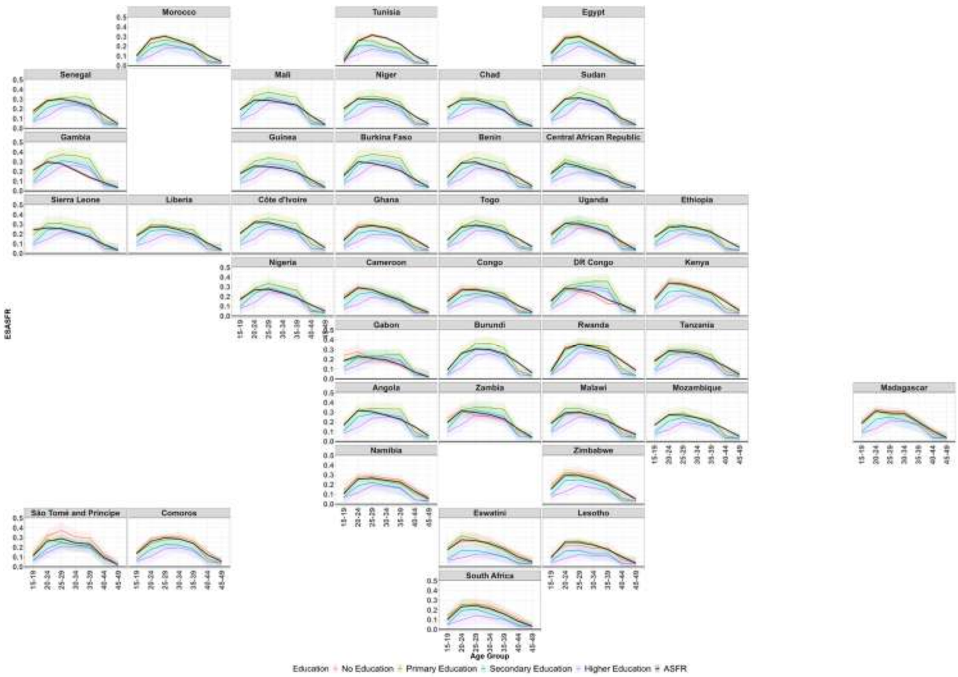
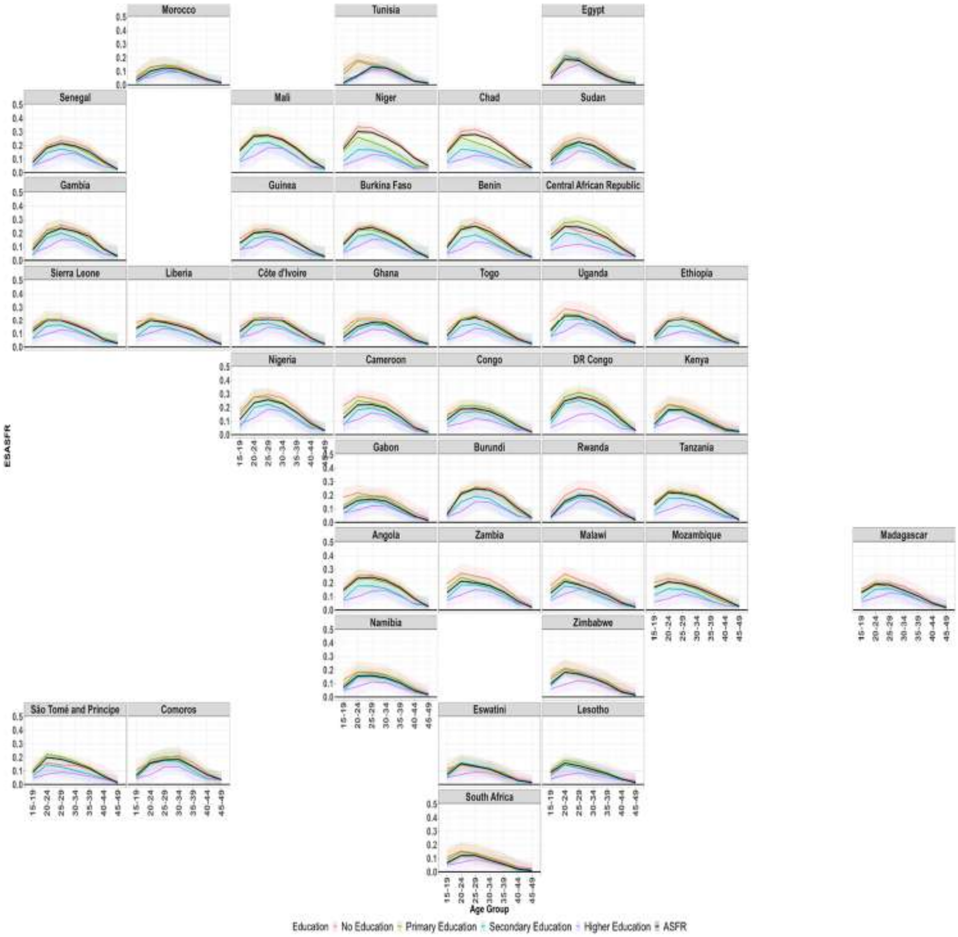


Figure 7: EASFR and ASFR, Africa, 2015–2020



In many African countries, overall ASFR and EASFR began to decline significantly from the 1980–1985 period (see supplementary material). On average, across all 5-year periods between 1970 and 2020, fertility among women with higher education peaked between the ages of 25 and 34 at around 0.2 to 0.3 children per woman. However, for women with secondary education, fertility was highest overall between the ages of 20 to 29 in all countries, at around 0.2 to 0.4 children per woman. For women with primary or no education, fertility peaked between the ages of 20–25, in general at around 0.3 to 0.4 children per woman.

Figure 8 shows the difference between 2015–2020 and 1970–1975 fertility rates for African countries. For many countries in the region the difference is negative, implying a decline in fertility rates. The biggest changes are among women with primary education, with a few exceptions. Among all levels of education other than no education, fertility rates have fallen for all ages (except ages 45–49) by around 0.1 children per woman, with the exception of Cameroon and the Central African Republic.

Figure 8: Difference between 2015–2020 and 1970–1975 for EASFR and ASFR, Africa

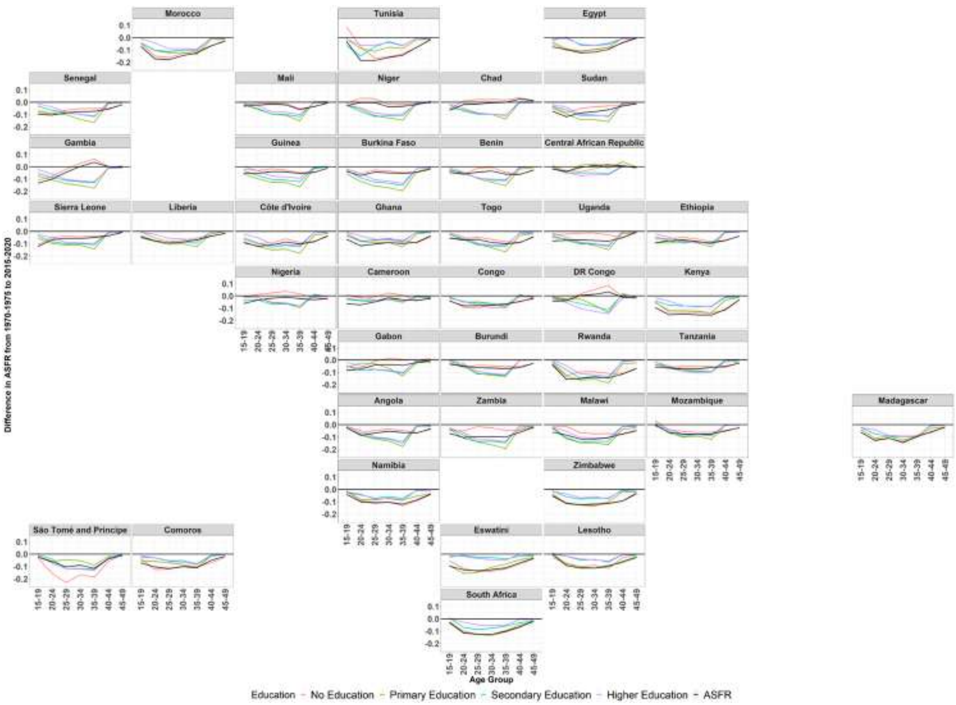
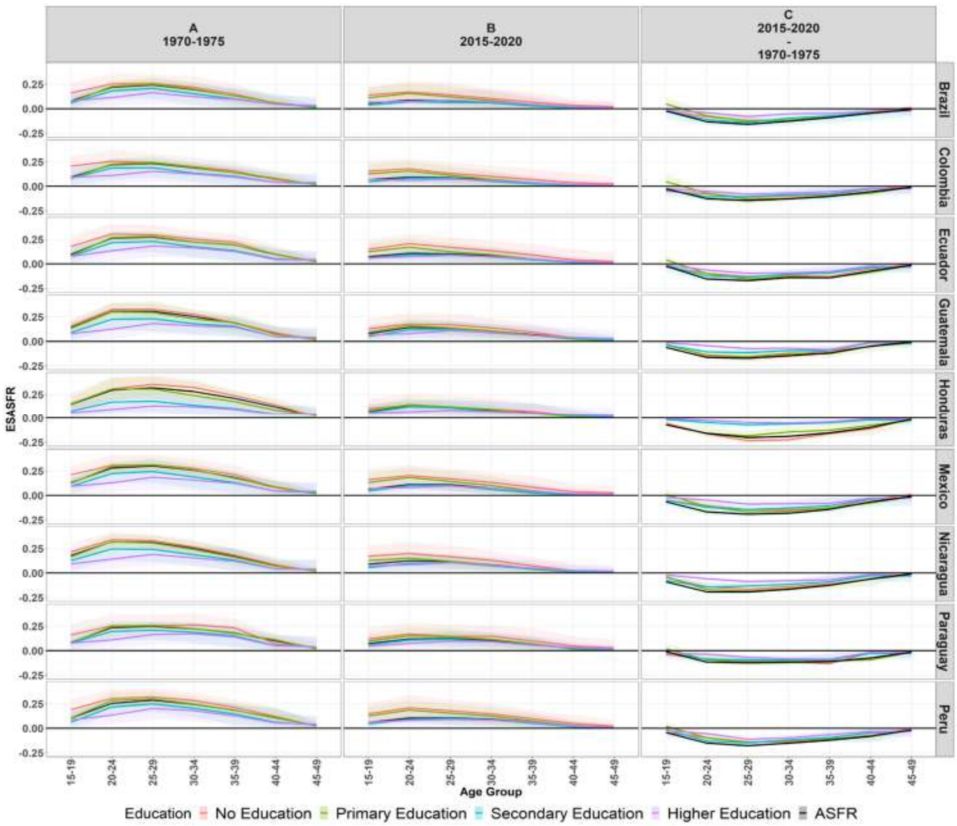


Figure 9 represents the 1970–1975 and 2015–2020 EASFR and the difference between the two periods for Latin America. The fertility transition and the education expansion occurred earlier in Latin America than in Africa (Bongaarts and Casterline 2013; Wils and Goujon 1998). As a result, compared with Africa the differentials in EASFR are larger in 1970–1975 and smaller in 2015–2020 (see Figure 9, panels A and

B, and Figure 8). Also, declines in overall ASFR and EASFR in Latin America are visible from the 1975–1980 period (see supplementary material). The overall ASFR appears to be largely influenced by the majority of women with primary education in Latin America from the 1970–1975 period until the 1990–1995 period, when the overall ASFR became more influenced by the large share of women with secondary and higher education. This is also reflected in Figure 9 for 1970–1975 and 2015–2020.

Figure 9: EASFR and ASFR, 1970–1975 (Panel A) and 2015–2020 (Panel B), and difference between 2015–2020 and 1970–1975 for EASFR and ASFR (Panel C), Latin American countries



In all the years under consideration, the fertility peak for women with higher education in Latin America occurred at 25 to 29 years old, at a level of around 0.2 children per woman in the 1975–1980 period (see supplementary material), declining to 0.07 children per woman in the 2015–2020 period. For women with secondary education, fertility peaked between the ages of 20 and 29, falling from around 0.25 children per woman in the 1975–1980 period to 0.06 children per woman in the 2010–2015 period. Women with primary education experienced a peak in fertility of around 0.32 children per woman in 1970–1975, declining to 0.14 children per woman in 2000–2005 at the ages of 20–24. Similarly, for women with no education at ages 20–24 we observed peak fertility rates of about 0.35 children in the 1990–1995 period, falling to 0.14 children in the 2015–2020 period.

Fertility rates in Latin America were significantly lower in the 2015–2020 period (Figure 9 panel B) than in the 1970–1975 period. Panel C shows differences in EASFR between 2015–2020 and 1970–1975. All age groups across all levels of education and countries saw their fertility rates decline. In Brazil, Colombia, Mexico, Ecuador, and Peru, women with secondary education between ages 15–19 and 20–24 saw a large decline of around 0.1 children per woman. In Peru and Brazil this decline reached almost 0.15 children per woman with a secondary education and was visible up until ages 25–29.

In all the studied countries, the smallest difference in fertility rates is observed among women with higher education. In Honduras, women with no education experienced a sharp drop in fertility rates, while in the other countries it was women with primary education who appeared to experience the largest decrease in fertility rates of all education groups.

5. Discussion and conclusion

Our analysis focuses on deriving and applying a methodology to estimate past total and age-specific fertility rates by level of education, for African and Latin American countries. Our aim is to fill a gap by providing a dataset for countries where historical good quality and consistent fertility data are scarce. We propose a Bayesian framework to reconstruct ASFRs by level of education and provide estimates for 41 African and 9 Latin American countries for 5-year periods from 1970 to 2020 using multiple data sources, combining more and less reliable datasets. Our estimates are UN WPP-consistent and provide complete age schedules by education level for past years which were previously not available for Africa and Latin America. However, a limitation is that it does not conscientiously model data quality by including the reliability of birth histories that happened a long time before the survey year. Another limitation is the inability to accurately validate the EASFRs, since these are the first estimates of their kind.

Nevertheless, we think these estimates of past fertility rates by age and level of education are essential for studying in detail the connection to the educational expansion and the role of education in the fertility transition. Furthermore, our estimates can be used to inform population projections. The analysis of the dataset resulting from the modeling confirms what has been demonstrated using existing data, as we will discuss below. These estimates support the general finding that women with higher education have lower fertility rates (e.g., Basu 2002; Bongaarts 2010; Weinberger, Lloyd, and Blanc 1989). The estimates across countries show that women with higher education have a relatively late onset of fertility as well as generally lower fertility rates. The model used can be expanded to other world regions within the DHS database to provide reliable estimates for other studies relating to fertility, population, and education.

Many countries in Africa (especially in sub-Saharan Africa) began their fertility transition in the 1980s (Bongaarts and Casterline 2013). Our estimate of education-specific rates in Africa before 1985 agrees with Cochrane (1979) and Martin (1995) in that in the least developed countries, where overall education levels were low, the achievement of some education (incomplete or completed primary) was sometimes reversely related to fertility. In countries like the Democratic Republic of the Congo, Gabon, and Nigeria, fertility rates for women with lower levels of education increased between the 1970–1975 and 2015–2020 periods. The analysis of the stalls in fertility decline show that they could in part be linked to the stalls in education progress (Goujon, Lutz, and KC 2015; Kebede, Goujon, and Lutz 2019). These countries have also experienced conflicts in the past, which can be followed by an increase in fertility in the early post-conflict period, as shown in the literature (e.g., Lindstrom and Berhanu 1999; Agadjanian and Prata 2002; Randall 2005).

Since Latin America was already undergoing the fertility transition at the beginning of our analysis, educational differentials in fertility rates were more visible. Our estimates in Latin America for periods before 2000–2005 are in line with the findings of Martin and Juarez (1995) concerning the relatively wide educational fertility differences. However, the differences appear to narrow rather drastically by the end of the period covered by our estimation exercise. The literature points out that the reduction in educational differentials and overall fertility rates could be attributed to government policies on public health and education, of women in particular (Rios-Neto and Guimarães 2014).

We also note a more pronounced decline in adolescent fertility rates by level of education when considering the differences between 1970–1975 and 2015–2020 in Latin American countries compared to African countries. The decrease in fertility differentials by educational level in many countries in Latin America and some countries in Africa supports the argument of Kravdal (2002), that as more women become educated and reduce their fertility in the community, women with lower levels of education tend to follow this behavior.

It is probable that education expansion has strongly contributed to the decrease in education-specific fertility rates between 1970–1975 and 2015–2020 in the countries under consideration. Santelli et al. (2017) report that Latin America and the Caribbean and sub-Saharan Africa regions spent about 4.4% and 4.6% of their GDP on education in 2012, respectively. This is 42% more than in 1990. As a result, on average the mean years of schooling for Latin American women aged 15 to 49 increased from 4.5 years in 1970 to 9.3 years in 2015. In Africa the increase started from much lower levels and was weaker, from 1.3 to 5.9 years (WIC 2018).

6. Acknowledgements

This paper is part of the BayesEdu Project (IF_2019_29_BayesEdu) at the Wittgenstein Centre for Demography and Global Human Capital (IIASA, VID/OeAW, University of Vienna), funded by the Innovation Fund Research, Science and Society of the Austrian Academy of Sciences (ÖAW). Most of the work was conducted at Vienna Institute of Demography of Austrian Academy of Sciences.

References

Abel, G.J. (2021). WCDE: Download data from the Wittgenstein human capital data explorer [R package version 0.0.2]. https://CRAN.R-project.org/package=wcde.

Agadjanian, V. and Prata, N. (2002). War, peace, and fertility in Angola. *Demography* 39(2): 215–231. doi:10.1353/dem.2002.0013.

Ahuja, V. and Filmer, D. (1996). Educational attainment in developing countries: New estimates and projections disaggregated by gender. *Journal of Educational Planning and Administration* 10(3): 229–254.

Al Zalak, Z. and Goujon, A. (2017). Assessment of the data quality in Demographic and Health Surveys in Egypt. (Vienna Institute of Demography Working Papers 06/2017).

Alkema, L., Raftery, A.E., Gerland, P., Clark, S.J., Pelletier, F., Buettner, T., and Heilig, G.K. (2011). Probabilistic projections of the total fertility rate for all countries. *Demography* 48(3): 815–839. doi:10.1007/s13524-011-0040-5.

Basu, A.M. (2002). Why does education lead to lower fertility? A critical review of some of the possibilities. *World Development* 30(10): 1779–1790. doi:10.1016/S0305-750X(02)00072-4.

Bijak, J. and Bryant, J. (2016). Bayesian demography 250 years after Bayes. *Population Studies* 70(1): 1–19. doi:10.1080/00324728.2015.1122826.

Bongaarts, J. (2010). The causes of educational differences in fertility in sub-Saharan Africa. *Vienna Yearbook of Population Research* 8: 31–50. doi:10.31899/pgy3.1023.

Bongaarts, J. and Casterline, J. (2013). Fertility transition: Is sub-Saharan Africa different? *Population and Development Review* 38(Suppl 1): 153–168. doi:10.1111/j.1728-4457.2013.00557.x.

Brand, J.E. and Davis, D. (2011). The impact of college education on fertility: Evidence for heterogeneous effects. *Demography* 48(3): 863–887. doi:10.1007/s13524-011-0034-3.

Bryant, J.R. and Graham, P.J. (2013). Bayesian demographic accounts: Subnational population estimation using multiple data sources. *Bayesian Analysis* 8(3): 591–622. doi:10.1214/13-BA820.

Cochrane, S.H. (1979). Fertility and education: What do we really know? (World Bank Staff Occasional Papers 26). New York: World Bank.

Ellison, J., Dodd, E., and Forster, J.J. (2020). Forecasting of cohort fertility under a hierarchical Bayesian approach. *Journal of the Royal Statistical Society Series A: Statistics in Society* 183(3): 829–856. doi:10.1111/rssa.12566.

Gebreselassie, T. and Shapiro, D. (2016). *Education and fertility in sub-Saharan Africa: What do we really know?* Paper presented at PAA Annual Meeting, Washington, DC, March 31–April 2, 2016.

Goujon, A., Lutz, W., and KC, S. (2015). Education stalls and subsequent stalls in African fertility: A descriptive overview. *Demographic Research* 33(47): 1281–1296. doi:10.4054/DemRes.2015.33.47.

Hilton, J., Dodd, E., Forster, J.J., Smith, P.W., and Bijak, J. (2019). Forecasting fertility with parametric mixture models. arXiv preprint arXiv:1909.09545.

ICF (2004–2017). Demographic and Health Surveys (various) [Datasets]. Funded by USAID. Rockville, Maryland: ICF.

Kebede, E., Goujon, A., and Lutz, W. (2019). Stalls in Africa's fertility decline partly result from disruptions in female education. *Proceedings of the National Academy of Sciences* 116(8): 2891–2896. doi:10.1073/pnas.1717288116.

Kravdal, Ø. (2002). Education and fertility in sub-Saharan Africa: Individual and community effects. *Demography* 39(2): 233–250. doi:10.1353/dem.2002.0017.

Lindstrom, D.P. and Berhanu, B. (1999). The impact of war, famine, and economic decline on marital fertility in Ethiopia. *Demography* 36(2): 247–261. doi:10.2307/2648112.

Lutz, W. and KC, S. (2011). Global human capital: Integrating education and population. *Science* 333(6042): 587–592. doi:10.1126/science.1206964.

Martin, T.C. (1995). Women's education and fertility: Results from 26 Demographic and Health Surveys. *Studies in Family Planning* 26(4): 187–202. doi:10.2307/2137845.

Martin, T.C. and Juarez, F. (1995). The impact of women's education on fertility in Latin America: Searching for explanations. *International Family Planning Perspectives* 21(2): 52–80. doi:10.2307/2133523.

Pullum, T. and Becker, S. (2014). Evidence of omission and displacement in DHS birth histories. (Methodological reports vol. 11). Rockville: ICF International.

Randall, S. (2005). The demographic consequences of conflict, exile and repatriation: A case study of Malian Tuareg. *European Journal of Population/Revue Européenne de Démographie* 21(2/3): 291–320. doi:10.1007/s10680-005-6857-0.

Rios-Neto, E.L. and Guimarães, R.D.M. (2014). *The educational gradient of low fertility in Latin America*. Paper presented at PAA Annual Meeting, Boston, MA, May 1–3, 2014.

Santelli, J.S., Song, X., Garbers, S., Sharma, V., and Viner, R.M. (2017). Global trends in adolescent fertility, 1990–2012, in relation to national wealth, income inequalities, and educational expenditures. *Journal of Adolescent Health* 60(2): 161–168. doi:10.1016/j.jadohealth.2016.08.026.

Schoumaker, B. (2011). *Omissions of births in DHS birth histories in sub-Saharan Africa: Measurement and determinants*. Paper presented at PAA Annual Meeting, Washington, DC, March 31–April 2, 2011.

Schoumaker, B. (2013). A Stata module for computing fertility rates and TFRs from birth histories: tfr2. *Demographic Research* 28(38): 1093–1144. doi:10.4054/DemRes.2013.28.38.

Schoumaker, B. (2014). Quality and consistency of DHS fertility estimates, 1990 to 2012. Rockville, MD: ICF International.

Ševčíková, H., Raftery, A.E., and Gerland, P. (2018). Probabilistic projection of subnational total fertility rates. *Demographic Research* 38(60): 1843–1884. doi:10.4054/DemRes.2018.38.60.

Speringer, M., Goujon, A., KC, S., Potančoková, M., Reiter, C., Jurasszovich, S., and Eder, J. (2019). Global reconstruction of educational attainment, 1950 to 2015: Methodology and assessment. (Vienna Institute of Demography Working Papers 02/2019). Vienna: VID.

The DHS Program – Available Datasets. (n.d.). https://dhsprogram.com/data/available-datasets.cfm.

United Nations, Department of Economic and Social Affairs, Population Division (2022a). World Population Prospects 2022: Data Sources. (UN DESA/POP/2022/DC/NO. 9).

United Nations, Department of Economic and Social Affairs, Population Division. (2022b). wpp2022: World population prospects 2022 (R package version 1.1-2) [Computer software]. The Comprehensive R Archive Network. Available from https://population.un.org/wpp.

Weinberger, M.B., Lloyd, C., and Blanc, A.K. (1989). Women's education and fertility: A decade of change in four Latin American countries. *International Family Planning Perspectives* 15(1): 4–28. doi:10.2307/2133273.

WIC – Wittgenstein Centre for Demography and Global Human Capital (2018). Wittgenstein Centre Data Explorer Version 2.0. Available at: https://www.wittgensteincentre.org/dataexplorer.

Wils, A. and Goujon, A. (1998). Diffusion of education in six world regions, 1960–90. *Population and Development Review* 24(2): 357–368. doi:10.2307/2807979.

Yildiz, D., Wisniowski, A., Brzozowska, Z., and Durowaa-Boateng, A. (2023). A flexible model to reconstruct education-specific fertility rates: Sub-Saharan Africa case study. (Vienna Institute of Demography Working Papers 02/2023). Vienna: VID. doi:10.1553/0x003e65e0.

Appendix A: List of countries and surveys

Table A-1: List of countries and their survey years

Country	Survey year											
Angola	2006	2011	2015									
Benin	1996	2001	2006	2011	2017							
Brazil	1986	1991	1996									
Burkina Faso	1992	1998	2003	2010	2014	2017						
Burundi	1987	2010	2012	2016								
Cameroon	1991	1998	2004	2011	2018							
Central African Republic	1994											
Chad	1996	2004	2014									
Colombia	1986	1995	2000	2004	2009	2015						
Comoros	1996	2012										
Congo	2005	2011										
Côte D'Ivoire	1994	1998	2011									
DR Congo	2007	2013										
Ecuador	1987											
Egypt	1988	1992	1995	2000	2003	2005	2008	2014				
Eswatini	2006											
Ethiopia	1992	1997	2003	2008								
Gabon	2000	2012										
Gambia	2013	2019										
Ghana	1988	1993	1998	2003	2008	2014	2016					
Guatemala	1987	1995	1998	2014								
Guinea	1999	2005	2012	2018								
Honduras	2005	2011										
Kenya	1988	1993	1998	2003	2008	2014	2015					
Lesotho	2004	2009	2014									
Liberia	1986	2006	2008	2011	2013	2016	2019					
Madagascar	1992	1997	2003	2008	2011	2013	2016					
Malawi	1992	2000	2004	2010	2012	2014	2015					
Mali	1987	1995	2001	2006	2012	2015	2018					
Mexico	1987											
Morocco	1987	1992	2003									
Mozambique	1997	1999	2003	2011	2015							
Namibia	1992	2000	2006	2013								
Nicaragua	1997	2001										
Niger	1992	1998	2006	2012								
Nigeria	1990	2003	2008	2010	2013	2015	2018					
Paraguay	1990											
Peru	1986	1991	1996	2000	2003	2009	2010	2011	2012			
Rwanda	1992	2000	2005	2007	2010	2013	2014	2017				
Sao Tome and Principe	2008											
Senegal	1986	1992	1997	2005	2010	2012	2014	2015	2016	2017	2018	2019
Sierra Leone	2008	2013	2016	2019								
South Africa	1998	2016										
Sudan	1989											
Tanzania	1991	1996	1999	2004	2009	2015	2017					
Togo	1988	1998	2013	2017								
Tunisia	1988	1999										
Uganda	1988	1995	2000	2006	2009	2011	2014	2016				
Zambia	1992	1996	2001	2007	2013	2018						
Zimbabwe	1988	1994	1999	2005	2010	2015						

Appendix B: Prior distributions and sensitivity analysis

We explored different characterizations of the parameter ν and how it changed the estimation of EASFRs for both the Latin America and Africa models. We keep the model described in Equations 2 to 4 and 8 to 10 and change the values in Equations 5 and 6.

We identify our model described in the text as the 'Main Model' for both the Latin American and African cases. In Model 1 we define the parameter ν as the standard deviation of the standard error of the glm estimates by country *c*, year *y*, and age group *a* and

$$\tau_{cya}^{easfr} \sim G(1/\sigma_{cya}^{easfr^2}, \sigma_{cya}^{easfr^2})$$

In Model 2 we express σ as the standard deviation of the standard error of the glm estimates by education level only. Then,

$$\tau_{e}^{easfr} \sim G(1/\sigma_{e}^{easfr^2}, \sigma_{e}^{easfr^2})$$

In Model 3, σ is the standard deviation of the standard error of the glm estimates by level of education only. However,

$$\tau_{e}^{easfr} \sim G(1/\alpha_{e}^{easfr}, \sigma_{e}^{easfr^2})$$

where α_{e}^{easfr} is the mean of the standard errors of the glm estimates by only level of education.

In Model 4, we define σ as the standard deviation of the standard errors of the glm estimates. The following are specified for Model 4:

$$\tau^{easfr} \sim G(1/\sigma^{easfr^2}, \sigma^{easfr^2})$$

We define the following in Model 5:

$$\tau \sim G(1/\sigma^{easfr^2}, \sigma^{easfr^2})$$

where σ is now the standard deviation of the estimates from the glm model in Step 1.

We describe Model 6 as:

$$\tau^{easfr} \sim G(\beta_1, \beta_2)$$

and β_1=$(\sigma/\alpha)^2$, where ν is the standard deviation of the standard errors from the glm estimates and α is the mean of the standard errors of the glm estimates. Similarly, β_2=$(\sigma)^2/\alpha$.

Finally, in Model 7 we estimate ν as the standard deviation of the standard errors of the glm estimates by each country, year, and education level. We specify the following;

$$\tau_{cye}^{easfr} \sim G(1/\sigma_{cye}^{easfr^2}, \sigma_{cye}^{easfr^2})$$

In Figures B-1 and B-2, we compare the estimates of the tfr2 module's DHS education-specific estimates against the described models to compare how close our models' estimations are to those of DHS.

Figure B-1: Comparison of results of different models for age-specific fertility rates by level of education for Africa

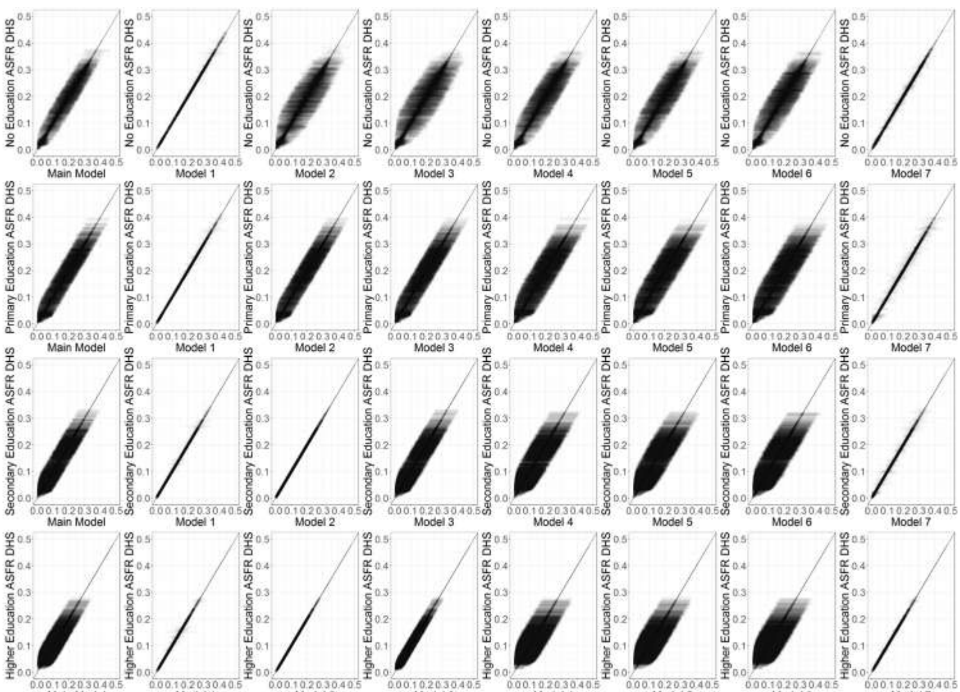
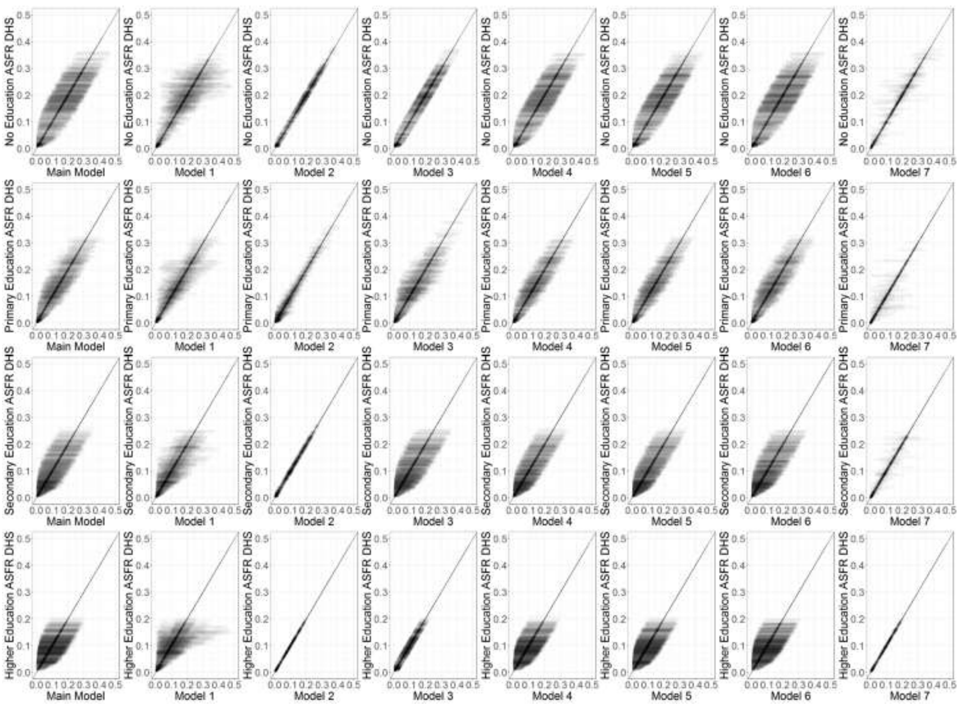


Figure B-2: Comparison of results of different models for age-specific fertility rates by level of education for Latin America



Appendix C: Differences between no education group and other levels of education

Figure C-1: Differences in education-specific, age-specific fertility rates, 1970–1975

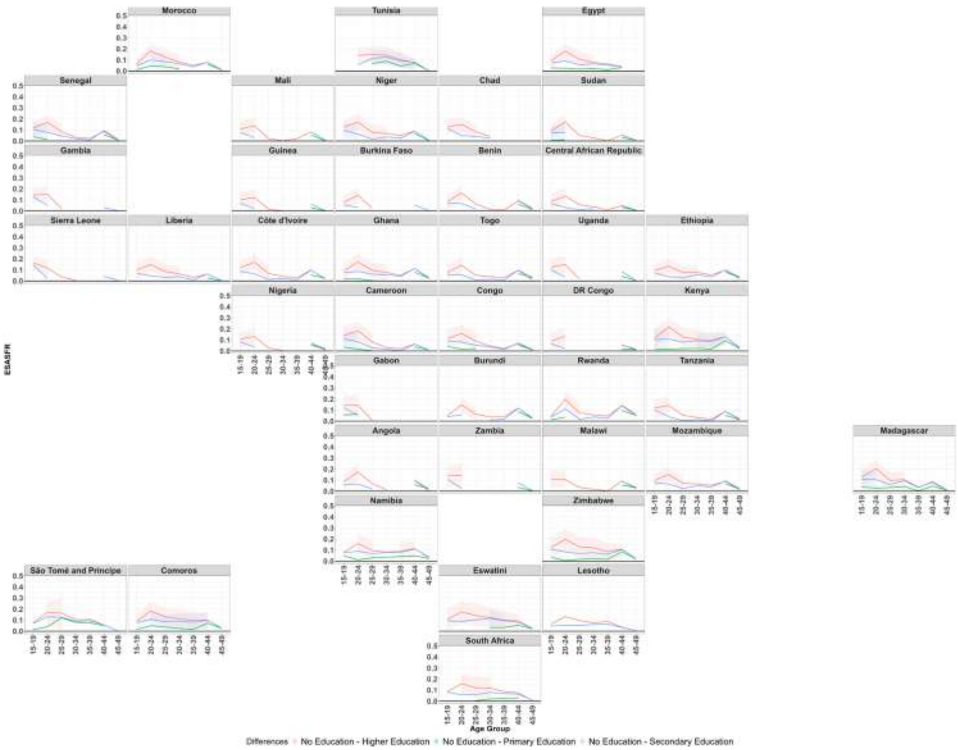


Figure C-2: Differences in education-specific, age-specific fertility rates, 1970–1975

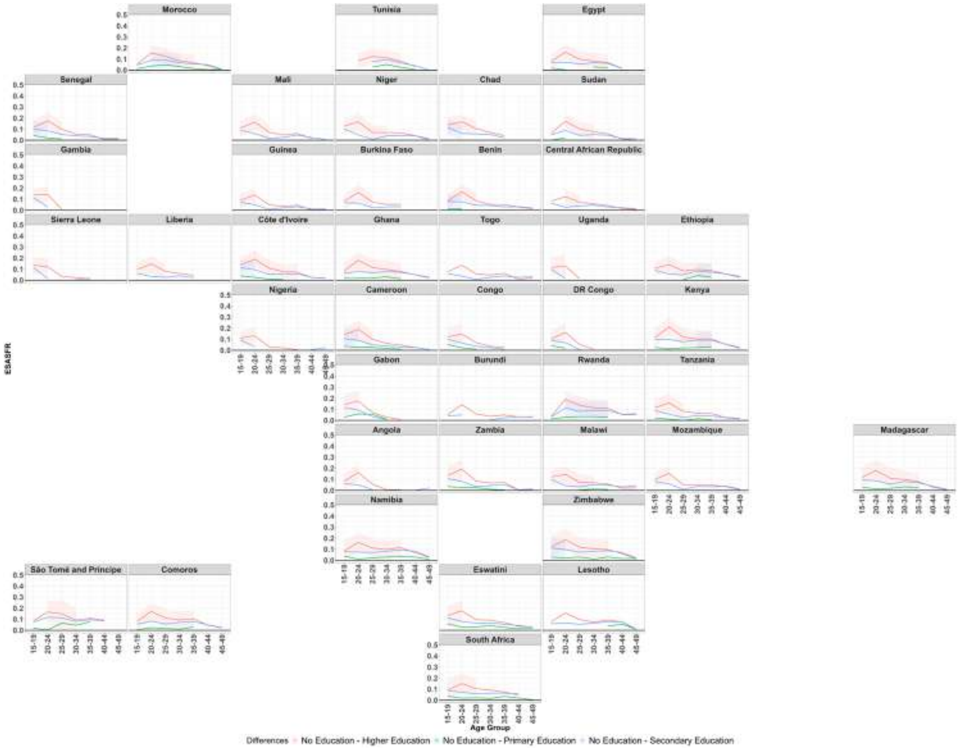


Figure C-3: Differences in education-specific, age-specific fertility rates, 1980–1985

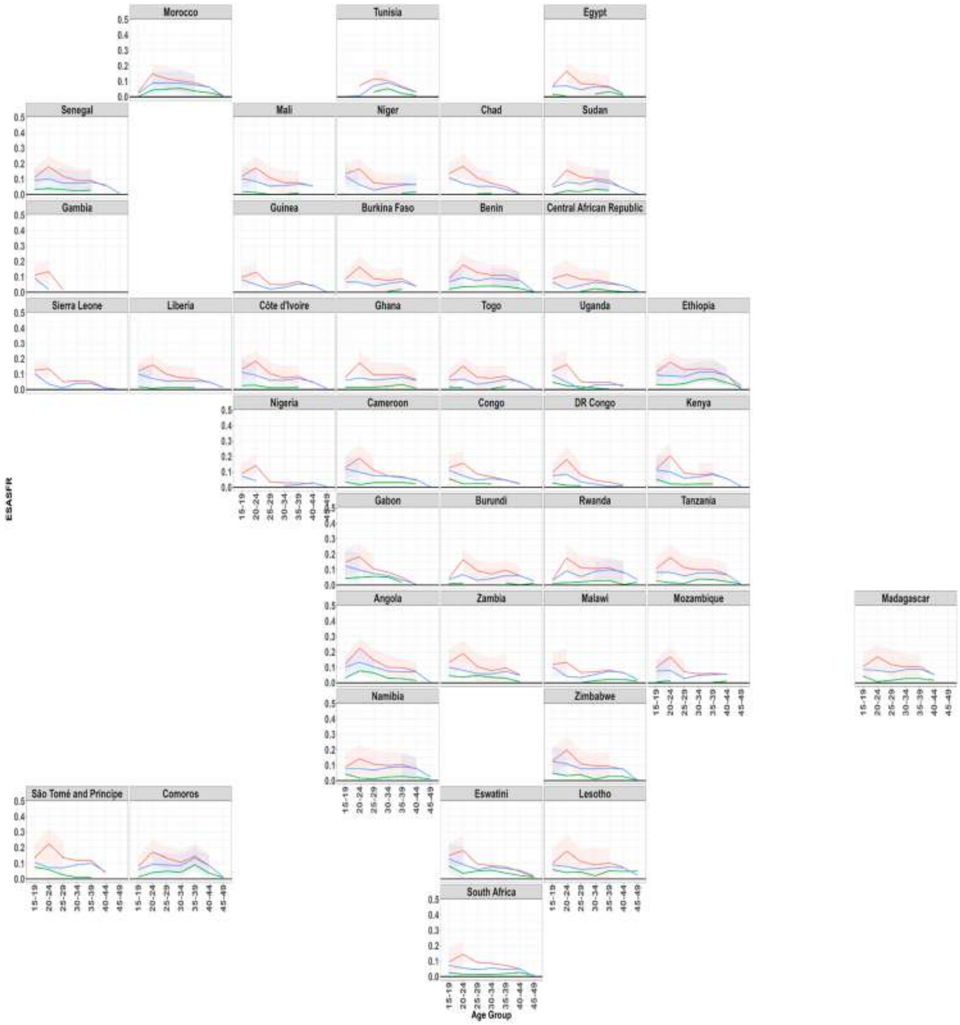


Figure C-4: Differences in education-specific, age-specific fertility rates, 1985–1990

Figure C-5: Differences in education-specific, age-specific fertility rates, 1990–1995

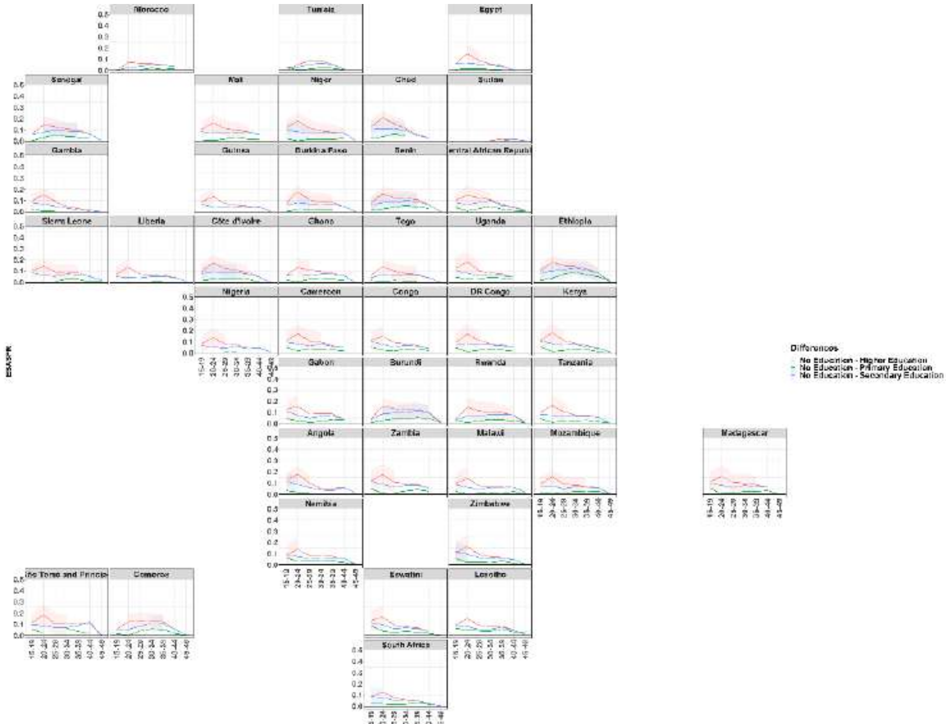


Figure C-6: Differences in education-specific, age-specific fertility rates, 1995–2000

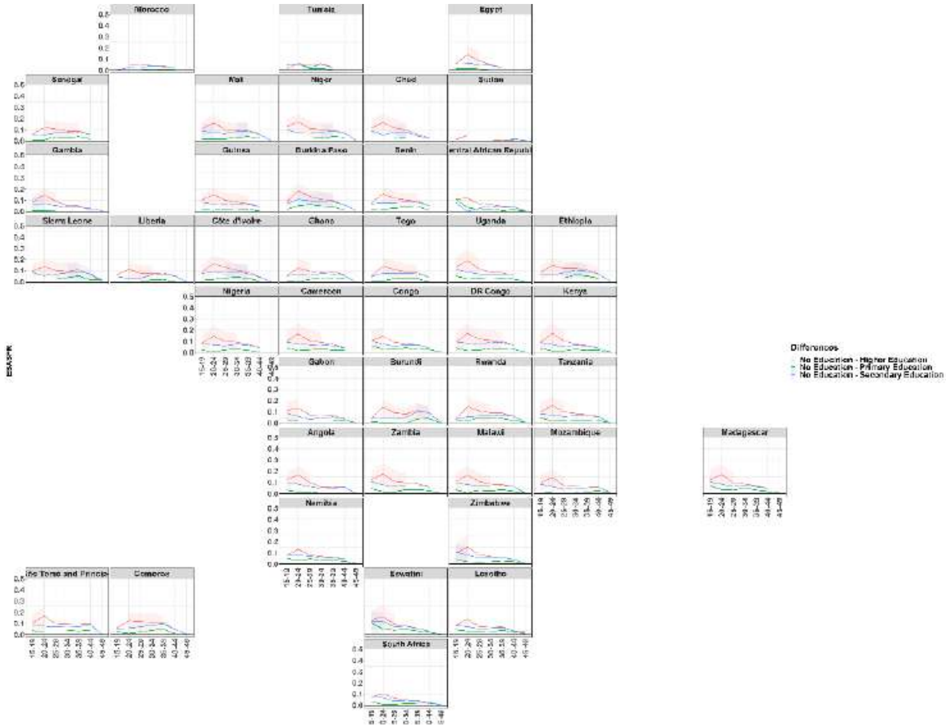


Figure C-7: Differences in education-specific, age-specific fertility rates, 2000–2005

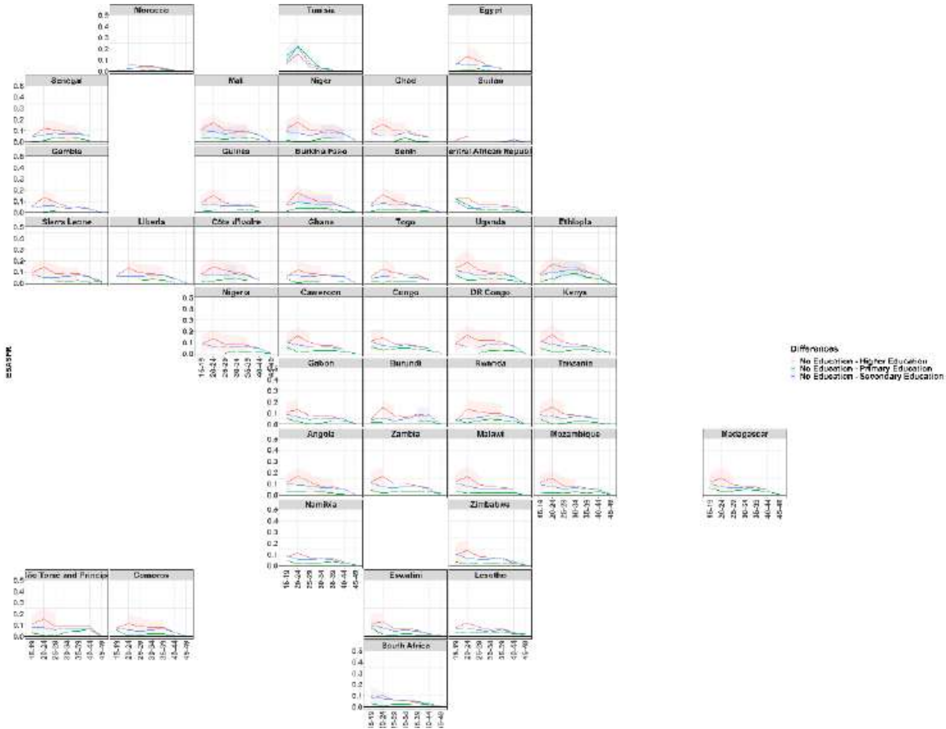


Figure C-8: Differences in education-specific, age-specific fertility rates, 2005–2010

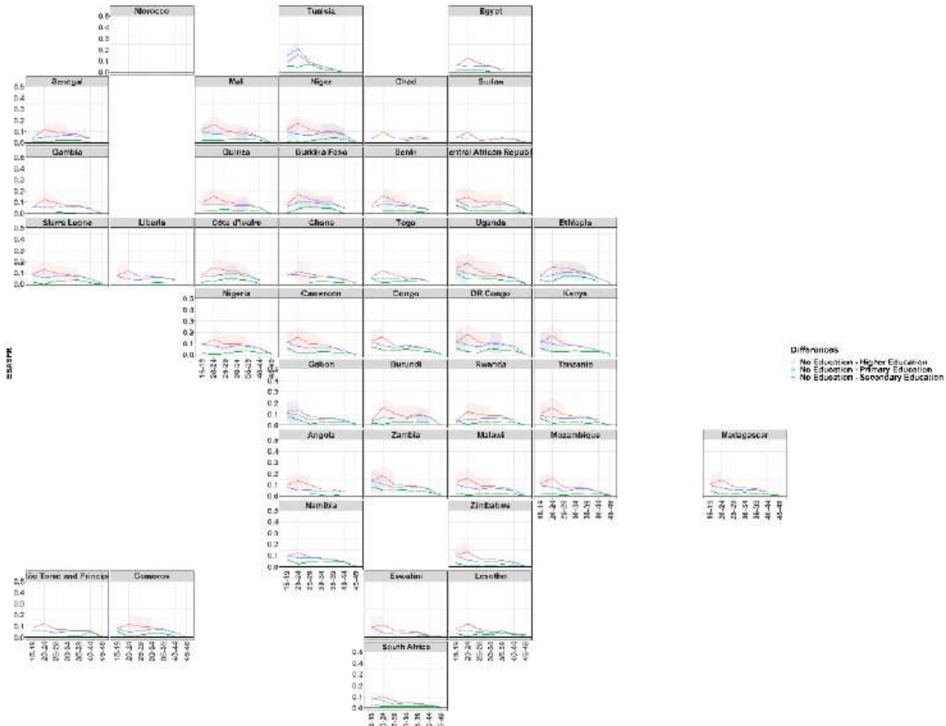


Figure C-9: Differences in education-specific, age-specific fertility rates, 2010–2015

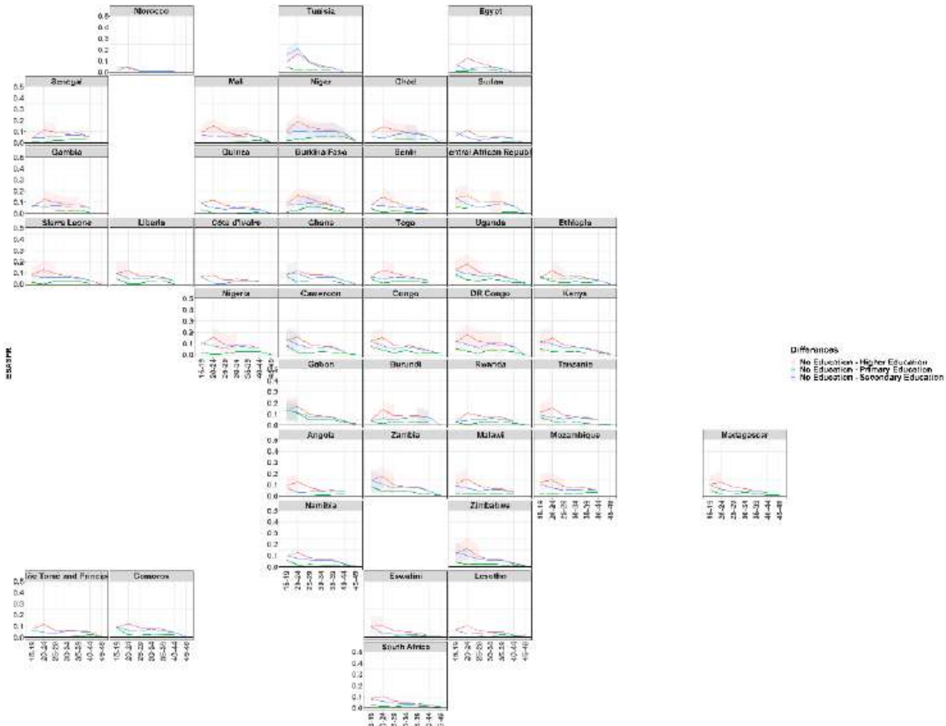


Figure C-10: Differences in education-specific, age-specific fertility rates, 2010–2015

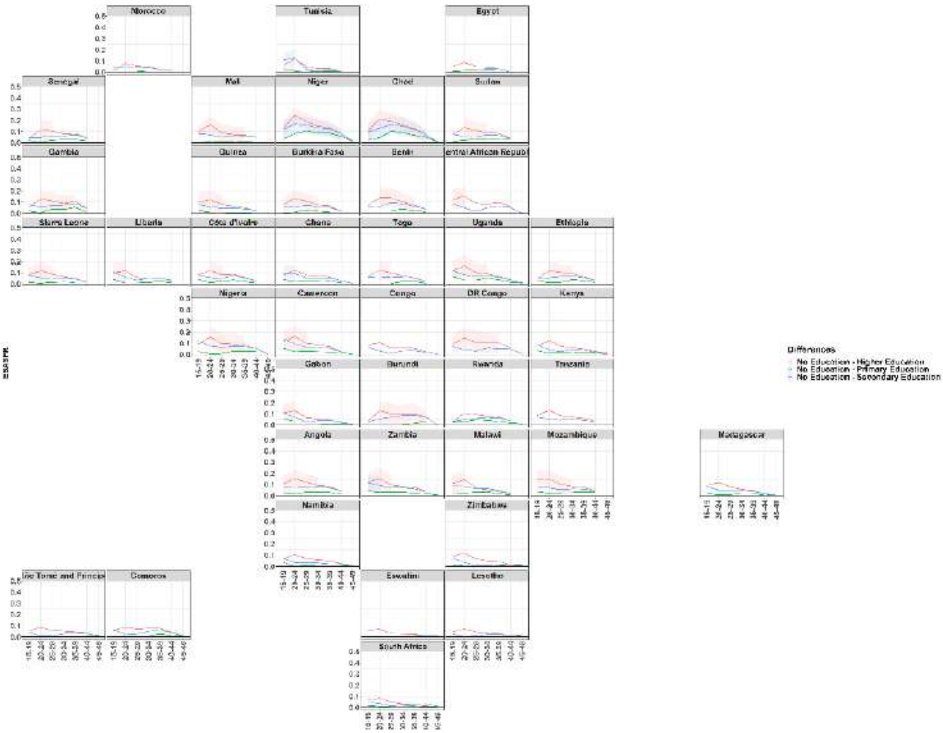
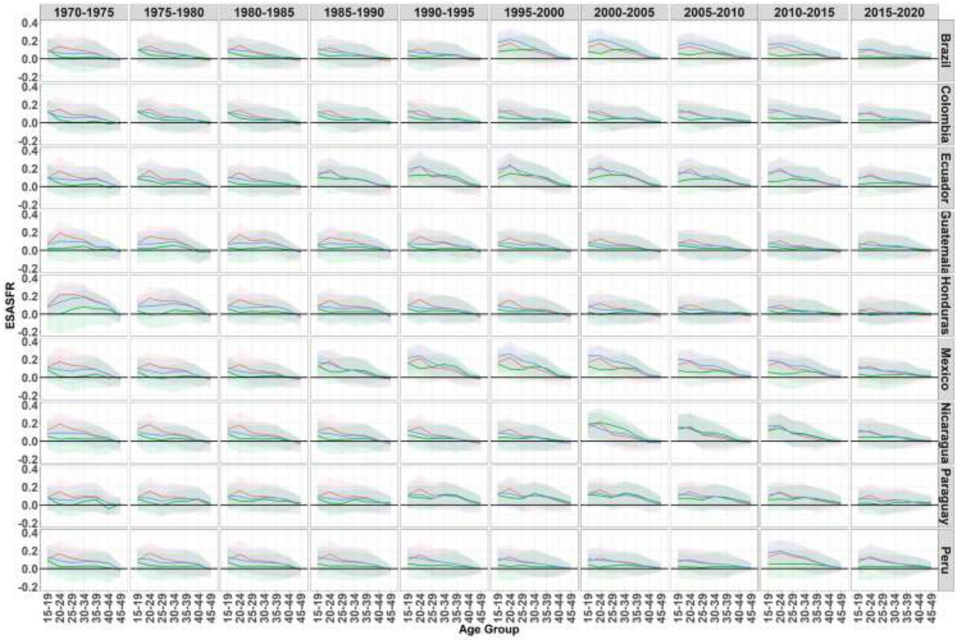


Figure C-11: Differences in education-specific, age-specific fertility rates



Appendix D: List of items in supplementary material

Name	Description
cc_y_edu_all_paper_models.csv	UN-consistent ESTFR estimates by Yildiz et al. (2023)
Cleaned_DHS_LA.xlsx	DHS estimates from tfr2 module for Latin American countries
Cleaned_DHS_Africa.xlsx	DHS estimates from tfr2 module for African countries
Bayesian model Africa.R	R code for Bayesian fertility rates for African countries
Bayesian model Latin America.R	R code for Bayesian fertility rates for Latin American countries
Glm code.R	Glm code for initial inputs for Bayesian estimation
glm_predict_all.xlsx	Glm estimates serving as initial values
Validation EASFR Africa.R	Model validation code for EASFRs Africa
Validation EASFR Latin America.R	Model validation code for EASFRs Latin America
UN_datasets3.xlsx	UN estimates from WPP 2022
WIC_datasets.xlsx	Female population estimates by age and education level from WIC
Validation_easfr Africa.pdf	PDF file containing graphs of validation exercise for Africa
Validation_easfr Latin America.pdf	PDF file containing graphs of validation exercise for Latin America

Note: Education-specific fertility estimates can be accessed at https://zenodo.org/record/8182960 and https://github.com/AfuaD-B/A-Bayesian-model-for-the-reconstruction-of-education--and-age-specific-fertility-rates