Summary

Educational attainment is a key social determinant of maternal and child health and is a large component of human capital[1–3]. Understanding relevant inequalities in educational attainment by area and between sexes is required to effectively pursue Sustainable Development Goal (SDG) 4: to ensure inclusive and equitable education for all. In addition to well-studied links between maternal education and child mortality[4–6], girls who achieve at least primary school have been shown to have higher earnings, increased agency and social capital, and increased health and well-being[7], making education a relevant tool for cross-cutting progress. Building on previous mapping of child health[8,9], and education[10], here we present subnational variation in educational attainment by predicting mean years of education and percent achievement of key schooling thresholds at a 5x5 km resolution for males and females in 121 low- and middle-income countries (LMICs) from 2000 to 2017. We show that while mean educational attainment has increased and disparities between sexes and geographies have improved over the duration, many women do not finish primary school, and a large proportion never start school. In many areas, this is despite men finishing school at a high rate. Pinpointing where this disadvantage is concentrated within countries serves as a tool for policy makers to understand where localized efforts to improve access to education should be implemented to change the landscape of opportunity, and its health implications, around the world.

Educational attainment, particularly for women of reproductive age, is a critical component of human capital and an important social determinant of maternal and child health[1,2]. Education is closely linked to several of the Sustainable Development Goals (SDGs) established by the United Nations in 2015[2]. Beyond SDG4, which explicitly focuses on education, a diverse body of research has documented the close association between increased schooling and improved gender equality (SDG5) and maternal and child health (SDG3)[3–5]. In a 2018 report focusing on the negative impacts of low educational attainment for girls, the World Bank stresses the multifaceted benefit of investment in education[6]. Drawing on data from a variety of low- and middle-income settings, numerous studies suggest that improving educational attainment boosts earnings and standards of living, reduces child marriage and early childbearing, and improves child health and well-being.

In 2016, aid to education reached its highest level since 2002, increasing by 13% since 2015[7]. Two-thirds of this increase was the result of a dramatic increase in aid to basic education. After years of de-prioritization in the global aid portfolio, the relative share of total aid attributable to education increased for the first time since 2009. Despite huge gaps in basic education being identified across Africa, however, the share of aid specifically allocated to countries in Sub-Saharan Africa continues to trend downward. Overall, only 22% of aid to basic education went to low-income countries in 2016 compared to 36% in 2002[8]. These trends suggest a continued problem with the distribution of aid not aligning with the distribution of need, even at the national level.

Policy and intervention research has called for an increased focus on mapping the relationships between the different SDGs, such as the effect of educational levels and equity on the goals related to child growth failure and mortality[9,10]. In determining risks to development during childhood, the educational attainment of mothers has been repeatedly identified as a critical determinant of the quality of child care, survival, and healthy growth trajectories[3–5,11–13]. This effect operates both on the quality of care sought and attained during pregnancy and post-partum, but also on the quality of care for children through the duration of breastfeeding and healthcare-seeking when children are ill[14,15]. Beyond an individual's education, a comprehensive multi-level study demonstrated that increases in the average educational attainment within communities can lead to improved nutrition and survival for all children in that community[16]. The relationship with health indicators is highly cyclical. Stunting, a common indicator of child growth failure, is linked to human capital formation through cognitive and learning outcomes, and new cohorts of children exposed to high rates of growth failure are less likely to attend and complete the level of schooling aspired to in SDG4[17]. Gender inequality is also recognized for its impact on child undernutrition through women's control of their time, household income, and mental health. UNESCO reports that around the world, girls are less likely than boys to attend and finish primary school[18]. Several studies from the World Bank also argue that gender inequality in education inhibits national economic growth[19–21]. This robust body of evidence, and a pattern of inequitable distribution of aid, suggests that precision estimates of educational attainment and inequality therein will provide a powerful tool for advocates and policymakers to achieve cross-cutting progress towards the SDGs.

Precision mapping and equity in education

The SDGs related to education, child mortality, and child nutrition are all framed around the importance of equity in progress towards targets across dimensions such as geography and gender[2,22]. Indicators such as under-5 mortality and stunting have seen dramatic improvement over the past few decades, but recent geospatial analyses find persistent disadvantage remains subnationally[23,24]. The global health agenda is increasingly focused on precision public health evidence illustrating the subnational distribution of disease and illness, but an agenda focused on equity for the future must integrate comparable evidence on the distribution of social determinants of health[24–26].

The present study seeks to expand on this body of precision evidence for benchmarking and targeting SDG programs by estimating the subnational distribution of educational attainment across all low- and middle-income countries (LMICs) from 2000-2017. Previous analyses have focused on geographic disparities in average attainment across Africa or for specific countries, but no analysis to our knowledge has examined the individual distribution of attainment across this panel at a high spatial resolution[27–29]. Synthesizing and geolocating subnational data sources on educational attainment, this analysis identifies precise inequalities across geography as well as within populations by estimating the proportion of men and women who have completed key levels of schooling. We further examine measures of gender parity to inform the SDG and UNESCO framework for gender equality in education.

Inequalities in attainment across region and gender

Figure 1 illustrates the mean years of attainment for men and women of reproductive age (15-49) in 2000 and 2017, as well as the difference between estimates across sex. Average educational attainment remains low in this age group across much of the Sahel region of Sub-Saharan Africa as shown previously[29], while we observe marked improvement across wide areas of South America and Asia, including India. In 2017 there is a large disparity between sexes in many regions, however, with men attaining higher average education across Central and Western Sub-Saharan Africa and South Asia. There also remains significant variation in 2017 between the highest- and lowest-performing administrative units within countries for average education among women of reproductive age. In Uganda, average attainment ranged from 1.1 (95% Uncertainty Interval 0.5-1.6) in rural Kotido to 10.9 (10.6-11.3) in Kampala, the capital city. A similar discrepancy is observed in Nigeria, where the average ranges from 1.9 (0.9-2.7) in Sokoto, a rural state in the northwest, to 10.7 (9.8-11.7) in Imo, an urban state in the southeast Niger Delta. By also examining the mean annualized rate of change in this indicator, we show that countries such as Ethiopia and Madagascar have experienced significant improvement in almost all subnational regions. However, the subnational discrepancies noted in 2017 within countries like Nigeria and Kenya have been very slow to change.

Figure 2 displays the proportion of men and women age 15-49 who have not completed primary school. By considering variation within population in different locations, these maps help to identify areas with large numbers of individuals in the vulnerable lower end of the attainment distribution. We observe large improvements in the proportion of men and women of reproductive age completing primary school in Mexico and China, but across much of the world women in this age group fail to complete primary school at a much higher rate than their male counterparts. There has been tremendous progress in subnational areas that had very high proportions failing to complete primary school in 2000. In the Xizang Zizhiqu province of China, part of the Tibetan Plateau, this proportion improved from 0.8 (0.8-0.8) in 2000 to 0.1 (0.1-0.3) in 2017. The Awdal province, the most westerly province of Somalia bordering Djibouti and Ethiopia, improved from 0.8 (0.7 – 1.0) to 0.4 (0.2-0.6). In the annualized rate of change we see extremely fast improvements since 2000 in southern sub-Saharan Africa and Western China. Across much of the world, trends in the proportion of men and women completing primary school have improved largely in parallel, meaning that the gap between genders on this indicator in 2000 is fairly similar to the gap observed in 2017.

Figure 3 further examines the population distribution by illustrating parity between genders (the ratio of proportions between men and women). This figure also highlights two additional advantages of this analytic framework. First, here we examine a smaller age group of 20-24 years. Though educational indicators in this age group are less directly relevant for maternal and child health than women in the full reproductive age range, this group allows us to capture how the landscape of education is shifting over time in successive cohorts and is thus more likely to pick up improvements to access and retention in education made since 2000. Second, we illustrate the probability that this estimated ratio is credibly different from 1 (parity between genders) given the full uncertainty estimated in our data and model (a more detailed explanation is provided in the Supplementary Information). We observe large variation in this indicator, with men achieving at a higher rate than women across much of the world (i.e. the probability that the parity ratio is greater than 1 is over 95%). This is largely true for both primary completion and secondary completion, but especially so in countries such as Burundi, Angola, Uganda, and Afghanistan.

In Figure 4 we summarize the number of women age 20-24 who have never attended school by multiplying our proportion estimates with high-resolution population estimates.[30] This provides a further perspective on the geographic distribution of this vulnerable population by focusing on the total individuals who have never attended school in an area rather than the rate. This map highlights the magnitude of women who aren't receiving any schooling and focuses on a younger population so as to more quickly pick up changes in successive generations due to new initiatives and programs to improve access to schooling. We find that many large subnational regions still have many women ages 20-24 receiving no education, especially in large population countries such as India, Ethiopia, and Nigeria. There is also significant concentration of this population geographically within these countries. In India, 45% of women receiving no education in 2016 live in Bihar, Uttar Pradesh, and Rajasthan. In Ethiopia, 61% of this population lives in Oromia or Amhara. While this total population is much smaller in other countries as a function of total population size, the geographic concentration can be extreme and is indicative of a specific type of marginalization. For example, in Cameroon 50% of young women in the country who are estimated to have completed no schooling in 2016 live in the Far North region. The same is true for the Southern region of Malawi (50%) and the Northern region of Ghana (40%).

Discussion and limitations

Here we present a comprehensive, comparable database of educational attainment estimates across all LMICs from 2000-2017. We have significantly built on previous modelling efforts that focused on the geographic distribution of education[29] by extending our estimation to the population distribution, highlighting not only average educational attainment but also disparities across the proportions of men and women completing specific levels of schooling that are central to policy efforts. By focusing on parity between new groups of young men and women over this period, our estimates will allow policy-makers and advocates to more closely track changes and whether improvements to education systems are being experienced equitably. As our estimates here demonstrate, throughout much of the world women still lag behind their male counterparts, and there is significant heterogeneity within subnational regions which may suggest unique social, economic, and cultural obstacles requiring further rigorous inquiry.

As we demonstrate in Figure 4, many young women across the world still face obstacles to attaining a basic level of education. Especially as larger proportions finish secondary school, as in South Africa, Peru, and Colombia, it will be important to focus on learning outcomes and quality of education. Many women across the world lack even basic attainment. The represents a critical missed opportunity for the global health agenda to reap the benefits of a well-studied determinant of maternal and child health. Even if there are only marginal returns to health in the short-term, particularly in places with extremely porous healthcare systems, at the very least communities will see on average increased social mobility, higher earnings, and less engagement in child marriage or early childbearing.

In terms of possible ways to intervene on these obstacles to basic attainment, a comprehensive 2018 World Bank report summarizes the most common reasons young girls drop out of school. These include poor learning outcomes and costs, failure at examinations, lack of nearby secondary school for rural communities, forced withdrawal from public school of married adolescents, never enrolling in school or enrolling late, and the influence of relatives concerning the demands on first daughters[30,3]. As stressed in the Commission on Social Determinants of Health, a critical step is acknowledging that commercialization in the area of education typically leads to higher inequity. Treating public education as a societal good by increasing access and removing fees, particularly in underserved rural communities, is a necessary step in the right direction, but generally there are no quick solutions. As the Commission states in their concluding report, that is the essence of investment in broader social determinants of health. Identifying areas that are stagnating or getting worse, particularly in the realm of basic education for young women across the world, is a first important step to targeted, long-term reform efforts that will ultimately have cross-cutting benefits within the SDG agenda for health and development.

Our analysis is not without several important limitations. First, it is extremely difficult to quantify quality of education on this scale in a comparable way. Quality is ultimately a large part of the SDG agenda and of utmost importance to achieving equity in opportunity for social mobility. However, many studies across diverse low- and middle-income settings have linked attainment, even very low levels, to measurable improvement in maternal and child health. As our analysis highlights with the proportional indicators, there are still many subnational regions across the world where large proportions do not complete primary school. A second limitation is that we are unable to measure or account for migration. A concept note released from the forthcoming GEMS Report 2019 focuses on how migration and displacement impacts schooling[33]. It's possible that geographic disparities reflect changes in population composition rather than changes in the underlying infrastructure or education system. Pathways for this change are complex and may be voluntary, such as those who do manage to receive an education in a low-attainment area having an increased ability to migrate and choosing to do so. They may also be involuntary, particularly in politically unstable areas where displacement may make geographic changes over time difficult to estimate. A shifting population composition is a general limitation of many longitudinal ecological analyses, but the spatially granular nature of the analyses employed here may be more sensitive to the effects of mobile populations. Lastly, our estimates cannot be seen as a replacement for proper surveillance systems, especially for tracking contemporaneous change. Our analysis of uncertainty at a high-resolution may be used to inform investment in more robust data systems and collection efforts, especially if the ultimate goal is to measure and track progress in the quality of schooling.

In 2008, the WHO Commission on Social Determinants of Health released their final report on what can be done globally to promote health equity[34]. The Commission had a large focus on early childhood development and conditions, noting the importance of education particularly among women of reproductive age. They also state that measurement of inequity within countries is critical to understanding and tracking the problem, and that geography is an increasingly important dimension of equity. Where people are born and raised greatly determines their life chances, and continuing to consider development and human capital formation on a national level is insufficient[34].

Looking forward to the next decade of the SDG agenda, it will be important to maintain the progress that has been made to reprioritize global investment in basic education systems. UNESCO notes that the recent increase in aid to education needs to be sustained for many years to make up for the stagnation since 2009. Despite recent progress, there also remains the problem of distributional accountability in aid, especially to basic education, where most funding is not going to the countries that need it most. Leaders in global health have noted the crucial need to invest in precise data systems and eliminate data gaps in order to continue effectively targeting resources, developing equitable policy, and tracking accountability.[35] Our analysis strives to provide the most robust evidence base possible for such decision-making and advocacy, as even if a country seems on track nationally there may be local communities that have seen no improvement. Decades of research on the role of basic education on maternal and child health outcomes in a diverse pool of low- and middle-income settings places this issue squarely in the purview of the global health agenda. Moving forward, it will be critical within the global health community to invest in long-term, sustainable improvement in the underlying distribution of human capital, as this is the only way to truly impact health equity across generations.

Methods

Overview

Employing a Bayesian model-based geostatistical framework and synthesizing geolocated data from 503 household and census datasets, this analysis provides 5x5 km estimates of mean years of education and proportion of the population attaining key levels of education for women of reproductive age (15-49), women age 20-24, and equivalent male age-bins between 2000-2016 in low- and middle-income countries. This includes 121 countries across all low- and middle-income countries. Countries were selected for inclusion in this analysis using the Socio-demographic Index (SDI) published in the GBD. The SDI is a measure of development that combines education, fertility, and poverty. All countries in the Middle, Lower-Middle, or Low SDI quintiles were included. Albania, Bosnia, and Moldova were excluded despite Middle SDI status due to geographic discontinuity with other included countries and lack of available survey data. Libya, Malaysia, Panama, and Turkmenistan were included despite Higher-Middle SDI status to create better geographic continuity. We do not estimate for American Samoa, Federated States of Micronesia, Fiji, Kiribati, Marshall Islands, Samoa, Solomon Islands, or Tonga, where no available survey data could be sourced. Analytical steps are described below and additional detail can be found in the Supplementary Information.

Data

We compiled a database of survey and census datasets in Africa that contained geocoding of subnational administrative boundaries or precise coordinates for sampled clusters. These included datasets from 503 sources (see Supplementary Table 2). We extracted demographic, education, and sample design variables. The coding of educational attainment varies across survey families. In many surveys, respondents can indicate their level of attainment on a continuous year scale. In others, respondents may only have several aggregate categories such as “Secondary completion”, “Primary completion”, or “less than primary”. When all that is known is that an individual completed a particular level of education, but it is not known if they continued onto the next level, a theoretical level of completion must be assigned to the individual in order to estimate summary statistics for the population such as mean years of educational attainment. For example, if the option “Primary completion” (6 years) is followed by “Secondary completion” (12 years), it can be assumed that an individual who selects the former has attained between 6 and 12 years of education. In previous literature examining trends in mean years of education, the assumption is made that all of these individuals have 6 years, or sometimes the midpoint of the feasible range (9)[32,3]. Trends in the single-year data demonstrate that this assumption introduces compositional bias in the estimation of attainment trends over time and space, as differences in true drop-out patterns or binning schema could lead to biased mean estimates.

For this analysis, we employed a recently developed method that selects a training subset of similar surveys across time and space to estimate the true single-year distribution of binned datasets[3]. This algorithmic approach significantly reduces bias in summary statistics estimated from datasets with binned coding schemes. The years in all coding schemes were mapped to the country- and year-specific references in the UNESCO International Standard Classification of Education (ISCED) for comparability[39]. We used a top coding of 18 years on all data; this is a common threshold in many surveys that have a cap and it is reasonable to assume that the importance of education for health outcomes (and other related SDGs) greatly diminishes after what is the equivalent of 2 to 3 years of graduate education in most systems.

Data were aggregated to mean years of education attained and the proportions achieving key levels of education. The levels chosen were proportion with zero years, proportion with less than primary school (1-5 years), proportion with at least primary school (6 – 11 years), and proportion achieving secondary school or higher (12 or more years). A subset of the data for a smaller age-bin (20-24) were also examined to more closely track temporal shifts. Equivalent age-bins were aggregated for both women and men in order to examine disparities in mean years of attainment by sex. Where precise coordinates were available, data were aggregated to a specific latitude/longitude assuming a simple-random-sample, as the cluster is the primary sampling unit for the stratified design survey families such as DHS and MICS. Where only geography information was available at the level of administrative units, data were aggregated according to their sample design. For aggregation to administrative units for which the survey was not sampled to be representative, design effects were re-estimated using a package for analyzing complex survey data in R[3].

Spatial covariates

In order to leverage strength from locations with observations to the entire spatial-temporal domain, we compiled several 5x5 km raster layers of possible socioeconomic and environmental correlates of education in Africa (see Supplementary Table 3 and Figure 5). Acquisition of temporally dynamic datasets, where possible, was prioritized in order to best match our observations and thus predict the changing dynamics of educational attainment. Of the 9 covariates included, 6 were temporally dynamic. The remaining 3 covariate layers were temporally static and were applied uniformly across all modelling years. More information, including plots of all covariates, can be found in the Supplementary Information.

Our primary goal is to provide educational attainment predictions across LMICs at a high (local) resolution and we have used methods to provide the best out-of-sample predictive performance at the sacrifice of inferential understanding. In order to select covariates and capture possible non-linear effects and complex interactions between them, an ensemble covariate modeling method was implemented[3]. For each region, three sub-models were fit to our dataset using all of our covariate data as explanatory predictors: generalized additive models, boosted regression trees, and lasso regression. Each sub-model was fit using five-fold cross validation to avoid overfitting and the out-of-sample predictions from across the five holdouts are compiled into a single comprehensive set of predictions from that model. Additionally, the same sub-models were also run using 100% of the data and a full set of in-sample predictions were created. The five sets of out-of-sample sub-model predictions were fed into the full geostatistical model as the explanatory covariates when performing the model fit. The in-sample predictions from the sub-models were used as the covariates when generating predictions using the fitted full geostatistical model. This methodology maximizes out-of-sample predictive performance at the expense of no longer being able to provide statistical inference on the relationships between the predictors and the outcome. A recent study has shown that this ensemble approach can improve predictive validity by up to 25% over an individual model[41]. More details on this approach can be found in the Supplementary Information.

Analysis

Geostatistical model

Gaussian and binomial data are modeled within a Bayesian hierarchical modeling framework using a spatially and temporally explicit hierarchical generalized linear regression model to fit mean years of education attainment and proportion of population achieving key bins of school in 14 regions across all LMICs as defined in the GBD study (see Extended Data Figure 3)[3]. This means we fit 14 independent models for each indicator (i.e., the proportion of women with zero years of schooling). GBD study design sought to create regions on the basis of two primary criteria: epidemiological homogeneity and geographic contiguity[42]. For each indicator (mean attainment, edu_i as Gaussian or the number of individuals attaining a certain level, C_i, as binomial) and GBD region we approximated the posterior distribution of our Bayesian model:

$$edu_i | \mu_i, \tau_i, s_i \sim Gaussian(\mu_i, \tau_i, s_i)$$

$$f_{edu_i|\mu_i,\tau_i,s_i}(edu_i) = \frac{\sqrt{\tau\, s_i}}{\sqrt{2\pi}} \exp(-\frac{1}{2}\tau\, s_i (edu_i - \mu_i)^2)$$

$$\mu_i = \beta_0 + \boldsymbol{X_i \beta} + \epsilon_{GP_i} + \epsilon_{ctry_i}\, \epsilon_i$$

$$C_i | p_i,\, N_i \sim Binomial(p_i,\, N_i)$$

$$logit(p_i) = \beta_0 + \boldsymbol{\beta X_i} + \epsilon_{GP_i} + \epsilon_{ctry_i} + \epsilon_i$$

$$\sum \beta = 1$$

$$\epsilon_i \sim N(0, \sigma^2_{nug})$$

$$\epsilon_{GP} | \Sigma_{space}, \Sigma_{time} \sim GP(0, \Sigma_{space} \otimes \Sigma_{time})$$

$$\Sigma_{space} = \frac{2^{1-v}}{\tau \times \Gamma(v)} \times (\kappa \boldsymbol{D})^v \times \mathrm{K}_v(\kappa \boldsymbol{D})$$

$$\Sigma_{time_{j,k}} = \rho^{|t_k - t_j|}$$

We model the mean years of attainment at cluster i as Gaussian data given precision τ and a fixed scaling parameter s_i. To account for the ordinal data structure of the binomial indicators and ensure all proportions sum to 1, we used a continuation-ratio modelling approach[43]. To do this, the proportion of population with zero years of education was modelled using a binomial model. The proportion with less than primary education was modelling as those with less than primary education of those that have more than zero years of education. The same method followed for the proportion of population completing primary education. The proportion achieving secondary school or higher was estimated as the complement of sum of the three binomial models. For each of three binomial indicators and for 14 regions, we modelled the number of people at cluster *i*, among a sample size, $\boldsymbol{N_i}$, who are subject to the indicator as binomial count data. We use the sample size in each cluster as our scaling parameter. We have suppressed the notation, but the means, $edui$, scaling parameters, si, predictions from the three submodels $\boldsymbol{Xi}$, and residual terms $\epsilon*$ are all indexed at a space-time coordinate. The means, $edui$

represent an individual's expected educational attainment given that they live at that particular location. Mean attainment and logit proportional attainment were modeled as a linear combination of the three sub-models (GAM, BRT, and lasso), $\boldsymbol{Xi}$, a correlated spatiotemporal error term, ϵGPi, and an independent nugget effect, ϵi. Coefficients, $\boldsymbol{\beta}$, on the sub-models represent their respective predictive weighting in the mean, while the joint error term, ϵGP, accounts for residual spatiotemporal autocorrelation between individual data points that remains after accounting for the predictive effect of the sub-model covariates and the nugget, ϵi, an independent error term. The residuals, ϵGP, are modeled as a three-dimensional Gaussian process in space-time centered at zero and with a covariance matrix constructed from a Kroenecker product of spatial and temporal covariance kernels. The spatial covariance, Σspace, is modeled using an isotropic and stationary Matérn function[3], and temporal covariance, Σtime, as an annual autoregressive (AR1) function over the 16 years represented in the model. This approach leveraged the data's residual correlation structure to more accurately predict prevalence estimates for locations with no data, while also propagating the dependence in the data through to uncertainty estimates[45]. The posterior distributions were fit using computationally efficient and accurate approximations in R-INLA (integrated nested Laplace approximation) with the stochastic partial differential equations (SPDE) approximation to the Gaussian process residuals[46]. Pixel-level uncertainty intervals (UIs) were generated from 1,000 draws (i.e., statistically plausible candidate maps)[47] created from the posterior-estimated distributions of modelled parameters.

To transform pixel level estimates into a range of information useful to a wide constituency of potential users, these estimates were aggregated from the 1,000 candidate maps up to district, provincial, and national levels using 5x5 km population data[4]. This aggregation also allowed for calibration of estimates to national GBD estimates for 2000-2016. This was achieved by calculating the ratio of the posterior mean national-level estimate from each candidate map draw in the analysis to the posterior mean national estimates from GBD, and then multiplying each cell in the posterior sample by this ratio. This method also enabled incorporating the calibration into the pixel level uncertainties and thus to the uncertainties at the different levels of aggregation. The median raking factors for mean attainment in women 15-49, men 15-49, women 20-24, and men 20-24 were 0.996 (interquartile range: 0.928-1.078), 0.973 (IQR: 0.900-1.032), 1.024 (interquartile range: 0.969-1.113), 1.028 (IQR: 0.973-1.090) respectively, indicating close agreement with GBD estimates. Scatter plots comparing national level estimates from this analysis with GBD estimates can be found in Supplementary Figures 24-27.

Although the model can predict at all locations covered by available raster covariates, all final model outputs for which land cover was classified as "barren or sparsely vegetated" were masked, on the basis of the most recently available MODIS satellite data (2013), as well as areas where the total population density was less than ten individuals per 1x1 km pixel in 2015. This step has led to improved understanding when communicating with data specialists and policy makers.

Model validation

Models were validated using spatially-stratified five-fold out-of-sample cross validation. In order to offer a more stringent analysis by respecting some of the spatial correlation in the data, holdout sets were created by combining sets of spatially contiguous data. Validation was performed by calculating bias (mean error), total variance (root-mean-square error), and 95% data coverage within prediction intervals, and correlation between observed data and predictions. All validation metrics were calculated on the out-of-sample predictions from the five-fold cross-validation. Where possible, estimates from these models were compared against other existing estimates. Furthermore, measures of spatial and temporal autocorrelation pre- and post-modeling were examined to verify correct recognition, fitting, and accounting for the complex spatiotemporal correlation structure in the data. All validation procedures and corresponding results are provided in the Supplementary Information.

Code Availability

Our study follows the Guidelines for Accurate and Transparent Health Estimates Reporting (GATHER). All code used for these analyses is available online at http://ghdx.healthdata.org/.

Data Availability

The findings of this study are supported by data available in public online repositories, data publicly available upon request of the data provider, and data not publicly available due to restrictions by the data provider, which were used under license for the current study, but may be available from the authors upon reasonable request and permission of the data provider. A detailed table of data sources and availability can be found in Supplementary Table 2.

Administrative boundaries were retrieved from the Global Administrative Unit Layers (GAUL) dataset, implemented by FAO within the CountrySTAT and Agricultural Market Information System (AMIS) projects[48]. Land cover was retrieved from the online Data Pool, courtesy of the NASA EOSDIS Land Processes Distributed Active Archive Center (LP DAAC), USGS/Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota[49]. Lakes were retrieved from the Global Lakes and Wetlands Database (GLWD), courtesy of the World Wildlife Fund and the Center for Environmental Systems Research, University of Kassel[50,51]. Populations were retrieved from WorldPop[30,52].

References

1. UNESCO. *Global Education Monitoring Report*. (2016).
2. United Nations (UN). *Transforming our World: the 2030 Agenda for Sustainable Development*. (2015).
3. Caldwell, J. C. How is greater maternal education translated into lower child mortality? *Source Heal. Transit. Rev.* **4,** 224–229 (1994).
4. Caldwell, J. C. Education as a Factor in Mortality Decline An Examination of Nigerian Data. *Popul. Stud. (NY).* **33,** 395 (1979).
5. Gakidou, E., Cowling, K., Lozano, R. & Murray, C. J. Increased educational attainment and its effect on child mortality in 175 countries between 1970 and 2009: A systematic analysis. *Lancet* **376,** 959–974 (2010).
6. Wodon, Q., Montenegro, C., Nguyen, H. & Onagoruwa, A. *Missed opportunities: The high cost of not educating girls*. (2018).
7. UNESCO. *UNESCO OPERATIONAL DEFINITION OF BASIC EDUCATION Thematic Framework*. (2007).
8. Report, U. U. I. for S. G. E. M. Aid to education: a return to growth? (2018).
9. Nilsson, M., Griggs, D. & Visbeck, M. Policy: Map the interactions between Sustainable Development Goals. *Nature* **534,** 320–322 (2016).
10. DiGirolamo, A. M., Stansbery, P. & Lung'aho, M. Advantages and challenges of integration: opportunities for integrating early childhood development and nutrition programming. *Ann. N. Y. Acad. Sci.* **1308,** 46–53 (2014).
11. Jejeebhoy, S. J. *Women's education, autonomy, and reproductive behaviour : experience from developing countries*. (Clarendon Press, 1995).
12. Desai, S. & Alva, S. Maternal Education and Child Health: Is There a Strong Causal Relationship? *Demography* **35,** 71 (1998).
13. Basu, A. M. & Stephenson, R. Low levels of maternal education and the proximate determinants of childhood mortality: a little learning is not a dangerous thing. *Soc. Sci. Med.* **60,** 2011–2023 (2005).
14. Ruel, M. T. & Alderman, H. Nutrition-sensitive interventions and programmes: how can they help to accelerate progress in improving maternal and child nutrition? *Lancet* **382,** 536–551 (2013).
15. LeVine, R. A., LeVine, S., Schnell-Anzola, B., Rowe, M. L. & Dexter, E. *Literacy and MotheringHow Women's Schooling Changes the Lives of the World's Children*. (Oxford University Press, 2012). doi:10.1093/acprof:oso/9780195309829.001.0001
16. Pamuk, E., Fuchs, R. & Lutz, W. Cmparing Relative Effects of Education and Economic Resources on Infant Mortality in Developing Countries. *Popul. Dev. Rev.* **37,** 637–664 (2011).
17. Perkins, J. M. *et al.* Understanding the association between stunting and child development in low- and middle-income countries: Next steps for research and intervention. *Soc. Sci. Med.* **193,** 101–109 (2017).
18. Report, U. U. I. for S. G. E. M. Meeting our commitments to gender equality in education. *Glob. Educ. Monit. Rep. 2017/18*
19. Klasen, S. Does Gender Inequality Reduce Growth and Development? Evidence from Cross-Country Regressions. (2000).
20. Klasen, S. Low Schooling for Girls, Slower Growth for All? Cross-Country Evidence on the Effect of Gender Inequality in Education on Economic Development. *World Bank Econ. Rev.* **16,** 345–373 (2002).
21. Klasen, S. & Lamanna, F. The Impact of Gender Inequality in Education and Employment on Economic Growth: New Evidence for a Panel of Countries. *Fem. Econ.* **15,** 91–132 (2009).

22. Horton, R. & Lo, S. Nutrition: a quintessential sustainable development goal. *Lancet* **382,** 371–372 (2013).
23. Osgood-Zimmerman, A. *et al.* Mapping child growth failure in Africa between 2000 and 2015. *Nature* **555,** 41–47 (2018).
24. Golding, N. *et al.* Mapping under-5 and neonatal mortality in Africa, 2000–15: a baseline analysis for the Sustainable Development Goals. *Lancet* **390,** 2171–2182 (2017).
25. Dowell, S. F., Blazes, D. & Desmond-Hellmann, S. Four steps to precision public health. *Nature* **540,** 189–191 (2016).
26. Osgood-Zimmerman, A. & Millear, A. Mapping child growth failure in Africa between 2000 and 2015. *Under Rev.* (2017).
27. Bosco, C. *et al.* Exploring the high-resolution mapping of gender-disaggregated development indicators. *J. R. Soc. Interface* **14,** (2017).
28. Roberts, D. A. *et al.* Benchmarking health system performance across regions in Uganda: a systematic analysis of levels and trends in key maternal and child health interventions, 1990-2011. *BMC Med.* **13,** 285 (2015).
29. Graetz, N. *et al.* Mapping local variation in educational attainment across Africa. *Nature* **555,** 48–53 (2018).
30. Tatem, A. J. WorldPop, open data for spatial demography. *Sci. Data* **4,** 170004 (2017).
31. Perlman, D., Adamu, F. & Wodon, Q. Why do adolescent girls drop out of school in Niger? A combined quantitative and qualitative analysis. *Marches Organ.* **32,** 179–94 (2018).
32. Perlman, D., Adamu, F. & Wodon, Q. *Understanding and ending child marriage: Insights from Hausa communities*. *Girls' Education and Child Marriage in West and Central Africa Notes Series* **Washington,** (2018).
33. Global Education Monitoring Report. *Concept note for the 2019 Global Education Monitoring Report on Education and Migration*. (2017).
34. Marmot, M., Friel, S., Bell, R., Houweling, T. A. & Taylor, S. Closing the gap in a generation: health equity through action on the social determinants of health. *Lancet* **372,** 1661–1669 (2008).
35. Annan, K. Data can help to end malnutrition across Africa. *Nature* **555,** 7–7 (2018).
36. Barro, R. & Lee, J.-W. *International Comparisons of Educational Attainment*. (1993). doi:10.3386/w4349
37. Barro, R. & Lee, J.-W. *A New Data Set of Educational Attainment in the World, 1950-2010*. (2010). doi:10.3386/w15902
38. Friedman, J., Graetz, N. & Gakidou, E. Improving the estimation of educational attainment: new methods for assessing average years of schooling from binned data. *Under Rev.* (2017).
39. UNESCO. ISCED Mappings. (2016). Available at: http://uis.unesco.org/en/isced-mappings. (Accessed: 2nd August 2016)
40. Thomas Lumley. Survey analysis in R. (2014).
41. Bhatt, S. *et al.* Improved prediction accuracy for disease risk mapping using Gaussian Process stacked generalisation. *J. R. Soc. Interface* **14,** (2017).
42. Murray, C. J. *et al.* GBD 2010: design, definitions, and metrics. *Lancet* **380,** 2063–2066 (2012).
43. Hosmer DW, Lemeshow S, S. R. Ordinal Logistic Regression Models. in *Applied logistic regression* 289–305 (Wiley, 2013).
44. Stein, M. L. *Interpolation of Spatial Data*. (Springer New York, 1999).
45. Waller, L. & Carlin, B. Disease Mapping. in *Handbook of Spatial Statistics* (eds. Gelfand, A., Diggle, P., Fuentes, M. & Guttorp, P.) 217–243 (CRC Press, 2010).
46. Rue, H., Martino, S. & Chopin, N. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J. R. Stat. Soc. Ser. B (Statistical Methodol.* **71,** 319–392 (2009).

47. Patil, A. P., Gething, P. W., Piel, F. B. & Hay, S. I. Bayesian geostatistics in health cartography: The perspective of malaria. *Trends Parasitol.* **27,** 246–253 (2011).
48. FAO-UN. The Global Administrative Unit Layers (GAUL). (2015). Available at: http://www.fao.org/geonetwork/srv/en/metadata.show?id=12691. (Accessed: 1st June 2017)
49. DAAC, N. L. Combined MODIS 5.1. Available at: https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table/mcd12q1. (Accessed: 1st June 2017)
50. World Wildlife Fund & Center for Environmental Systems Research, University of Kassel, G. Global Lakes and Wetlands Database, Level 3. (2004). Available at: https://www.worldwildlife.org/pages/global-lakes-and-wetlands-database. (Accessed: 1st June 2017)
51. Lehner, B. & Döll, P. Development and validation of a global database of lakes, reservoirs and wetlands. *J. Hydrol.* **296,** 1–22 (2004).
52. World Pop. Get data. Available at: http://www.worldpop.org.uk/data/get_data/. (Accessed: 7th July 2017)

End Notes

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was primarily supported by grant OPP1132415 from the Bill & Melinda Gates Foundation.

Author contributions S.I.H. and N.G. conceived and planned the study. K.W. and J.H. extracted, processed, and geo-positioned these data. L.W. and N.G. carried out the statistical analyses. All authors provided intellectual inputs into aspects of this study. N.G., L.W., J.H., and L.E. prepared figures and tables. N.G. wrote the first draft of the manuscript, and all authors contributed to subsequent revisions.

Author information The authors declare no competing financial interests.

Figures

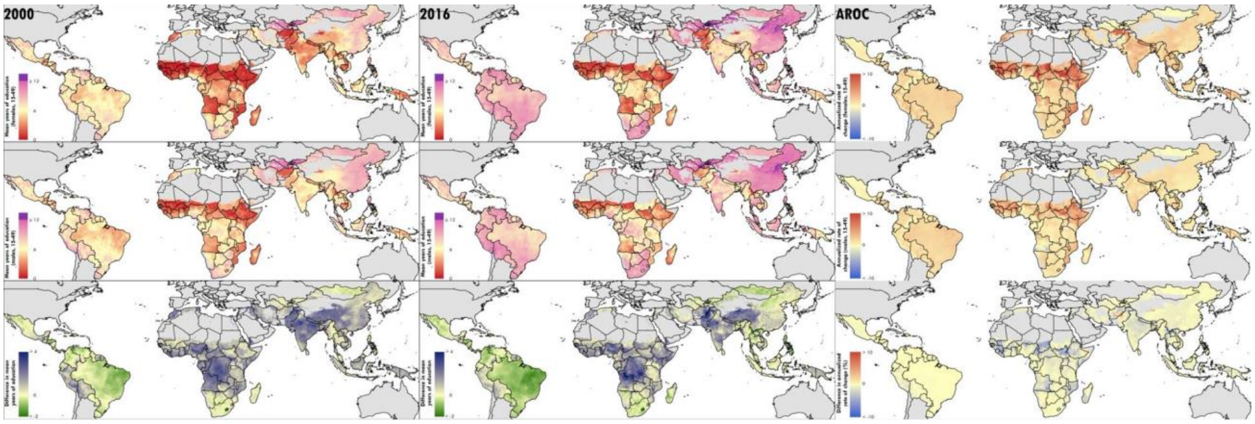

Figure 1. **Average educational attainment for and absolute difference between women and men aged 15-49 in 2000 and 2017.**

a-d, Average educational attainment for women (**a, b**) and men (**c, d**) aged 15–49 in 2000 (**a, c**) and 2017 (**b, d**). **e, f**, The absolute difference in average educational attainment between men and women aged 15-49 in 2000 (**e**) and 2017 (**f**). Maps reflect administrative boundaries, land cover, lakes and population; pixels with fewer than ten people per 1 × 1 km and classified as “barren or sparsely vegetated” are colored in grey.

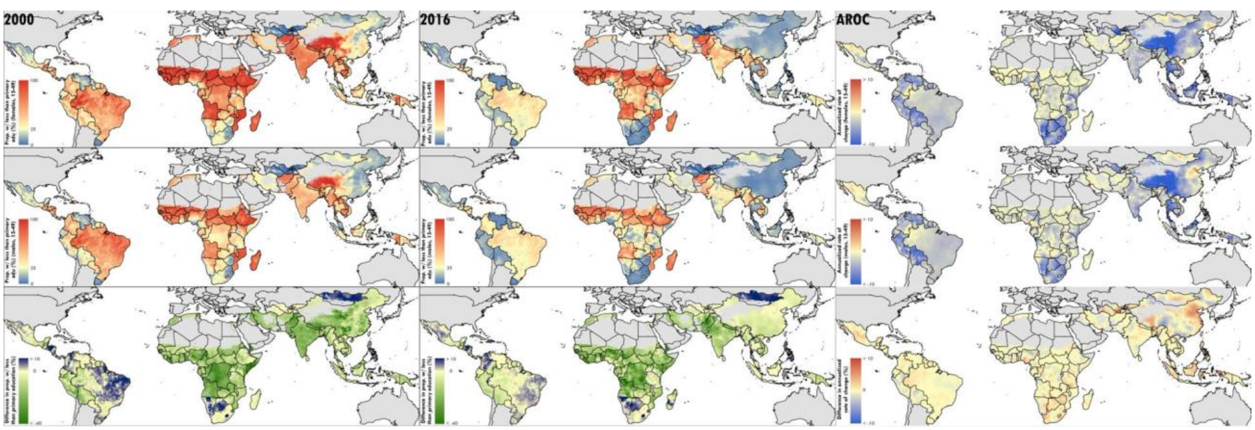

Figure 2. Proportion with no primary school and difference in proportions between women and men aged 15-49 in 2000 and 2017.

a–d, Proportion with no primary school for women (**a, b**) and men (**c, d**) aged 15–49 in 2000 (**a, c**) and 2017 (**b, d**). **e, f**, The absolute difference in proportion with no primary school between men and women aged 15–49 in 2000 (**e**) and 2017 (**f**). Maps reflect administrative boundaries, land cover, lakes and population; pixels with fewer than ten people per 1 × 1 km and classified as “barren or sparsely vegetated” are colored in grey.

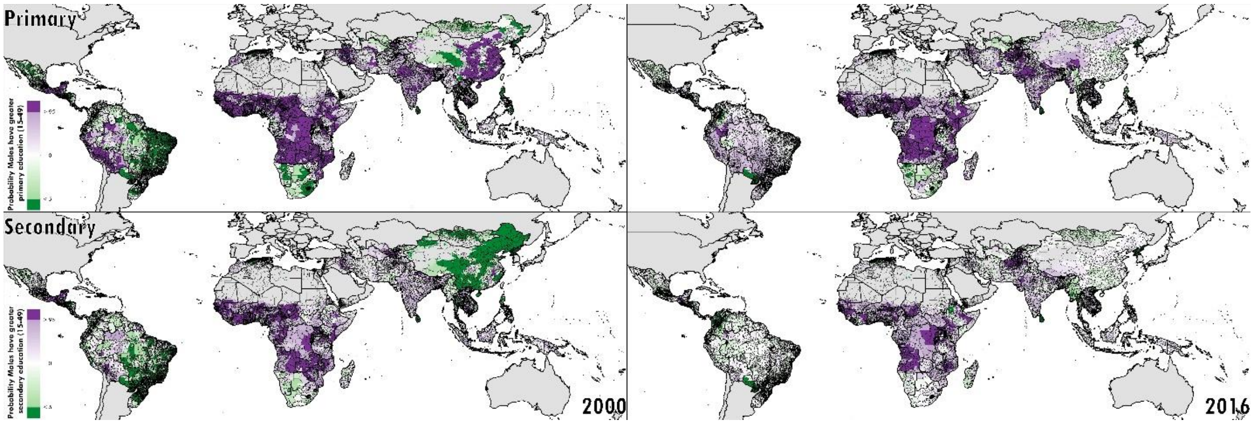

Figure 3. Probability that ratio of men and women aged 20-24 attaining primary and secondary education is >1 in 2000 and 2017.

a–d, Ratio of men to women aged 15-49 attaining primary education (**a, b**) and secondary education (**c, d**) in 2000 (**a, c**) and 2017 (**b, d**). Maps reflect administrative boundaries, land cover, lakes and population; pixels with fewer than ten people per 1 × 1 km and classified as "barren or sparsely vegetated" are colored in grey.

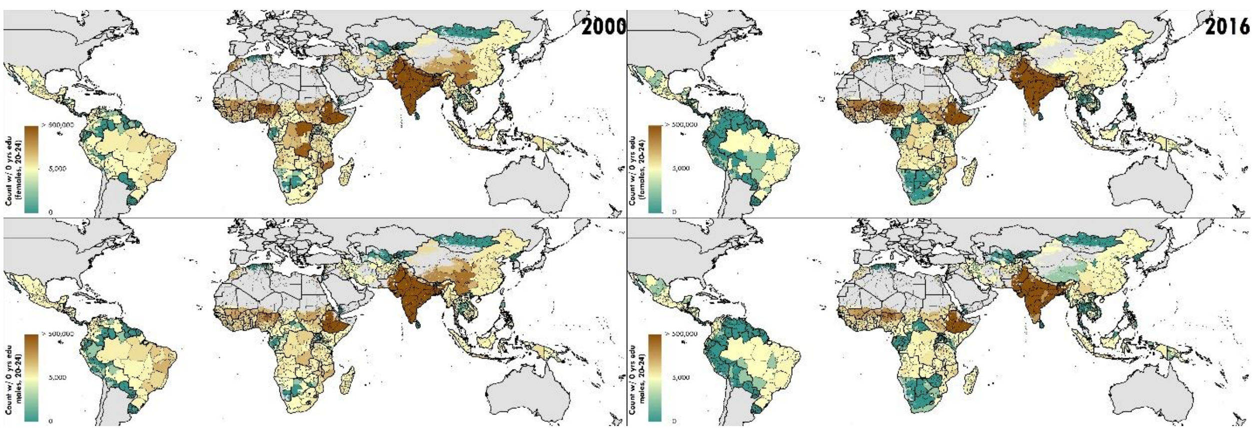

Figure 4: Number of individuals with no primary education by first administrative units for men and women aged 20-24 in 2000 and 2017, male/female, sex difference.

a–d, Number of individuals aged 20-24 with no primary education among women (**a, b**) and men (**c, d**) in 2000 (**a, c**) and 2017 (**b, d**). Maps reflect administrative boundaries, land cover, lakes and population; pixels with fewer than ten people per 1 × 1 km and classified as "barren or sparsely vegetated" are colored in grey.

Extended Data

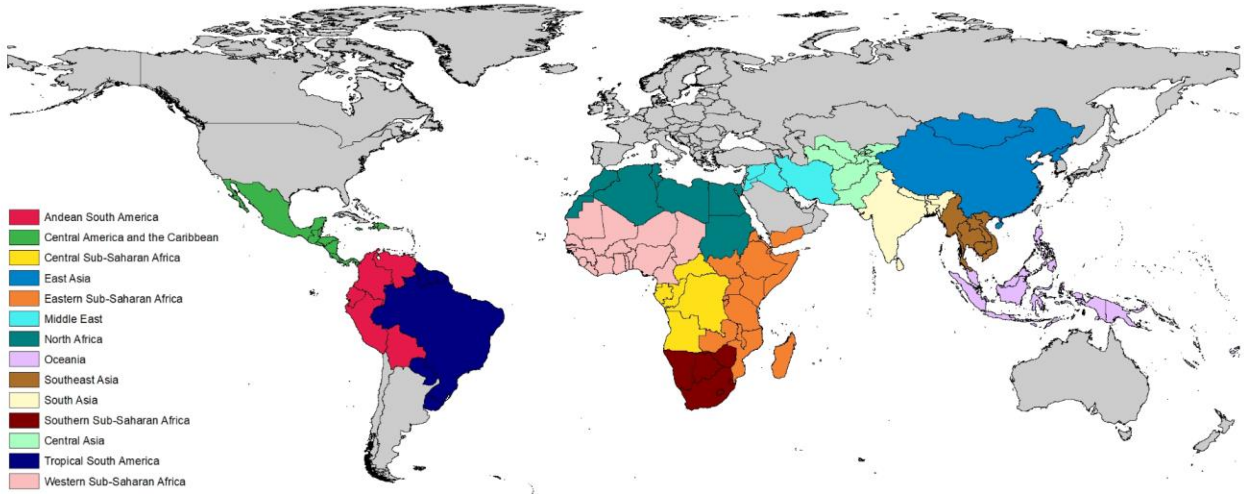

Figure 1: Modelling regions based on geographical and Socio-Demographic Index (SDI) regions from the Global Burden of Disease.
Minor changes were made to preserve geographic contiguity.

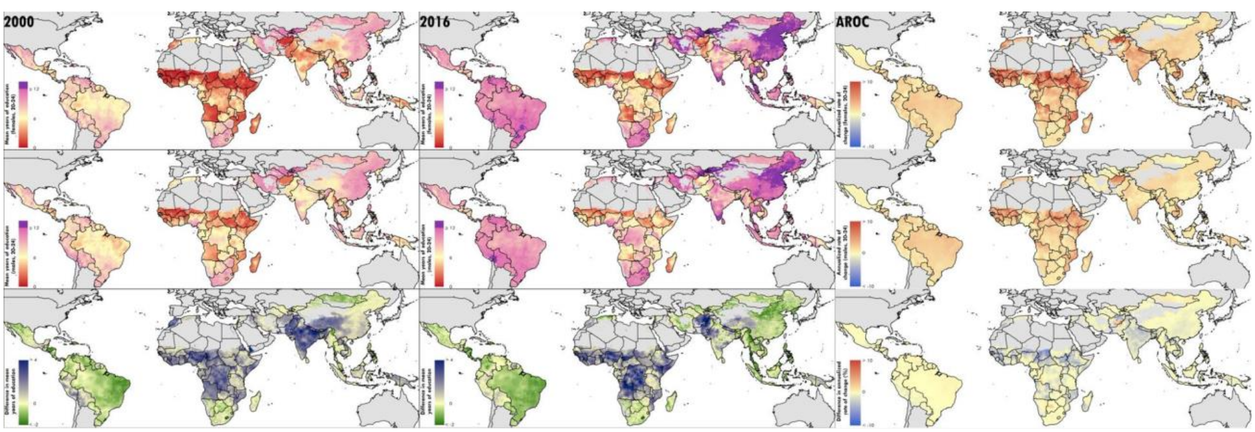

Figure 2. **Average educational attainment for and absolute difference between women and men aged 20-24 in 2000 and 2017.**
a-d, Average educational attainment for women (**a, b**) and men (**c, d**) aged 20-24 in 2000 (**a, c**) and 2017 (**b, d**). **e, f**, The absolute difference in average educational attainment between men and women aged 15-49 in 2000 (**e**) and 2017 (**f**). Maps reflect administrative boundaries, land cover, lakes and population; pixels with fewer than ten people per 1 × 1 km and classified as "barren or sparsely vegetated" are colored in grey. Interactive visualization tools containing all results are available at https://vizhub.healthdata.org/lbd/education.

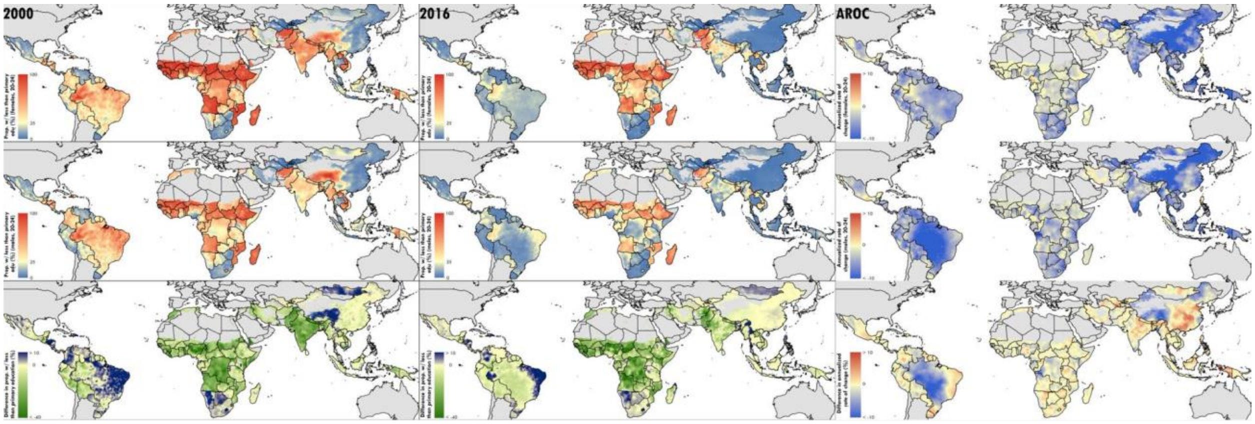

Figure 3. Proportion with no primary school and difference in proportions between women and men aged 20-24 in 2000 and 2017.

a–d, Proportion with no primary school for women (**a, b**) and men (**c, d**) aged 20–24 in 2000 (**a, c**) and 2017 (**b, d**). **e, f**, The absolute difference in proportion with no primary school between men and women aged 15–49 in 2000 (**e**) and 2017 (**f**). Maps reflect administrative boundaries, land cover, lakes and population; pixels with fewer than ten people per 1 × 1 km and classified as "barren or sparsely vegetated" are colored in grey.