

A harmonized dataset of education- and age-specific fertility rates for DHS countries

Received: 11 November 2025

Accepted: 26 March 2026

Cite this article as: Durowaa-Boateng, A., Yildiz, D., Goujon, A. A harmonized dataset of education- and age-specific fertility rates for DHS countries. *Sci Data* (2026). <https://doi.org/10.1038/s41597-026-07136-z>

Afua Durowaa-Boateng, Dilek Yildiz & Anne Goujon

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

A harmonized dataset of education- and age-specific fertility rates for DHS countries

Afua Durowaa-Boateng¹ ✉ afua.durowaa-boateng@oeaw.ac.at, Dilek Yildiz² ✉ yildiz@iiasa.ac.at, and Anne Goujon^{1,2}

¹Vienna Institute of Demography of the Austrian Academy of Sciences, Wittgenstein Center for Demography and Global Human Capital (VID/OeAW, IIASA, University of Vienna), Vienna, Austria

²International Institute for Applied Systems Analysis, Wittgenstein Center for Demography and Global Human Capital (IIASA, OeAW, University of Vienna), Laxenburg, Austria

Abstract

Education and fertility levels of individuals and communities are closely interlinked. Education-specific fertility rates are useful to understand how differences in educational attainment influence reproductive behaviour. These estimates also serve as foundations for designing and developing population projections under varying education and fertility scenarios. However, consistent series of historical education- and age-specific fertility rates are rarely available outside countries with strong vital registration systems. Here, we estimate education- and age-specific fertility rates for 78 countries within the Demographic and Health Surveys (DHS) database, spanning the period from 1970 to 2020, in five-year intervals. Using a Bayesian approach, we combine data from multiple sources to estimate UN-consistent education- and age-specific fertility rates.

Background & Summary

Education plays a vital role in the demographic transition and particularly the transition from high to low fertility¹. The role of education in accelerating this process encompasses increased autonomy², access to and utilization of contraceptives, delayed childbearing that inadvertently lowers fertility levels, and the rising opportunity costs associated with expanded economic opportunities among others^{3–8}. The reverse also holds, where declines in fertility rates at younger ages (especially adolescent and teenage pregnancies), foster and increase the chances of higher educational attainment^{9,10}. Given the well-established relationship between education and female fertility, it is imperative to develop historically comparable fertility rates by educational attainment. These estimates would be useful, for instance, to track the impacts of past educational policies on fertility rates and project the future trajectories of both fertility rates and population sizes.

The age-specific fertility rate (ASFR) is calculated by dividing the number of live births to women in a specific age group during a given year by the total number of women in that same age group. The total fertility rate (TFR) is then calculated by summing the ASFRs across all reproductive ages (typically 15 to 49) and multiplying this sum by the width of the age interval (usually five years when ASFRs are grouped, or one year when using single-year ages). Data to calculate fertility rates are ideally obtained from complete vital registration systems, which provide reliable and annual data on births by the mother's single years of age. However, such complete registration systems are still lacking in many Global South countries. Even in Global North countries with vital registrations, the data availability is often limited to the most recent decades. In the absence of a valid registration system, data from population and housing censuses and representative demographic surveys are used as alternative sources. Censuses collect detailed information on lifetime fertility and the date of the last child born alive¹¹ but are prone to misreporting and underestimation of fertility rates over time. Moreover, not all countries—particularly Global South countries—conduct censuses regularly, which limits the availability and comparability of fertility data.

The Demographic and Health Surveys (DHS) are the main source of demographic data in low- and middle-income countries. Demographic and Health Surveys provide data for multiple surveys and consist of about 90 countries with surveys spanning from 1984 to 2025 (ongoing), with the most recent available surveys from 2023–2024¹². The birth history module of the standard DHS collects retrospective information from the female respondents in reproductive age groups, including the details of their pregnancies, births, and children. The Multiple Indicator Cluster Surveys programme (MICS)¹³, by the United Nations Children's Fund (UNICEF), provides data for over 120 countries since 1990. The survey modules, much like the DHS, comprise birth histories and other indicators of women¹⁴. The Malaria Indicators Survey (MIS), conducted in over 30 countries starting from the 2000s (and ongoing), also implemented by the DHS program, combines modules within the standard DHS and MICS¹². For Europe, the Gender and Generation Survey (GGS), a cross-national panel survey focusing on the life-course and family dynamics, has been collecting data on fertility for more than 20 countries since 2004¹⁵.

Multiple global databases combine fertility data from vital registrations, censuses, and demographic surveys. However, no source provides harmonized data on education- and age-specific fertility for a large number of Global South countries spanning multiple decades. Historical ASFRs are more widely available as various international organisations collect and analyse such data. The United Nations World Population Prospects (UN WPP)¹⁶, produced by the UN Department of Economic and Social Affairs, provides the most widely used global and long-term series of demographic indicators, including TFRs and ASFRs. These indicators are derived from censuses, surveys, and vital registration systems, analysed vigorously and harmonized using Bayesian hierarchical models¹⁷. The UN WPPs are updated approximately

every two years and offer data for 237 countries and areas from 1950. Complementary estimates are available from other UN bodies such as the UN Economic Commission for Latin America and the Caribbean (ECLAC)¹⁸, which publishes regional fertility data, together with additional indicators. For Europe, Eurostat, the statistical institute of the European Union, publishes age-specific fertility rates and number of live births by mothers' age and educational attainment, as reported by the national statistical institutes of 46 countries for the 2007-2023 period¹⁹. The United States (US) Census Bureau International Database compiles and estimates TFRs and ASFRs as well as age-specific births for 227 countries and equivalent areas starting from 1950²⁰.

In addition, several research institutes and research teams compile and harmonize historical fertility measures. At the global level, the Wittgenstein Centre for Demography and Global Human Capital (WIC) and the Institute for Health Metrics and Evaluation (IHME)²¹ provide historical population counts by age and sex, as well as demographic indicators, including fertility rates²². WIC's historical population series is modelled on the UN WPP and disaggregated by educational attainment; however, its historical fertility estimates are not education-specific, even though WIC produces projections of education-specific fertility rates from 2020 until 2100. IHME's historical fertility rates are estimated using mixed-effects regression models combined with spatiotemporal Gaussian process regression²³. For developed countries, the Human Fertility Database²⁴ provides historical and recent fertility data for 33 countries and areas with full coverage of vital statistics for different time periods, starting as early as the 1890s for Sweden²⁵. Similar to the Human Fertility Database, the Cohort Fertility and Education (CFE) Database publishes data mostly for the European and developed countries in other regions. The CFE Database, being one of the few databases that provides education-specific fertility data, publishes completed cohort fertility and parity distribution by level of education for about 45 countries using data from 2001 to 2011 censuses²⁶. For Global South countries, the African Fertility Data Explorer (AFD)²⁷ provides harmonized education-specific fertility rates from 1970 to 2020 for 41 countries in Africa. In this paper, we build on the methodology of AFD using Bayesian models and extend the education-specific ASFR (EAFR) and TFR (ETFR) data set to 78 countries for which DHS survey data are available^{6,7}.

Several studies and reports have found that the quality of DHS surveys varies over time and across countries due to recall error, age heaping, omission of events, and sampling error^{28,29}. To address these challenges and borrow information from multiple data sources, Bayesian models have been employed for demographic estimation, particularly for fertility, mortality, and migration when registration systems are incomplete. At the global scale, Bayesian hierarchical models are used in the United Nations probabilistic population projections and in their under-five mortality estimations³⁰⁻³². At the national, sub-national, and regional levels, several studies have applied Bayesian models and other methodologies to estimate both education-specific and overall mortality, fertility, and migration rates^{6,7,33-43}.

In the remainder of this article, we first outline our methodology and highlight the differences with the previous studies. We illustrate education-specific ASFRs and TFRs over time. We then provide some guidance for potential users of the dataset and present the results of technical validation.

Methods

Data Sources

To estimate five-year *EAFR* and *ETFR* for 78 countries between 1970-1975 and 2015-2020 period, we obtained and cleaned data from the DHS, the UN WPP 2022, and the Wittgenstein Centre Human Capital Data Explorer (WIC)^{12,17,44}. Although the UN WPPs have a 2024 round of projections, this study employs the 2022 rounds to ensure consistency across the input data. The WIC education estimates are calibrated to the UN WPP 2022 rounds of revision. Updating the benchmark from the 2022 to the 2024 revision would change the final estimates, because the model benchmarks the overall ASFR schedule to the UN WPP. The *EAFR* estimates are for five-year groups from 15-19 to 45-49, spanning the standard reproductive ages for women. In both *EAFR* and *ETFR*, education represents the highest educational attainment category for the respondent, as indicated by the DHS⁴⁵. The categories are “No Education”, “Primary Education”, “Secondary Education” and “Higher Education”.

All input data are publicly available and freely downloadable; however, the DHS microdata require free registration. The UN WPP fertility estimates can be accessed on the UN WPP site at <https://population.un.org/wpp/downloads?folder=Standard%20Projections&group=Fertility>. The 2022 round of revisions was used for this methodology and was accessed via the wpp2022 R package⁴⁶.

The education-specific birth histories from the DHS microdata can be accessed free of charge after user registration with the DHS program at <https://www.dhsprogram.com/data/new-user-registration.cfm>¹². To derive the estimates for the GLM in Equation 1 below, input data are obtained using the Individual/women's recode (IR) file. The recode files are processed in STATA using the "tfr2" module and follow the example code structure below:

```
use CCIR01FL.DTA
by v106, sort: tabexp, len(30) trend(5) ageg(5) cy rates
```

The above code, which is an example replica for the estimates used, generates the birth histories for the 30 years before the survey from the survey wave by level of education. The file name above is an example of the general file name of the DHS data, which is country code or abbreviation (CC), file type, in this case individual recode (IR), survey wave (usually numeric), followed by FL.DTA. The study draws on data from 343 DHS surveys, incorporating all available survey waves up to 2024¹². This process is repeated for all countries and survey waves listed in Table 1. The resulting data is cleaned in R, where the reported birth count exceeding 30 is removed for women who have higher education and other implausible point estimates after visualizing the data. The cleaned data form the basis of the GLM, and the predicted estimates are saved as the initial values for the Bayesian model.

Table 1: Country list with survey years

Survey Year	Countries
1985	El Salvador
1986	Brazil, Colombia, Dominican Republic, Liberia, Peru, Senegal
1987	Burundi, Ecuador, Guatemala, Indonesia, Mali, Mexico, Morocco, Sri Lanka, Thailand, Trinidad and Tobago
1988	Egypt, Ghana, Kenya, Togo, Tunisia, Uganda, Zimbabwe
1989	Bolivia, Sudan
1990	Jordan, Nigeria, Pakistan, Paraguay
1991	Brazil, Cameroon, Dominican Republic, Indonesia, Peru, Tanzania, Yemen
1992	Burkina Faso, Egypt, India, Madagascar, Malawi, Morocco, Namibia, Niger, Rwanda, Senegal, Zambia
1993	Bangladesh, Bolivia, Ghana, Kenya, Philippines, Turkey
1994	Central African Republic, Côte d'Ivoire, Haiti, Indonesia, Zimbabwe
1995	Colombia, Egypt, Guatemala, Kazakhstan, Mali, Uganda
1996	Bangladesh, Benin, Brazil, Chad, Comoros, Dominican Republic, Nepal, Peru, Tanzania, Uzbekistan, Zambia
1997	Indonesia, Jordan, Kyrgyzstan, Madagascar, Mozambique, Nicaragua, Senegal, Vietnam
1998	Bolivia, Burkina Faso, Cameroon, Côte d'Ivoire, Ghana, Guatemala, India, Kenya, Niger, Philippines, South Africa, Togo, Turkey
1999	Bangladesh, Dominican Republic, Guinea, Kazakhstan, Mozambique, Tanzania, Tunisia, Zimbabwe
2000	Armenia, Cambodia, Colombia, Egypt, Ethiopia, Gabon, Haiti, Malawi, Namibia, Peru, Rwanda, Uganda
2001	Benin, Mali, Nepal, Nicaragua, Zambia
2002	Dominican Republic, Indonesia, Jordan, Vietnam
2003	Bolivia, Burkina Faso, Egypt, Ghana, Kenya, Madagascar, Morocco, Mozambique, Nigeria, Peru, Philippines, Turkey
2004	Bangladesh, Cameroon, Chad, Colombia, Lesotho, Malawi, Tanzania
2005	Armenia, Cambodia, Congo, Egypt, Ethiopia, Guinea, Guyana, Haiti, Honduras, India, Moldova, Rwanda, Senegal, Vietnam, Zimbabwe
2006	Angola, Azerbaijan, Benin, Eswatini, Liberia, Mali, Namibia, Nepal, Niger, Pakistan, Uganda
2007	Bangladesh, Democratic Republic of Congo, Indonesia, Jordan, Rwanda, Ukraine, Zambia
2008	Albania, Bolivia, Egypt, Ghana, Kenya, Liberia, Madagascar, Nigeria, Philippines, Sao Tome and Principe, Sierra Leone, Turkey
2009	Colombia, Guyana, Jordan, Lesotho, Maldives, Peru, Tanzania, Timor-Leste, Uganda

2010	Armenia, Burkina Faso, Burundi, Cambodia, Malawi, Nigeria, Peru, Rwanda, Senegal, Zimbabwe
2011	Angola, Bangladesh, Benin, Cameroon, Congo, Côte d'Ivoire, Ethiopia, Honduras, Liberia, Madagascar, Mozambique, Nepal, Peru, Uganda
2012	Burundi, Comoros, Gabon, Guinea, Haiti, Indonesia, Jordan, Kyrgyzstan, Malawi, Mali, Niger, Pakistan, Peru, Senegal, Tajikistan
2013	Democratic Republic of Congo, Gambia, Guatemala, Liberia, Madagascar, Namibia, Nigeria, Philippines, Rwanda, Sierra Leone, Togo, Turkey, Yemen, Zambia
2014	Bangladesh, Burkina Faso, Cambodia, Chad, Egypt, Ghana, Kenya, Lesotho, Malawi, Rwanda, Senegal, Uganda
2015	Afghanistan, Angola, Armenia, Colombia, India, Kenya, Malawi, Mali, Mozambique, Myanmar, Nigeria, Senegal, Tanzania, Zimbabwe
2016	Burundi, Ethiopia, Ghana, Haiti, Liberia, Madagascar, Maldives, Nepal, Papua New Guinea, Senegal, Sierra Leone, South Africa, Timor-Leste, Uganda
2017	Albania, Bangladesh, Benin, Burkina Faso, Indonesia, Jordan, Pakistan, Philippines, Rwanda, Senegal, Tajikistan, Tanzania, Togo
2018	Cameroon, Guinea, Mali, Nigeria, Senegal, Uganda, Zambia
2019	Ethiopia, Gambia, Ghana, Liberia, Senegal, Sierra Leone, Turkey
2020	Rwanda
2021	Burkina Faso, Côte d'Ivoire, Gabon, Guinea, India, Madagascar, Mali, Mauritania, Niger, Nigeria
2022	Bangladesh, Cambodia, Cameroon, Kenya, Liberia, Nepal, Philippines, Tanzania
2023	Ghana, Jordan, Mozambique, Senegal
2024	Lesotho

Education-specific population sizes used in the Bayesian model are collected from the Wittgenstein Centre for Demography and Global Human Capital using the R package `wcde`, specifically the 2018 WIC estimates denoted by `wcde-v2` in the package⁴⁴.

For sub-Saharan African countries, previous estimates of Bayesian education-specific fertility rates were collected. The UN-consistent Bayesian education-specific total fertility rates for sub-Saharan African countries (ETFR DHS UN) can be publicly downloaded at Zenodo (DOI:[10.5281/zenodo.18185201](https://doi.org/10.5281/zenodo.18185201))⁴⁷.

Model structure

The main aim of the proposed methodology is to disaggregate total and age-specific fertility rates by level of education. The approach relies on a single source for overall fertility rates (the UN WPP), a primary source for education-specific fertility rates (the DHS), and a source for population size by level of education (the WIC). Additionally, education-specific fertility estimates from a previous harmonisation effort were included in the modelling framework for consistency⁷.

Following Durowaa-Boateng et al.⁶, we employ a two-step approach. In the first step, initial EAFRs are generated using a Generalized Linear Model (GLM) to complete age schedules that are only partially observed in the DHS microdata. These estimates are then used as initial values in the second step, where they inform the hierarchical Bayesian model. The main difference between our model and previous research is the joint estimation of fertility rates. This improvement allows the model to share information across regions, rather than modeling them separately. The region parameter follows the UN regions: sub-Saharan Africa, North Africa, West Asia, Europe, Central Asia, South & Southeast Asia, and Latin America.

Step 1: Modelling initial values from DHS (GLM)

To estimate the initial values, we use a generalized linear model (GLM) following the equation,

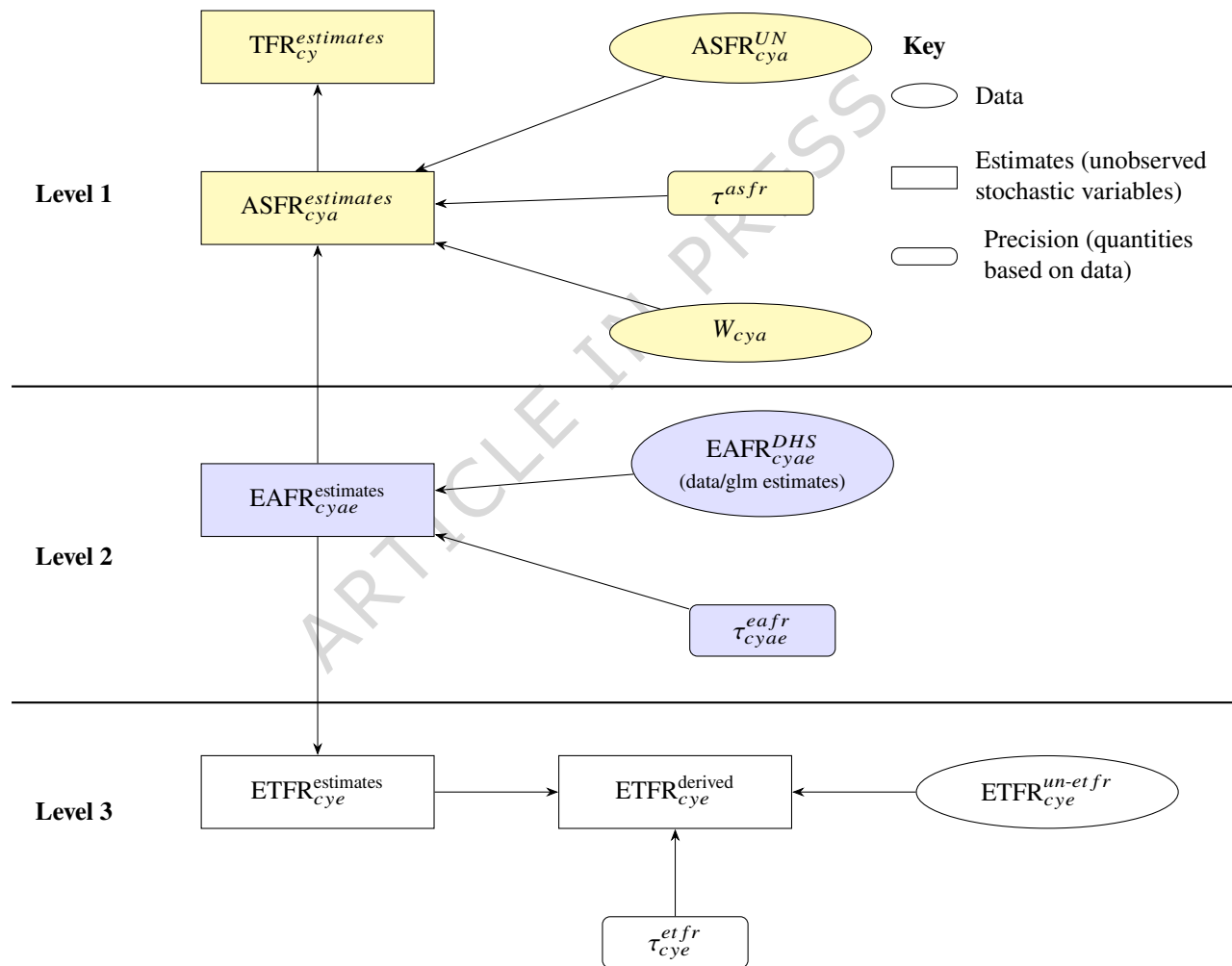
$$\begin{aligned}
 EAFR^{DHS} \sim & \text{Education} + \text{Region} + \text{Age Group} + \text{Country} + \text{Year} \\
 & + \text{Education} \cdot \text{Age Group} + \text{Year} \cdot \text{Age Group} \\
 & + \text{Region} \cdot \text{Age Group} + \text{Region} \cdot \text{Year} + \text{Region} \cdot \text{Country}.
 \end{aligned} \tag{1}$$

In Equation 1, $EAFR^{DHS}$ represents the education- and age-specific fertility rates from the Demographic and Health Surveys. Due to the joint estimation across regions, a *Region* parameter is included in the GLM, which was not considered in Durowaa-Boateng et al.⁶.

Step 2: Three-level Bayesian Hierarchical Model

The second step implements a three-level hierarchical Bayesian framework, as shown in Figure 1. In this step, in addition to the primary data sources, we also include previously published fertility estimates by education level for DHS countries in the Bayesian model. This helps ensure consistency for countries that were part of earlier harmonisation efforts.

Figure 1: Graphical presentation of the Bayesian model framework.



Level 1 benchmarks age-specific fertility rates to the UN WPP ASFRs, Level 2 estimates EAFR, and Level 3 produces ETFR. The fertility estimates are for each country c , five-year period y , education level e and five-year age-group a , where necessary. These categories remain similar to those in the GLM in Equation 1.

Harmonizing education-specific fertility rates

Level 1 of the framework ensures the consistency of the estimated $EAFR$ values with the UN WPP¹⁷ $ASFR$. By applying education- and age-specific female population weights (WIC pop data) to the rates, this level ensures that the weighted education-specific fertility rates are in alignment with the UN WPP's age-specific fertility rates.

The mathematical formulations of Level 1 in Figure 1 follow:

$$TFR_{cy}^{estimates} = \sum_{a=15-19}^{45-49} \left(ASFR_{cya}^{estimates} \right) \cdot 5 \quad (2)$$

$$ASFR_{cya}^{UN} \sim N_+ \left(ASFR_{cya}^{estimates}, \tau^{asfr} \right) \quad (3)$$

$$ASFR_{cya}^{estimates} = \sum_{e=1}^4 \left(EAFR_{cyae}^{estimates} \cdot w_{cyae} \right) \quad (4)$$

$$\tau^{asfr} = 1/\sigma^{asfr^2} \quad (5)$$

Quantities with the superscript “estimates” are the unknown parameters that are estimated within the posterior sampling. The variable $ASFR^{UN}$ is the age-specific fertility rate from the UN WPP. The parameter σ^{asfr^2} is the variance of the UN WPP's age-specific fertility rates. τ^{asfr} represents the precision of the ASFR estimates in Equation 3. The variable w_{cyae} denotes the weights derived from the WIC population by education level.

The output of the GLM model ($EAFR_{cyae}^{DHS}$ (data/glm estimates)) enters the modelling framework in Level 2 as initial values to estimate $EAFR_{cyae}^{estimates}$. As shown in Equation 6, the EAFRs are modelled using a truncated normal distribution. The distribution is centred on the GLM estimates ($EAFR_{cyae}^{DHS}$) generated in Equation 1. The quantity σ_e^{eafr} represents the education-specific standard deviations of the standard errors derived from the GLM model.

The equations for Level 2 are:

$$EAFR_{cyae}^{estimates} \sim N_+ \left(EAFR_{cyae}^{DHS}, \tau_{cyae}^{eafr} \right), \quad (6)$$

$$\tau_{cyae}^{eafr} \sim G(1/\sigma_e^{eafr}, 2\sigma_e^{eafr}) \quad (7)$$

Finally, Level 3 follows:

$$ETFR_{cye}^{derived} = \left(\sum_{a=15-19}^{45-49} EAFR_{cyae}^{estimates} \right) \cdot 5 \quad (8)$$

$$ETFR_{cye}^{estimates} \sim \begin{cases} N_{+(0,10)} \left(ETFR_{cye}^{derived}, \mu_{cye}^{etfr} \right), & \text{for all regions} \\ ETFR_{cye}^{un-etfr} \sim N_{+(0,10)} \left(ETFR_{cye}^{derived}, \mu_{cye}^{etfr} \right), & \text{for sub-Saharan Africa} \end{cases} \quad (10)$$

$$\tau_{cye}^{etfr} \sim G(1/\sigma_{e4}^{etfr}, 2 \cdot \sigma_{e4}^{etfr^2}) \quad (11)$$

The quantities σ^{etfr} and σ^{etfr^2} are the standard deviation and variance, respectively, of the education-specific total fertility rates for the highest education level (Higher Education). In level 3, the estimates only for the sub-Saharan African countries are in place to ensure some degree of consistency between the education-specific TFR estimates in Yildiz et al.⁷, which were also used as a benchmark in Durowaa-Boateng et al.⁶ and the current estimates.

Data Overview

In Figures 2, 3, 4, 5 and 6, the median estimates of the EAFRs (posterior estimates) for all DHS countries are presented together with 95% credible intervals (CI) and grouped by sub-regions for each education level. These results point to an overall decline in EAFRs and ETFRs across all countries and education levels. Across all regions, estimates for the education levels “Secondary Education” and “Higher Education” are generally lower than “No Education” and “Primary Education”.

The estimates, in terms of the curvature of age-specific fertility rates, show a well-documented shift in age schedules both across the years and education levels (postponement)^{3–5}. The results and estimated patterns also show that in some countries, particularly in sub-Saharan Africa, primary-education-specific fertility rates were higher than those for women with no education in the earlier years (1970-1975). This phenomenon has been previously established and observed in other studies for that period, as women moved away from traditional birth-spacing practices (shortening postpartum abstinence and relying more on bottle feeding/shorter breastfeeding), thereby reducing the postpartum non-susceptible period and increasing the pace of childbearing relative to women with no education^{8,48,49}.

Furthermore, it is well documented that conflicts influence fertility rates at different times. In areas of conflict or natural disasters, the immediate response is usually high mortality and subsequently declining fertility rates; however, periods succeeding these periods are marked by high fertility rates^{50,51}. While the effects of conflicts on fertility rates are observed in the TFRs, they influence the timing of fertility (ASFRs). There are various mechanisms through which conflicts, as well as other events, affect the ages at which women have children. These events could directly or indirectly cause postponement among age groups—particularly in periods of active conflicts—thus leading to shifts in the age-specific fertility curve and overall TFR^{52–55}.

It is important to note that the fertility schedules observed in different regions and countries may have different reasons, and those listed here are not the sole reasons for the shape of the fertility curve. For some periods, particularly the 1970s in sub-Saharan Africa, the nature of the fertility curve and overall fertility rates could be due to a variety of events, including conflicts, such as early and near universal marriages for women, reduced female education, which has been linked to lower fertility rates and postponements, polygamy and a high desired number of children among others⁵⁶. Similarly, each country has its own age at which fertility rates are highest. This happens in different years at different age groups. The country-specific plots can be found on the GitHub repository <https://github.com/AfuaD-B/Bayesian-methodology-on-fertility-reconstruction>.

In Figure 6, the estimates reflect period fluctuations observed in the individual countries. For instance, the distinct dip detected in the ETFR of no and primary education in Cambodia from 1975 to 1980 is indicative of the actual overall TFR of the country and has been recorded in other studies^{57,58}. The cause of the sharp decline in Cambodia has in part been due to the political events during the “Khmer Rouge” regime⁵⁹. In addition to Cambodia in South and Southeast Asia, Timor-Leste’s fertility pattern is attributed to previous years of conflicts⁶⁰. Generally, women with no and primary education in lower-educated populations drive overall fertility rates due to their larger share in the composition of the female populations⁶¹. As with the case of Cambodia, the model output reflects patterns consistent with events and other contextual factors that may have influenced fertility rates at that time.

Figure 2: No Education- and age-specific fertility rates

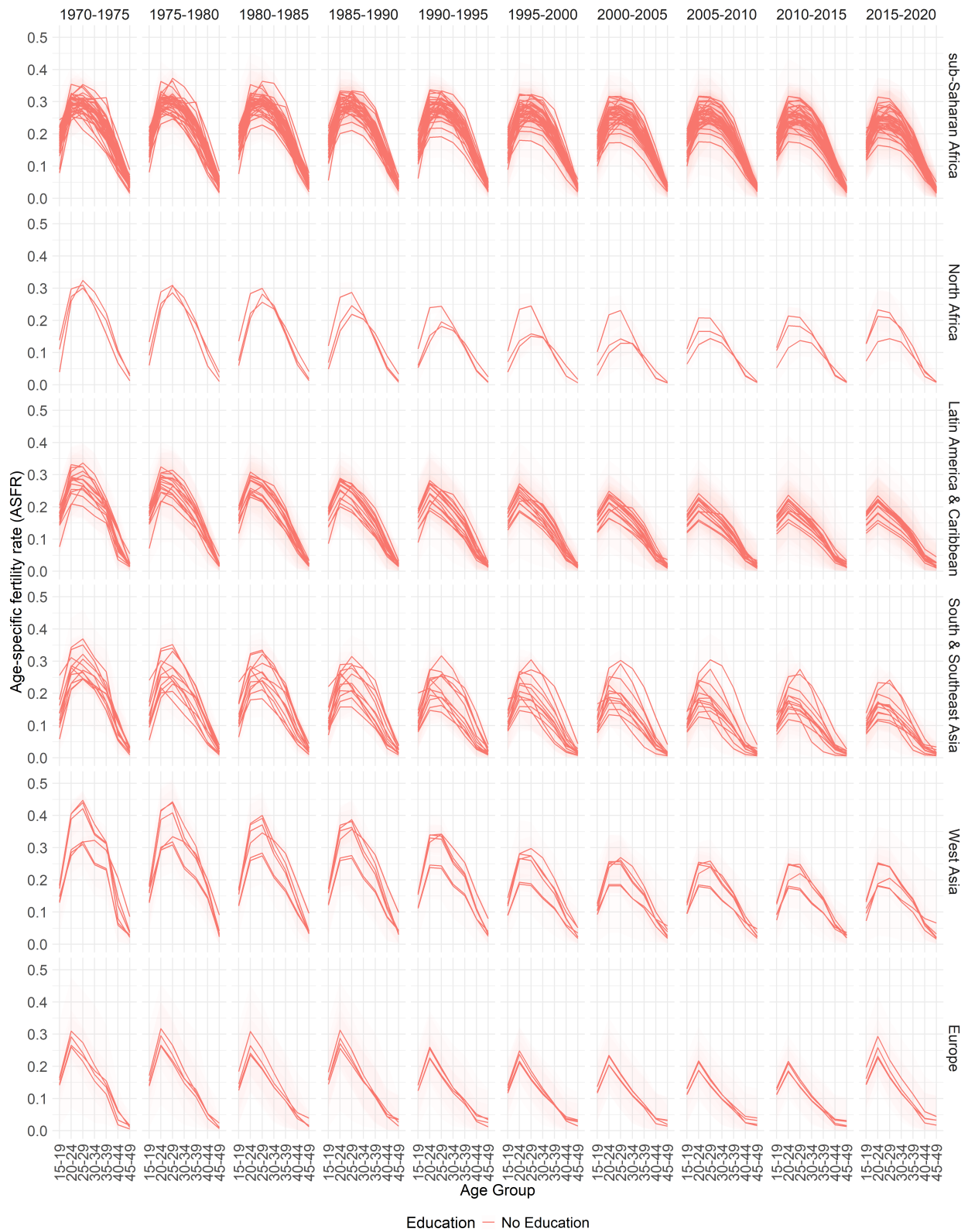


Figure 3: Primary Education- and age-specific fertility rates

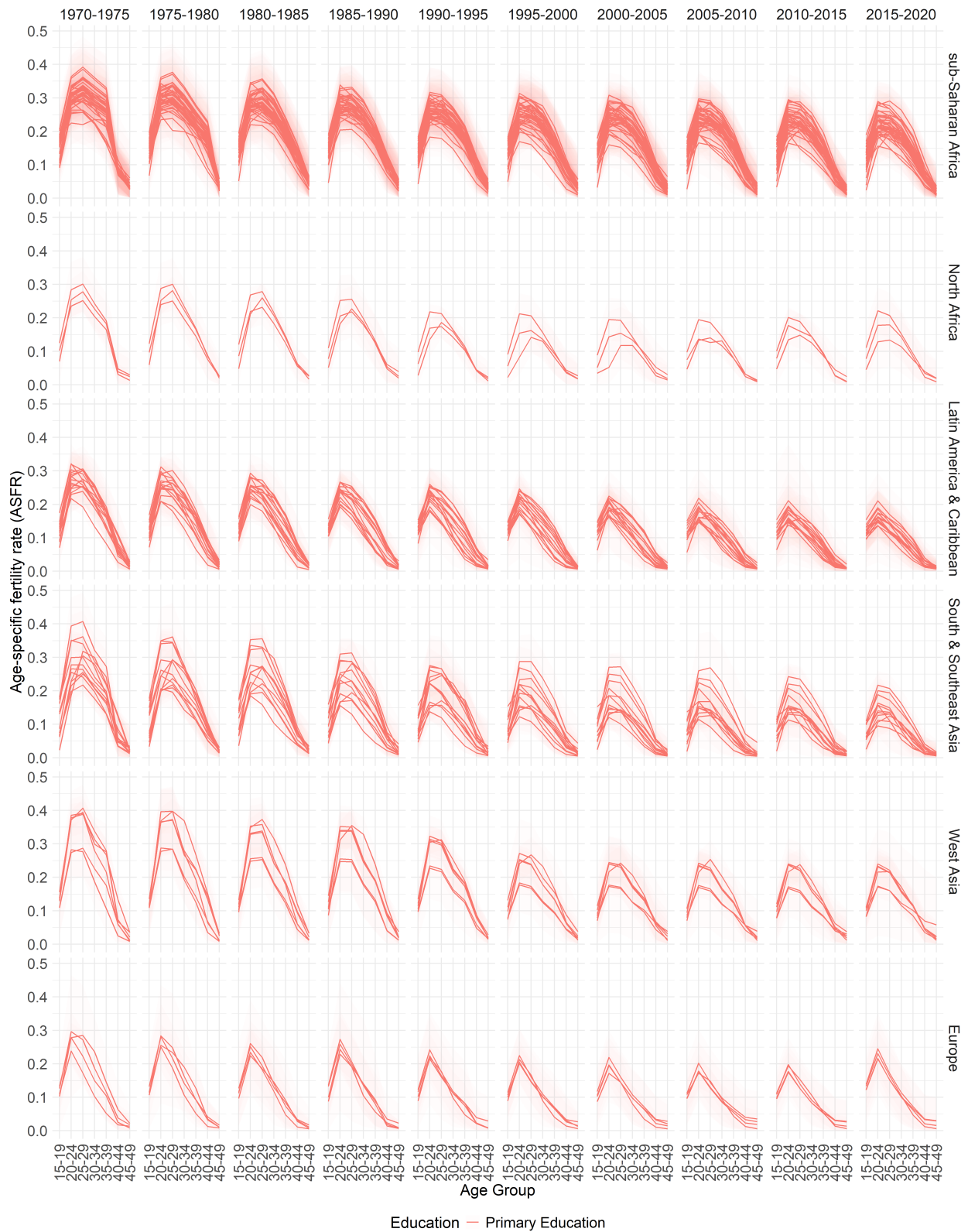


Figure 4: Secondary Education- and age-specific fertility rates

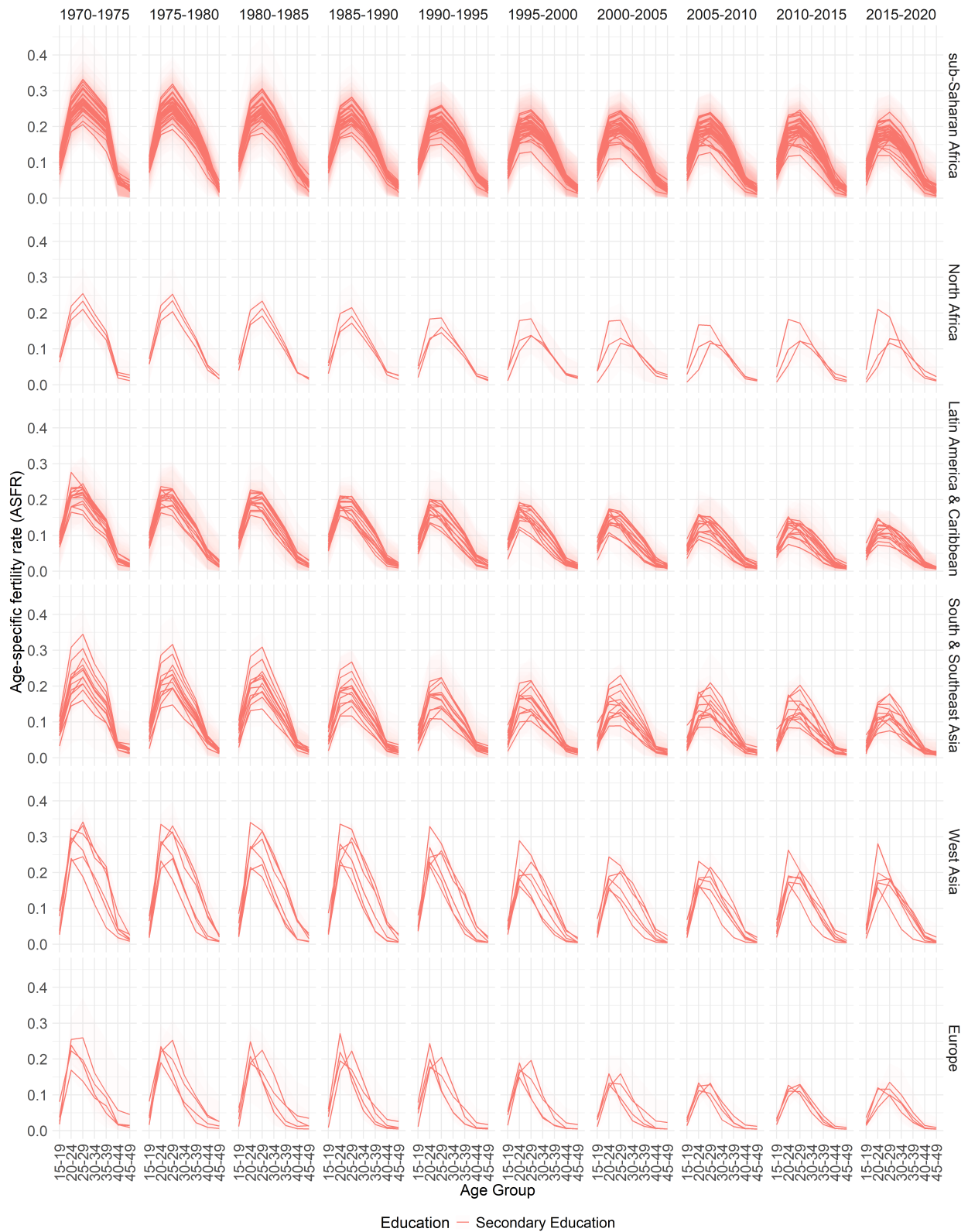


Figure 5: Higher Education- and age-specific fertility rates

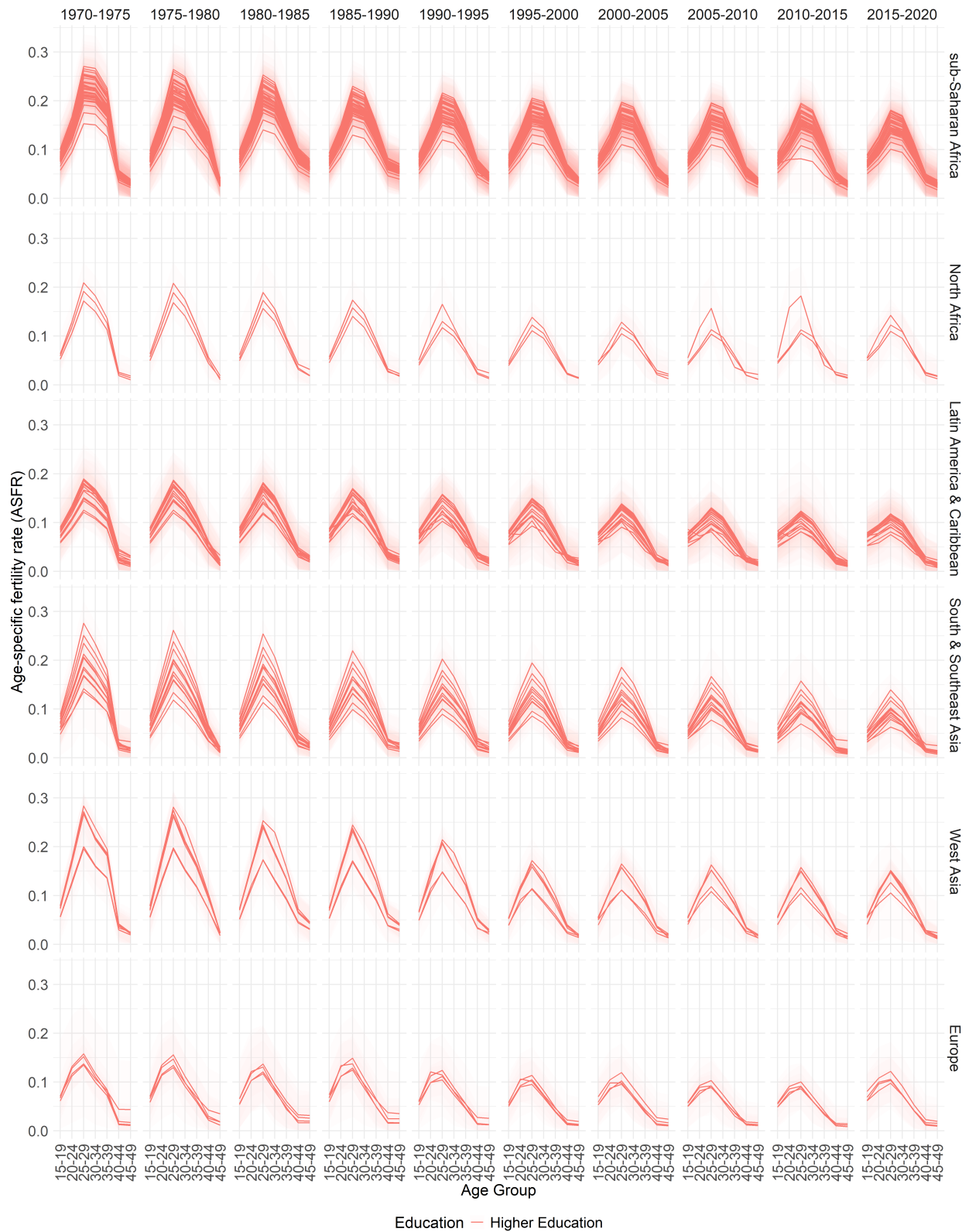
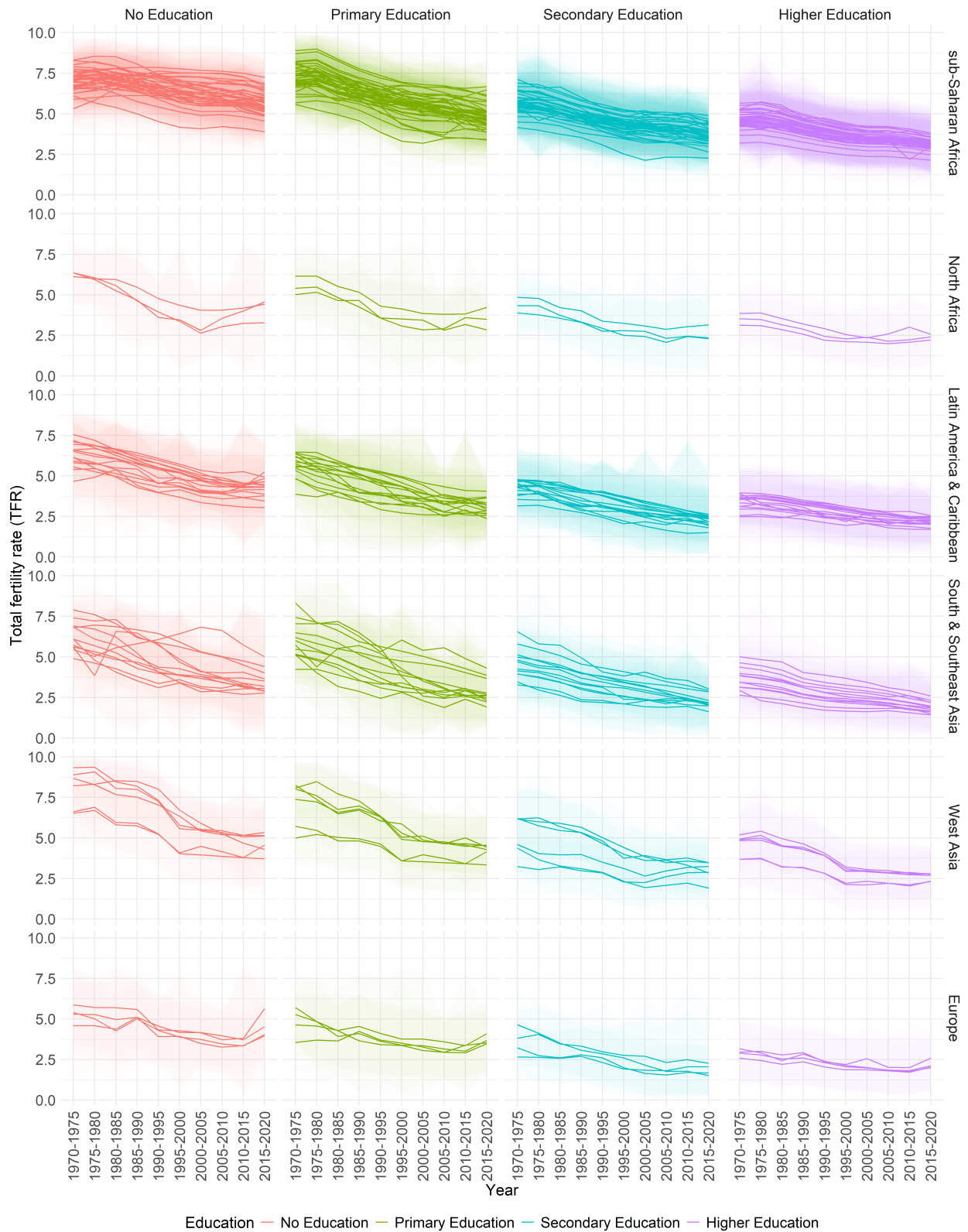


Figure 6: Education-specific total rates



Education — No Education — Primary Education — Secondary Education — Higher Education

Data Records

The dataset containing the education-specific estimates for the five-year age groups from 1970 to 2020 is available at Zenodo (DOI:10.5281/zenodo.18359482)⁶² and mirrored on GitHub under the repository <https://github.com/AfuaD-B/Bayesian-methodology-on-fertility-reconstruction>. The Zenodo contains the main dataset generated from this model, “BESFR_estimates_global_south.xlsx”. This Excel workbook has the median estimates together with the credible intervals of the education-specific fertility estimates. The file contains the columns “Country”, “Age Group”, “Education”, “Year”, “Upper_CI”, “Lower_CI”, and “Median”. The columns “Upper_CI”, “Lower_CI” represent the 95% credible intervals. Only in the GitHub repository are the separate country-specific EAFRs provided in a PDF titled “final_results_plot.pdf”. The Zenodo record also contains additional output files, including the Excel workbooks “edu_spec_sd_init_prior.xlsx”, which contains the Bayesian priors from the main model, and “edu_diff.xlsx”, which contains the estimated differences in education- and age-specific fertility rates across education levels. Code files used to fit the Bayesian models, generate the estimates and figures, and run the sensitivity and validation analyses are available in both Zenodo and GitHub, and are described in the Code Availability section.

Technical Validation

In this model estimation, we ran two validation models, both of which are random holdout validation approaches. In the first approach, 5%, 10%, 15%, 20%, and 30% of the UN *ASFR* estimates were randomly left out of the model (Figure 7). We then refit the model for the remainder of the dataset. The second version omits data from five countries: Albania, Bangladesh, Colombia, Ghana, and the Philippines (Figure 8). The data were randomly removed only from the UN WPP’s *ASFR*. Since the model is benchmarked to the UN *ASFR*s, we removed some UN *ASFR* values and refit the model to see how well it can recover the missing benchmark values from the remaining ones. This tests how sensitive the model is to missing values in the UN benchmark *ASFR* schedule.

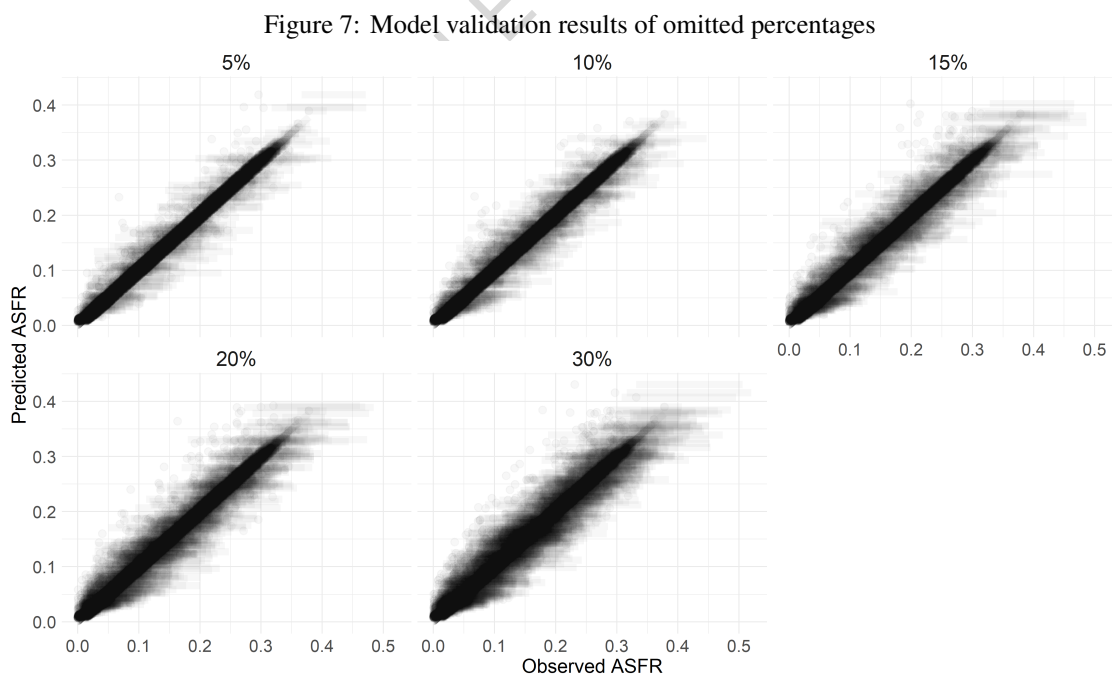
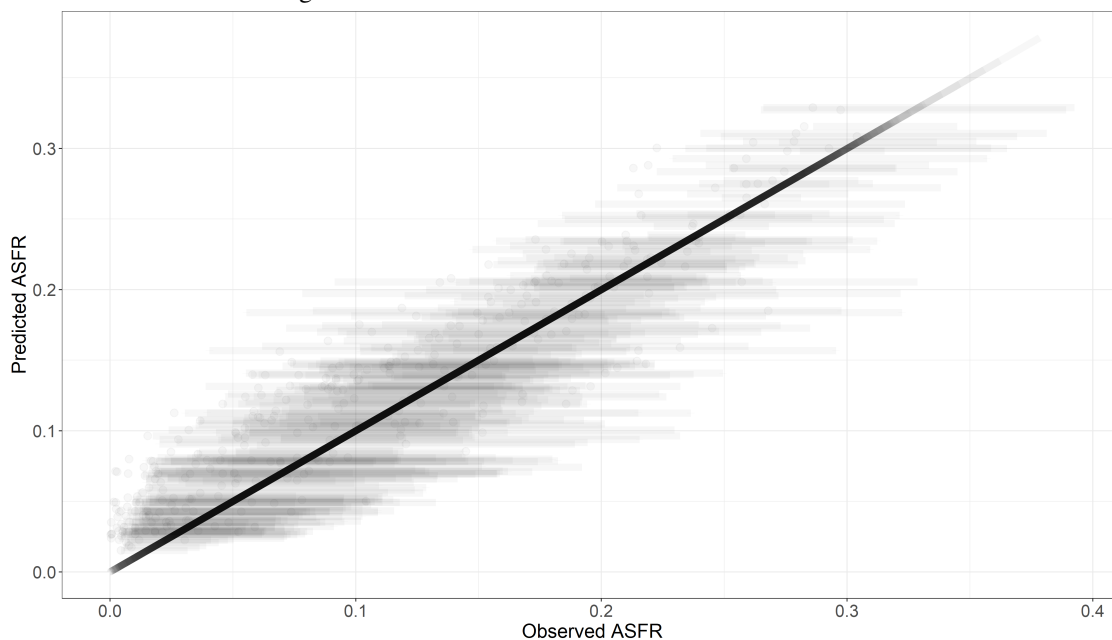


Table 2 shows the percentages of observations that fall within 50%, 80%, and 90% credible intervals by the country omitted.

Table 2: Credible interval coverage of validation by Country.

Country omitted	50% CI	80% CI	90% CI
Albania	66.0%	82.2%	87.3%
Bangladesh	66.0%	82.1%	87.1%
Colombia	66.0%	81.9%	86.9%
Ghana	66.4%	82.5%	87.2%
Philippines	66.4%	82.4%	87.2%

Figure 8: Model validation results of omitted countries



Similarly, Table 3 shows the coverage by percentages omitted.

Table 3: Credible interval coverage of validation by percentages omitted.

Percentage omitted	50% CI	80% CI	90% CI
5%	64.6%	80.9%	86.2%
10%	63.0%	79.8%	84.9%
15%	61.0%	78.2%	83.7%
20%	60.2%	76.8%	82.7%
30%	55.8%	73.6%	80.4%

In our results, we find that the credible interval coverage declines as more benchmark data are removed, which is expected because the model has less information to anchor the age pattern. When only a small share of the benchmark is removed (for instance, 5-10%), the 80% intervals contain about 80% of the held-out values. When a larger share is removed (30% in this case), coverage falls, particularly, the 90% credible intervals contain fewer than 90% of the held-out values in these tests (about 80-87% across Tables 2 to 3). These validation results imply that the model becomes less precise when the benchmark is sparse. These checks suggest the model can reproduce the UN benchmark ASFR schedule reasonably well when some benchmark values are missing, but it becomes less precise as more benchmark data are removed. Due to the unavailability of education-specific fertility rates for all the countries estimated, the validation of the EAFRs themselves was not conducted.

Additionally, we ran sensitivity analyses using different priors. These codes are available in the GitHub repository.

The sensitivity analysis comprised seven different prior models for the precision parameter τ_{cyae}^{eaf} . Table 4 shows the different prior distributions used in the sensitivity analysis.

Table 4: Different prior distributions for EAFR precision.

Model	Distribution	Source
Final model	$\tau_{cyae}^{eaf} \sim G(1/\sigma_e^{eaf}, 2\sigma_e^{eaf})$	standard deviations of standard errors by education level (GLM)
Model 1	$\tau \sim G(\alpha, \beta)$	relative variance, variance-mean ratio of standard errors (GLM)
Model 2	$\tau \sim G(1/\mu_e^{eaf}, \sigma_e^{eaf})$	mean and variance of standard errors by education level (GLM)
Model 3	$\tau \sim G(1/\sigma_e^{eaf}, \sigma_e^{eaf})$	variance of standard errors by education level (GLM)
Model 4	$\tau \sim G(1/\sigma_e^{eaf}, \sigma_e^{eaf})$	variance of standard errors (GLM)
Model 5	$\tau \sim G(1/\sigma_e^{eaf}, \sigma_e^{eaf})$	variance of estimates (GLM)
Model 6	$\tau_{cyae}^{eaf} \sim G(1/\sigma_{a,e}^{eaf}, \sigma_{a,e}^{eaf})$	variance of standard errors by age and education level (GLM)
Model 7	$\tau_{cyae}^{eaf} \sim G(1/\sigma_e^{eaf}, \sigma_e^{eaf})$	variance of standard deviation by education level (GLM)

We selected the final model because its estimated ASFRs align most closely with the UN WPP ASFRs among the prior specifications considered.

Usage Notes

The methodology described in this article is flexible and adaptable to reconstruct or estimate education-specific rates, or fertility rates in general, in regions or areas where data are sparse and/or come from multiple data sources, not limited to the DHS. For future studies, the estimates generated from this methodology can be downloaded freely for population projections by level of education, policy evaluation, and other such studies. While the study employed the 2022 revision rounds of the UN WPP, a comparison between the Bayesian ASFRs from this exercise and the 2024 rounds showed that about 85% of the UN WPP 2024 ASFR values (country-period-age combinations) fell within the 95% credible intervals of the current model estimates. In addition, 47.1% of the ASFR values fall within the 50% CI, 72.1% within the 80% CI, and 80.4% fell within the 90% CI.

Data Availability

The resulting outputs can be accessed on Zenodo⁶² and GitHub at <https://github.com/AfuaD-B/Bayesian-methodology-on-fertility-reconstruction> in the Excel workbook “BESFR_estimates_global_south.xlsx”. All other input data sources used to generate these outputs are described in the Methods section.

Code Availability

The codes used to fit the Bayesian models, run sensitivity and validation analyses, and generate the output files and figures are available at Zenodo (DOI:10.5281/zenodo.18359482)⁶² and mirrored on GitHub at <https://github.com/AfuaD-B/Bayesian-methodology-on-fertility-reconstruction>. All codes except the “tfr2” STATA module were run in RStudio software. The code for the final Bayesian model is in the file “Bayesian model global_south.R”. The validation code files are “Bayesian model global_south_validation_country_omission.R” and “Bayesian model global_south_validation.R”. The other models for the sensitivity analysis are in the code files, “glm input data global south model_alpha_beta.R” (Model 1 in Table 4), “glm input data global south model_edu_spec_mean.R” (Model 2 in Table 4), “glm input data global south model_edu_spec_sd.R” (Model 3 in Table 4), “glm input data global south model_gamma_47.R” (Model 4 in Table 4), “glm input data global south model_pred_prec.R” (Model 5 in Table 4), “glm input data global south model_se_sd_c_y_a.R” (Model 6 in Table 4) and “glm input data global south model_se_sd_c_y_e.R” (Model 7 in Table 4).

Funding

This work was supported by the Austrian Academy of Sciences' Innovation Fund Research, Science and Society(IF_2019_29). Open-access publication was supported in part by the Open-Access Fund of the Austrian Academy of Sciences.

Author Contributions

D.Y. originally developed an earlier version of the methodological framework. A.D.B. and D.Y. adapted and designed the current model and wrote the manuscript. A.D.B. performed model analysis and validation. A.G. revised and improved the manuscript.

Competing Interests

The authors declare no competing interests.

References

1. Notestein, F. W. Population-The long view. *Food for the World.*, 36–57 (1945).
2. Jejeebhoy, S. J. *Women's education, autonomy, and reproductive behaviour: Experience from developing countries* (Oxford University Press, 1995).
3. Sobotka, T. Post-transitional fertility: the role of childbearing postponement in fuelling the shift to low and unstable fertility levels. *Journal of Biosocial Science* **49**, S20–S45 (2017).
4. May, J. F. & Rotenberg, S. A Call for Better Integrated Policies to Accelerate the Fertility Decline in sub-Saharan Africa. *Studies in Family Planning* **51**, 193–204 (2020).
5. Ní Bhrolcháin, M. & Beaujouan, É. Fertility postponement is largely due to rising educational enrolment. *Population Studies* **66**, 311–327 (2012).
6. Durowaa-Boateng, A., Yildiz, D. & Goujon, A. A Bayesian model for the reconstruction of education-and age-specific fertility rates. *Demographic Research* **49**, 809–848 (2023).
7. Yildiz, D., Wiśniowski, A., Brzozowska, Z. & Durowaa-Boateng, A. A flexible model for reconstructing education-specific fertility rates: The case of sub-Saharan Africa. *Demographic Research* **54**, 371–404. <https://www.demographic-research.org/articles/volume/54/12/> (2026).
8. Schoumaker, B. D. & Sánchez-Páez, D. A. Disruptions in educational progress and fertility dynamics by educational level: Unraveling the link between education and fertility stalls in Sub-Saharan Africa. *Population and Development Review* **50**, 59–85 (2024).
9. Stevenson, A. J., Genadek, K. R., Yeatman, S., Mollborn, S. & Menken, J. A. The impact of contraceptive access on high school graduation. *Science Advances* **7**, eabf6732 (2021).
10. Lloyd, C. B. & Mensch, B. S. Marriage and childbirth as factors in dropping out from school: an analysis of DHS data from sub-Saharan Africa. *Population studies* **62**, 1–13 (2008).
11. United States Census Bureau. *Measuring fertility from a census: Select topics in international censuses* Feb. 2019.
12. *The DHS program* <https://dhsprogram.com/data/data-processing.cfm#:~:text=For%5C%20the%5C%20DHS%5C%20surveys%5C%2C%5C%20this,process%5C%20survey%5C%20and%5C%20census%5C%20data..>
13. *MICS Plus datasets and reports | UNICEF MICS* <https://mics.unicef.org/mics-plus/mics-plus-results>.
14. Khan, S. & Hancioglu, A. Multiple indicator cluster surveys: delivering robust data on children and women across the globe. *Studies in family planning* **50**, 279–286 (2019).
15. <https://www.ggp-i.org/generations-and-gender-survey/>.
16. <https://population.un.org/dataportal/home?df=8dd98658-ca8a-46ad-9690-07ff75759040>.

17. United Nations, Department of Economic and Social Affairs, Population Division. *World Population Prospects 2024: Methodology of the United Nations population estimates and projections* UN DESA/POP/2024/DC/NO (2024).
18. UN Economic Commission for Latin America and the Caribbean. *CEPALSTAT Portal de Datos y Publicaciones Estadísticas — statistics.cepal.org* <https://statistics.cepal.org/portal/cepalstat/dashboard.html?theme=1&lang=en>. [Accessed 25-09-2025]. 2025.
19. Eurostat. *Data Browser* https://doi.org/10.2908/DEMO_FAEDUC. [Accessed 25-09-2025].
20. *United States Census Bureau* <https://www.census.gov/data-tools/demo/idb/>.
21. Institute for Health Metrics and Evaluation. <https://www.healthdata.org/>.
22. KC, S. *et al.* Updating the shared socioeconomic pathways (SSPs) global population and human capital projections (2024).
23. Liu, B. *et al.* *The global, regional, and national burden of pancreatitis in 204 countries and territories, 1990–2021: A systematic analysis for the global burden of disease study 2021* (2024).
24. Max Planck Institute for Demographic Research (Germany) and Vienna Institute of Demography (Austria). *Human Fertility Database* <https://www.humanfertility.org/>.
25. Jasilioniene, A. *et al.* Data resource profile: the human fertility database. *International Journal of Epidemiology* **45**, 1077–1078e (2016).
26. Zeman, K., Brzozowska, Z., Sobotka, T., Beaujouan, E. & Matysiak, A. *Cohort Fertility and Education Database: Methods Protocol* Accessed 2025-09-25. 2017. <https://www.cfe-database.org> (2025).
27. *African Fertility Data Explorer* <https://populationafrica.org/>.
28. Schoumaker, B. *Quality and consistency of DHS fertility estimates, 1990 to 2012* (ICF International Rockville, Maryland, 2014).
29. Al Zalak, Z. & Goujon, A. *Assessment of the data quality in Demographic and Health Surveys in Egypt* tech. rep. (Vienna Institute of Demography Working Papers, 2017).
30. Raftery, A. E., Li, N., Ševčíková, H., Gerland, P. & Heilig, G. K. Bayesian probabilistic population projections for all countries. *Proceedings of the National Academy of Sciences* **109**, 13915–13921 (2012).
31. Alkema, L. *et al.* Probabilistic projections of the total fertility rate for all countries. *Demography* **48**, 815–839 (2011).
32. Alkema, L. & New, J. R. Global estimation of child mortality using a Bayesian B-spline bias-reduction model. *The Annals of Applied Statistics*, 2122–2149 (2014).
33. Raymer, J., Wiśniowski, A., Forster, J. J., Smith, P. W. & Bijak, J. Integrated modeling of European migration. *Journal of the American Statistical Association* **108**, 801–819 (2013).
34. Wiśniowski, A., Forster, J. J., Smith, P. W., Bijak, J. & Raymer, J. Integrated modelling of age and sex patterns of European migration. *Journal of the Royal Statistical Society Series A: Statistics in Society* **179**, 1007–1024 (2016).
35. Ellison, J., Dodd, E. & Forster, J. J. Forecasting of cohort fertility under a hierarchical Bayesian approach. *Journal of the Royal Statistical Society Series A: Statistics in Society* **183**, 829–856 (2020).
36. Bryant, J. R. & Graham, P. J. Bayesian demographic accounts: Subnational population estimation using multiple data sources (2013).
37. Dharamshi, A., Alexander, M., Winant, C. & Barbieri, M. Jointly estimating subnational mortality for multiple populations. *Demographic Research* **52**, 71–110 (2025).
38. Utazi, C. *et al.* A spatial regression model for the disaggregation of areal unit based data to high-resolution grids with application to vaccination coverage mapping. *Statistical methods in medical research* **28**, 3226–3241 (2019).
39. Bryant, J. & Zhang, J. L. *Bayesian demographic estimation and forecasting* (Chapman and Hall/CRC, 2018).
40. Alexander, M., Zagheni, E. & Barbieri, M. A flexible Bayesian model for estimating subnational mortality. *Demography* **54**, 2025–2041 (2017).
41. Greulich, A. & Toulemon, L. Measuring the educational gradient of period fertility in 28 European countries. *Demographic Research* **49**, 905–968 (2023).
42. Nisén, J. *et al.* Educational differences in cohort fertility across sub-national regions in Europe. *European Journal of Population* **37**, 263–295 (2021).
43. Tymicki, K. Cultural capital and fertility trends in Poland throughout the socio-economic transformation: estimation of TFR based on birth histories from a longitudinal survey. *Genus* **81**, 26 (2025).
44. Wittgenstein Centre Human Capital Data Explorer. *Wittgenstein Centre Human Capital Data Explorer* <https://dataexplorer.wittgensteincentre.org/wcde-v2/>.

45. ICF. Health Surveys Standard Recode Manual for DHS7. *The Demographic and Health Surveys Program* (2018).
46. United Nations Population Division. *wpp2022: World Population Prospects 2022* R package version 1.1-2 (2022). <http://population.un.org/wpp>.
47. Yildiz, D., Wiśniowski, A., Brzozowska, Z. & Durowaa-Boateng, A. *A flexible model to reconstruct education-specific fertility rates: Sub-Saharan Africa case study* (Zenodo, Nov. 2025). <https://doi.org/10.5281/zenodo.18185201>.
48. Jain, A. K. & Nag, M. Importance of female primary education for fertility reduction in India. *Economic and Political Weekly*, 1602–1608 (1986).
49. Cochrane, S. H. *Fertility and Education. What do we really know?* **26** (1979).
50. Nobles, J., Frankenberg, E. & Thomas, D. The effects of mortality on fertility: population dynamics after a natural disaster. *Demography* **52**, 15–38 (2015).
51. Agadjanian, V. & Prata, N. War, peace, and fertility in Angola. *Demography* **39**, 215–231 (2002).
52. Torrisi, O. Armed conflict and the timing of childbearing in Azerbaijan. *Population and Development Review* **46**, 501–556 (2020).
53. Urdal, H. & Che, C. P. War and gender inequalities in health: the impact of armed conflict on fertility and maternal mortality. *International Interactions* **39**, 489–510 (2013).
54. Mugisho, G. How conflict shapes the impact of infant and child mortality on fertility in the DR Congo. *Advances in Life Course Research*, 100692 (2025).
55. Verdiyeva, N. How the population of the Republic of Azerbaijan is ageing: causes and potential for social and economic development. *Population and Economics* **3**, 23–42 (2019).
56. Tabutin, D., Schoumaker, B. & Rabenoro, M. The Demography of Sub-Saharan Africa from the 1950s to the 2000s. *Population* **59**, 455–555 (2004).
57. Sokteang, S., Ou, P., Pirtea, P. & Ziegler, D. d. Fertility Rates in Cambodia: A Bumpy Road. *Fertility & Reproduction* **6**, 156–165 (2024).
58. Hukin, E. Cambodia's fertility transition: the dynamics of contemporary childbearing. *Population and Development Review* **40**, 605–628 (2014).
59. Heuveline, P. & Poch, B. The phoenix population: Demographic crisis and rebound in Cambodia. *Demography* **44**, 405–426 (2007).
60. Kiernan, B. The demography of genocide in Southeast Asia: The death tolls in Cambodia, 1975-79, and East Timor, 1975-80. *Critical Asian Studies* **35**, 585–597 (2003).
61. Bongaarts, J. Completing the fertility transition in the developing world: The role of educational differences and fertility preferences. *Population studies* **57**, 321–335 (2003).
62. Durowaa-Boateng, A., Yildiz, D. & Goujon, A. *Bayesian methodology on reconstructing education- and age-specific fertility rates* 2026. <https://doi.org/10.5281/zenodo.18359482>.