Mapping child health: statistical applications for high-resolution estimation of child mortality and healthcare utilization

Roy Burstein

A dissertation

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2018

Reading Committee:

Simon I Hay, Chair

Emmanuela Gakidou

Haidong Wang

Robert C Reiner

Program Authorized to Offer Degree:

Global Health

University of Washington

Abstract

Mapping child health: statistical applications for high-resolution estimation of child mortality and healthcare utilization

Chair of the Supervisory Committee:

Professor Simon Iain Hay, DPhil

Global Health

This dissertation is comprised of three distinct chapters which each aim to improve understanding of child health and survival in developing countries through novel applications of statistical modelling. In each case, modelling was used to make estimates at more refined levels than previously existed, either geographically or in age categories, with the goal of improving the evidence base upon which trends and inequalities in child health could be monitored and assessed.

In the first chapter, *Mapping under-5 and neonatal mortality in Africa, a baseline analysis for the Sustainable Development Goals*, a geostatistical model was developed to estimate the child mortality rate for each 5 x 5 kilometer piece of land in 46 countries in Africa, from 2000 to 2015. The study utilized geographically referenced data from 235 household surveys and censuses and a suite of geospatially disaggregated covariates to make high resolution estimates. Resultant estimates on a gridded surface were also aggregated to the district and province level in each country, to provide a full range of estimates at useful spatial resolutions. Despite large declines in mortality rates overall during the study period, there was substantial heterogeneity in both the absolute levels and rates of change in child mortality, both across and within countries. While declines in many areas met or surpassed the 2015 UN Millennium Development Goal for a 4.4% annual reduction, the hard threshold of 25 deaths per 1000 livebirths set for the 2030 Sustainable Development goal would require much of Africa to reduce mortality at unprecedented rate during the coming years.

The second chapter, *Development and validation of a new method for indirect estimation of neonatal, infant, and under-5 mortality using summary birth histories,* dealt with the issue arising in child mortality estimation wherein the dominant source of data, summary birth histories (SBHs), do not on their own provide enough information to estimate age-specific trends in child mortality. To address this, a discrete hazards model was trained on complete birth history data from 243 surveys in 76 countries, representing over 8 million births. A novel approach was developed for prediction at the level of the hypothetical child, each weighted by their probability of birth. Three validation and verification approaches were developed: survey-wise cross validation, comparison against existing indirect methods, and external validation employed through the application of the new method to an addition 243 SBH-only data sources. The new method was found to produce results which are comparable to current best methods for under-5 mortality estimation while additionally producing valid age-specific estimates. Use of this method could allow researchers to utilize a massive amount of SBH data for estimation of trends in neonatal and infant mortality.

The third chapter, *Geographic accessibility and utilization of facility-based care in Zambia: a geostatistical analysis*, explored the predictive ability of geographic factors, including travel time to nearest health facility, to explain healthcare utilization in Zambia, and developed on geostatistical model in order to predict a gridded surface representing treatment seeking rates for diarrhea and febrile illness in children across the country. This analysis overcame a number of important data and methodological limitations present in previous research, including using exact household locations, using flexible splines to represent the decay function over travel time, and using a full probabilistic model and providing uncertainty around all estimates. Results indicate that while at least three quarters of Zambians live within an hour of a health facility, small differences in travel time to healthcare are independently associated with large declines in utilization rates within the first hour. The decision to seek care at a health facility is a complex process that is not easily reduced to geography. As such, a univariate model based solely on geographic accessibility is not sufficient for accurate prediction of utilization across a gridded surface. Improved prediction using a probabilistic model is possible but uncertainty remains high.

Acknowledgements

I am grateful to *a lot* people who have supported me and my work on this dissertation.

I would like to thank my advisor Simon Hay for your guidance and support over the years. For allowing me independence, and giving me opportunities to explore and grow as a researcher and take on a role of leadership. You have been flexible, open, understanding, and have always kept your door open for me.

I would like to thank my committee members. Haidong Wang, for your guidance, friendship and intellectual support in all things Demography related. Bobby Reiner for being so generous with your time; you are a natural teacher and our conversations helped to solidify many concepts for me. To Emmanuel Gakidou for encouraging me over the years and enabling so many of the opportunities that have formed me into the researcher I have become. For helping me through indecision, and for seeing my potential and giving me the opportunities with which to grow.

I would like to thank my family for being my consistent base of support. You have instilled in me the important values of education and persistence. Thank you to my parents for sacrificing so that it didn't have to be a hard decision for me to decide to pursue a PhD.

I would like to thank Clare Ortblad for filling my life with love and unconditional support. For being the best listener, best friend, and study buddy. For encouraging me when I've been down and for showing me I can be proud of myself.

I would like to thank my peers in the PhD program in Global Health. I'll always cherish our times in the 'nice office' and in the 'cave'. To David Phillips for paving the way, sagely advice, and solid friendship all the way through. To Sarah Wulf Hanson for always insisting on self-care and community. To Grégoire Lurton, I feel lucky to have had you to go through this with me, your thoughtful perspectives have always challenged my thinking. To Gloria, William, Matt, Wisal, Reed, Emily, Hong, Xiaochen, Caroline, Orvalho, and all the others for your support and commiseration.

I would like to thank all the colleagues and collaborators who have contributed to this research. Special thanks to Nick Graetz, Nat Henry, Michael Collison, Molly Biehl, Annie Browne, Josh Longbottom, Laura Dwyer-Lindgren, Chloe Shields, Ian Davis, Chloe Morozoff, Drew Dolgert, Chun-Wei Yuan, Amber Sligar, Bob Thompson, Lucas Earl, and many others. Thank you to Aaron Osgood-Zimmerman: you were the friend and confidant I needed when you joined our team. Thanks for teaching me all that statistics, and for your patience with all my questions. I've enjoyed sharing my office space, research questions, and countless laughs with you. Thank you to Nancy Fullman for being such a constant positive source of support and energy, and for helping me with me so much with my writing. Thank you to Felix Masiye for giving me the opportunity to work in Zambia, which was a formative experience for me, and for encouraging me to develop research using the ZHHEUS data. Thank you to Nick Golding for your valued collaboration on mapping child mortality. Thanks to Mike Hanlon for being there from the start as a source of candid advice and continued mentorship. Thank you to Jon Wakefield for taking an interest in me and my work and lending so much of your time and intellectual support; I have greatly appreciated your friendship and mentorship.

I would like to thank IHME, and in particular Kelsey Bannon, Michael MacIntyre, and Chris Murray for giving me a place to grow professionally and giving me the institutional support that every PhD student deserves.

I would like to thank the Bill and Melinda Gates Foundation for funding and for giving me the opportunity as an intern to experience how my research can have value on practical grounds.

I would like to thank Dane Boog and Céline Abell for doing all the hard program management work to make this whole thing happen.

I would like to thank Ralphie, you are a wonderful companion.

Thanks to all my dear friends for all the inspiration, and all the love and fun times we've shared.

Table of Contents

Chapter 1: Mapping under-5 and neonatal mortality in Africa, 2000-2015: a baseline analysis for the Sustainable Development Goals

(Reprinted from https://doi.org/10.1016/S0140-6736(17)31758-0)

Abstract

During the Millennium Development Goal (MDG) era, many countries in Africa achieved marked reductions in under-5 and neonatal mortality. Yet the pace of progress toward these goals substantially varied at the national level, demonstrating an essential need for tracking even more local trends in child mortality. With the adoption of the Sustainable Development Goals (SDGs) in 2015, which established ambitious targets for improving child survival by 2030, optimal intervention planning and targeting will require understanding of trends and rates of progress at a higher spatial resolution. In this study, we aimed to generate high-resolution estimates of under-5 and neonatal all-cause mortality across 46 countries in Africa.

We assembled 235 geographically resolved household survey and census data sources on child deaths to produce estimates of under-5 and neonatal mortality at a resolution of 5 × 5 km grid cells across 46 African countries for 2000, 2005, 2010, and 2015. We used a Bayesian geostatistical analytical framework to generate these estimates, and implemented predictive validity tests. In addition to reporting 5 × 5 km estimates, we also aggregated results obtained from these estimates into three different levels—national, and subnational administrative levels 1 and 2—to provide the full range of geospatial resolution that local, national, and global decision makers might require.

Amid improving child survival in Africa, there was substantial heterogeneity in absolute levels of under-5 and neonatal mortality in 2015, as well as the annualised rates of decline achieved from 2000 to 2015. Subnational areas in countries such as Botswana, Rwanda, and Ethiopia recorded some of the largest decreases in child mortality rates since 2000, positioning them well to achieve SDG targets by 2030 or earlier. Yet these places were the exception for Africa, since many areas, particularly in central and western Africa, must reduce under-5 mortality rates by at least 8.8% per year, between 2015 and 2030, to achieve the SDG 3.2 target for under-5 mortality by 2030.

In the absence of unprecedented political commitment, financial support, and medical advances, the viability of SDG 3.2 achievement in Africa is precarious at best. By producing under-5 and neonatal mortality rates at multiple levels of geospatial resolution over time, this study provides key information for decision makers to target interventions at populations in the greatest need. In an era when precision public health increasingly has the potential to transform the design, implementation, and impact of health programmes, our 5 × 5 km estimates of child mortality in Africa provide a baseline against which local, national, and global stakeholders can map the pathways for ending preventable child deaths by 2030.

Introduction

Improvement of child survival is a long-standing international priority [1-3], and as shown in the last few decades, substantial progress has been accomplished in reducing child mortality and absolute inequalities in rates of child death across countries worldwide [4,5]. Yet by the conclusion of the Millennium Development Goals (MDGs), which aimed to reduce under-5 mortality by two-thirds from 1990 to 2015, only 57 of 195

countries and territories worldwide met or exceeded the pace of progress required to achieve MDG 4 (ie, a 4.4% annualised rate of decline) during that period [5]. Additionally, despite narrowing disparities over time, geographic inequalities persisted among countries with the lowest and highest child mortality rates. In sub-Saharan Africa, for example, this divergence in 2015 spanned from 15.6 deaths per 1000 livebirths in Botswana to 135.0 deaths per 1000 livebirths in the Central African Republic [5]. National mortality rates, although useful for macro-level comparisons [4,6,7], obscure variations in child survival at lower administrative units (eg, districts), the levels at which most health programme planning and implementation occur. Without advancing the aims of precision public health, which includes robust subnational monitoring of child mortality levels and trends, health authorities face sizeable challenges to optimally funding and targeting interventions for the populations who most need them [8-11]. Ending all preventable child deaths by 2030 is the bold aim set forth by the Sustainable Development Goals (SDGs) [3], and an ambition that requires a much better understanding of where exactly the largest gaps remain in improving child survival.

Advances in both data availability and statistical methods have facilitated subnational assessments of under-5 mortality in several sub-Saharan African countries, including provinces in South Africa [4], regions in Tanzania [12], states in Nigeria [13], counties in Kenya [4], and districts in Ghana, Mozambique, Uganda, and Zambia [14-16]. Such work has unveiled the initial magnitude of subnational disparities in all-cause child deaths, but it is likely that much more heterogeneity remains within formal administrative units. In an ideal setting, national vital registration or health information systems would routinely capture local data on deaths and births, and estimates of child survival could be generated at a similar resolution on the basis of vital registration data. However, few countries in sub-Saharan Africa have complete or high-quality vital registration systems [17-19], and thus often rely on household surveys and periodic censuses to assess their demographic and health profiles [20]. These data sources often include geolocated cluster-level or administrative-area-level identification. They can therefore be analysed with spatially explicit methods such as model-based geostatistics, which quantify spatial differences in variables from geolocated data. Use of model-based geostatistics allows for the synthesis of disparate geographical data into gridded maps and thus yields comparable high-resolution estimates over larger study areas [21,22]. Previous studies have used model-based geostatistics methods to produce gridded estimates of infant mortality and child mortality in Mali [23], and advancements in computational statistics now allow for high-resolution estimation of health outcomes and related indicators at the continental scale [24-27].

Recent analyses [28,29] have sought to quantify under-5 mortality with greater geospatial resolution in sub-Saharan Africa, including one study that produced 10 × 10 kilometre (km) estimates of all-cause under-5 mortality in 28 countries [29]. However, these studies feature several data and methodological shortcomings. First, previous studies exclusively drew from the Demographic and Health Survey (DHS) series for their data sources [28,29], constraining their analyses to a limited subset of surveys with both complete birth histories and GPS-identified survey clusters. Second, popular spatial interpolation methods such as kernel-density estimation can be overly sensitive to data variations that result from small numbers found at the survey cluster rather than true subnational differences in child survival. Without a more stable estimation approach that accounts for spatial and temporal correlations in the data, misleading or implausible results can arise (eg, localities experiencing increases in under-5 mortality that exceeded an average of 10% per year from the 1980s to 2000s) [29]. Third, recent analyses do not have the temporal specificity or timeliness that is of greatest demand from policy makers; for instance, a study that presented decade-wise estimates for the 1980s, 1990s, and 2000s [29] could not detail the potential effects of the MDGs on under-5 mortality trends. Last, previous studies have not calibrated the aggregation of geospatial estimates to externally validated national estimates [4,5,30], a key step to ensuring the internal consistency of subnational results. Since high-resolution estimates of child mortality could serve as a crucial input for local health programme funding and deployment, it is of equally high importance to address these outstanding data and analytic challenges.

In this study, we aimed to advance geospatial analysis of child mortality in Africa by using a larger set of data sources and types than previously published, as well as applying advanced modelling techniques [22,24,27,31] in order to generate high-resolution estimates of under-5 and neonatal all-cause mortality in 46 countries in Africa in 2000, 2005, 2010, and 2015.

Methods

Overview

Our analysis provides estimates of under-5 mortality (the probability of death before age 5 years per 1000 livebirths) and neonatal mortality (probability of death within first month of life per 1000 livebirths) for each 5 × 5 km cell in 46 countries in Africa. These countries accounted for 54% of global under-5 deaths in 2015 [5], and they fully overlap with the African countries included in the EQUitable Impact Sensitive Tool (EQUIST), a UNICEF-supported initiative that aims to maximise the effect of health policies for children who reside in low-income countries. Analytical steps are described below and in the appendix. Our study follows the Guidelines for Accurate and Transparent Health Estimates Reporting (GATHER). The GATHER checklist appears in the appendix.

Data

We extracted data on child mortality and geographical locations from censuses and several household survey series, including the DHS, UNICEF Multiple Indicator Cluster Surveys, and other country-specific surveys. We included data sources if they had summary or complete birth history modules and subnational geographical information. When available at the cluster level (ie, a group of neighbouring households or a village associated with a set of geolocated data), data were extracted together with cluster GPS or located on the basis of precise location names (n=59 279). In the absence of geographic coordinates, we extracted names of the smallest available administrative area units (n= 6111). The appendix lists all data sources (pp 7–22).

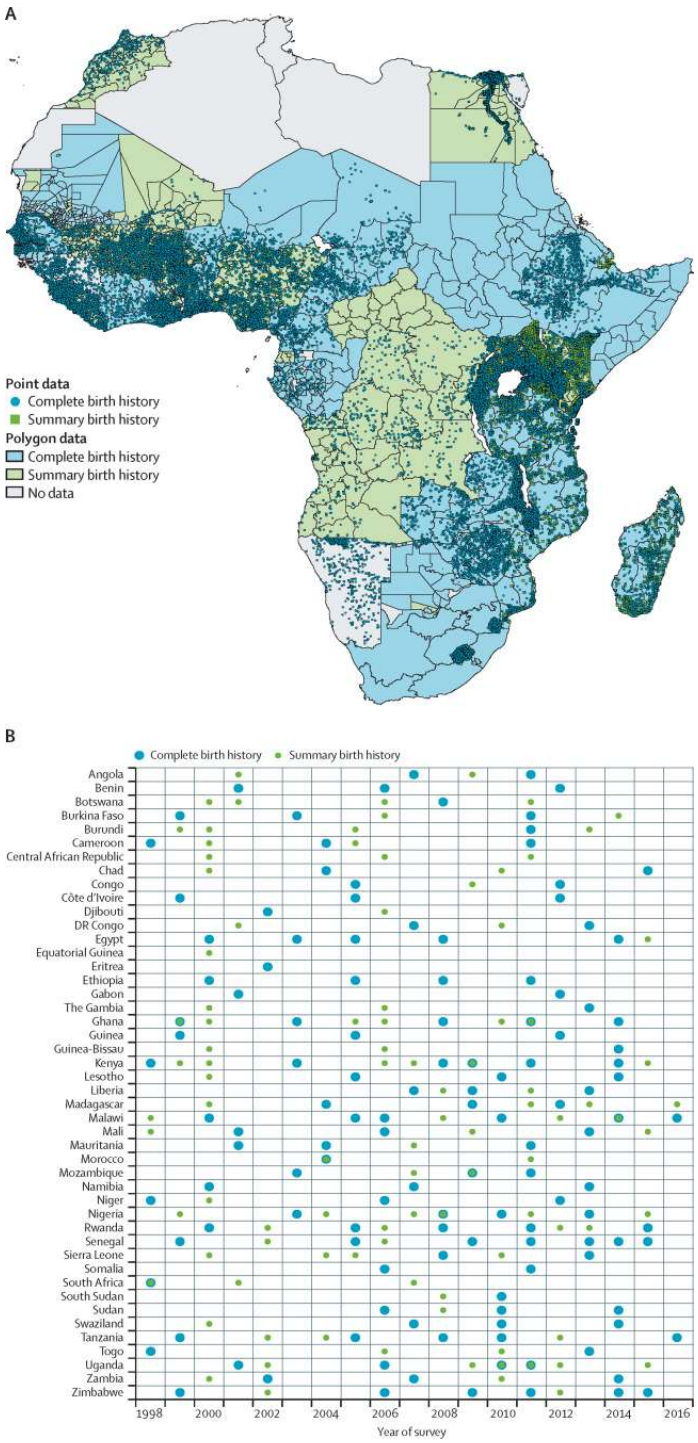


Figure 1. Data availability by type and country, 2000–15. All data are mapped (A), and shown by country and year of survey (B). Surveys can contribute mortality data up to 17 years before the time of the survey. (A) Complete birth history data are displayed in preference to summary birth histories when both have been used in that location. Cluster locations are mapped as points, and polygon data as shapes where available. (B) Data on summary birth histories are green and complete birth histories data are blue.

We estimated child mortality rates in 5-year periods, such that mortality rates for each period of time represented probabilities of death for a synthetic cohort. Thus, mortality rates at each point in time

reflected average mortality rates over the corresponding years. For the estimates, we refer to 1998–2002 as 2000, 2003–07 as 2005, 2008–12 as 2010, and 2013–17 as 2015. These estimates were disaggregated by monthly probabilities of death for the age groups of first month of life, 1–11 months, 12–35 months, and 36–59 months. As such, exposures and deaths could fall into one of 16 age-group–time period bins. A month counted as an exposure if the child was alive at the beginning of the month. Probability of death before reaching age 5 years for a given period can then be calculated directly as 1 minus the product of age-bin-specific survival probabilities (see appendix) [32]. Only livebirths were captured and we did not estimate stillbirths.

We extracted two types of child mortality data, namely complete birth histories and summary birth histories. Data availability by type and country is shown in figure 1, and is further detailed in the appendix. Records of complete birth histories contain dates of birth and death, as applicable, of all children of sampled women. We summed the total exposure months and death events within each age group and time period at the most precise available geographic level (ie, either cluster points or administrative polygons). Summary birth history records provide considerably less information than complete birth histories. They only include mothers' reports of the total number of children ever born and those who died, and do not include children's age or date of death. However, summary birth history data are relatively easy to collect and are widely available in Multiple Indicator Cluster Surveys and censuses. After we extended methodological advancements in model-based approaches to adjust data on summary birth histories [33], we developed a model using data in which both complete and summary birth histories were available, to partition summary birth history data into age-specific and period-specific mortality probabilities, and applied this relationship to datasets in which summary birth history data were available. Additional detail on our summary birth histories-adjustment model is in the appendix. Finally, bias ratios calculated for GBD 2016 [5], which are specific to data type, source, and year, were applied to both complete birth histories and summary birth histories data (see appendix).

When cluster-level geographic coordinates were not available for a given country, we matched reported administrative units with first-level, second-level, or third-level administrative divisions from the Global Administrative Unit Layers database or the Database of Global Administrative Areas. We refer to the geographical data that only contained administrative boundaries as polygons. If data could not be mapped to administrative units (polygons) from these sources, we matched them using other sources or modified available polygons on the basis of the available information about polygon boundaries. In several cases, for example, we disaggregated an administrative unit into city boundaries and the surrounding rural area to match the disaggregation of the available mortality data. We resampled polygon data into geographically dispersed points and weighted them on the basis of population. In terms of sample size, polygons contributed 41% of the total data (see appendix).

To inform our model, we compiled several layers of sociodemographic and health-related covariates at the 5×5 km pixel level (value estimates for each 5×5 km grid cell in the study area), all of which have shown some relationship with child survival or overall child health outcomes. The covariates were average years of educational attainment of women aged 15–49 years [34,35], prevalence of wasting and stunting in children younger than 5 years [36], *Plasmodium* falciparum parasite rate [24], a proxy index of fertility (based on the ratio of women aged 15–49 years and children under 5 years) per pixel [37,38], and total population [39]. Additionally, we included several covariate layers that were reflective of potential environmental and infrastructural factors related to overall development and thus child survival (ie, an enhanced vegetation index [40], daytime land-surface temperature [40], proportion of land under irrigation [41], urban-rural distinction [42], brightness of night-time light from the Defense Meteorological Satellite Program [43,44], and accessibility to cities with populations greater than 50 000) [45]. When several years' worth of data were available, we either took the synoptic mean from available years in each estimation period or used the mid-period-year estimate. Two covariate layers, irrigation and accessibility, were only available for

the year 2000, and therefore did not vary over time in our analysis. The appendix contains more information on our spatial covariates, including plots of all covariates.

We then used an ensemble method, stacked generalisation, to select covariates, capture possible non-linear effects, and to account for the complex interactions between them [31]. For each age group and year, we fit four sub-models, namely generalised additive models, boosted regression trees, lasso regression, and ridge regression, and predicted for each cluster in the data using five-fold cross-validation. The four sub-model predictions were included as covariates when fitting the full geostatistical model described below. This approach removes covariates that are not predictive and identifies optimal combinations of covariates that are predictive, and thus is expected to improve overall predictive performance over a singular model [31]. Because of their heightened predictive validity, ensemble modelling approaches are increasingly used in population health measurement [46], as well as other fields [47,48]. More details on this approach are in the appendix.

Analysis

We fitted four separate Bayesian model-based geostatistics models (ie, one for each age group) to estimate the monthly probability of mortality in any given pixel. We modelled covariate effects using the ensemble approach discussed above, and the model explicitly accounted for spatiotemporal autocorrelation by modelling the covariance of data residuals in space and time. This allowed us to leverage the correlation structure of the data to more accurately predict mortality within locations where data on child death were absent. Pixel-level uncertainty intervals (UIs) were generated from 1000 draws (ie, candidate maps) [22] that were created from the posterior distributions of modelled parameters. The appendix includes additional detail of our model and estimation process.

We then aggregated pixel-level estimates from the 1000 candidate maps up to two subnational administrative units and national levels [49]. Such aggregation allowed us to further calibrate estimates of under-5 and neonatal mortality to national GBD estimates for the 2000, 2005, 2010, and 2015 periods. [5] We achieved this by calculating the ratio of the population-weighted posterior mean national-level estimate from our analysis to mean national estimates for the same time period from GBD, [5] and then multiplying each cell in the posterior sample by this ratio. The median for these ratios was 1.01 (IQR 0.95–1.07), indicating generally close agreement with GBD estimates. The appendix includes scatterplots comparing our national-level estimates from this analysis with GBD estimates.

For reported results, we masked all final model outputs for which land cover was classified as barren or sparsely vegetated, on the basis of MODIS satellite data in 2013 [50], the most recent year of available data, as well as areas in which total population density was less than ten individuals per 1 × 1 km pixel in 2015 [39].

We validated our models using spatially stratified five-fold out-of-sample cross validation and report bias (mean error), total variance (root-mean-square error), and 95% cluster-level data coverage within prediction intervals. We stratified space either by the first or second subnational administrative unit. By aggregating predictions and observations to administrative units, we could increase sample sizes to a level high enough for better determining model fit; this was not feasible at the cluster level because of data noisiness due to very small sample sizes. Generally, across our four models and at both levels of aggregation, we found out-of-sample mean error to be very close to zero, indicating no systematic bias. Correlations between aggregated out-of-sample model fit and aggregated holdout data ranged from 0.81 to 0.94 for administrative level 1, and from 0.62 to 0.87 for administrative level 2. Root-mean-square error relative to mean estimated probability ranged from 17.7% to 33.3% for administrative level 1 and 32.7% to 50.0% for administrative level 2. These metrics indicate relatively good model fits; however, these metrics are sensitive

to sample sizes at aggregation and average probabilities at different age groups, and thus should be interpreted within that context. The 95% prediction interval at the cluster level covered 94% of the out-of-sample data points, indicating that our models' fit could reproduce out-of-sample data within the specified level of uncertainty. Detail on validation procedures and full results is in the appendix.

Data sharing

Data are available at http://ghdx.healthdata.org/.

Results

Although under-5 mortality rates decreased throughout Africa between 2000 and 2015 (figure 2), stark disparities endured across the continent and within national borders. On the basis of pixel-level estimates, we found that in 2000, most of sub-Saharan Africa recorded under-5 mortality rates exceeding 138 deaths per 1000 livebirths, while large swathes of Nigeria, Niger, Sierra Leone, and Mali, along with other countries in western and central Africa, surpassed 200 deaths per 1000 livebirths. By 2015, half of sub-Saharan Africa had under-5 mortality rates below 72 deaths per 1000 livebirths, and increasingly more places in Africa neared or had fewer than 25 deaths per 1000 livebirths, the SDG3.2 target for 2030. Nonetheless, 118 locations at the second administrative level in Chad, Mali, Burkina Faso, the Central African Republic, and Nigeria still faced average under-5 mortality rates higher than 170 per 1000 livebirths in 2015. Further, sizeable within-country inequalities remained. For instance, Nigeria had a national under-5 mortality rate of 115.4 deaths (95% uncertainty interval [UI] 99.6–132.2) per 1000 livebirths in 2015 [5], yet at the local government area level (administrative level 2), under-5 mortality rates ranged from 54.8 (48.2–62.9) deaths per 1000 livebirths in the Osun local governmental area of Osogbo to 214.6 (190.2–240.6) deaths per 1000 livebirths in the Bauchi local governmental area of Ningi.

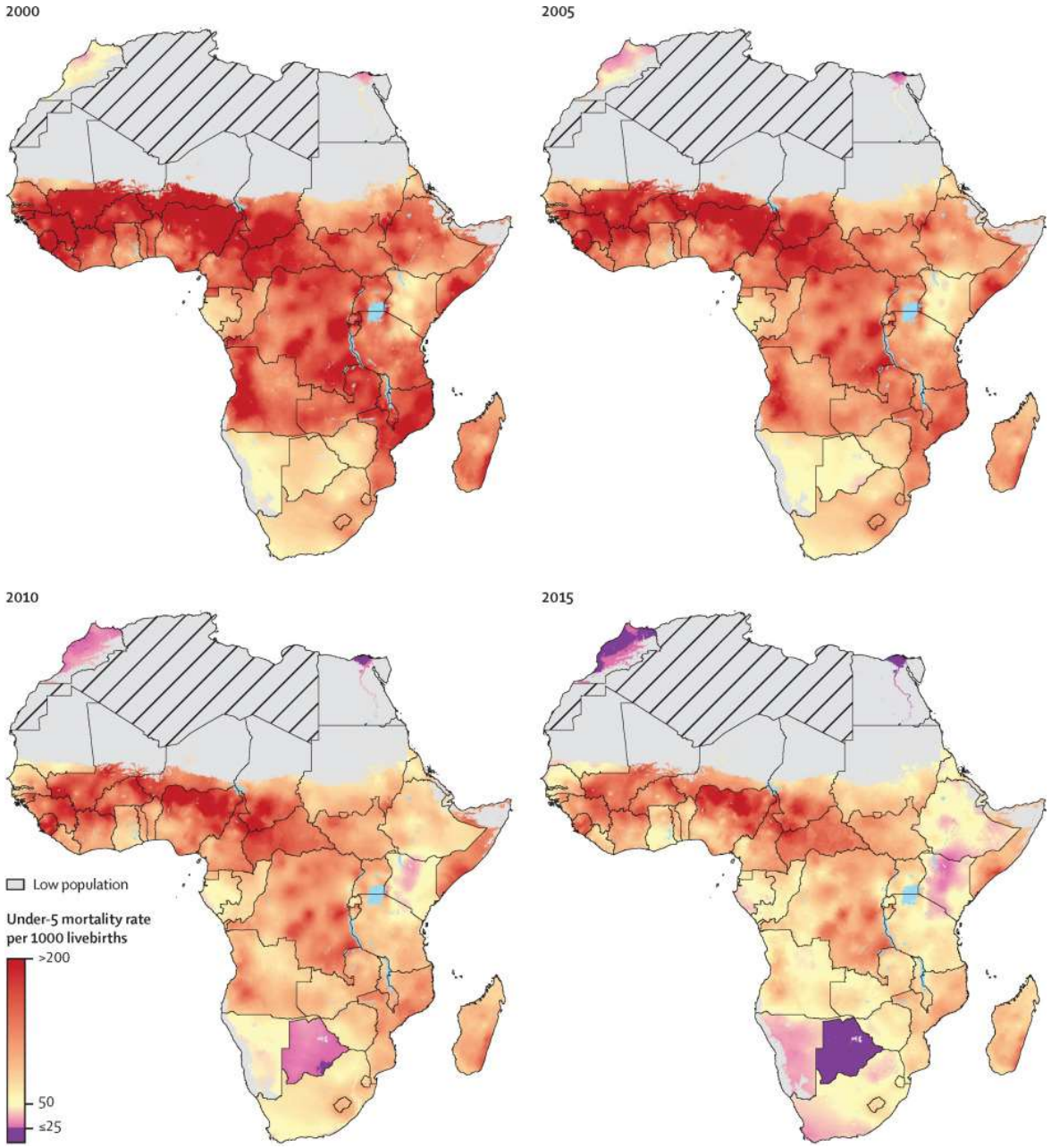


Figure 2. Under-5 mortality rates at the 5 × 5 km resolution in 2000, 2005, 2010, and 2015. Data are at 5 × 5 km resolution. All pixels with an under-5 mortality rate equal to or fewer than 25 deaths per 1000 livebirths (the Sustainable Development Goal 3.2 target for under-5 mortality) are coloured purple. Pixels with fewer than ten people and classified as barren or sparsely vegetated are coloured in grey. Grey areas with diagonal lines are not included in this analysis. km=kilometre.

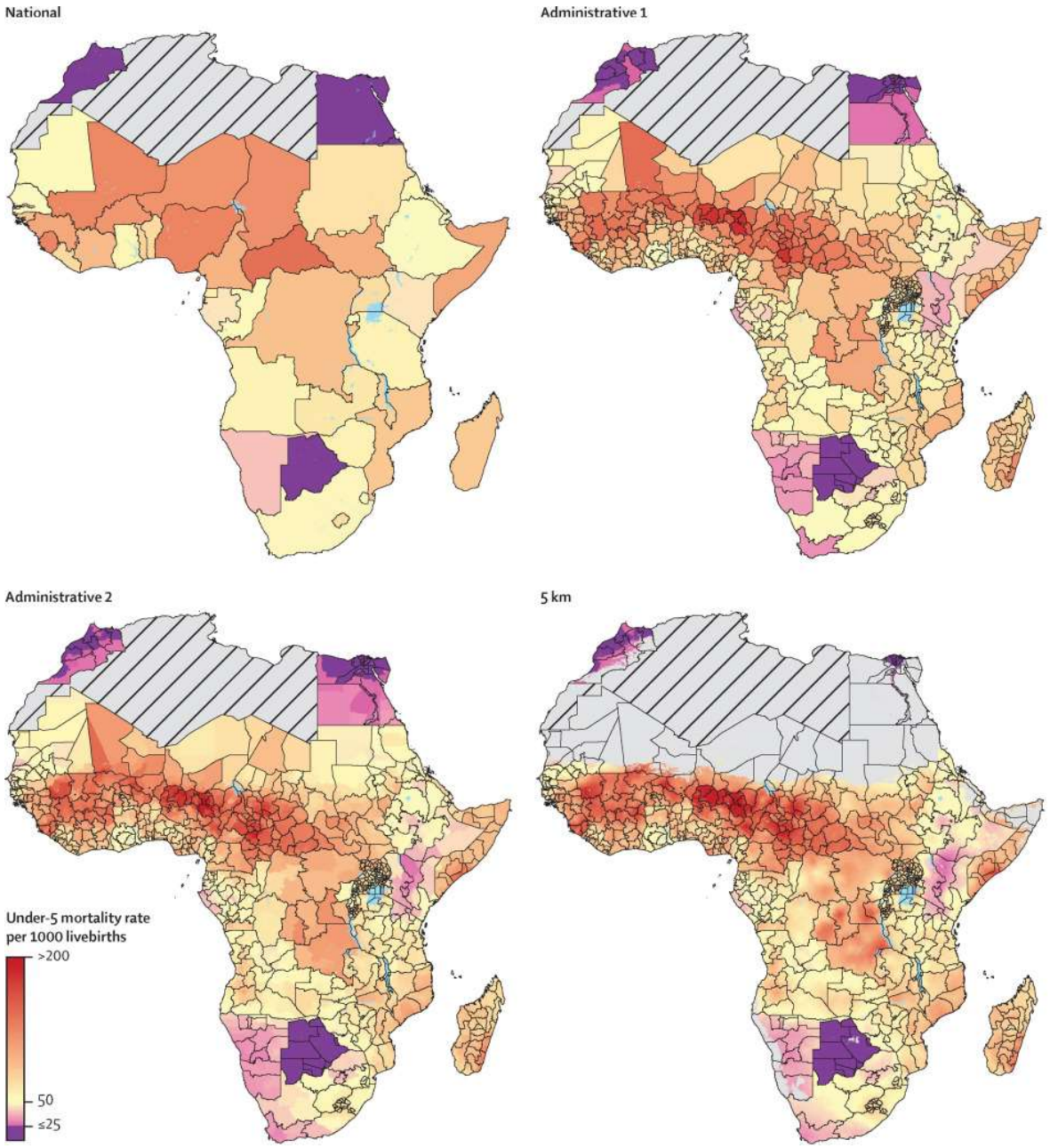


Figure 3. Under-5 mortality rates at the national, first and second administrative, and 5 × 5 km levels in 2015. All locations with a mortality rate equal to or fewer than 25 deaths per 1000 livebirths (the Sustainable Development Goal 3.2 target for under-5 mortality) are coloured purple. Pixels with fewer than ten people and classified as barren or sparsely vegetated are coloured in grey. Grey areas with diagonal lines are not included in this analysis. km=kilometre.

Somewhat similar geographic patterns emerged for neonatal mortality as were found for under-5 mortality (appendix), with Botswana and Egypt having large areas meeting the SDG target (ie, 12 deaths per 1000 livebirths). In 2015, Côte d'Ivoire had one of the widest gaps between districts, ranging from 29.6 (95% UI 23.9–35.7) neonatal deaths per 1000 livebirths in Cavally in Montagnes district to 52.9 (43.2–64.1) neonatal

deaths per 1000 livebirths in Bagoue in Savanes district. Côte d'Ivoire, Mali, and Nigeria all had second administrative units with mean neonatal mortality rates greater than 50 deaths per 1000 livebirths. More information on neonatal mortality can be found in the appendix.

Figure 3 illustrates the effects of disaggregating under-5 mortality estimates across levels of geospatial granularity (national, first and second administrative levels, and the 5 × 5 km grid), and how inequalities in child survival can be masked by geographic aggregation. In 2015, Egypt and Morocco achieved the SDG3.2 target for under-5 mortality, at 20.6 deaths per 1000 livebirths in Egypt and 22.9 deaths per 1000 livebirths in Morocco [5]. Still, approximately 25% of each country's population lived in areas with mortality rates higher than the SDG threshold. Botswana was the only other African nation that met the SDG 3.2 target by 2015, with an under-5 mortality rate of 15.6 deaths per 1000 livebirths [5]. The appendix includes full geographical disaggregation of neonatal mortality rates by administrative level.

Figure 4 compares dimensions of under-5 mortality, from low to high, against relative uncertainty, as measured by the ratio of the 95% UI range to the mean, in 2015. Senegal, Gambia, Ghana, Kenya, Rwanda, Tanzania, and Zimbabwe had large areas with relatively low under-5 mortality and low uncertainty, whereas most of the Central African Republic and Somalia had both relatively high under-5 mortality and uncertainty. There were geographical areas of high mortality and low uncertainty in northern and eastern Nigeria, Mali, Burkina Faso, northern Cameroon, and southwestern Chad. By contrast, most of Botswana, Namibia, eastern Ethiopia, eastern Angola, and South Africa had relatively low under-5 mortality rates, but these estimates were accompanied by relatively high uncertainty. More plots of uncertainty in predictions are available in the appendix.

In many countries, under-5 mortality decreased by more than 4.4% per year from 2000 to 2015 (figure 5) [5], a rate that exceeded the pace of progress established under MDG 4 (ie, a two-thirds reduction by 2015). Further, average annualised rates of decline in Botswana, Ethiopia, Liberia, Rwanda, and Angola exceeded 6.0%. Many other countries, including Burundi, Malawi, Togo, Uganda, and Tanzania, had second-level administrative areas that had a mixture of annualised rates of decline between 2.0% and 4.4%, as well as those that exceeded a 4.4% decrease each year; results for administrative levels 1 and 2 are in the appendix. By contrast, areas throughout 17 countries, including the Central African Republic, South Sudan, Lesotho, and Madagascar, had no second-level administrative areas that achieved an average annualised rate of decline exceeding 4.4% from 2000 to 2015. Nigeria had wide disparities in terms of progress, with annualised rates of change spanning from 0.7% annual decline (95% UI 0.2% to −1.6%) to a 5.0% annual decline (−4.1 to −5.9%). On the basis of pixel-level annualised rates of decline achieved from 2000 to 2015, and projections of these rates through 2030 (figure 5), several countries could have localities achieving SDG 3.2 if past rates of decline in under-5 mortality are sustained over the next 15 years. These locations were primarily in northern, southern, and eastern Africa, but also included areas in Senegal, Liberia, and Ghana. 20 of the 46 countries had second-level administrative areas that, on the basis of average current trajectories, could reduce under-5 mortality rates to 25 deaths per 1000 livebirths. However, only 26.1% of second administrative-level areas had an annualised rate of decline from 2000 to 2015 that was faster than the MDG 4 target of 4.4% per year. At least 60% of these locations need to match or surpass the rates of progress achieved from 2000 to 2015 to meet the SDG3.2 target for under-5 mortality by 2030 (figure 5). For instance, within the Central African Republic, Mali, Sierra Leone, Niger, Chad, and Burkina Faso, the majority of populations live in areas where annual declines of 8.8% or more are needed to achieve SDG 3.2.

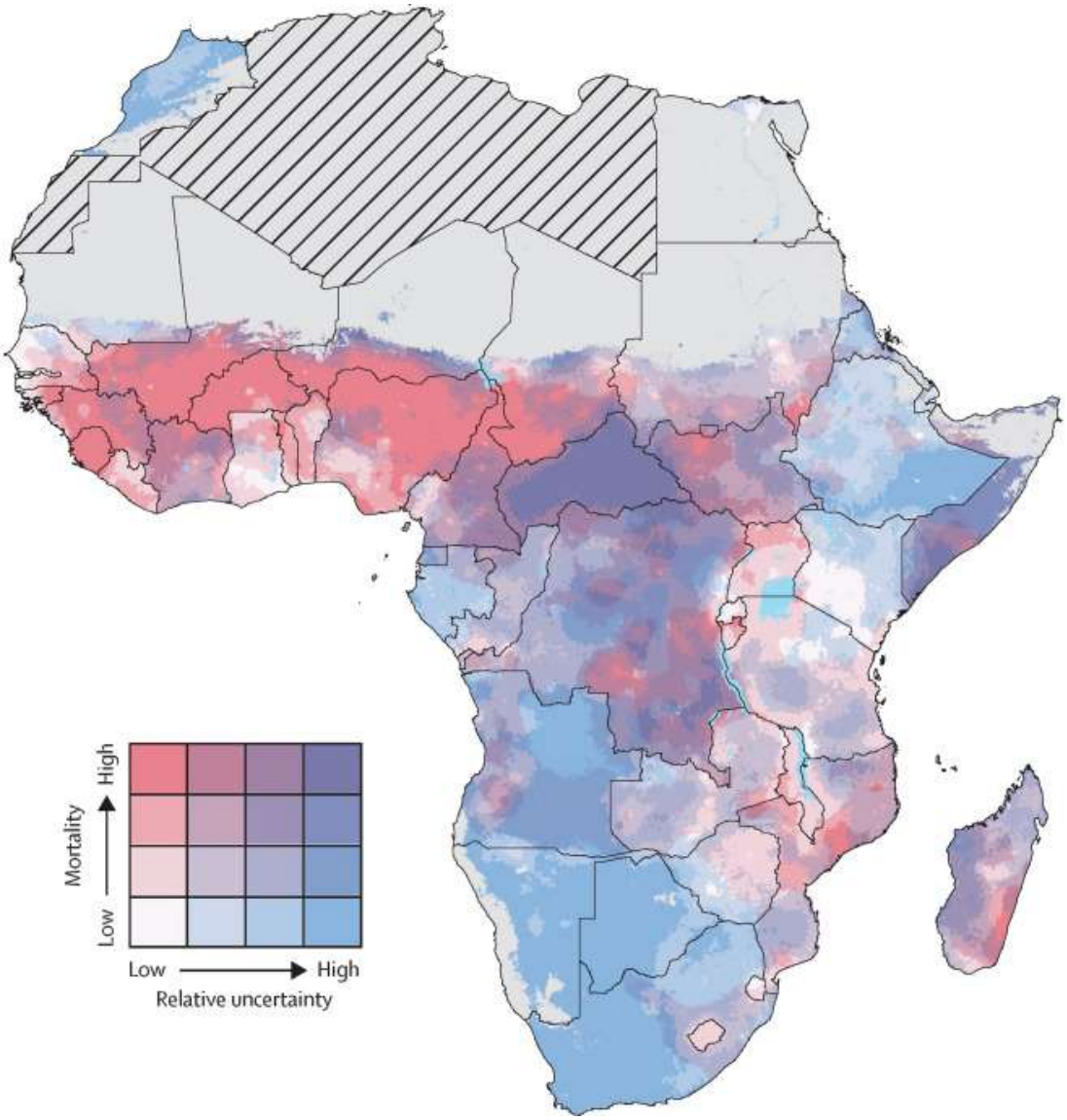


Figure 4. Overlapping population-weighted quartiles of under-5 mortality and relative uncertainty in 2015. Under-5 mortality rate quartile cutoff points were 56, 80, and 102 deaths per 1000 livebirths. Relative uncertainty was computed as the ratio of the 95% uncertainty intervals and under-5 mortality rate for each pixel. Cutoff points for uncertainty were 29%, 35%, and 41%. The lowest quartile of mortality is white, and the highest is dark pink. The lowest quartile for uncertainty is white and the highest is blue. These colours overlap such that areas coloured purple have both high under-5 mortality rates and high relative uncertainty. Pixels with fewer than ten people and classified as barren or sparsely vegetated are coloured in grey. Grey areas with diagonal lines are not included in this analysis.

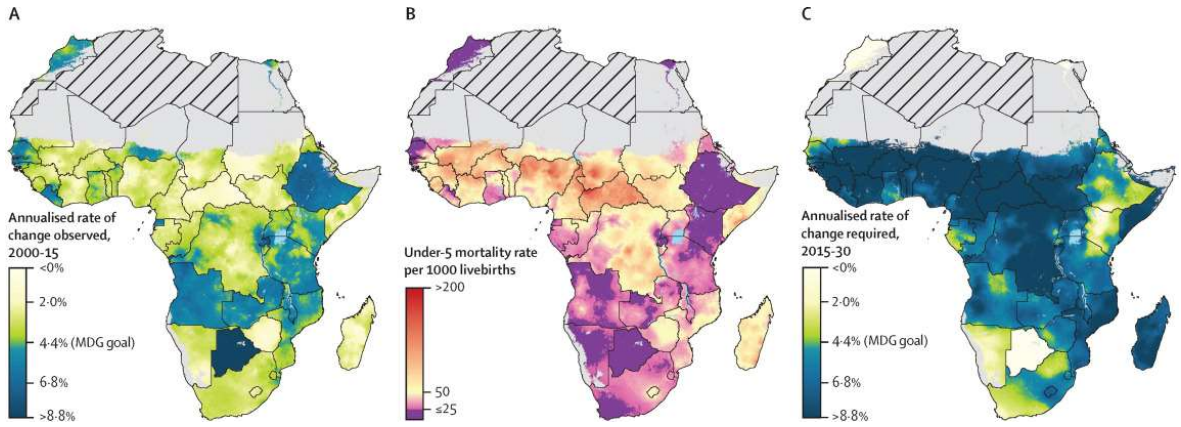


Figure 5. Annualised rates of decline in under-5 mortality during the MDG era, with projections to 2030, and needed rates of decline to reach the SDG target. 4.4% is the annualised rate of decline that was equivalent to the pace of progress required to meet Millennium Development Goal 4. (A) Annualised rates of decline for under-5 mortality from 2000 to 2015. Pixels coloured blue exceeded the annualised rate of decline between 2000 and 2015, whereas pixels coloured green to yellow had a slower rate of annualised decline during this time. (B) Predicted under-5 mortality rates in 2030, based on annualised rates of decline achieved between 2000 and 2015. Pixel-level under-5 mortality rates were predicted for 2030 on the basis of annualised rates of decline achieved from 2000 to 2015. Based on this prediction, pixels for which under-5 mortality rates equalled or were less than 25 deaths per 1000 livebirths in 2030 are coloured purple. (C) Rates of decline required to reach the SDG 3.2 target for under-5 mortality by 2030 (25 deaths per 1000 livebirths). Pixels coloured blue will need to achieve a 4.4% or greater decline per year from 2015 to 2030 to achieve the SDG 3.2 target for under-5 mortality. Pixels coloured green to yellow can meet the SDG 3.2 target by 2030 at a pace slower than a 4.4% reduction per year from 2015 to 2030. Pixels with fewer than ten people and classified as barren or sparsely vegetated are coloured in grey. Grey areas with diagonal lines are not included in this analysis. SDG=Sustainable Development Goal.

Discussion

To the best of our knowledge, our study offers the first quantification of 5 × 5 km estimates of under-5 and neonatal mortality in 46 African countries, highlighting a mixture of impressive gains and enduring disparities in child survival across the continent. By 2015, nearly all locations had a reduction in under-5 mortality rates from 2000, with many areas of Ethiopia, Botswana, and Rwanda recording particularly large reductions. Yet in Chad, Mali, the Central African Republic, and Sierra Leone, more than half of populations live in places where under-5 mortality rates still exceeded 120 deaths per 1000 livebirths in 2015. Despite achieving notable rates of decline between 2000 and 2015, most of Africa must substantially accelerate reductions in under-5 mortality to meet the SDG 3.2 target of 25 deaths per 1000 livebirths by 2030. These results underscore the crucial importance of tracking geospatially granular patterns in child survival, particularly if all countries aim to end all preventable child deaths by 2030.

Charting 5 × 5 km trends in child mortality from 2000 to 2015 provides the foundation from which local achievements and challenges during the MDG era can be better understood. Although the duration of MDG assessment spanned from 1990 to 2015, its greatest catalytic effects on political, social, and financial commitments to improving child health in Africa mainly occurred from 2000—when the MDGs were established—to 2015. Nationally, 19 countries in Africa met or exceeded the MDG4 rate of reduction (4.4% each year) between 2000 and 2015 [5]. Although several factors probably influenced this progress during the 2000s, the confluence of escalated development assistance focused on child health [51], the rapid scale-up of multiple interventions that target childhood illnesses (eg, vaccination, malaria control, and HIV prevention) [25,52-56], and heightened overall socioeconomic development undoubtedly contributed to improving child survival in many African countries. By contrast, it is an unlikely coincidence that many places with slower gains (ie, much of the Central African Republic, Chad, and Somalia) also received less international funding for newborn and child health [26] and had some of Africa's lowest levels of overall coverage for key maternal and child health interventions [55]. In Nigeria, for example, where annualised declines in under-5 mortality ranged from 0.7% each year to 5.0% per year since 2000, large differences in

state-level trends for various maternal and child health interventions also occurred during that time [13]. National case studies have explored drivers of MDG 4 progress in sub-Saharan Africa [54,57-60], but few delve into more local factors and their association with changes in under-5 and neonatal survival at a high geographic resolution. In-depth analyses that link our 5 × 5 km estimates of child mortality to more granular measures of intervention coverage and other indicators, akin to a recent study on the effects of malaria control in Africa [25], could strengthen inputs into local health policy and resource allocation planning.

In the transition from MDG 4 to SDG 3.2, child survival targets changed from achieving relative rates of progress to attaining absolute levels by 2030. This shift is heralded by many [61-63], because setting specific thresholds could encourage a greater focus on the places which bear the highest toll of child deaths. Effectively directing such attention might be a challenge in Africa, however, especially since more than 75% of the continent's children live in areas where annual declines in under-5 mortality must exceed the pace of reductions they achieved from 2000 to 2015 to meet the SDG 3.2 target by 2030. From 2000 to 2015, only 26.1% of second administrative level areas recorded an annualised rate exceeding the MDG target of 4.4%, whereas at least 60% of these locations will need to at least match their current pace to make the SDG 3.2 target for under-5 mortality a reality. Of particular concern are the areas encompassing nearly 27% of Africa's population that must at least double the MDG 4 rate to achieve the SDG 3.2 target by 2030, a pace that is unprecedented in the last few decades. Envisaging such a feat is difficult without the occurrence of substantial medical breakthroughs, such as a fully effective malaria vaccine [64], or considerably extending access to high-quality health care, eliminating risk factors that account for a large proportion of child deaths [36], and bolstering socioeconomic factors that directly affect child survival [7], or all of the above. That is not to say achieving SDG3.2 is impossible in Africa, since there have been several instances in which the introduction and scale-up of cost-effective interventions swiftly reduced child mortality in many countries [45], such as the expansion of measles immunisation, screening and treatment of maternal syphilis during antenatal care, and the provision of oral rehydration therapy for severe diarrhoea. At a time when precision public health could offer transformative power for local intervention design and implementation [9,10], these results and initiatives, such as EQUIST, which account for subnational variations in health, intervention effectiveness, and costs, are vital going forward in the SDG era. However, staying the current course and failing to address more systemic barriers to improving child survival will not be sufficient to meet the SDG 3.2 target for most of Africa.

This study offers the analytical framework from which we aim to extend geospatial modelling of child mortality to an increased number of locations, with a heightened focus on estimating specific causes with greater temporal resolution. By expanding our high-resolution estimation of child mortality to additional high-burden countries outside of Africa, we aim to generate estimates in locations that represent 95% of under-5 deaths globally [5]. We also intend to disaggregate estimates by year rather than by 5-year intervals. Ultimately, our goal is to generate cause-specific mortality estimates at the 5 × 5 km resolution globally, but this undertaking necessitates improved encoding or identification of specific causes of death by precise locations. A recent analysis [24] produced 5 × 5 km estimates of malaria mortality in Africa by age group [24], including children younger than 5 years, which served as an initial foray into this kind of high-resolution, cause-specific mortality mapping. Generating pixel-level estimates of mortality for children younger than 5 years from other causes that disproportionately affect children, such as diarrhoeal diseases and lower respiratory infections, are also future analytical priorities.

The production of more geospatially granular estimates of key child health outcomes hinges upon heightened accessibility to and collection of geo-referenced data. Increased data availability would both facilitate cause-specific mortality estimation and reduce uncertainty in all-cause estimates, which particularly affect locations with substantive geographical or temporal data gaps. In an ideal setting, agencies involved in local data collection and management would work with the array of data users to identify best practices for sharing geo-referenced data and thus creating global goods, while maintaining

privacy and respecting data use agreements. Increasing the availability of geolocated data, from surveys to censuses to facility records, would greatly strengthen the precision of local monitoring of child health needs. By highlighting geographical areas where data gaps are most strikingly pronounced, we hope to encourage more collaboration between data users and providers.

Although household survey and census data offer good information on child deaths, they are inferior substitutes to high-quality, fully representative vital registration systems for providing timely, continuous, and complete subnational information on births and deaths. Vital registration systems should remain the gold standard for routine monitoring of national and local child mortality, and in recent years both political and financial investments in improving vital registration have increased [65-67]. Moreover, SDG indicator 17.19.2 explicitly outlines targets for birth and death registration completeness [3], and efforts such as the Bloomberg Data for Health Initiative seek to swiftly strengthen existing vital registration systems and to help build vital statistics infrastructure in places where they are inadequate or non-existent. Nonetheless, massive disparities persist in birth and death registration levels, and Africa has some of the largest gaps in the establishment, coverage, and quality of vital registration [68]. Using household survey and census data to inform child mortality estimation is likely to be a necessity in many parts of the world for the immediate future, but as vital registration systems continue to improve, future analyses should involve developing methods to integrate vital registration data with survey and census data within the model-based geostatistics framework.

Our findings should be interpreted within the context of some methodological limitations. First, we assumed no migration, which implied that all recorded births, deaths, and exposures were assumed to occur in the survey location. The ability to properly measure and incorporate indices of migration has been an enduring challenge for large-scale demographic and epidemiological studies [4,5,69,70], and though innovative efforts such as the WorldPop project [71] are improving the quantification of population movement, they do not yet provide the temporal and geographical resolution necessary for our analysis. Continued data collation and methodological advances are required to appropriately account for migration in mortality estimation. Second, we modelled death probabilities in age groups separately, because of software and computation limitations. This allowed us to avoid potential age-composition bias in small clusters, but ultimately meant ignoring a high level of correlation in these data. Future work is needed to develop methods that enable computationally efficient estimation of the under-5 survival curve simultaneously rather than approximating survival in a piece-wise manner. Third, the included set of spatial covariates does not represent the full universe of potential correlates for drivers of under-5 mortality (eg, exposure to unsafe water and sanitation) because of an absence of high-resolution spatial data or layers for these particular indicators. As these measures become available, future studies should incorporate them into modelling approaches. Fourth, to include a massive amount of polygon data in our spatially continuous model, we re-sampled polygon data to points. It is possible that this procedure introduced over-smoothing, although these effects are probably minimal given their agreement with other subnational mortality models (see appendix). Future research will need to develop computational methods for scaling up geostatistical integration of point and polygon data to continental-scale mapping studies [72]. Last, the potential effects of urban slums on child mortality were not explicitly quantified. Although our study offers the highest geospatial resolution of child mortality to date, the 5 × 5 km pixel-level remains too coarse to fully account for intra-city slums and disparities.

Amid impressive overall gains in decreasing under-5 mortality rates in Africa, sizeable populations within the continent have yet to experience such improvements in child survival. The SDG 3.2 targets for under-5 and neonatal mortality demand extraordinary progress in child health for Africa—and without substantial, sustained commitments to financing better access to improved health care and targeting interventions to high-burden areas, we are likely to fall short of these aims. Monitoring high-resolution trends in child mortality is a vital tool for swiftly recognising local child health needs and prioritising resources accordingly. With continued and expanded mapping efforts, collectively we can garner greater recognition of and attention to communities in which child survival remains tenuous.

References

1. UN. World Declaration on the Survival, Protection and Development of Children. New York: United Nations, 1990.
2. UN. Millennium Declaration. New York: United Nations, 2000.
3. UN. Transforming our World: the 2030 Agenda for Sustainable Development. New York: UN, 2015.
4. GBD 2015 Child Mortality Collaborators. Global, regional, national, and selected subnational levels of stillbirths, neonatal, infant, and under-5 mortality, 1980–2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet 2016; 388: 1725–74.
5. GBD 2016 Mortality Collaborators. Global, regional, and national under-5 mortality, adult mortality, age-specific mortality, and life expectancy, 1970–2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet 2017; 390: 1084–150.
6. Liu L, Oza S, Hogan D, et al. Global, regional, and national causes of under-5 mortality in 2000–15: an updated systematic analysis with implications for the Sustainable Development Goals. Lancet 2016; 388: 3027–35.
7. GBD 2015 SDG Collaborators. Measuring the health-related Sustainable Development Goals in 188 countries: a baseline analysis from the Global Burden of Disease Study 2015. Lancet 2016; 388: 1813–50.
8. Bhutta ZA. Mapping the geography of child mortality: a key step in addressing disparities. Lancet Glob Health 2016; 4: e877–78.
9. Dowell SF, Blazes D, Desmond-Hellmann S. Four steps to precision public health. Nat News 2016; 540: 189.
10. Desmond-Hellmann S. Progress lies in precision. Science 2016; 353: 731–31.
11. Bill & Melinda Gates Foundation. Annual Letter from Bill Gates. 2013. http://www.gatesfoundation.org/Who-We-Are/Resourcesand-Media/Annual-Letters-List/Annual-Letter-2013 (accessed June 26, 2017).
12. Mercer LD, Wakefield J, Pantazis A, Lutambi AM, Masanja H, Clark S. Space-time smoothing of complex survey data: small area estimation for child mortality. Ann Appl Stat 2015; 9: 1889–905.
13. Wollum A, Burstein R, Fullman N, Dwyer-Lindgren L, Gakidou E. Benchmarking health system performance across states in Nigeria: a systematic analysis of levels and trends in key maternal and child health interventions and outcomes, 2000–2013. BMC Med 2015; 13: 208.
14. Gavi Full Country Evaluations Team. Gavi Full Country Evaluations: 2015 annual dissemination report. Seattle: Institute for Health Metrics and Evaluation, 2015.
15. Dwyer-Lindgren L, Kakungu F, Hangoma P, et al. Estimation of district-level under-5 mortality in Zambia using birth history data, 1980–2010. Spat Spatiotemporal Epidemiol 2014; 11: 89–107.
16. Arku RE, Bennett JE, Castro MC, et al. Geographical inequalities and social and environmental risk factors for under-five mortality in ghana in 2000 and 2010: Bayesian spatial analysis of census data. PLoS Med 2016; 13: e1002038.
17. Phillips DE, Lozano R, Naghavi M, et al. A composite metric for assessing data on mortality and causes of death: the vital statistics performance index. Popul Health Metr 2014; 12: 14.
18. Mikkelsen L, Phillips DE, AbouZahr C, et al. A global assessment of civil registration and vital statistics systems: monitoring data quality and progress. Lancet 2015; 386: 1395–406.
19. Murray CJL, Rajaratnam JK, Marcus J, Laakso T, Lopez AD. What can we conclude from death registration? improved methods for evaluating completeness. PLoS Med 2010; 7: e1000262.
20. Fabic MS, Choi Y, Bird S. A systematic review of Demographic and Health Surveys: data availability and utilization for research. Bull World Health Organ 2012; 90: 604–12.
21. Diggle P, Ribeiro PJ. Model-based geostatistics. New York: Springer, 2009.
22. Patil AP, Gething PW, Piel FB, Hay SI. Bayesian geostatistics in health cartography: the perspective of malaria. Trends Parasitol 2011; 27: 246–53.
23. Gemperli A, Vounatsou P, Kleinschmidt I, Bagayoko M, Lengeler C, Smith T. Spatial patterns of infant mortality in Mali: the effect of malaria endemicity. Am J Epidemiol 2004; 159: 64–72.
24. Gething PW, Casey DC, Weiss DJ, et al. Mapping Plasmodium falciparum mortality in Africa between 1990 and 2015. N Engl J Med 2016; 375: 2435–45.
25. Bhatt S, Weiss DJ, Cameron E, et al. The effect of malaria control on Plasmodium falciparum in Africa between 2000 and 2015. Nature 2015; 526: 207–11.

26. 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 Stat Methodol 2009; 71: 319–92.
27. Lindgren F, Rue H, Lindström J. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. J R Stat Soc Ser B Stat Methodol 2011; 73: 423–98.
28. Pezzulo C, Bird T, Utazi EC, et al. Geospatial modeling of child mortality across 27 countries in Sub-Saharan Africa. Rockville: ICF International, 2016.
29. Burke M, Heft-Neal S, Bendavid E. Sources of variation in under-5 mortality across sub-Saharan Africa: a spatial analysis. Lancet Glob Health 2016; 4: e936–45.
30. Liu L, Oza S, Hogan D, et al. Global, regional, and national causes of under-5 mortality in 2000–15: an updated systematic analysis with implications for the Sustainable Development Goals. Lancet 2016; 388: 3027–35.
31. Bhatt S, Cameron E, Flaxman SR, Weiss DJ, Smith DL, Gething PW. Improved prediction accuracy for disease risk mapping using Gaussian Process stacked generalisation. http://arxiv.org/ abs/1612.03278 (accessed Jan 5, 2017).
32. Ahmad OB, Lopez AD, Inoue M. The decline in child mortality: a reappraisal. Bull World Health Organ 2000; 78: 1175–91.
33. Rajaratnam JK, Tran LN, Lopez AD, Murray CJL. Measuring underfive mortality: validation of new low-cost methods. PLoS Med 2010; 7: e1000253.
34. Gakidou E, Cowling K, Lozano R, Murray CJ. Increased educational attainment and its effect on child mortality in 175 countries between 1970 and 2009: a systematic analysis. Lancet 2010; 376: 959–74.
35. Wang H, Liddell CA, Coates MM, et al. Global, regional, and national levels of neonatal, infant, and under-5 mortality during 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet 2014; 384: 957–79.
36. Forouzanfar MH, Afshin A, Alexander LT. Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet 2016; 388: 1659–724.
37. Tatem AJ, Garcia AJ, Snow RW, et al. Millennium development health metrics: where do Africa's children and women of childbearing age live? Popul Health Metr 2013; 11: 11.
38. Canning D, Günther I, Linnemayr S, Bloom D. Fertility choice, mortality expectations, and interdependent preferences— an empirical analysis. Eur Econ Rev 2013; 63: 273–89.
39. Linard C, Gilbert M, Snow RW, Noor AM, Tatem AJ. Population distribution, settlement patterns and accessibility across Africa in 2010. PLoS One 2012; 7: e31743.
40. Weiss DJ, Atkinson PM, Bhatt S, Mappin B, Hay SI, Gething PW. An effective approach for gap-filling continental scale remotely sensed time-series. ISPRS J Photogramm Remote Sens 2014; 98: 106–18.
41. FAO. Global map of irrigation areas, version 5.0. 2013. http://www.fao.org/nr/water/aquastat/irrigationmap/index10.stm (accessed June 26, 2017).
42. Dijkstra L, Poelmann H. A harmonised definition of cities and rural areas: the new degree of urbanization. 2014. http://ec.europa.eu/ regional_policy/sources/docgener/work/2014_01_new_urban.pdf (accessed June 26, 2017).
43. Chen X. Explaining subnational infant mortality and poverty rates: what can we learn from night-time lights? Spat Demogr 2015; 3: 27–53.
44. Ebener S, Murray C, Tandon A, Elvidge CC. From wealth to health: modelling the distribution of income per capita at the sub-national level using night-time light imagery. Int J Health Geogr 2005; 4: 5.
45. Nelson A. Travel time to major cities: a global map of accessibility. 2008. http://forobs.jrc.ec.europa.eu/products/gam/ (accessed June 26, 2017).
46. Foreman KJ, Lozano R, Lopez AD, Murray CJ. Modeling causes of death: an integrated approach using CODEm. Popul Health Metr 2012; 10: 1.
47. Gneiting T, Raftery AE. Weather forecasting with ensemble methods. Science 2005; 310: 248–49.
48. Bell RM, Koren Y. Lessons from the Netflix Prize Challenge. SIGKDD Explor Newsl 2007; 9: 75–79.
49. Gething PW, Patil AP, Hay SI. Quantifying aggregated uncertainty in Plasmodium falciparum malaria prevalence and populations at risk via efficient space-time geostatistical joint simulation. PLoS Comput Biol 2010; 6: e1000724.
50. Channan S, Collins K, Emanuel WR. Global mosaics of the standard MODIS land cover type data. College Park: University of Maryland and the Pacific Northwest National Laboratory, 2014.

51. Dieleman JL, Schneider MT, Haakenstad A, et al. Development assistance for health: past trends, associations, and the future of international financial flows for health. Lancet 2016; 387: 2536–44.
52. Ng M, Colson KE, Fullman N, et al. Assessing the contribution of malaria vector control and other maternal and child health interventions in reducing all-cause under-five mortality in Zambia. Am J Trop Med Hyg 2016; published online Feb 15. DOI:10.4269/ajtmh.15-0315.
53. Bhutta ZA, Chopra M, Axelson H, et al. Countdown to 2015 decade report (2000–10): taking stock of maternal, newborn, and child survival. Lancet 2010; 375: 2032–44.
54. UNICEF, World Health Organization (WHO). Countdown to 2015: a decade of tracking progress for maternal, newborn, and child survival. Geneva: UNICEF and WHO, 2015.
55. Lim SS, Allen K, Bhutta ZA, et al. Measuring the health-related Sustainable Development Goals in 188 countries: a baseline analysis from the Global Burden of Disease Study 2015. Lancet 2016; 388: 1813–50.
56. Jones G, Steketee RW, Black RE, Bhutta ZA, Morris SS. How many child deaths can we prevent this year? Lancet 2003; 362: 65–71.
57. EPHI. Countdown to 2015: Ethiopia's progress towards reduction in under-five child mortality. Addis Abba: Ethiopian Public Health Institute, 2014.
58. Kanyuka M, Ndawala J, Mleme T, et al. Malawi and Millennium Development Goal 4: a Countdown to 2015 country case study. Lancet Glob Health 2016; 4: e201–14.
59. Afnan-Holmes H, Magoma M, John T, et al. Tanzania's Countdown to 2015: an analysis of two decades of progress and gaps for reproductive, maternal, newborn, and child health, to inform priorities for post-2015. Lancet Glob Health 2015; 3: e396–409.
60. Moucheraud C, Owen H, Singh NS, et al. Countdown to 2015 country case studies: what have we learned about processes and progress towards MDGs 4 and 5? BMC Public Health 2016; 16: 794.
61. Glass RI, Guttmacher AE, Black RE. Ending preventable child death in a generation. JAMA 2012; 308: 141–12.
62. Murray CJL. Shifting to Sustainable Development Goals— implications for global health. N Engl J Med 2015; 373: 1390–93.
63. UNICEF. Committing to child survival: a promise renewed. Progress report 2015. New York: United Nations International Children's Emergency Fund, 2015.
64. Thera MA, Plowe CV. Vaccines for malaria: how close are we? Annu Rev Med 2012; 63: 345–57.
65. Lopez AD, Setel PW. Better health intelligence: a new era for civil registration and vital statistics? BMC Med 2015; 13: 73.
66. Lo S, Horton R. Everyone counts—so count everyone. Lancet 2015; 386: 1313–14.
67. AbouZahr C, de Savigny D, Mikkelsen L, et al. Civil registration and vital statistics: progress in the data revolution for counting and accountability. Lancet 2015; 386: 1373–85.
68. Mikkelsen L, Phillips DE, AbouZahr C, et al. A global assessment of civil registration and vital statistics systems: monitoring data quality and progress. Lancet 2015; 386: 1395–406.
69. WHO. WHO methods and data sources for country-level causes of death 2000–2015. Geneva: World Health Organization, 2016.
70. WHO and MCEE. MCEE-WHO methods and data sources for child causes of death 2000–2015. Geneva: World Health Organization, 2016.
71. Tatem AJ. WorldPop, open data for spatial demography. Sci Data 2017; 4: 170004.
72. Diggle PJ, Moraga P, Rowlingson B, Taylor BM. Spatial and spatio-temporal log-Gaussian Cox processes: extending the geostatistical paradigm. Stat Sci 2013; 28: 542–63.

Supplementary Information for Chapter 1

1. GATHER Checklist

Item #	Checklist item	Reference
Objectives and funding		
1	Define the indicator(s), populations (including age, sex, and geographic entities), and time period(s) for which estimates were made.	Main text (Methods)
2	List the funding sources for the work.	Bill and Melinda Gates Foundation
Data Inputs		
For all data inputs from multiple sources that are synthesized as part of the study:		
3	Describe how the data were identified and how the data were accessed.	Main text (Methods)
4	Specify the inclusion and exclusion criteria. Identify all ad-hoc exclusions.	Main text (Methods); Methods appendix (Sections 2 and 4)
5	Provide information on all included data sources and their main characteristics. For each data source used, report reference information or contact name/institution, population represented, data collection method, year(s) of data collection, sex and age range, diagnostic criteria or measurement method, and sample size, as relevant.	Methods appendix (Supplementary Table 1) and available through ghdx.healthdata.org/
6	Identify and describe any categories of input data that have potentially important biases (e.g., based on characteristics listed in item 5).	Main text (Methods); Methods appendix (Section 6)
For data inputs that contribute to the analysis but were not synthesized as part of the study:		
7	Describe and give sources for any other data inputs.	Methods appendix (Section 10.1)
For all data inputs:		
8	Provide all data inputs in a file format from which data can be efficiently extracted (e.g., a spreadsheet rather than a PDF), including all relevant meta-data listed in item 5. For any data inputs that cannot be shared because of ethical or legal reasons, such as third-party ownership, provide a contact name or the name of the institution that retains the right to the data.	Available through ghdx.healthdata.org/ or on request
Data analysis		
9	Provide a conceptual overview of the data analysis method. A diagram may be helpful.	Main text (Methods);
10	Provide a detailed description of all steps of the analysis, including mathematical formulae. This description should cover, as relevant, data cleaning, data pre-processing, data adjustments and weighting of data sources, and mathematical or statistical model(s).	Main text (Methods); methods appendix
11	Describe how candidate models were evaluated and how the final model(s) were selected.	Main text (Analysis); methods appendix section 11.6)
12	Provide the results of an evaluation of model performance, if done, as well as the results of any relevant sensitivity analysis.	Main text (Analysis); methods appendix (Supplementary tables

		9 and 10, Supplementary Figure 13)
13	Describe methods for calculating uncertainty of the estimates. State which sources of uncertainty were, and were not, accounted for in the uncertainty analysis.	Main text (Analysis); methods appendix (Section 11)
14	State how analytic or statistical source code used to generate estimates can be accessed.	Code is provided in an online repository
Results and Discussion		
15	Provide published estimates in a file format from which data can be efficiently extracted.	Raster files for spatial data and csvs of administrative level 1 and 2 estimates to be made available at ghdx.healthdata.org. PDF tables of administrative level 1 and 2 estimates in appendix.
16	Report a quantitative measure of the uncertainty of the estimates (e.g. uncertainty intervals).	Main text (Results; Figure 4); Appendix (Supplementary figures 13 and 14); Appendix tables
17	Interpret results in light of existing evidence. If updating a previous set of estimates, describe the reasons for changes in estimates.	Main text (Research in Context)
18	Discuss limitations of the estimates. Include a discussion of any modelling assumptions or data limitations that affect interpretation of the estimates.	Main text (Limitations)

2. Geographic Inclusion

This analysis aimed to provide high-resolution under-5 and neonatal mortality rate maps across 46 countries in Africa considered among priority countries for monitoring child mortality by the Countdown to 2015 for Maternal, Newborn, and Child Survival.[1] To this list we added Namibia so that the resulting maps also had continuous coverage of sub-Saharan Africa, a region over which mortality rates and other development metrics are often aggregated.

It was not possible to extend the analysis to include Algeria, Tunisia, Libya, or Western Sahara due to a lack of subnational mortality data from these countries. Cape Verde, Comoros, Mauritius, and São Tomé and Príncipe are each geographically isolated from included countries and cover very small geographic areas, so including these countries in the analysis would provide little advance over existing national estimates. The map of 46 included countries is shown in Supplementary Figure 1.

Supplementary Figure 6. Countries included in mapping under-5 and neonatal mortality rates.

3. Mortality data sources

*Supplementary Table 1. Sources of mortality data used in the model. The datasets and reports from which complete birth history (CBH) and summary birth history (SBH) were extracted for each country are detailed under Citation, and the institutions that provided these data are given under Source. Further details are given marked for those sources labelled † as follows: ICF International. 2004-2012. Demographic Health Surveys (various). Calverton, Maryland: ICF International, 2012. CBH surveys marked with * included both CBH and SBH data and were used in the SBH adjustment model described in section 6. Admin = Administrative level*

Country	Year	Data Type	Source	Geographic level	Citation
Angola	2011	CBH*	DHS Program†	Point	Cosep Consultoria, Consaúde & ICF International. Angola Malaria Indicator Survey 2011 [Dataset] AOBR61DT. (ICF International [Distributor], 2011, Calverton, Maryland, USA, 2011).
Angola	2009	SBH	Discussion Forum of the Fifth National Conference of Civil Society (Angola)	Admin 1	Ministerio do Planeamento. Integrated Survey on the Welfare of Population 218pp (Luanda, Angola, 2011).
Angola	2007	CBH	DHS Program†	Point	Consultoria de Serviços e Pesquisas–COSEP Lda, Consultoria de Gestão e Administração em Saúde–Consaúde Lda (Angola) & Macro International Inc. Angola Malaria Indicator Survey 2006-07 [Dataset] AOIR51DT. 112pp (ICF International [Distributor], Calverton, Maryland, USA, 2007).

Country	Year	Data Type	Source	Geographic level	Citation
Angola	2001	SBH	UNICEF	Admin 1	National Institute of Statistics & United Nations Children's Fund. MICS Multiple Indicator Cluster Survey - Assessing the Situation of Angolan Children and Women at the Beginning of the Millennium [Dataset]. 142pp (Luanda, Angola, 2003).
Benin	2012	CBH*	DHS Program†	Point	Institut National de la Statistique et de l'Analyse Économique & ICF International. Enquête Démographique et de Santé du Bénin 2011-2012 [Dataset] BJBR61DT. (ICF International [Distributor], Calverton, Maryland, USA, 2013)
Benin	2006	CBH*	DHS Program†	Admin 1	Institut National de la Statistique et de l'Analyse Économique & Macro International Inc. Enquête Démographique et de Santé (EDSB-III) - Bénin 2006 [Dataset] BJBR51DT. (ICF International [Distributor], Calverton, Maryland, USA, 2007).
Benin	2001	CBH*	DHS Program†	Point	Institut National de la Statistique et de l'Analyse Économique & ORC Macro. Enquête Démographique et de Santé au Bénin 2001 [Dataset] BJBR41DT. (ICF International [Distributor], Calverton, Maryland, USA, 2002).
Botswana	2011	SBH	IPUMS INTERNATIONAL	Admin 1	Central Statistics Office (Botswana) & Minnesota Population Center. Botswana Population and Housing Census 2011 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. (University of Minnesota, Minneapolis, USA, 2017).
Botswana	2008	CBH	Central Statistics Office (Botswana)	Admin 2	Central Statistics Office (Botswana). Botswana Family Health Survey 2007-2008 [Dataset]. (Gaborone, Botswana, 2009).
Botswana	2006	SBH	Central Statistics Office (Botswana)	Admin 2	Central Statistics Office (Botswana). Botswana Demographic Survey 2006 [Dataset]. (Gaborone, Botswana, 2006).
Botswana	2001	SBH	IPUMS INTERNATIONAL	Admin 1	Central Statistics Office (Botswana) & Minnesota Population Center. Botswana Population and Housing Census 2001 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. (University of Minnesota, Minneapolis, USA, 2017).
Botswana	2000	SBH	UNICEF	Admin 1	Central Statistics Office (Botswana) & United Nations Children's Fund (UNICEF). Botswana Multiple Indicator Cluster Survey 2000 [Dataset]. (New York, USA, 2015).
Burkina Faso	2014	SBH	DHS Program†	Point	Institut National de la Statistique et de la Démographie, Programme National de Lutte contre le Paludisme Ouagadougou Burkina Faso & ICF International. Enquête sur les Indicateurs du Paludisme au Burkina Faso (EIPBF) 2014. 170pp (Rockville, Maryland, USA, 2015).
Burkina Faso	2011	CBH*	DHS Program†	Point	Institut National de la Statistique et de la Démographie & ICF International. Enquête Démographique et de Santé et à Indicateurs Multiples du Burkina Faso 2010 [Dataset] BFBR62DT. (ICF International [Distributor], Calverton, Maryland, USA, 2012).
Burkina Faso	2006	SBH	IPUMS INTERNATIONAL	Admin 3	Minnesota Population Center & National Institute of Statistics and Demography (Burkina Faso). Burkina Faso Population and Housing Census 2006 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. (University of Minnesota, Minneapolis, USA, 2013).
Burkina Faso	2006	SBH	UNICEF	Point	United Nations Children's Fund & Institut National de la Statistique et la Demographie. Burkina Faso Enquete par Grappes a Indicateurs Muliples 2006 [Dataset]. (United Nations Children's Fund, 2008).
Burkina Faso	2003	CBH*	DHS Program†	Point	Institut National de la Statistique et de la Démographie & ORC Macro. Enquête Démographique et de Santé du Burkina Faso 2003 [Dataset] BFBR43DT. (ICF International [Distributor], Calverton, Maryland, USA, 2004).
Burkina Faso	1999	CBH*	DHS Program†	Point	Institut National de la Statistique et de la Démographie & Macro International Inc. Enquête Démographique et de Santé, Burkina Faso 1998-1999 [Dataset] BFBR31DT. (ICF International [Distributor], Calverton, Maryland, USA, 2000).
Burundi	2013	SBH	DHS Program†	Point	Institut de Statistiques et d'Études Économiques du Burundi, Ministère de la Santé Publique et de la Lutte contre le Sida (Burundi) & ICF International. Enquête sur les Indicateurs du Paludisme Burundi 2012 [Dataset] BUHR6HSV. 138pp (ICF International [Distributor], Bujumbura, Burundi, 2013).
Burundi	2011	CBH*	DHS Program†	Point	Institut de Statistiques et d'Études Économiques du Burundi, Ministère de la Santé Publique et de la Lutte contre le Sida (Burundi) & ICF International. Enquête Démographique et de Santé Burundi 2010 [Dataset] BUBR61DT. (ICF International [Distributor], Bujumbura, Burundi, 2012).
Burundi	2005	SBH	UNICEF	Point	Institut de Statistiques et d'Études Économiques du Burundi. Enquête Nationale d'Évaluation des Conditions de vie de l'Enfant et de la Femme au Burundi-2005 [Dataset]. 192pp (Bujumbura, Burundi, 2008).
Burundi	2000	SBH	UNICEF	Admin 1	Buzingo, D., Habimana, F. & Nduwabika, N. Enquête Nationale d'Evaluation des Conditions de vie de l'Enfant et de la Femme au Burundi (ENECEF-BURUNDI 2000) [Dataset]. 44pp (Burundi, 2001).

Country	Year	Data Type	Source	Geographic level	Citation
Burundi	1999	SBH	World Bank	Admin 1	Vincent, R. Enquete Prioritaire 1998 Etude Nationale sur les Conditions de vie des Populations. 119pp (Bujumbura, Burundi, 2001).
Cameroon	2011	CBH*	DHS Program†	Point	Institut National de la Statistique & ICF. International. Enquête Démographique et de Santé et à Indicateurs Multiples du Cameroun 2011 [Dataset] CMBR60DT. (ICF International [Distributor], Calverton, Maryland, USA, 2012).
Cameroon	2005	SBH	IPUMS INTERNATIONAL	Admin 3	Minnesota Population Center, National Institute of Statistics (Cameroon) & Central Bureau of the Census and Population Studies (Cameroon). Cameroon Population and Housing Census 2005 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. (University of Minnesota, Minneapolis, USA, 2013).
Cameroon	2004	CBH*	DHS Program†	Point	Institut National de la Statistique & ORC Macro. Enquête Démographique et de Santé du Cameroun 2004 [Dataset] CMBR44DT. (ICF International [Distributor], Calverton, Maryland, USA, 2004).
Cameroon	2000	SBH	UNICEF	Admin 1	Fonds des Nations Unies pour l'Enfrance & Ministère de l'Economie et des Finances Gouvernement du Cameroun. Enquête à Indicateurs Multiples (MICS) au Cameroun 2000 [Dataset]. 45 (Cameroun, 2002).
Cameroon	1998	CBH*	DHS Program†	Admin 1	Fotso, M. et al. Enquête Démographique et de Santé, Cameroun 1998 [Dataset] CMBR31FL. (Bureau Central des Recensements et des Études de Population & Macro International Inc [Distributor], Calverton, Maryland, USA, 1999).
Central African Republic	2011	SBH	UNICEF	Admin 1	Institut Centrafricain des Statistiques et des Études Économiques et Sociales. Enquête par grappes à indicateurs multiples – MICS couplée avec la sérologie VIH, RCA, 2010 Rapport final [Dataset]. 229pp (Bangui, République Centrafricaine, 2012).
Central African Republic	2006	SBH	UNICEF	Admin 1	Institut Centrafricain des Statistiques et des Etudes Economiques et Sociales. Résultats de l'enquête nationale à indicateurs multiples couplée avec la sérologie VIH et anémie en RCA 2006 [Dataset]. 378pp (Bangui, République Centrafricaine, 2009).
Central African Republic	2000	SBH	UNICEF	Admin 1	United Nations Children's Fund & Ministere du Plan et de la Cooperation Internationale. Enquete a Indicateurs Multiples - MICS 2000 Rapport Final [Dataset]. 304pp (Bangui République Centrafricaine, 2001).
Chad	2015	CBH*	DHS Program†	Point	National Institute of Statistical Economic and Demographic Studies (Chad) & ICF International. Chad Demographic and Health Survey 2014-2015 [Dataset] TDBR71DT. (ICF International [Distributor], Fairfax, USA, 2016).
Chad	2010	SBH	UNICEF	Admin 2	Institut national de la statistique des études économiques et démographiques. Enquête par grappes à indicateurs multiples Tchad 2010 [Dataset]. 364pp (Chad, 2011).
Chad	2004	CBH*	DHS Program†	Admin 1	Ouagadjio, B. et al. Enquête Démographique et de Santé Tchad 2004 [Dataset] TDBR41DT. (ICF International [Distributor], Calverton, Maryland, USA, 2004).
Chad	2000	SBH	UNICEF	Admin 1	Direction de la Statistique des Etudes Economiques et Démographiques. Enquete par grappes a indicateurs multiples - Rapport complet [Dataset]. 117pp (Chad, 2001).
Congo	2012	CBH*	DHS Program†	Admin 1	Centre Nationale de la Statistique et des Études Économiques (Congo) & ICF International. Enquête Démographique et de Santé du Congo (EDSC-II) 2011-2012 [Dataset] CGBR60DT. (ICF International [Distributor], Calverton, Maryland, USA 2013).
Congo	2009	SBH	DHS Program†	Admin 1	Centre National de la Statistique et des Études Économiques & ICF Macro. Enquête de Séroprévalence et sur les Indicateurs du Sida – Congo 2009 [Dataset] CDBR61DT. (ICF Macro [Distributor], Calverton, Maryland, USA, 2009).
Congo	2005	CBH*	DHS Program†	Admin 1	Centre National de la Statistique et des Études Économiques & ORC Macro. Enquête Démographique et de Santé du Congo 2005 [Dataset] CGBR51DT. (ICF International [Distributor], Calverton, Maryland, USA 2006).
Cote d'Ivoire	2012	CBH*	DHS Program†	Point	Institut National de la Statistique & ICF International. Enquête Démographique et de Santé et à Indicateurs Multiples de Côte d'Ivoire 2011-2012 [Dataset] CIBR61DT. (ICF International [Distributor], Calverton, Maryland, USA, 2012).
Cote d'Ivoire	2005	CBH	DHS Program†	> Admin 1	Institut National de la Statistique, Ministère de la Lutte contre le Sida (Côte d'Ivoire) & Macro, O. Enquête sur les Indicateurs du Sida, Côte d'Ivoire 2005 [Dataset] CIBR50DT. 283pp (ICF International [Distributor], Calverton, Maryland, USA, 2006).
Cote d'Ivoire	1999	CBH*	DHS Program†	Point	Institut National de la Statistique (Côte d'Ivoire) & ORC Macro. Enquête Démographique et de Santé, Côte d'Ivoire 1998-1999 [Dataset] CIBR3ADT. (ICF International [Distributor], Calverton, Maryland, USA, 2001).

Country	Year	Data Type	Source	Geographic level	Citation
Djibouti	2012	CBH	Pan Arab Project for Family Health (PAPFAM)	Admin 1	Department of Statistics and Demographic Studies (Djibouti), League of Arab States, Ministry of Health (Djibouti), Pan Arab Project for Family Health (PAPFAM). Djibouti Family Health Survey 2012 [Dataset]. (2002).
Djibouti	2006	SBH	UNICEF	Point & Admin 2	Ministry of Economy Finance and Planning in charge of Privatization (Djibouti), Ministry of Health (Djibouti) & United Nations Children's Fund. Djibouti Multiple Indicator Cluster Survey 2006 [Dataset]. 210pp (New York, USA, 2006).
Djibouti	2002	CBH	Pan Arab Project for Family Health (PAPFAM)	National	Department of Statistics and Demographic Studies (Djibouti), League of Arab States, Ministry of Health (Djibouti) & Pan Arab Project for Family Health. Djibouti Family Health Survey 2002 [Dataset]. (2002).
DR Congo	2013	CBH*	DHS Program†	Point	Ministère du Plan et Suivi de la Mise en œuvre de la Révolution de la Modernité, Ministère de la Santé Publique & ICF International. Enquête Démographique et de Santé en République Démocratique du Congo 2013-2014 [Dataset] CDBR61DT. (ICF International [Distributor], Rockville, Maryland, USA, 2014).
DR Congo	2010	SBH	UNICEF	Admin 1	Institut National de la Statistique & Fonds des Nations Unies pour l'Enfrance. Enquete par Grappes a Indicateurs Multiples en Republique du Congo (MICS-RDC 2010) [Dataset]. 384pp (Republique Democratique du Congo, 2011).
DR Congo	2007	CBH*	DHS Program†	Point	Ministère du Plan & Macro International. Enquête Démographique et de Santé, République Démocratique du Congo 2007 [Dataset] CDBR50DT. (ICF International [Distributor], Calverton, Maryland, USA, 2008).
DR Congo	2001	SBH	UNICEF	Admin 1	Ministère du Plan et de la Reconstruction, Fonds des Nations Unies pour l'Enfance & Agence des Etats-Unis pour le Développement International. Enquête Nationale Sur la Situation des Enfants et Des Femmes MICS2/2001 [Dataset]. 258pp (Kinshasa, Democratic Republic of the Congo, 2002).
Egypt	2015	SBH	DHS Program†	Admin 1	El-Zanaty and Associates, ICF International, Ministry of Health and Population (Egypt) & National Population Council (Egypt). Egypt Special Demographic and Health Survey 2015 [Dataset] EGIQ73DT. (ICF International [Distributor], Fairfax, USA, 2015).
Egypt	2014	CBH*	DHS Program†	Point	Ministry of Health and Population (Egypt), El-Zanaty and Associates (Egypt) & ICF International. Egypt Demographic and Health Survey 2014 [Dataset] EGBR61DT. (ICF International [Distributor], Cairo, Egypt and Rockville, Maryland, USA, 2015).
Egypt	2008	CBH*	DHS Program†	Point	El-Zanaty, F. & Way, A. Egypt Demographic and Health Survey 2008 [Dataset] EGBR5ADT. (ICF International [Distributor], Cairo, Egypt, 2009).
Egypt	2005	CBH*	DHS Program†	Point	El-Zanaty, F. & Way, A. Egypt Demographic and Health Survey 2005 [Dataset] EGBR51DT. (ICF International [Distributor], Cairo, Egypt, 2006).
Egypt	2003	CBH*	DHS Program†	Point	El-Zanaty, F. & Way, A. A. 2003 Egypt Interim Demographic and Health Survey [Dataset] EGBR4ADT. (ICF International [Distributor], Cairo, Egypt, 2004).
Egypt	2000	CBH*	DHS Program†	Point	El-Zanaty, F. & Way, A. Egypt Demographic and Health Survey 2000 [Dataset] EGBR42DT. (ICF International [Distributor], Calverton, Maryland, USA, 2001).
Equatorial Guinea	2000	SBH	UNICEF	Admin 1	Ministry of Planning Economic Development and Public Investment (Equatorial Guinea) & United Nations Children's Fund. Encuesta de Indicadores Multiples (MICS 2000) [Dataset]. 79pp (Malabo, Equatorial Guinea, 2001).
Eritrea	2002	CBH*	DHS Program†	Admin 1	National Statistics and Evaluation Office (Eritrea) & ORC Macro. Eritrea Demographic and Health Survey 2002 [Dataset] 331pp (Calverton, Maryland, USA, 2003).
Ethiopia	2011	CBH*	DHS Program†	Point	Central Statistical Agency (Ethiopia) & ICF International. Ethiopia Demographic and Health Survey 2011 [Dataset] ETBR61DT. (ICF International [Distributor], Addis Ababa, Ethiopia and Calverton, Maryland, USA, 2012).
Ethiopia	2008	CBH	Ministry of Health (Ethiopia)	Admin 2	Ethiopian Health and Nutrition Research Center (EHNRI), Macro International, Inc, Ministry of Health (Ethiopia). Ethiopia Global Fund Household Health Coverage Survey 2008 [Dataset]. (2008).
Ethiopia	2007	SBH	IPUMS INTERNATIONAL	Point & Admin 3	Central Statistical Agency (Ethiopia) & Minnesota Population Center. Ethiopia Population and Housing Census 2007 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. (University of Minnesota, Minneapolis, USA, 2015).
Ethiopia	2005	CBH	DHS Program†	Point	Central Statistical Agency (Ethiopia) & ORC Macro. Ethiopia Demographic and Health Survey 2005 [Dataset] ETBR51DT. (ICF International [Distributor], Addis Ababa, Ethiopia and Calverton, Maryland, USA, 2006).
Ethiopia	2000	CBH	DHS Program†	Point	Central Statistical Authority (Ethiopia) & ORC Macro. Ethiopia Demographic and Health Survey 2000 [Dataset] ETBR41DT. (ICF International [Distributor], Addis Ababa, Ethiopia and Calverton, Maryland, USA, 2001).

Country	Year	Data Type	Source	Geographic level	Citation
Gabon	2012	CBH*	DHS Program†	Point	Direction Générale de la Statistique & ICF International. Enquête Démographique et de Santé du Gabon 2012 [Dataset] GABR60DT. (ICF International [Distributor], Calverton, Maryland, USA and Libreville, Gabon, 2013).
Gabon	2001	CBH*	DHS Program†	Admin 2	Direction Générale de la Statistique, Des Études Économiques (Gabon) & ORC Macro. Enquête Démographique et de Santé Gabon 2000 [Dataset] GABR41DT. 385pp (ICF International [Distributor], Calverton, Maryland, USA, 2001).
Gambia	2013	CBH*	DHS Program†	Admin 2	The Gambia Bureau of Statistics & ICF International. The Gambia Demographic and Health Survey 2013 [Dataset] GMBR60DT. (ICF International [Distributor], Banjul, The Gambia and Rockville, Maryland, USA, 2014).
Gambia	2006	SBH	UNICEF	Admin 2	Gambia Bureau of Statistics & United Nations Children's Fund. The Gambia Multiple Indicator Cluster Survey 2005-2006 [Dataset]. 263pp (Bakau, The Gambia, 2007).
Gambia	2000	SBH	UNICEF	Admin 1	Central Statistics Department (Gambia) & United Nations Children's Fund. The Gambia Multiple Indicator Cluster Survey Report, 2000 [Dataset]. 116pp (New York, USA, 2002).
Ghana	2014	CBH*	DHS Program†	Point	Ghana Statistical Service, Ghana Health Service & ICF International. Ghana Demographic and Health Survey 2014 [Dataset] GHBR71DT. 530pp (ICF International [Distributor], Rockville, Maryland, USA, 2015).
Ghana	2011	CBH*	UNICEF	Point	Ghana Statistical Service. Ghana Multiple Indicator Cluster Survey with an Enhanced Malaria Module and Biomarker, 2011 [Dataset]. (Accra, Ghana, 2012).
Ghana	2011	SBH	UNICEF	Point	Institute of Statistical Social and Economic Research & United Nations Children's Fund. Ghana-Accra, Multiple Indicator Cluster Survey in 5 High Densly Populated Localities, 2010-2011 [Dataset]. 273pp (Institute of Statistical Social and Economic Research, Accra, Ghana, 2012).
Ghana	2010	SBH	IPUMS INTERNATIONAL	Admin 2	Ghana Statistical Service & Minnesota Population Center. Ghana Census 2010 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. (University of Minnesota, Minneapolis, USA).
Ghana	2008	CBH	DHS Program†	Admin 4	Ghana Statistical Service, Ghana Health Service & Macro International. Ghana Maternal Health Survey 2007 [Dataset] GHIQ51DT. 243pp (ICF International [Distributor], Calverton, Maryland, USA, 2009).
Ghana	2008	CBH*	DHS Program†	Point	Ghana Statistical Service, Ghana Health Service & ICF Macro. Ghana Demographic and Health Survey 2008 [Dataset] GHBR5ADT. (ICF International [Distributor], Accra, Ghana, 2009).
Ghana	2006	SBH	UNICEF	Admin 1	Ministry of Health (Ghana), Ghana Statistical Service, and United Nations Children's Fund. Ghana Multiple Indicator Cluster Survey 2006 [Dataset]. United Nations Children's Fund, New York, USA, pp 273
Ghana	2005	SBH	Ghana Statistical Service	Admin 1	Ghana Statistical Service. Ghana Living Standards Measurement Survey 2005-2006 [Dataset]. (Accra, Ghana, 2006).
Ghana	2003	CBH*	DHS Program†	Point	Ghana Statistical Service, Noguchi Memorial Institute for Medical Research & ORC Macro International. Ghana Demographic and Health Survey 2003 [Dataset] GHBR4BDT. (ICF International [Distributor], Calverton, Maryland, USA, 2004).
Ghana	2000	SBH	IPUMS INTERNATIONAL	Admin 2	Ghana Statistical Service & Minnesota Population Center. Ghana Population and Housing Census 2000 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. (University of Minnesota, Minneapolis, USA).
Ghana	1999	CBH*	DHS Program†	Point	Ghana Statistical Service & Macro International Inc. Ghana Demographic and Health Survey 1998 [Dataset] GHBR41DT. (ICF International [Distributor], Calverton, Maryland, USA, 1999).
Ghana	1999	SBH	Ghana Statistical Service	Admin 2	Ghana Statistical Service. Ghana Living Standards Survey 1998-1999 [Dataset]. (1999).
Guinea	2012	CBH*	DHS Program†	Point	Institut National de la Statistique (Guinée) & ICF International. Enquête Démographique et de Santé et à Indicateurs Multiples (EDS-MICS) [Dataset] GNBR61DT. (ICF International [Distributor], Calverton, Maryland, USA, 2013).
Guinea	2005	CBH*	DHS Program†	Point	Direction Nationale de la Statistique (Guinée) & ORC Macro. Enquête Démographique et de Santé, Guinée 2005 [Dataset] GNBR52DT. (ICF International [Distributor], Calverton, Maryland, USA, 2006).
Guinea	1999	CBH*	DHS Program†	Point	Direction Nationale de la Statistique (Guinée) & Macro International Inc. Enquête Démographique et de Santé, Guinée 1999 [Dataset] GNBR41DT. (ICF International [Distributor], Calverton, Maryland, USA, 2000).

Country	Year	Data Type	Source	Geographic level	Citation
Guinea Bissau	2014	CBH*	UNICEF	Admin 1	National Statistics Institute (Guinea-Bissau) & United Nations Children's Fund (UNICEF). Guinea-Bissau Multiple Cluster Indicator Survey 2014 [Dataset]. (New York, USA, 2016).
Guinea Bissau	2006	SBH	UNICEF	Admin 1	Ministère de l`Economie - Secrétariat d`Etat du Plan et à l`Intégration Régionale. Enquête par Grappes à Indicateurs Multiples, Guinée-Bissau, 2006 [Dataset]. 259pp (Bissau, Guinée-Bissau, 2006).
Guinea Bissau	2000	SBH	UNICEF	Admin 1	Secretary State of Planning, National Institute of Statistics and Census & United Nations Children's Fund. Guinea-Bissau Multiple Indicator Cluster Survey 2000 [Dataset]. 31pp (New York, USA, 2000).
Kenya	2015	SBH	DHS Program†	Point	National Malaria Control Programme, Ministry of Health, Kenya National Bureau of Statistics & ICF International. Kenya Malaria Indicator Survey 2015 [Dataset] KEIR7HDT. 165pp (ICF International [Distributor], Nairobi, Kenya, and Rockville, Maryland, USA, 2016).
Kenya	2014	CBH*	DHS Program†	Point	Kenya National Bureau of Statistics et al. Kenya Demographic and Health Survey 2014 [Dataset] KEBR70DT. (ICF International [Distributer], Rockville, Maryland, USA, 2015).
Kenya	2014	CBH*	UNICEF	Point	Kenya National Bureau of Statistics, Population Studies and Research Institute, University of Nairobi (Kenya), United Nations Children's Fund (UNICEF). Kenya - Bungoma County Multiple Indicator Survey 2013-2014 [Dataset]. (New York, USA, 2015).
Kenya	2014	CBH*	UNICEF	Point	Kenya National Bureau of Statistics, Population Studies and Research Institute, University of Nairobi (Kenya), United Nations Children's Fund (UNICEF). Kenya - Kakamega County Multiple Indicator Survey 2013-2014 [Dataset]. (New York, USA, 2015).
Kenya	2014	CBH*	UNICEF	Point	Kenya National Bureau of Statistics, Population Studies and Research Institute, University of Nairobi (Kenya), United Nations Children's Fund (UNICEF). Kenya - Turkana County Multiple Indicator Survey 2013-2014 [Dataset]. (New York, USA, 2015).
Kenya	2011	CBH*	UNICEF	Point	Kenya National Bureau of Statistics. Nyanza Province Multiple Indicator Cluster Survey 2011 [Dataset]. (Kenya National Bureau of Statistics, Nairobi, Kenya, 2013).
Kenya	2009	CBH*	DHS Program†	Point	Kenya National Bureau of Statistics & ICF Macro. Kenya Demographic and Health Survey 2008-09 [Dataset] KEBR52DT. (ICF International [Distributor], Calverton, Maryland, USA, 2010).
Kenya	2009	SBH	IPUMS INTERNATIONAL	Admin 1	Minnesota Population Center & Kenya National Bureau of Statistics. Kenya Population Census 2009 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. (University of Minnesota, Minneapolis, USA, 2013).
Kenya	2009	SBH	Kenya National Bureau of Statistics	Admin 2	Kenya National Bureau of Statistics, USAID, United Nations Population Fund (UNFPA), United States Census Bureau. Kenya Population and Housing Census 2009. (Kenya National Bureau of Statistics, Nairobi, Kenya 2010).
Kenya	2008	CBH	UNICEF	Point	United Nations Children's Fund & Kenya National Bureau of Statistics. Kenya - Eastern Province Multiple Indicator Cluster Survey 2008 [Dataset]. (Kenya National Bureau of Statistics, Nairobi, Kenya, 2009).
Kenya	2007	SBH	Kenya National Bureau of Statistics	Point	National AIDS/STI Control Programme. Kenya AIDS Indicator Survey: Final Report. 384 (Nairobi, Kenya, 2009).
Kenya	2007	SBH	Kenya National Bureau of Statistics	Point	Centers for Disease Control and Prevention (CDC), KEMRI Wellcome Trust Research Programme (KWTRP), Kenya National Bureau of Statistics, Ministry of Public Health and Sanitation (Kenya), National Coordinating Agency for Population and Development (Kenya), Population Services International (PSI). Kenya Malaria Indicator Survey 2007 [Dataset].
Kenya	2007	SBH	UNICEF	Point	United Nations Children's Fund & Kenya National Bureau of Statistics. Kenya - North Eastern Province Multiple Indicator Cluster Survey 2007 [Dataset]. (Kenya National Bureau of Statistics, Nairobi, Kenya).
Kenya	2006	SBH	Central Bureau of Statistics (Kenya)	Point	Central Bureau of Statistics (Kenya), UK Department for International Development (DFID), United States Agency for International Development (USAID), European Union (EU), Danish International Development Agency (DANIDA), World Bank (WB), et al. Kenya Integrated Household Budget Survey 2005-2006 [Dataset]. (Nairobi, Kenya).
Kenya	2003	CBH*	DHS Program†	Point	Central Bureau of Statistics (Kenya), Ministry of Health (Kenya) & ORC Macro. Kenya Demographic and Health Survey 2003 [Dataset] KEBR42DT. (ICF International [Distributor], Calverton, Maryland, USA, 2004).

Country	Year	Data Type	Source	Geographic level	Citation
Kenya	2000	SBH	UNICEF	Point	Central Bureau of Statistics (Kenya) & United Nations Children's Fund. Kenya Multiple Indicator Cluster Survey 2000 [Dataset]. 31pp (New York, USA, 2000).
Kenya	1999	SBH	IPUMS INTERNATIONAL	Admin 2	Central Bureau of Statistics (Kenya) & Minnesota Population Center. Kenya Population and Housing Census 1999 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. (University of Minnesota, Minneapolis, USA, 2001).
Kenya	1999	SBH	Central Bureau of Statistics (Kenya)	Point	Central Bureau of Statistics (Kenya), UK Department for International Development (DFID), United Nations Development Programme (UNDP), United Nations Population Fund (UNFPA), United States Agency for International Development (USAID). Kenya Population and Housing Census 1999 [Dataset].
Kenya	1998	CBH*	DHS Program†	Point	National Council for Population and Development, Central Bureau of Statistics, Office of the Vice President and Ministry of Planning and National Development (Kenya) & Macro International Inc. Kenya Demographic and Health Survey 1998 [Dataset] KEBR3ADT. (ICF International [Distributor], Calverton, Maryland, USA, 1999).
Lesotho	2014	CBH*	DHS Program†	Point	Ministry of Health and Social Welfare (Lesotho) & ICF International. Lesotho Demographic and Health Survey 2014 [dataset] LSBR71DT. (ICF International [Distributor], Fairfax, USA, 2016).
Lesotho	2010	CBH*	DHS Program†	Point	Ministry of Health and Social Welfare (Lesotho) & ICF Macro. Lesotho Demographic and Health Survey 2009 [Dataset] LSBR60DT. (ICF International [Distributor], Maseru, Lesotho, 2010).
Lesotho	2005	CBH*	DHS Program†	Point	Ministry of Health and Social Welfare (Lesotho), Bureau of Statistics (Lesotho) & ORC Macro. Lesotho Demographic and Health Survey 2004 [Dataset] LSBR41DT. (ICF International [Distributor], Calverton, Maryland, USA, 2005).
Lesotho	2000	SBH	UNICEF	Admin 1	Bureau of Statistics (Lesotho) & United Nations Children's Fund. 2000 End Decade Multiple Indicator Cluster Survey (EMICS) [Dataset]. 100pp (New York, USA, 2002).
Liberia	2013	CBH*	DHS Program†	Point	Liberia Institute of Statistics and Geo-Information Services, Ministry of Health and Social Welfare (Liberia), National AIDS Control Program (Liberia) & ICF International. Liberia Demographic and Health Survey 2013 [Dataset] LBBR6ADT. (ICF International [Distributor], Monrovia, Liberia, 2014).
Liberia	2011	SBH	DHS Program†	Point	National Malaria Control Program (Liberia), Ministry of Health and Social Welfare, Liberia Institute of Statistics and Geo-Information Services & ICF International. Liberia Malaria Indicator Survey 2011 [Dataset] LBBR61DT. 124pp (ICF International [Distributor], Monrovia, Liberia, 2012).
Liberia	2009	CBH*	DHS Program†	Point	National Malaria Control Program (Liberia), Ministry of Health and Social Welfare, Liberia Institute of Statistics and Geo-Information Services & ICF Macro. Liberia Malaria Indicator Survey 2009 [Dataset] LBBR5ADT. (ICF International [Distributor], Monrovia, Liberia, 2009).
Liberia	2008	SBH	IPUMS INTERNATIONAL	Admin 2	Liberia Institute for Statistics and Geo-information Services & Minnesota Population Center. Liberia Census 2008 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. (University of Minnesota, Minneapolis, USA).
Liberia	2007	CBH*	DHS Program†	Point	Liberia Institute of Statistics and Geo-Information Services (Liberia), Ministry of Health and Social Welfare (Liberia), National AIDS Control Program (Liberia) & Macro International Inc. Liberia Demographic and Health Survey 2007 [Dataset] LBBR51DT. (ICF International [Distributor], Monrovia, Liberia, 2008).
Madagascar	2016	SBH	DHS Program†	Admin 1	ICF International, Ministry of Public Health (Madagascar), National Institute of Statistics (Madagascar), National Program for the Fight Against Malaria (PNLP) (Madagascar), Pasteur Institute of Madagascar (IPM). Madagascar Malaria Indicator Survey 2016 [Dataset]. (ICF International [Distributor], Fairfax, USA, 2017).
Madagascar	2013	SBH	DHS Program†	Point	Institut National de la Statistique, Programme National de lutte contre le Paludisme, Institut Pasteur de Madagascar & ICF International. Enquête sur les Indicateurs du Paludisme (EIPM) 2013 [Dataset] MDBR6HDT. 179pp (ICF International [Distributor], Calverton, Maryland, USA, 2013).
Madagascar	2012	CBH*	UNICEF	Point	Institut National de la Statistique & Fonds des Nations Unies pour l'enfance. Madagascar Sud Enquête par Grappes à Indicateurs Multiples (MICS) 2012 [Dataset]. (Fonds des Nations Unies pour l'enfance (UNICEF), New York, USA, 2013).
Madagascar	2011	SBH	DHS Program†	Point	Institut National de la Statistique, Programme National de lutte contre le Paludisme, Institut Pasteur de Madagascar & ICF International. Enquête sur les Indicateurs du Paludisme à Madagascar (EIPMD) 2011 [Dataset] MDBR61DT. (ICF International [Distributor], Calverton, Maryland, USA, 2012).
Madagascar	2009	CBH*	DHS Program†	Point	Institut National de la Statistique & ICF Macro. Enquête Démographique et de Santé de Madagascar 2008-2009 [Dataset] MDBR51DT. (ICF International [Distributor], Antananarivo, Madagascar, 2010).

Country	Year	Data Type	Source	Geographic level	Citation
Madagascar	2004	CBH*	DHS Program†	Admin 2	Mariko, S. & Rabeza, V. Enquête de Base sur la Santé de la Reproduction et la Survie des Enfants dans les zones d'intervention USAID, à Madagascar - EBSRSE 2003-2004 [Dataset] MDBR41DT. (ICF International [Distributor], Calverton, Maryland, USA 2005).
Madagascar	2000	SBH	UNICEF	Admin 1	Institut National de la Statistique & Fonds des Nations Unies pour l'Enfrance. Enquete a Indicateurs Multiples MICS 2000 Madagascar [Dataset]. 237pp (New York, USA, 2001).
Malawi	2016	CBH	DHS Program†	Point	ICF International, Ministry of Health (Malawi), National Statistical Office of Malawi. Malawi Demographic and Health Survey 2015-2016 [Dataset]. (ICF International [Distributor], Fairfax, USA, 2017).
Malawi	2014	CBH*	UNICEF	Admin 2	National Statistical Office. Malawi MDG Endline Survey 2014. 684pp (Zomba, Malawi, 2015).
Malawi	2014	SBH	DHS Program†	Point	National Malaria Control Programme (Malawi) & ICF International. Malawi Malaria Indicator Survey (MIS) 2014 [Dataset] MWIR71DT. 124pp (ICF International [Distributor], Lilongwe, Malawi, and Rockville, Maryland, USA, 2014).
Malawi	2012	SBH	DHS Program†	Point	National Malaria Control Programme (Malawi) & ICF International. Malawi Malaria Indicator Survey (MIS) 2012 [Dataset] MWBR6HDT. 115pp (ICF International [Distributor], Lilongwe, Malawi, and Calverton, Maryland, USA, 2012).
Malawi	2010	CBH*	DHS Program†	Point	National Statistical Office & ICF Macro. Malawi Demographic and Health Survey 2010 [Dataset] MWBR61DT. (ICF International [Distributor], Zomba, Malawi, and Calverton, Maryland, USA, 2011).
Malawi	2008	SBH	IPUMS INTERNATIONAL	Admin 2	National Statistical Office (Malawi) & Minnesota Population Center. Malawi Population and Housing Census 2008 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. (University of Minnesota, Minneapolis, USA, 2011).
Malawi	2006	CBH*	UNICEF	Admin 2	National Statistical Office & United Nations Children's Fund. Malawi Multiple Indicator Cluster Survey 2006 [Dataset]. (National Statistical Office & United Nations Children's Fund, Lilongwe, Malawi, 2008).
Malawi	2005	CBH*	DHS Program†	Point	National Statistical Office (Malawi) & ORC Macro. Malawi Demographic and Health Survey 2004 [Dataset] MWBR4DDT. (ICF International [Distributor], Calverton, Maryland, USA, 2005).
Malawi	2000	CBH*	DHS Program†	Point	National Statistical Office (Malawi) & ORC Macro. Malawi Demographic and Health Survey 2000 [Dataset] MWBR41DT. (ICF International [Distributor], Zomba, Malawi and Calverton, Maryland, USA, 2001).
Malawi	1998	SBH	IPUMS INTERNATIONAL	Admin 2	National Statistical Office (Malawi) & Minnesota Population Center. Malawi Population and Housing Census 1998 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. (University of Minnesota, Minneapolis, USA, 2011).
Mali	2015	SBH	DHS Program†	Point	Programme National de Lutte contre le Paludisme, Institut National de la Statistique, INFO-STAT, Institut National de la Recherche en Santé Publique & ICF International. Enquête sur les Indicateurs du Paludisme au Mali (EIPM) 2015 [Dataset] MLIR70SV. 180pp (ICF International [Distributor], Rockville, Maryland, USA, 2016).
Mali	2013	CBH*	DHS Program†	Point	Cellule de Planification et de Statistique, Institut National de la Statistique , INFO-STAT & ICF International. Enquête Démographique et de Santé au Mali 2012-2013 [Dataset] MLBR6HDT. (ICF International [Distributor], Rockville, Maryland, USA, 2014).
Mali	2009	SBH	IPUMS INTERNATIONAL	Admin 3	Central Census Bureau (Mali) & Minnesota Population Center. Mali Census 2009 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. (University of Minnesota, Minneapolis, USA, 2009).
Mali	2006	CBH*	DHS Program†	Point	Cellule de Planification et de Statistique du Ministère de la Santé, Direction Nationale de la Statistique et de l'Informatique du Ministère de l'Économie de l'Industrie et du Commerce & Macro International Inc. Enquête Démographique et de Santé du Mali 2006 [Dataset] MLBR53DT. (ICF International [Distributor], Calverton, Maryland, USA, 2007).
Mali	2001	CBH*	DHS Program†	Point	Cellule de Planification et de Statistique du Ministère de la Santé, Direction Nationale de la Statistique et de l'Informatique & ORC Macro. Enquête Démographique et de Santé au Mali 2001 [Dataset] MLBR41DT. (ICF International [Distributor], Calverton, Maryland, USA, 2002).
Mali	1998	SBH	IPUMS INTERNATIONAL	Admin 3	Central Census Bureau (Mali) & Minnesota Population Center. Mali General Population and Housing Census 1998 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. (University of Minnesota, Minneapolis, USA).

Country	Year	Data Type	Source	Geographic level	Citation
Mauritania	2011	CBH*	UNICEF	Admin 3	Office National de la Statistique (Mauritania) & Fonds des Nations unies pour l'Enfance Mauitanie Enquête par grappes à indicateurs multiples 2011 [Dataset]. (New York, USA, 2014).
Mauritania	2007	SBH	UNICEF	Admin 3	Office National de la Statistique (Mauritania) & Fonds des Nations Unies pour l'Enfrance. Mauritanie Enquête par Grappes à Indicateurs Multiples 2007 [Dataset]. 221pp (New York, USA, 2008).
Mauritania	2004	CBH*	DHS Program†	Admin 1	Isselmou & Ould, A. Enquête sur la Mortalité Infantile et le Paludisme (EMIP) 2003-2004 [Dataset] MRIQ4AFL. (ICF International [Distributor], Calverton, Maryland, USA, 2004).
Mauritania	2001	CBH*	DHS Program†	Admin 1	Office National de la Statistique (Mauritanie) & Macro, O. Enquête Démographique et de Santé Mauritanie 2000-2001 [Dataset]. (ICF Internations; [Distributor], Calverton, Maryland, USA, 2001).
Morocco	2011	SBH	PAPFAM	Admin 2	Ministry of Health (Morocco), Pan Arab Project for Family Health (PAPFAM), United Nations Children's Fund (UNICEF), United Nations Population Fund (UNFPA), World Health Organization (WHO). Morocco National Survey on Population and Family Health 2010-2011 [Dataset].
Morocco	2004	CBH*	DHS Program†	Point	Ministère de la Santé (Maroc), ORC Macro & Ligue des États Arabes. Enquête sur la Population et la Santé Familiale (EPSF) 2003-2004 [Dataset] MABR43DT. (ICF International [Distributor], Calverton, Maryland, USA, 2005).
Morocco	2004	SBH	IPUMS INTERNATIONAL	Admin 2	Minnesota Population Center & High Commission for Planning (Morocco). Morocco Population and Housing Census 2004 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. (University of Minnesota, Minneapolis, USA, 2012).
Mozambique	2011	CBH*	DHS Program†	Point	Ministerio da Saude, Instituto Nacional de Estatística & ICF International. Moçambique Inquérito Demográfico e de Saúde 2011 [Dataset] MZBR62DT. (ICF International [Distributor], Calverton, Maryland, USA, 2013).
Mozambique	2009	CBH	UNICEF	Admin 1	De Araujo, S. N. et al. Final Report of the Multiple Indicator Cluster Survey, 2008 [Dataset]. (Maputo, Mozambique, 2009).
Mozambique	2009	SBH	DHS Program†	Point	Instituto Nacional de Saúde, Instituto Nacional de Estatística & ICF Macro. Inquérito Nacional de Prevalência, Riscos Comportamentais e Informação sobre o HIV e SIDA em Moçambique 2009 [Dataset] MZIR51DT. (ICF International [Distributor], Calverton, Maryland, USA, 2010).
Mozambique	2007	SBH	IPUMS INTERNATIONAL	Admin 3	Minnesota Population Center & National Institute of Statistics and Demography (Mozambique). Integrated Public Use Microdata Series, International: [Machine-readable database]. (University of Minnesota, Minneapolis, USA, 2007).
Mozambique	2003	CBH*	DHS Program†	Admin 1	Instituto Nacional de Estatística (Moçambique) & ORC Macro. Mocambique Inquérito Demográfico e de Saúde 2003 [Dataset] MZBR41DT. (ICF International [Distributor], Calverton, Maryland, USA, 2005).
Namibia	2013	CBH*	DHS Program†	Point	The Namibia Ministry of Health and Social Services & ICF International. The Namibia Demographic and Health Survey 2013 [Dataset] NMBR61DT. (ICF International [Distributor], Windhoek, Namibia, and Rockville, Maryland, USA, 2014).
Namibia	2007	CBH*	DHS Program†	Point	Ministry of Health and Social Services (Namibia) & Macro International Inc. Namibia Demographic and Health Survey 2006-07 [Dataset] NMBR51DT. (ICF International [Distributor], Windhoek, Namibia and Calverton, Maryland, USA, 2008).
Namibia	2000	CBH*	DHS Program†	Point	Ministry of Health and Social Services (Namibia). Namibia Demographic and Health Survey 2000 [Dataset] NMBR41DT. (ICF International [Distributor], Windhoek, Namibia, 2003).
Niger	2012	CBH*	DHS Program†	Admin 1	Institut National de la Statistique & ICF International. Enquête Démographique et de Santé et à Indicateurs Multiples du Niger 2012 [Dataset] NIBR61DT. (ICF International [Distributor], Calverton, Maryland, USA 2013).
Niger	2006	CBH*	DHS Program†	Admin 1	Institut National de la Statistique & Macro International Inc. Enquête Démographique et de Santé et à Indicateurs Multiples du Niger 2006 [Dataset] NIBR51DT. (ICF International [Distributor], Calverton, Maryland, USA 2007).
Niger	2000	SBH	UNICEF	Admin 1	Republique du Niger & Fonds des Nations Unies pour l'Enfrance. Enquete a indicateurs multiples 2000 (MICS2) [Dataset]. 175pp (Niamey, Niger, 2000).
Niger	1998	CBH*	DHS Program†	Point	Sabine, A., Seroussi, M., Kourguéni, A. I., Koché, H. & Barrère, B. Enquête Démographique et de Santé, Niger 1998 [Dataset] NIBR31DT. (ICF International [Distributor], Calverton, Maryland, USA, 1998).

Country	Year	Data Type	Source	Geographic level	Citation
Nigeria	2015	SBH	DHS Program†	Point	ICF International, National Bureau of Statistics (Nigeria), National Malaria Control Programme (Nigeria), National Population Commission of Nigeria. Nigeria Malaria Indicator Survey 2015 [Dataset]. (ICF International [Distributor], Fairfax, USA, 2016)
Nigeria	2013	CBH*	DHS Program†	Point	National Population Commission (Nigeria) & ICF International. Nigeria Demographic and Health Survey 2013 [Dataset] NGBR6ADT. (ICF International [Distributor], Abuja, Nigeria, and Rockville, Maryland, USA, 2014).
Nigeria	2011	SBH	UNICEF	Admin 1	National Bureau of Statistics. Nigeria Multiple Indicator Cluster Survey 2011 Main Report [Dataset]. 420pp (Abuja, Nigeria, 2011).
Nigeria	2010	CBH*	DHS Program†	Point	National Population Commission (Nigeria), National Malaria Control Programme (Nigeria) & ICF International. Nigeria Malaria Indicator Survey 2010 [Dataset] NGBR61DT. (ICF International [Distributor], Abuja, Nigeria, 2012).
Nigeria	2008	CBH*	DHS Program†	Point	National Population Commission (Nigeria) & ICF Macro. Nigeria Demographic and Health Survey 2008 [Dataset] NGBR53DT. (ICF International [Distributor], Abuja, Nigeria, 2009).
Nigeria	2008	SBH	National Bureau of Statistics (Nigeria)	Admin 1	Central Bank of Nigeria, National Bureau of Statistics (Nigeria), Nigerian Communications Commission (NCC). Nigeria General Household Survey 2008 [Dataset].
Nigeria	2007	SBH	UNICEF	Admin 1	National Bureau of Statistics. Nigeria Multiple Indicator Cluster Survey 2007 Final Report [Dataset]. 294pp (Abuja, Nigeria, 2007).
Nigeria	2004	SBH	National Bureau of Statistics (Nigeria)	Admin 1	National Bureau of Statistics - Federal Government of Nigeria. Nigeria Living Standards Survey 2003-2004. (Nigeria, 2004).
Nigeria	2003	CBH*	DHS Program†	Point	National Population Commission (Nigeria) & ORC Macro. Nigeria Demographic and Health Survey 2003 [Dataset] NGBR4BDT. (ICF International [Distributor], Calverton, Maryland, USA, 2004).
Nigeria	1999	SBH	UNICEF	Admin 1	United Nations Children's Fund & National Bureau of Statistics (Nigeria). Multiple Indicator Cluster Survey (1999) Nigeria [Dataset]. (2002).
Rwanda	2015	CBH*	DHS Program†	Point	National Institute of Statistics of Rwanda, ICF International & Ministry of Health (Rwanda). Rwanda Demographic and Health Survey 2014-2015 [Dataset] RWBR70DT. (ICF International [Distributor], Fairfax, USA, 2016).
Rwanda	2013	SBH	DHS Program†	Admin 1	Malaria and Other Parasitic Diseases Division (MAL & OPD Division-RBC) [Rwanda] & ICF International. Rwanda Malaria Indicator Survey 2013 [Dataset] RWHR6QDT. 103pp (ICF International [Distributor], 2014, Rockville, Maryland, USA, 2014).
Rwanda	2012	SBH	National Institute of Statistics of Rwanda	Admin 2	National Institute of Statistics of Rwanda. Rwanda Population and Housing Census 2012 [Dataset]. (Kigali, Rwanda, 2015).
Rwanda	2011	CBH*	DHS Program†	Point	National Institute of Statistics of Rwanda & ICF International. Rwanda Demographic and Health Survey 2010 [Dataset] RWBR61DT. (ICF International [Distributor], Calverton, Maryland, USA, 2012).
Rwanda	2008	CBH*	DHS Program†	Point	Ministère de la Santé, Institut National de la Statistique du Rwanda & ICF Macro. Enquête Intermédiaire sur les indicateurs Démographiques et de Santé, Rwanda 2007-2008 [Dataset] RWBR5ADT. (ICF International [Distributor], Calverton, Maryland, USA, 2009).
Rwanda	2006	SBH	National Institute of Statistics of Rwanda	Admin 2	National Institue of Statistics Rwanda. Integrated Living Conditions Survey 2005/6. 57pp (Kigali, Rwanda, 2006).
Rwanda	2005	CBH*	DHS Program†	Point	Institut National de la Statistique du Rwanda & ORC Macro. Rwanda Demographic and Health Survey 2005 [Dataset] RWBR53DT. (ICF International [Distributor], Calverton, Maryland, USA, 2006).
Rwanda	2002	SBH	IPUMS INTERNATIONAL	Admin 1	National Census Commission (Rwanda) & Minnesota Population Center. Rwanda Population and Housing Census 2002 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. (University of Minnesota, Minneapolis, USA).
Rwanda	2000	CBH*	DHS Program†	Admin 1	Office National de la Population (Rwanda) & ORC Macro. Enquête Démographique et de Santé, Rwanda 2000 [Dataset] RWBR41DT. (ICF International [Distributor], Kigali, Rwanda et Calverton, Maryland, USA, 2001).

Country	Year	Data Type	Source	Geographic level	Citation
Senegal	2015	CBH*	DHS Program†	Admin 1	Cheikh Anta Diop University, ICF International, National Agency of Statistics and Demography (Senegal). Senegal Continuous Demographic and Health Survey 2015 [Dataset]. (ICF International [Distributor], Fairfax, USA, 2016).
Senegal	2014	CBH*	DHS Program†	Admin 1	Agence Nationale de la Statistique et de la Démographie (Sénégal) & ICF International. Sénégal : Enquête Démographique et de Santé Continue (EDS-Continue 2012-14), Rapport Régional [Dataset] SNBR70DT. (ICF International [Distributor], Rockville, Maryland, USA, 2015).
Senegal	2013	CBH*	DHS Program†	Point	Agence Nationale de la Statistique et de la Démographie (Sénégal) & ICF International. Enquête Démographique et de Santé Continue (EDS-Continue 2012-2013) [Dataset] SNBR6CDT. (ICF International [Distributor], Calverton, Maryland, USA, 2013).
Senegal	2011	CBH*	DHS Program†	Point	Agence Nationale de la Statistique et de la Démographie (Sénégal) & ICF International. Enquête Démographique et de Santé à Indicateurs Multiples au Sénégal (EDS-MICS) 2010-2011 [Dataset] SNBR61DT. (ICF International [Distributor], Calverton, Maryland, USA, 2012).
Senegal	2009	CBH*	DHS Program†	Point	Ndiaye, S. & Ayad, M. Enquête Nationale sur le Paludisme au Sénégal 2008-2009 [Dataset] SNBR5HDT. (ICF International [Distributor], Calverton, Maryland, USA, 2009).
Senegal	2006	SBH	DHS Program†	Admin 1	Ndiaye, Salif & Mohamed Ayad. Enquête Nationale sur le Paludisme au Sénégal 2006 [Dataset] SNIR50DT. (Macro International Inc [Distributor], Calverton, Maryland, USA, 2007).
Senegal	2005	CBH*	DHS Program†	Point	Ndiaye, S. & Ayad, M. Enquête Démographique et de Santé au Sénégal 2005 [Dataset] SNBR4HDT. (ICF International [Distributor], Calverton, Maryland, USA, 2006).
Senegal	2002	SBH	IPUMS INTERNATIONAL	Admin 2	Directorate of Forecasting and Statistics (Senegal) & Minnesota Population Center. Senegal General Population and Housing Census 2002 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. (University of Minnesota, Minneapolis, USA).
Senegal	1999	CBH*	DHS Program†	Admin 1	Groupe SERDHA, Macro International, Inc, Ministry of Health and Prevention (Senegal). Senegal Demographic and Health Survey 1999 [Dataset]. (Macro International, Inc. [Distributor], Calverton, USA).
Sierra Leone	2013	CBH*	DHS Program†	Point	Statistics Sierra Leone & ICF International. Sierra Leone Demographic and Health Survey 2013 [Dataset] SLBR61DT. (ICF International [Distributor], Freetown, Sierra Leone and Rockville, Maryland, USA, 2014).
Sierra Leone	2010	SBH	UNICEF	Admin 2	Statistics Sierra Leone & United Nations Children's Fund. Sierra Leone Multiple Indicator Cluster Survey 2010, Final Report [Dataset]. 253pp (Freetown, Sierra Leone, 2011).
Sierra Leone	2008	CBH*	DHS Program†	Point	Statistics Sierra Leone & ICF Macro. Sierra Leone Demographic and Health Survey 2008 [Dataset] SLBR51DT. (ICF International [Distributor], Calverton, Maryland, USA, 2009).
Sierra Leone	2005	SBH	UNICEF	Admin 2	Statistics Sierra Leone & United Nations Children's Fund. Sierra Leone Multiple Indicator Cluster Survey 2005, Final Report [Dataset]. 261pp (Freetown, Sierra Leone, 2007).
Sierra Leone	2004	SBH	IPUMS INTERNATIONAL	Point & Admin 3	Minnesota Population Center. Sierra Leone Census 2004 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. (University of Minnesota, Minneapolis, USA).
Sierra Leone	2000	SBH	UNICEF	Admin 1	Central Statistics Office (Sierra Leone) & United Nations Children's Fund. A Household Survey Report (MICS-2) [Dataset]. 144pp (New York, USA, 2000).
Somalia	2011	CBH*	UNICEF	Point	United Nations Children's Fund Somalia & Somaliland Ministry of Planning and National Development. Somaliland Multiple Indicator Cluster Survey 2011 [Dataset]. (Nairobi, Kenya, 2014).
Somalia	2011	CBH*	UNICEF	Admin 1	United Nations Children's Fund Somalia & Ministry of Planning and International Cooperation. Northeast Zone Multiple Indicator Cluster Survey 2011, [Dataset]. (Nairobi, Kenya, 2014).
Somalia	2006	CBH*	UNICEF	Admin 1	Pan Arab Project for Family Health & United Nations Children's Fund. Somalia Multiple Indicator Cluster Survey 2006 [Dataset]. (New York, USA, 2006).
South Africa	2011	CBH	DataFirst	Admin 1	Southern Africa Labour and Development Research Unit. National Income Dynamics Study 2010-2011, Wave 2 [Dataset]. (Cape Town, South Africa, 2012).

Country	Year	Data Type	Source	Geographic level	Citation
South Africa	2007	SBH	IPUMS INTERNATIONAL	Admin 3	Statistics South Africa & Minnesota Population Center. South Africa Community Survey 2007 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. (University of Minnesota, Minneapolis, USA).
South Africa	2001	SBH	IPUMS INTERNATIONAL	Admin 3	Statistics South Africa & Minnesota Population Center. South Africa Census 2001 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. (University of Minnesota, Minneapolis, USA).
South Africa	1998	CBH*	DHS Program†	Admin 1	Department of Health (South Africa), Macro International Inc & South African Medical Research Council. South Africa Demographic and Health Survey 1998 [Dataset] ZABR31DT. 420pp (Macro International, Inc [Distributor], Calverton, USA, 1998).
South Africa	1998	SBH	International Food Policy Research Institute (IFPRI)	Admin 1	International Food Policy Research Institute (IFPRI), University of Natal UoW, Data Research Africa (DRA), Policy and Praxis, Southern Africa Labour Development Research Unit (SALDRU) S, chool of Economics, et al. South Africa KwaZulu-Natal Income Dynamics Study 1998 [Dataset]. (Durban, South Africa).
South Sudan	2010	CBH	UNICEF	Admin 1	Ministry of Health & National Bureau of Statistics. South Sudan Household Survey 2010, [Dataset]. (Juba, South Sudan, 2013).
South Sudan	2008	SBH	IPUMS INTERNATIONAL	Admin 2	National Population Census Council (Sudan), Central Bureau of Statistics (Sudan), Southern Sudan Centre for Census, Statistics and Evaluation & Minnesota Population Center. Sudan Population and Housing Census 2008 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. (University of Minnesota, Minneapolis, USA, 2011).
Sudan	2014	CBH*	UNICEF	Admin 1	Central Bureau of Statistics & United Nations Children's Fund Sudan. Multiple Indicator Cluster Survey 2014 of Sudan [Dataset]. (United Nations Children's Fund and Central Bureau of Statistics [Distributor], Khartoum, Sudan, 2016).
Sudan	2010	CBH	UNICEF	Admin 1	Federal Ministry of Health & Central Bureau of Statistics. Sudan Household and Health Survey - Round 2, 2010, [Dataset]. (Khartoum, Republic of Sudan, 2012).
Sudan	2008	SBH	IPUMS INTERNATIONAL	Admin 3	National Population Census Council (Sudan), Central Bureau of Statistics (Sudan), Southern Sudan Centre for Census, Statistics and Evaluation & Minnesota Population Center. Sudan Population and Housing Census 2008 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. (University of Minnesota, Minneapolis, USA, 2011).
Sudan	2006	CBH	Pan Arab Project for Family Health (PAPFAM)	Admin 1	Ministry of Health (Southern Sudan), Federal Ministry of Health (Sudan), Southern Sudan Centre for Census, Statistics and Evaluation (SSCCSE), Central Bureau of Statistics (Sudan). Sudan Family Health Survey 2006 [Dataset].
Swaziland	2014	CBH	UNICEF	Admin 1	Central Statistical Office (Swaziland), United Nations Children's Fund (UNICEF), United Nations Educational, Scientific and Cultural Organization (UNESCO), United Nations Population Fund (UNFPA). Swaziland Multiple Indicator Cluster Survey 2014 [Dataset]. (New York, USA, 2016).
Swaziland	2010	CBH*	UNICEF	Admin 1	Central Statistical Office & United Nations Children's Fund. Swaziland Multiple Indicator Cluster Survey 2010 [Dataset]. (Mbabane, Swaziland, 2011).
Swaziland	2007	CBH*	DHS Program†	Point	Central Statistical Office (Swaziland) & Macro International Inc. Swaziland Demographic and Health Survey 2006-07 [Dataset] SZBR51DT. (ICF International [Distributor], Mbabane, Swaziland, 2008).
Swaziland	2000	SBH	UNICEF	Admin 1	Central Statistical Office (Swaziland) & United Nations Children's Fund. Swaziland Multiple Indicator Cluster Survey 2000 [Dataset]. 101pp (New York, USA, 2000).
Tanzania	2016	CBH*	DHS Program†	Point	ICF International, Ministry of Health (Zanzibar), Ministry of Health CD, Gender, Elderly and Children (MoHCDEC) (Tanzania), National Bureau of Statistics (Tanzania), Office of Chief Government Statistician (OCGS-Zanzibar). Tanzania Demographic and Health Survey 2015-2016 [Dataset]. (ICF International [Distributor], Fairfax, USA, 2016).
Tanzania	2012	SBH	IPUMS INTERNATIONAL	Admin 2	National Bureau of Statistics (Tanzania) & Minnesota Population Center. Tanzania Population and Housing Census 2012 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. (University of Minnesota, Minneapolis, USA, 2017).
Tanzania	2012	SBH	DHS Program†	Point	Tanzania Commission for AIDS (TACAIDS), Zanzibar AIDS Commission (ZAC), National Bureau of Statistics (NBS), Office of the Chief Government Statistician (OCGS) & ICF International. Tanzania HIV/AIDS and Malaria Indicator Survey 2011-12 [Dataset] TZBR6ASV. 16pp (ICF International [Distributor], Dar es Salaam, Tanzania, 2013).
Tanzania	2010	CBH*	DHS Program†	Point	National Bureau of Statistics (Tanzania) & ICF Macro. Tanzania Demographic and Health Survey 2010 [Dataset] TZBR63DT. (ICF International [Distributor], Dar es Salaam, Tanzania, 2011).

Country	Year	Data Type	Source	Geographic level	Citation
Tanzania	2008	CBH	DHS Program†	Point	Tanzania Commission for AIDS, Zanzibar AIDS Commission, National Bureau of Statistics, Office of the Chief Government Statistician & Macro International Inc. Tanzania HIV/AIDS and Malaria Indicator Survey 2007-08 [Dataset] TZBR51DT. (ICF International [Distributor], Dar es Salaam, Tanzania, 2008).
Tanzania	2005	CBH*	DHS Program†	Admin 1	National Bureau of Statistics (Tanzania) & ORC Macro. Tanzania Demographic and Health Survey 2004-05 [Dataset] TZBR4IDT. (ICF International [Distributor], Dar es Salaam, Tanzania, 2005).
Tanzania	2004	SBH	DHS Program†	Point	Tanzania Commission for AIDS (TACAIDS), National Bureau of Statistics (NBS) & ORC Macro. Tanzania HIV/AIDS Indicator Survey 2003-04 [Dataset] TZIR4ADT. (ORC Macro [Distributor], Calverton, Maryland, USA, 2005).
Tanzania	2002	SBH	IPUMS INTERNATIONAL	Admin 2	National Bureau of Statistics (Tanzania) & Minnesota Population Center. Tanzania Population and Housing Census 2002 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. (University of Minnesota, Minneapolis, USA).
Tanzania	1999	CBH*	DHS Program†	Point	National Bureau of Statistics (Tanzania) & Macro International Inc. Tanzania Reproductive and Child Health Survey 1999 [Dataset] TZBR41DT. (ICF International [Distributor], Calverton, Maryland, USA, 2000).
Togo	2013	CBH*	DHS Program†	Point	Ministère de la Planification du Développement et de l'Aménagement du Territoire, Ministère de la Santé & ICF International. Enquête Démographique et de Santé au Togo 2013-2014 [Dataset] TGBR61DT. (ICF International [Distributor], Rockville, Maryland, USA, 2015).
Togo	2010	SBH	UNICEF	Admin 1	Direction Générale de la Statistique et de la Comptabilité Nationale. Enquête par grappes à indicateurs multiples MICS Togo, 2010, Rapport final [Dataset]. 244pp (New York, USA, 2012).
Togo	2006	SBH	UNICEF	Admin 1	Direction Générale de la Statistique et de la Comptabilité Nationale. Résultats de l'enquête nationale à indicateurs multiples Togo 2006 [Dataset]. 304pp (New York, USA, 2007).
Togo	1998	CBH*	DHS Program†	Point	Anipah, K. et al. Enquete Demographique et de Sante, Togo 1998 [Dataset] TGBR31DT. (ICF International [Distributor], Calverton, Maryland, USA, 1999).
Uganda	2015	SBH	DHS Program†	Point	Uganda Bureau of Statistics & ICF International. Uganda Malaria Indicator Survey 2014-15 [Dataset] UGIR72DT. 139pp (ICF International [Distributor], Kampala, Uganda, and Rockville, Maryland, USA, 2015).
Uganda	2012	SBH	World Bank	Point	Uganda Bureau of Statistics. Uganda Living Standards Measurement Survey - Integrated Survey on Agriculture 2011-2012 [Dataset]. (Kampala, Uganda, 2013).
Uganda	2011	CBH*	DHS Program†	Point	Uganda Bureau of Statistics & ICF International Inc. Uganda Demographic and Health Survey 2011 [Dataset] UGBR60DT. (ICF International [Distributor], Kampala, Uganda and Calverton, Maryland, USA, 2012).
Uganda	2011	SBH	DHS Program†	Point	Centers for Disease Control and Prevention (CDC), ICF International, Ministry of Health (Uganda), Uganda Bureau of Statistics, Uganda Viral Research Institute. Uganda AIDS Indicator Survey 2011 [Dataset]. (ICF International [Distributor], Calverton, USA, 2012).
Uganda	2011	SBH	World Bank	Point	Uganda Bureau of Statistics. The Uganda National Panel Survey 2010/11 [Dataset]. (Kampala, Uganda, 2013)
Uganda	2010	CBH*	DHS Program†	Point	Uganda Bureau of Statistics & ICF Macro. Uganda Malaria Indicator Survey 2009 [Dataset] UGBR5HDT. (ICF International [Distributor], Calverton, Maryland, USA, 2010).
Uganda	2009	SBH	World Bank	Point	Uganda Bureau of Statistics. Uganda Living Standards Measurement Survey - Integrated Survey on Agriculture 2009-2010. (World Bank, Washington DC, USA).
Uganda	2006	CBH*	DHS Program†	Point	Uganda Bureau of Statistics & Macro International Inc. Uganda Demographic and Health Survey 2006 [Dataset] UGBR52DT. (ICF International [Distributor], Calverton, Maryland, USA, 2007).
Uganda	2002	SBH	IPUMS INTERNATIONAL	Admin 2	Uganda Bureau of Statistics & Minnesota Population Center. Uganda Population and Housing Census 2002 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. (University of Minnesota, Minneapolis, USA).
Uganda	2001	CBH*	DHS Program†	Point	Uganda Bureau of Statistics & ORC Macro. Uganda Demographic and Health Survey 2000-2001 [Dataset] UGBR41DT. (ICF International [Distributor], Calverton, Maryland, USA, 2001).

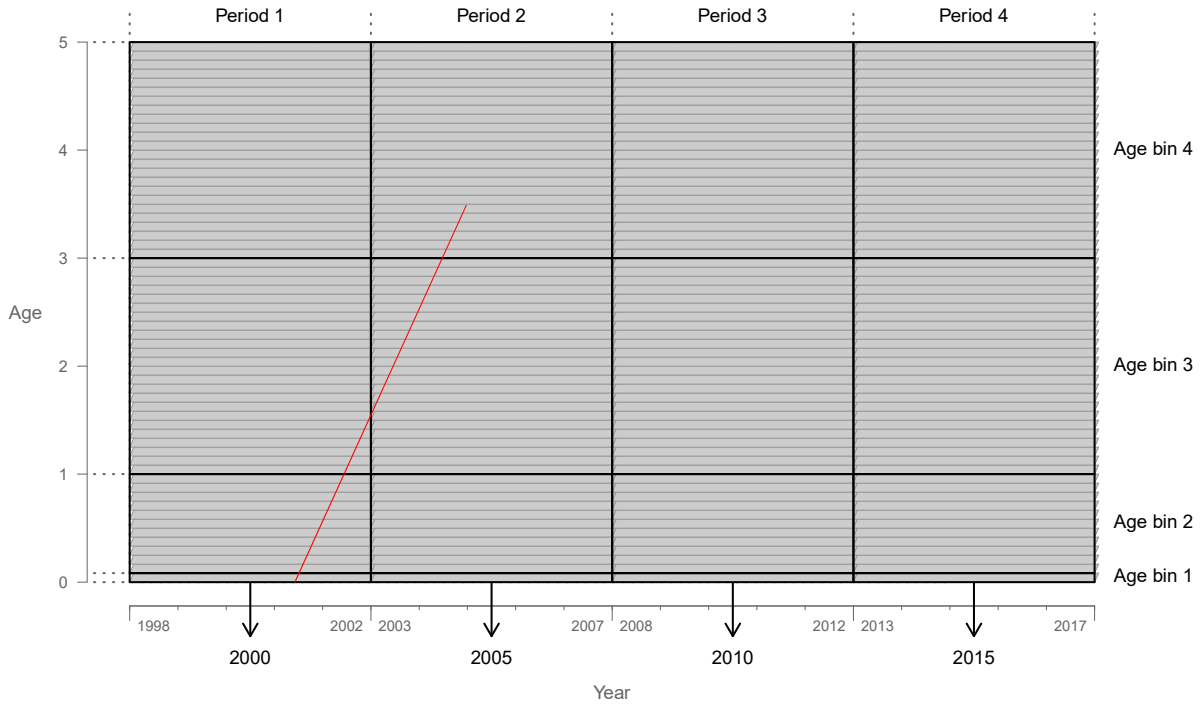
Country	Year	Data Type	Source	Geographic level	Citation
Zambia	2014	CBH*	DHS Program†	Point	Central Statistical Office (Zambia), Ministry of Health (Zambia) & ICF International. Zambia Demographic and Health Survey 2013-14 [Dataset] ZMBR61DT. (ICF International [Distributor], Rockville, Maryland, USA, 2014).
Zambia	2010	SBH	IPUMS INTERNATIONAL	Admin 3	Central Statistical Office (Zambia) & Minnesota Population Center. Zambia Census 2010 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. (University of Minnesota, Minneapolis, USA).
Zambia	2007	CBH*	DHS Program†	Point	Central Statistical Office, Ministry of Health, Tropical Diseases Research Centre, University of Zambia & Macro International Inc. Zambia Demographic and Health Survey 2007 [Dataset] ZMBR51DT. (ICF International [Distributor], Calverton, Maryland, USA, 2009).
Zambia	2002	CBH*	DHS Program†	Admin 1	Central Statistical Office (Zambia), Central Board of Health (Zambia) & ORC Macro. Zambia Demographic and Health Survey 2001-2002 [Dataset] ZMBR42DT. (ICF International [Distributor], Calverton, Maryland, USA, 2003).
Zambia	2000	SBH	IPUMS INTERNATIONAL	Admin 3	Central Statistical Office (Zambia) & Minnesota Population Center. Zambia Census 2000 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. (University of Minnesota, Minneapolis, USA).
Zimbabwe	2015	CBH*	DHS Program†	Point	Zimbabwe National Statistics Agency & ICF International. Zimbabwe Demographic and Health Survey 2015 [Dataset] ZWBR70DT. 575pp (ICF International [Distributor], Rockville, Maryland, USA, 2016).
Zimbabwe	2014	CBH*	UNICEF	Admin 1	Zimbabwe National Statistics Agency. Zimbabwe Multiple Indicator Cluster Survey 2014 [Dataset]. (Harare, Zimbabwe, 2015).
Zimbabwe	2012	SBH	Zimbabwe National Statistics Agency	Point & Admin 2	Zimbabwe National Statistics Agency. Zimbabwe Population Census 2012. 152pp (Harare, Zimbabwe, 2013).
Zimbabwe	2011	CBH*	DHS Program†	Point	Zimbabwe National Statistics Agency & ICF International. Zimbabwe Demographic and Health Survey 2010-11 [Dataset] ZWBR62DT. (ICF International [Distributor], Calverton, Maryland, USA, 2012).
Zimbabwe	2009	CBH*	UNICEF	Admin 1	Zimbabwe National Statistics Agency & United Nations Children's Fund. Zimbabwe Multiple Indicator Monitoring Survey (MIMS) 2009 [Dataset]. (Harare, Zimbabwe, 2010).
Zimbabwe	2006	CBH*	DHS Program†	Point	Central Statistical Office (Zimbabwe) & Macro International Inc. Zimbabwe Demographic and Health Survey 2005-06 [Dataset] ZWBR51DT. (ICF International [Distributor], Calverton, Maryland, USA, 2007).
Zimbabwe	2002	SBH	Zimbabwe National Statistics Agency	Admin 1	Central Statistical Office (Zimbabwe). Zimbabwe Population and Housing Census 2002. 173pp (Zimbabwe, 2004).
Zimbabwe	1999	CBH*	DHS Program†	Point	Central Statistical Office (Zimbabwe) & Macro International Inc. Zimbabwe Demographic and Health Survey 1999 [Dataset] ZWBR42DT. (ICF International [Distributor], Calverton, Maryland, USA, 2000).

4. Data outliers

Several surveys were identified for inclusion but were ultimately excluded from the final analysis for concerns about data quality. The 2007 Ethiopia census[2] was dropped due to 80% missingness in SBH data and unrealistic geographic distribution of mortality risk. A CBH dataset was dropped if the percent of children who died missing either birth or death date exceeded 10%, or exceeded 5% and had reported data quality issues. Such surveys were South Africa NIDS 2011, subnational MICS surveys from Kenyan counties Bungoma, Turkana, Kakamega (2013-2014), and Somalia NE Zone (2011), and the Djibouti PAPFAM survey 2012. Finally, we dropped data points from Diffa region (2012 Niger DHS) and Jonglei and Unity states (South Sudan 2010 MICS) as they were highly discordant with data points from the same surveys in nearby regions with similar sociodemographic profiles.

5. Tabulating CBH data

For each survey cluster we tabulated monthly exposures and deaths for each of four age bins over four separate five-year periods, as illustrated in Supplementary Figure 2. This tabulation process is similar to that used to compute under-5 mortality estimates for GBD 2016.[3] The exception to this is that we use four component bins: month 0, months 1-11, months 12-35, and months 36–59, whereas GBD uses six bins: 0 months, 1–11 months, 12–23 months, 24–36 months, 36–47 months, and 48–59 months. For areas larger than survey clusters, survey weights were used in the tabulation step to make weighted mortality rates for each of the component bins, which were then applied to the observed tabulated exposure months to estimate an equivalent number of deaths. Estimates for the 2015 period may be slightly overestimated due to the relatively smaller amount of data in the later years of the period.



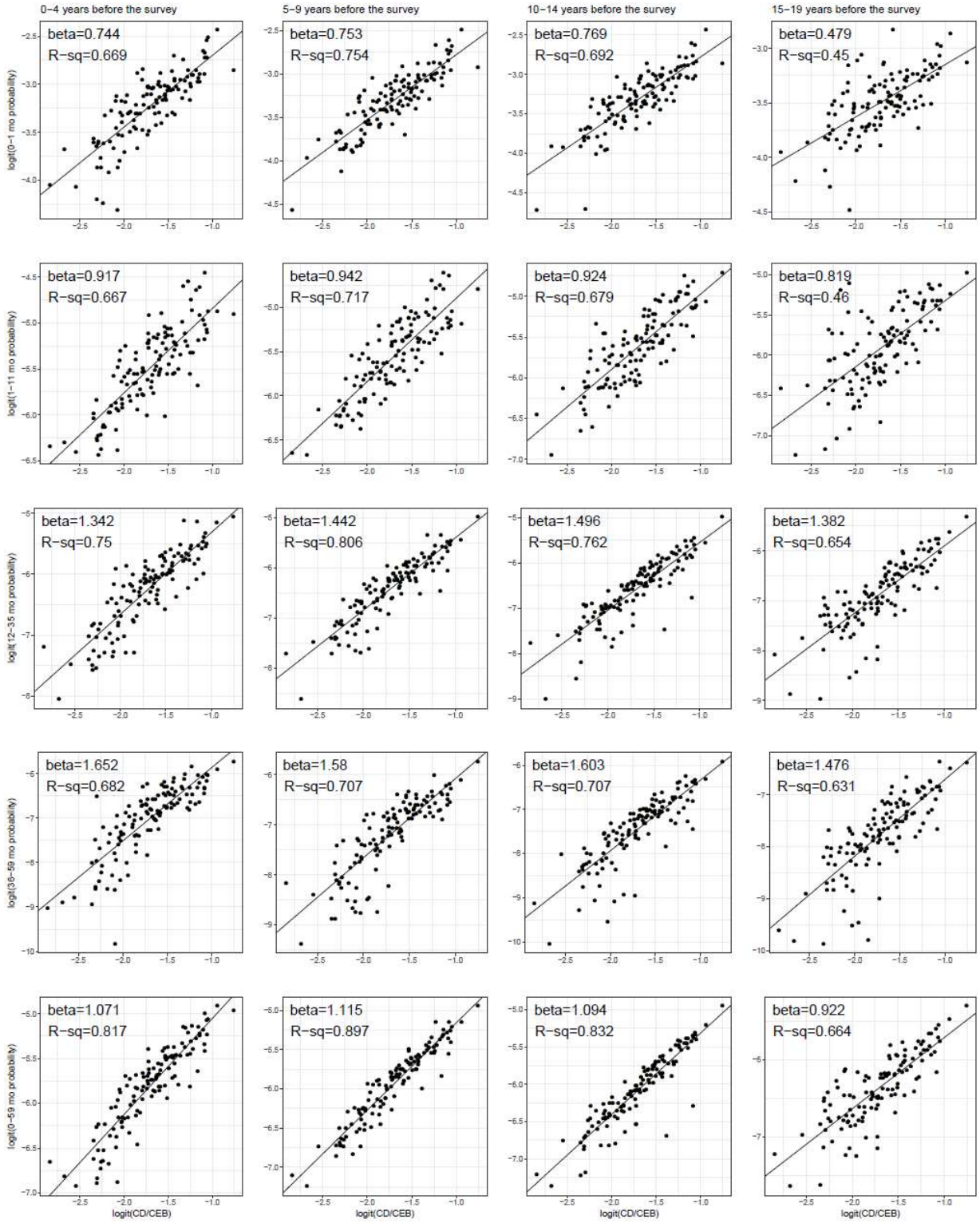
Supplementary Figure 7: Lexis diagram illustrating periods and age bins within which CBH data were tabulated. For each month (grey parallelograms) within each age bin/period pair, (rectangles bounded by solid black lines) we tabulated the number of exposures (number of children alive and entering the month of life within the given period) and number of deaths (exposures who died during that month) for each survey cluster, or geographic area comprising multiple clusters, where cluster-specific data were unavailable. For example, if we observed a living child who was 3.5 years old in a January 2004 survey, their life history would be represented with the red line in the figure, extending back to midway through 2001 when they were born. This child would have contributed exposure months in the cells (month/period combinations) crossed by their life history: (age bin 4/period 2005: 6 months; 3/2005: 18 months; 3/2000: 6 months, 2/2000: 12 months; and 1/2000, 1 month). If the same child had died at age 3.5 in January 2004, they would have contributed the same number of exposure months, but also a death event in age bin 4/period 2005.

6. Adjustment of summary birth history data

In a summary birth history (SBH) survey module, a woman is asked a minimum of two questions: the number of children ever born (CEB) to them and the number of those who have died (CD). While providing much less information than the full account of life histories available in CBH data, SBH is widely available. Of the 235 surveys for which we had microdata, 104 contained only summary birth histories, most notably in many Unicef MICS and census datasets.

To be usable for child mortality estimation, SBH needs to be adjusted such that they are representative of age- and time-specific mortality. This involves an adjustment of the rate as well as localizing the mortality rates in time.[4] Recent model-based improvements have been made on classical indirect methods[4–6] by relating the CBH- and SBH-derived rates from surveys where both are available, along with information on parity and maternal age.[7]

SBH and CBH survey modules capture different information on child births and deaths. For CBH data we can tabulate information for each child to directly estimate deaths and exposure months for any desired age bin and time period, while in SBH, the only mortality data available is the CD/CEB ratio for each mother interviewed. Ultimately, for this analysis, we wish to approximate CBH-style binomial samples from available information in SBH. The strong linear relationship between the logit of the CD/CEB ratio and the logit under-5 probability of death (5q0) has been noted in the past and used as motivation for a model-based approach to adjusting SBH to match CBH data.[7] In this analysis, we had the added challenge of dealing with data that need to be child-age disaggregated, which has not been well explored in the literature. Using surveys with both SBH and CBH modules, we found evidence that logit of CD/CEB had a strong relationship with the monthly probabilities of death tabulated in our analytical age groups from CBH data. The plots in Supplementary Figure 3 show the bivariate relationships across ages and lag-times between in surveys that collect both CBH and SBH data. These plots give strong motivation that a modelling approach should work for smaller age bins as well.



Supplementary Figure 8: Comparison of SBH and CBH mortality. Using surveys that collect both SBH and CBH mortality data, we find a strong relationship between the logit of the monthly probability of death for all of our analytical age bins, as well as with the full under-5 bin, with the logit of CD/CEB for each survey. The logit of CD/CEB is highly predictive of CBH tabulated age-specific monthly mortality probability going back up to 19 years before the survey. Lines drawn show a linear regression line, with a beta coefficient and R-squared provided. SBH = Summary birth history; CBH = Complete birth history; CD = Children died; CEB = Children ever born.

6.1 Data preparation

First, we collated data on over four million children from 119 datasets (mostly DHS) collected since 1998 (the beginning of our study period) that contain both SBH and CBH data, these are marked with an asterisk on Supplementary table 1. We tabulated the CBH deaths and exposure months at the age-period level as described in the section above. Thus, each survey produced up to 16 rows of data, one for each child age bin (month 0, months 1–11, months 12–35, and months 36–59) and time period bin (centred at 2000, 2005, 2010, 2015), each with the number of CBH-tabulated deaths and exposures and SBH-tabulated CD and CEB for the cluster.

We will refer to the CBH-SBH combined dataset as the training set, and the SBH-only data which we aim to adjust as the prediction dataset. The prediction dataset will be aggregated to the level of the smallest possible geography – either survey cluster or some identifiable administrative unit. We will refer from now on to this smallest geography simply as cluster.

In order to use a logit transformation of the CD/CEB ratio in the modelling step (described in the following subsection), we had to ensure the CD data did not include zeroes since the logit of 0 is negative infinity. To do so, we fitted the following Bayesian hierarchical shrinkage model to adjust death counts in surveys that contained clusters with no deaths recorded:

(1)

$$CD_i \sim Binomial(q_i, CEB_i)$$
$$logit(q_i) = \alpha + \gamma_i$$
$$\gamma_i \sim N(0, \sigma^2)$$

where *i* is cluster and γ is a normal independent random effect by cluster. Models were fit separately for each survey. A minimally informative prior, $\log\left(\frac{1}{\sigma^2}\right) \sim loggamma(1, 0.00005)$, was placed on σ. The effect of this adjustment was extremely small, correlation between original CD and adjusted was greater than 0.9999. We then used $\widehat{q_i}$ in place of the CD/CEB ratio in the full SBH adjustment model, described in the next section.

6.2 SBH adjustment models

We developed a two-stage regression model for cross-walking SBH data into age and period bin approximate binomial samples to match the CBH input data. In the first model we estimate $\hat{p}$, the bin specific monthly probability of death. In the second model we estimate $\widehat{N}$, the number of exposure months in that bin. From these two quantities we then simulate bin-specific deaths $\widehat{N^+}$. $\widehat{N}$ and $\widehat{N^+}$ are used as training data at SBH clusters in the geostatistical model described in detail later in this document. 10,000 $\widehat{N^+}$ draws are simulated from draws taken from the predictive posterior distributions of $\widehat{N}$ and $\hat{p}$ and the resulting variance is used to down-weight these observations when fitting the geostatistical model. We describe these models in detail below.

The $\hat{p}$ model was fit to the training set at the survey, age bin, period, and level, and then used to predict to SBH-only data at the smallest identifiable geography.

We fit the following Bayesian hierarchical logistic regression model to the training data:

(2)

$$N^{+}_{s,a,p} \sim Binomial\left(p_{s,a,p},\ N_{s,a,p}\right)$$

$$log\left(p_{s,a,p}/(1-p_{s,a,p})\right) = \beta_0 + logit(\hat{q}_s)\left(\beta_1 + Age\beta_2 + Lag_{s,p}\beta_3 + Age * Lag_{s,p}\beta_4\right) + \boldsymbol{X}_{s,a,p}\boldsymbol{\beta_{fe}} + \gamma_s + \gamma_{country}$$

$$\gamma_s \sim N(0, \sigma_s^2)$$

$$\gamma_{country} \sim N\left(0, \sigma^2_{country}\right)$$

where $N_{s,a,p}$ are the number of monthly exposures recorded in a survey *s*, age bin *a,* and time period *p*, and $N^{+}_{s,a,p}$ are the number of deaths resulting from all monthly exposures. $N_{s,a,p}$ and $N^{+}_{i,a,p}$ both come from direct tabulations of the CBH data. α_0 is a global intercept term. $\hat{q}_i$ is the zero-adjusted CD/CEB ratio described above in equation (1). The logit of CD/CEB is modified by interactions with child age bin, lag, and the interaction of child age bin and lag. Lag represents the number of periods between the survey to time of mortality risk. We further included fixed effects for all individual terms from the interactions as well as for mean maternal age, year, parity ratios between 15-19 and 20-24 year olds, parity ratios between 20-24 and 25-29 year olds, the proportions of mothers in the survey in various maternal age bins (15-19, 20-24, and 25-39), and the number of years of coverage the survey had in the period bin. We represent the values of these fixed effect covariates with the bolded matrix element $\boldsymbol{X}_{s,a,p}$ and their associated coefficients in the vector $\boldsymbol{\beta_{fe}}$. Survey weights were used to make estimates of the independent variables at areas larger than clusters in the prediction data set.

We fit another model to estimate $\widehat{N}$, the expected number of exposure months of life in each child age, period bin. For this second model, the target of inference was a number and not a ratio, so we chose to fit the model on the training set aggregated to the survey-admin1 level, and thus more similar to numbers seen in the prediction dataset. This model was very similar to the model for $\hat{p}$ described in equation (2) but with several key differences: 1) we fit to $N_{i,a,p}$ using a Poisson likelihood; 2) we used log(CEB) instead of logit(CD/CEB) as the main effect driving the adjustment; 3) we added a fixed effects term for log(CD) and a random effect on survey-admin1; and 4) we replaced the interaction with lag with an interaction with number of years in the period covered by the survey.

In addition, two prediction set censuses (Zimbabwe 2002, and Zimbabwe 2012) were only available in tabulated form. As such, for these three surveys we trained and predicted slightly modified $\hat{p}$ and $\widehat{N}$ models, using proportion of women instead of proportion of mothers fixed effects.

All models were fitted using R-INLA. The following minimally informative priors (http://www.r-inla.org/models/latent-models) were used: $\beta \sim N(0, 1000)$, $\log\left(\frac{1}{\sigma^2}\right) \sim loggamma(1, 0.00005)$.

6.2.1 Model fits of SBH adjustment models

Supplementary Table 2: SBH Adjustment model results for the p model. SBH=Summary birth history. Reference levels for factor variables lag and child age bin are lag 0 and child age bin 1.

Variable	Mean coef. estimate	2.5% UI	97.5% UI
Intercept	-4.3226	-5.629	-3.0253
Variable	**Mean coef. estimate**	**2.5% UI**	**97.5% UI**
CD/CEB	0.1168	0.026	0.2071
lag 1	0.2435	-0.0841	0.5709
lag 2	0.5205	0.1469	0.8936

lag 3	0.1055	-0.4127	0.6232
child age bin 2	-1.6537	-1.7677	-1.5399
child age bin 3	-1.5178	-1.6369	-1.3987
child age bin 4	-2.2989	-2.4504	-2.1474
mean maternal age	-0.0075	-0.0322	0.0173
period	0.123	0.0991	0.1465
proportion mothers 15-19	3.5498	1.1685	5.909
proportion mothers 20-24	1.5733	0.1497	2.9987
proportion mothers 25-39	0.9012	-0.3119	2.1217
P(15-19)/P(20-24)	-1.3741	-2.079	-0.6674
P(20-24)/P(25-25)	0.6768	0.0632	1.2865
years in period	-0.041	-0.051	-0.0309
CD/CEB * lag 1	0.1887	0.1317	0.2456
CD/CEB * lag 2	0.3763	0.3123	0.4403
CD/CEB * lag 3	0.3049	0.2217	0.388
CD/CEB * child age bin 2	0.414	0.3499	0.4781
CD/CEB * child age bin 3	0.9688	0.9002	1.0374
CD/CEB * child age bin 4	0.9449	0.8569	1.0331
lag 1 * child age bin 2	-0.1006	-0.2422	0.0409
lag 2 * child age bin 2	-0.0961	-0.2549	0.0625
lag 3 * child age bin 2	0.3137	0.1049	0.5224
lag 1 * child age bin 3	0.0649	-0.0832	0.2128
lag 2 * child age bin 3	-0.081	-0.2488	0.0866
lag 3 * child age bin 3	-0.2431	-0.4618	-0.0245
lag 1 * child age bin 4	-0.2064	-0.4000	-0.0129
lag 2 * child age bin 4	-0.4383	-0.6594	-0.2174
lag 3 * child age bin 4	-0.4778	-0.7686	-0.1871
lag 1 * mean maternal age	0.0086	-0.0017	0.0189
lag 2 * mean maternal age	0.0123	0.0000	0.0238
lag 3 * mean maternal age	0.0203	0.0041	0.0365
CD/CEB * lag 1 * child age bin 2	-0.076	-0.155	0.003
CD/CEB * lag 2 * child age bin 2	-0.1101	-0.1971	-0.0232
CD/CEB * lag 3 * child age bin 2	0.0212	-0.0885	0.1308
CD/CEB * lag 1 * child age bin 3	0.003	-0.0817	0.0876
CD/CEB * lag 2 * child age bin 3	-0.1392	-0.233	-0.0455
CD/CEB * lag 3 * child age bin 3	-0.3275	-0.4439	-0.2111
CD/CEB * lag 1 * child age bin 4	-0.0792	-0.1908	0.0323
CD/CEB * lag 2 * child age bin 4	-0.2944	-0.4183	-0.1706
CD/CEB * lag 3 * child age bin 4	-0.4218	-0.5768	-0.2667
Variance of country random effect	0.0101	0.0052	0.0212
Variance of survey random effect	0.0056	0.0037	0.0090

Supplementary Table 3: SBH Adjustment model results for the N model. SBH=Summary birth history. Reference levels for factor variables lag and child age bin are lag 0 and child age bin 1.

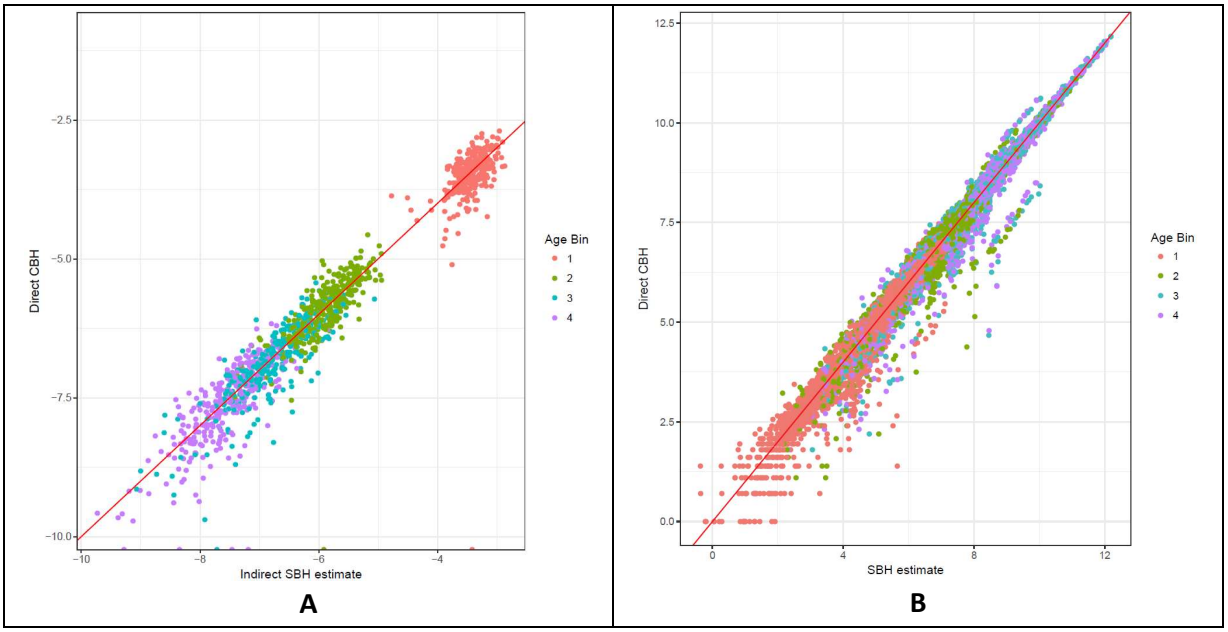
Variable	**Mean coef. estimate**	**2.5% UI**	**97.5% UI**

Intercept	4.6750	3.5904	5.7556
log(CEB)	0.9499	0.9069	0.9928
lag 1	-0.8378	-0.9235	-0.7523
lag 2	-1.7154	-1.8859	-1.5452
lag 3	-2.7182	-2.9740	-2.4630
child age bin 2	2.5561	2.5188	2.5933
child age bin 3	3.3709	3.3344	3.4074
child age bin 4	3.3548	3.3182	3.3913
mean maternal age	-0.1140	-0.1255	-0.1026
period	-0.1662	-0.2511	-0.0813
proportion mothers 15-19	-2.2269	-2.8797	-1.5749
proportion mothers 20-24	-1.5814	-2.0317	-1.1316
proportion mothers 25-39	-0.8190	-1.1570	-0.4815
P(15-19)/P(20-24)	-1.2373	-3.6473	1.1637
P(20-24)/P(25-25)	-3.4624	-5.8158	-1.1069
years in period	0.5652	0.5573	0.5730
log(CD)	-0.0578	-0.0955	-0.0201
log(CEB) * years in period	-0.0039	-0.0049	-0.0029
log(CEB) * child age bin 2	-0.0154	-0.0203	-0.0105
log(CEB) * child age bin 3	-0.0302	-0.0350	-0.0254
log(CEB) * child age bin 4	-0.0279	-0.0327	-0.0231
child age bin 2 * years in period	-0.0476	-0.0558	-0.0394
child age bin 3 * years in period	-0.0744	-0.0824	-0.0664
child age bin 4 * years in period	-0.0923	-0.1003	-0.0842
log(CEB) * child age bin 2 * years in period	0.0032	0.0022	0.0043
log(CEB) * child age bin 3 * years in period	0.0063	0.0053	0.0074
log(CEB) * child age bin 4 * years in period	0.0054	0.0044	0.0065
lag 1 * mean maternal age	0.0096	0.0092	0.0099
lag 2 * mean maternal age	0.0363	0.0359	0.0367
lag 3 * mean maternal age	0.0621	0.0615	0.0627
Variance for country random effect	0.0355	0.0110	0.1446
Variance for admin1 random effect	0.1501	0.1408	0.1600
Variance for survey	0.1942	0.1407	0.2760

6.3 Out of sample validation of SBH Adjustment Model

We used five-fold cross validation to test how well these models predicted *p* and *N* out-of-sample. Using the two training datasets, we fit each of the models five times, each time holding out the response variable from random subset of one-fifth of the surveys. The plot below shows the out-of-sample fits by child age bin, and the table below shows summary predictive validity metrics for the two models for each child age-year bin.

Supplementary Figure 9. Out of sample fits for SBH adjustment models. SBH = summary birth history. X axis: Out of sample data at from the SBH adjustment model. Y axis: observed truth from CBH tabulations. 4A shows out of sample predictions from the *p* model for monthly probability of death. 4B shows the out of sample predictions from the *N* model for monthly exposures. Plotted in modelled logit and log scales, respectively. Colours indicate child-age groups. Red line indicates unity. SBH=Summary birth history.



Supplementary Table 4. Out-of-sample predictive validity metrics for SBH adjustment models in probability and exposure month space, respectively. Calculated using five holdout folds.

Age Bin	Year	Mean *p* (observed)	mean $\widehat{p}$	RMSE	ME	Mean *N* (observed)	Mean $\widehat{N}$	RMSE	ME
1	**2015**	0.02534	0.02446	0.00996	-0.00089	181.5	188.6	66.5	7.1
2	**2015**	0.00166	0.00204	0.00075	0.00038	2068.0	2104.4	685.5	36.4
3	**2015**	0.00067	0.00077	0.00028	0.00010	4299.3	4247.1	1428.4	-52.2
4	**2015**	0.00041	0.00037	0.00020	-0.00004	4138.4	4128.8	1347.0	-9.6
1	**2010**	0.02866	0.02670	0.00716	-0.00196	407.5	433.8	94.4	26.3
2	**2010**	0.00236	0.00239	0.00075	0.00004	4279.0	4469.9	756.3	190.9
3	**2010**	0.00095	0.00096	0.00036	0.00002	8843.8	8934.3	1152.2	90.5
4	**2010**	0.00044	0.00044	0.00020	0.00000	8354.2	8123.0	1156.8	-231.2
1	**2005**	0.03120	0.03126	0.00723	0.00006	421.9	418.1	95.4	-3.8
Age Bin	**Year**	**Mean *p* (observed)**	**mean $\widehat{p}$**	**RMSE**	**ME**	**Mean *N* (observed)**	**Mean $\widehat{N}$**	**RMSE**	**ME**
2	**2005**	0.00324	0.00302	0.00090	-0.00022	4346.7	4307.7	815.6	-39.0
3	**2005**	0.00132	0.00133	0.00040	0.00001	8572.6	8559.1	1313.4	-13.5
4	**2005**	0.00059	0.00060	0.00020	0.00001	7691.9	7659.0	1108.8	-33.0

1	**2000**	0.03482	0.03633	0.00795	0.00152	386.6	373.9	73.5	-12.7
2	**2000**	0.00399	0.00379	0.00110	-0.00020	3915.9	3835.7	609.8	-80.2
3	**2000**	0.00174	0.00178	0.00049	0.00005	7637.5	7608.4	1157.5	-29.1
4	**2000**	0.00080	0.00079	0.00030	-0.00001	6592.9	6750.3	1424.1	157.3

6.4 Inverse-variance weighting of SBH observations

To account for the fact that the adjusted SBH approximated binomial data points were modelled, we chose to weight SBH observations in the prediction proportionally to the uncertainty in each prediction from the two-stage model. These weights are then used when fitting the full geostatistical model.

For each observation of SBH-only data, we took 10,000 draws from the predictive posterior distributions of $\hat{p}$ and $\widehat{N}$. For each draw *m* we then simulated $\widehat{N^+}_m$ from a binomial distribution with parameters $\widehat{p_m}$ and $\widehat{N_m}$, giving us 10,000 corresponding draws of $\widehat{N^+}$. We then took the weights as the inverse of the ratio of the simulated variance of $\widehat{N^+}$and the expected binomial variance $E[\widehat{N}]E[\,\hat{p}](1 - E[\,\hat{p}])$. As defined, the weights represent the proportion of excess variance in these observations induced by the two-stages of *N* and *p* modelling.

In 7.2% of our predicted SBH observations, weights exceeded 1 because small N and small monthly death probabilities resulted in 10,000 draws of nearly all zeros. This caused very low variance of observed deaths across the draws (lower than expected) and inflated weights. We view this as a computational issue in managing to capture low binomial variance via simulation in situations where we have high confidence that that the outcome should be 0. Taking more draws from the predictive posterior distributions of $\hat{p}$ and $\widehat{N}$ is computationally expensive and for these observations we set the weight to 1. Weights ranged from 0.00013 to 1.0, with 25% falling below 0.36, a median weight of 0.75, and a 75th quantile of 0.91.

7. Applying GBD bias correction

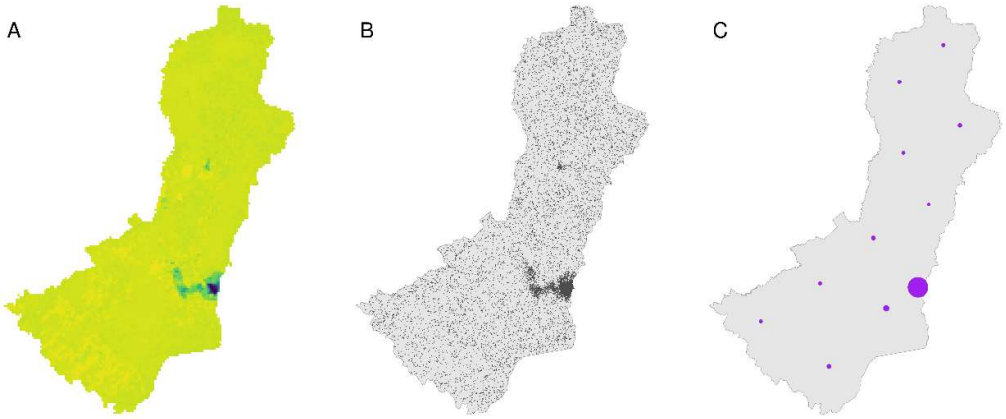
Prior to geostatistical modelling, we adjusted mortality data by applying the bias correction ratios used by Wang and colleagues for GBD 2016.[3] Briefly, raw under-5 mortality data may differ due to various source-specific biases, sampling-related or non-sampling, and it is thus desirable to account for them when possible in order to generate a consistent time series of under-5 mortality. To adjust for this, a non-linear mixed effects model with source-type specific fixed effects and source-specific random effects nested within a location was fit by Wang and colleagues.[3] In addition, one high-quality reference source is chosen as the reference based on expert opinion and general data quality from specific survey series for each country. In the case of Africa, this is typically complete birth histories from DHS. For all non-reference sources, data are adjusted based on the difference between the combination of source-type fixed effects and source specific random effects between the source of interest and the reference source. The ratio of the unadjusted mortality to the adjusted from the model are thus country-source-year specific. We converted these to monthly ratios and multiplied the number of deaths in each bin-period-cluster by them for all matching records from non-reference data sources.

8. Spatial integration over polygon records

Supplementary Figure 5, shown below, illustrates the k-means clustering method used to generate spatial integration points for areal mortality estimates. For each geographic area, we extracted a raster layer of population counts with cell area of approximately 5x5 km from an Africa-wide population raster for 2010 from the WorldPop project (panel A).[8] We then sampled 10,000 point locations with replacement from each

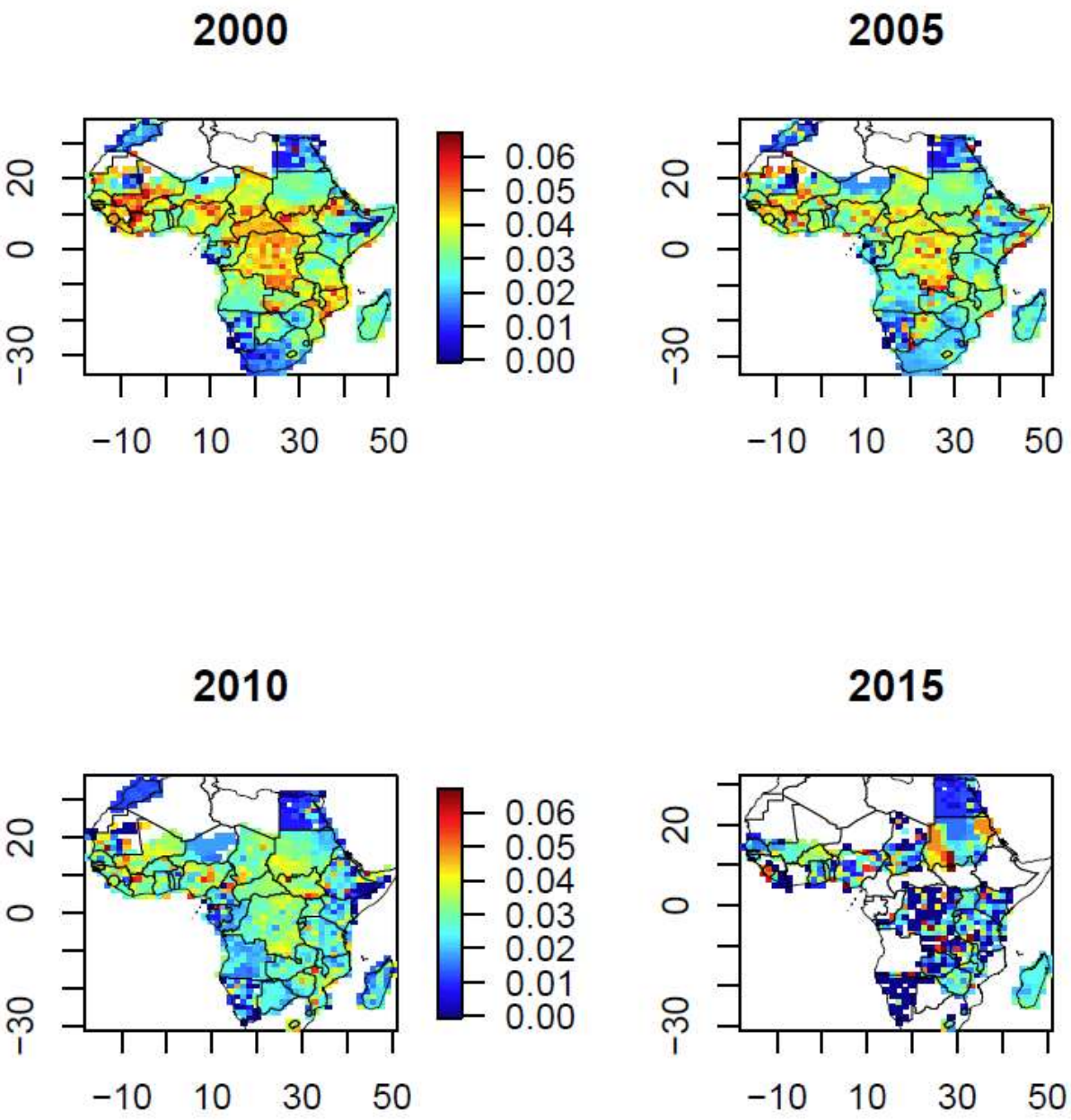
area, with sampling probability proportional to cell population, to simulate a random sample of the population of the area (panel B). Finally, we applied k-means clustering to the latitudes and longitudes of these locations, with J clusters, such that there was one cluster per 100 raster cells within each polygon area. For each of the J cluster centres, we computed the integration weight as the proportion of the representative points falling within this cluster (being geographically closer to that cluster centre than any other cluster centre) (panel C). These weighted points (pseudo-clusters, representing areal estimates) are then combined with true geographically referenced clusters to create the dataset to train the geostatistical model.

Supplementary Figure 10: Illustration of k-means clustering for selecting spatial integration points and weights for polygon mortality data, applied to the district of Makonde, Zimbabwe. A) Spatial raster of cell-level population counts (blue cells indicating larger populations), B) locations of 10,000 population-representative point locations (with added spatial jitter for illustrative purposes), C) 11 spatial integration points selected by k-means clustering, with the size of each point proportional to its integration weight.

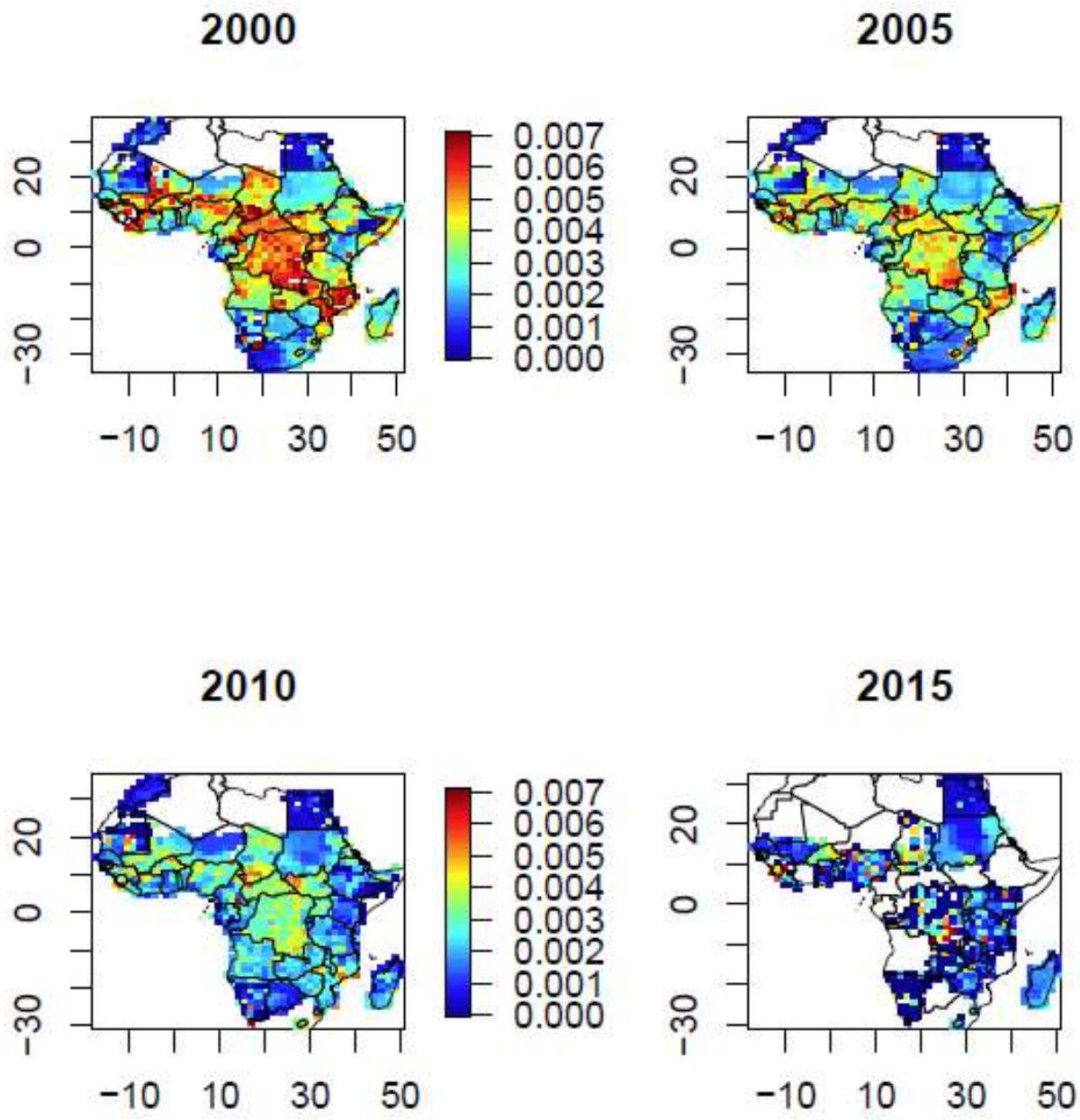


9. Raw data plots

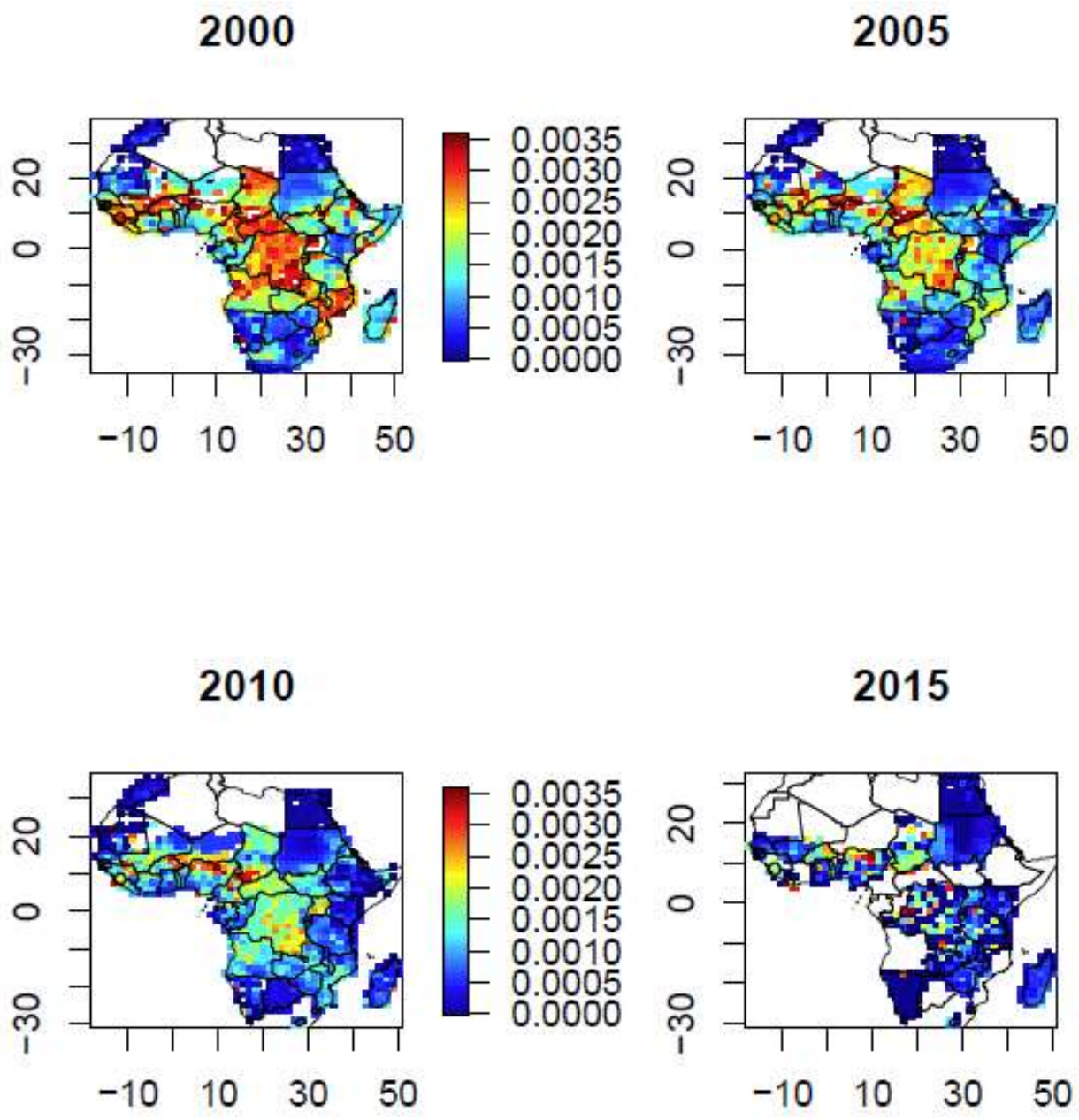
The plots in Supplementary Figures 6 through 10 show the model input data for the four age bins over time. Monthly probabilities of death are plotted over a 48 x 48 pixel lattice over Africa. Pixels represent average mortality probabilities from points falling within each pixel. Plots are unweighted by sample size and are meant only to give a general idea of data coverage and values.



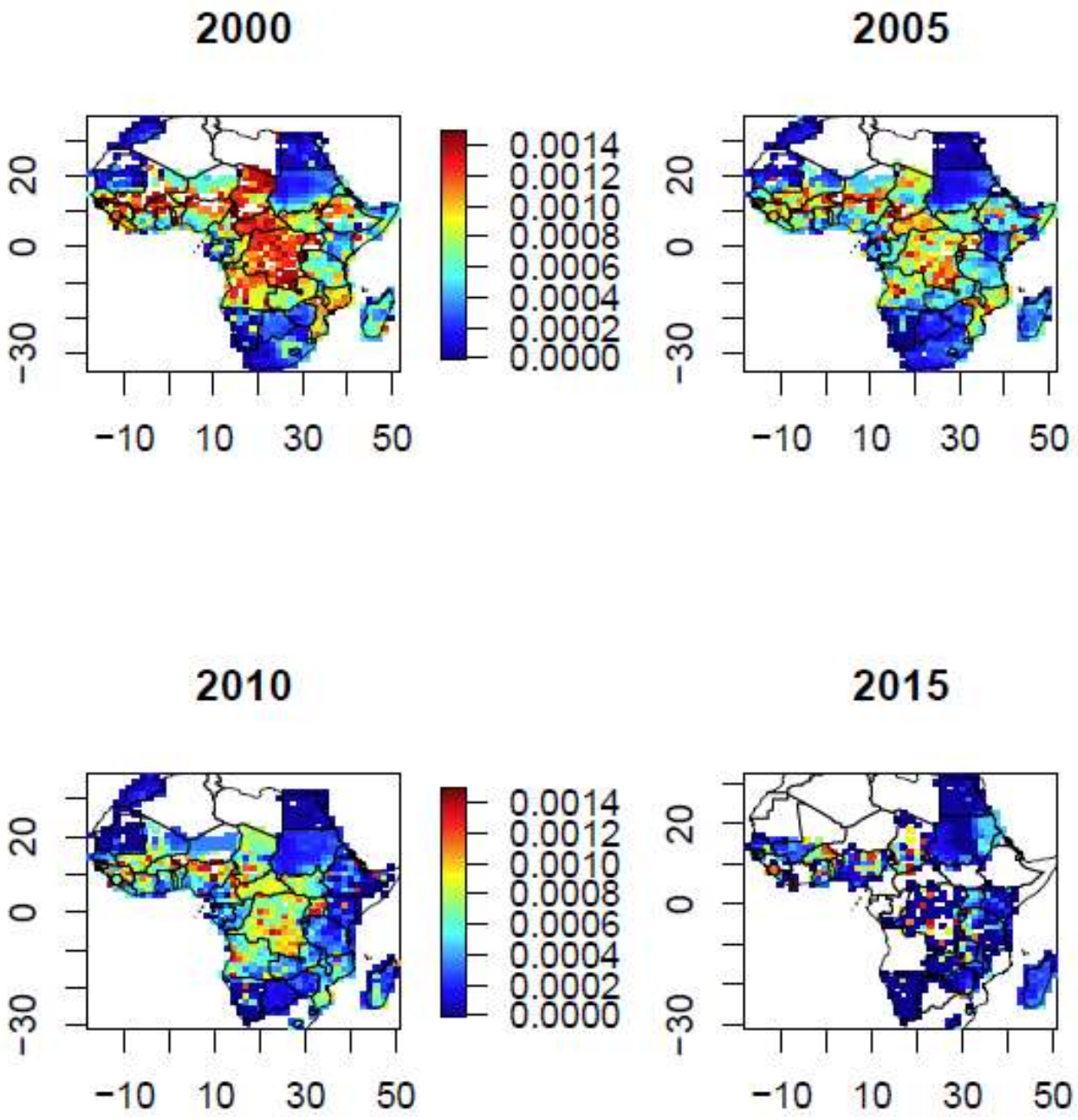
Supplementary Figure 11: Model input data for age bin 1 (neonatal), values indicate the mean monthly probability of death among clusters in each grid cell.



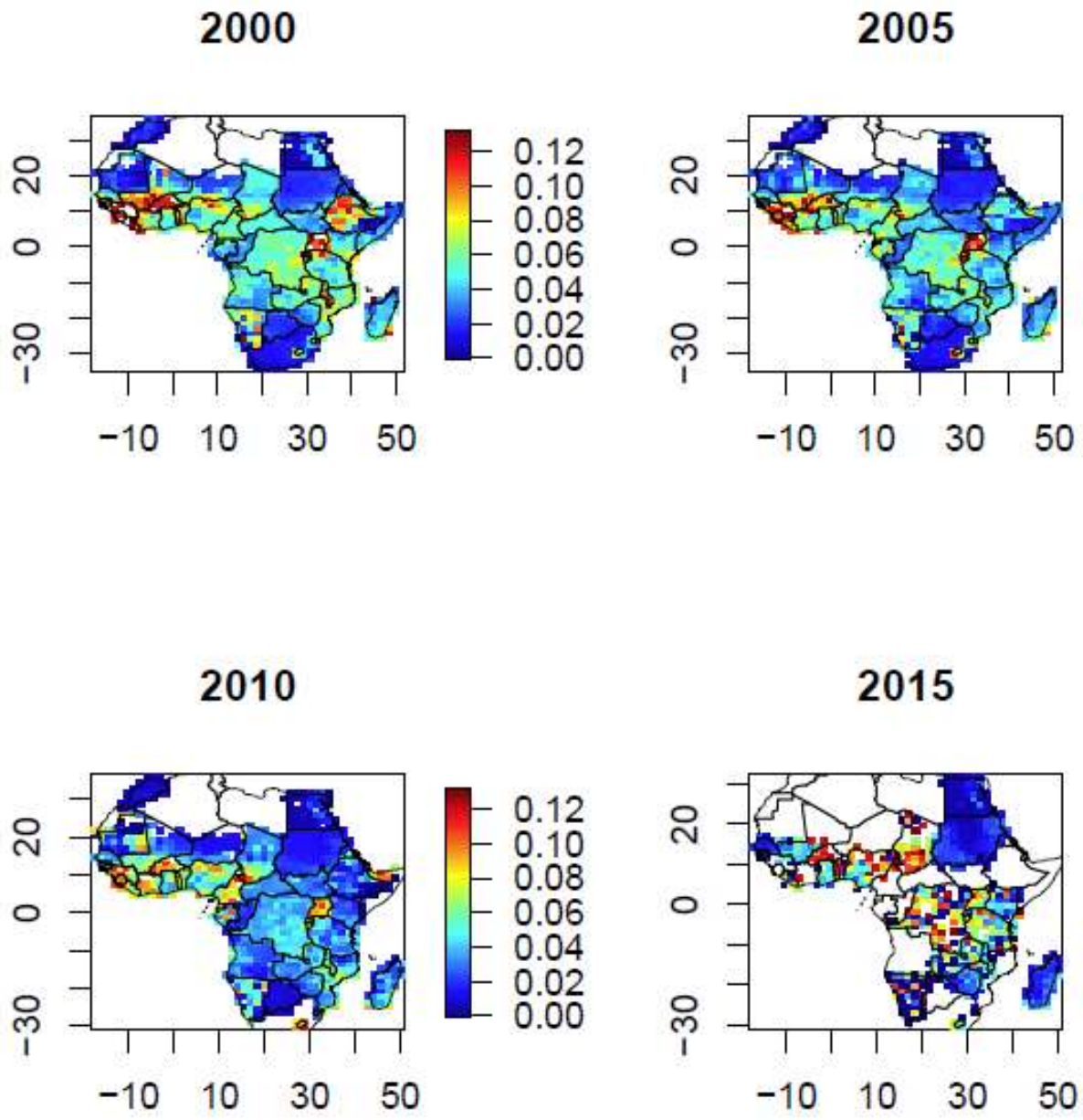
Supplementary Figure 12: Model input data for age bin 2 (1 – 11 months), values indicate the mean monthly probability of death among clusters in each grid cell.



Supplementary Figure 13: Model input data for age bin 3 (12 – 35 months), values indicate the mean monthly probability of death among clusters in each grid cell.



Supplementary Figure 14: Model input data for age bin 4 (36 – 59 months), values indicate the mean monthly probability of death among clusters in each grid cell.



Supplementary Figure 15: Combined input data for age bins1-4, approximating under-5 mortality.

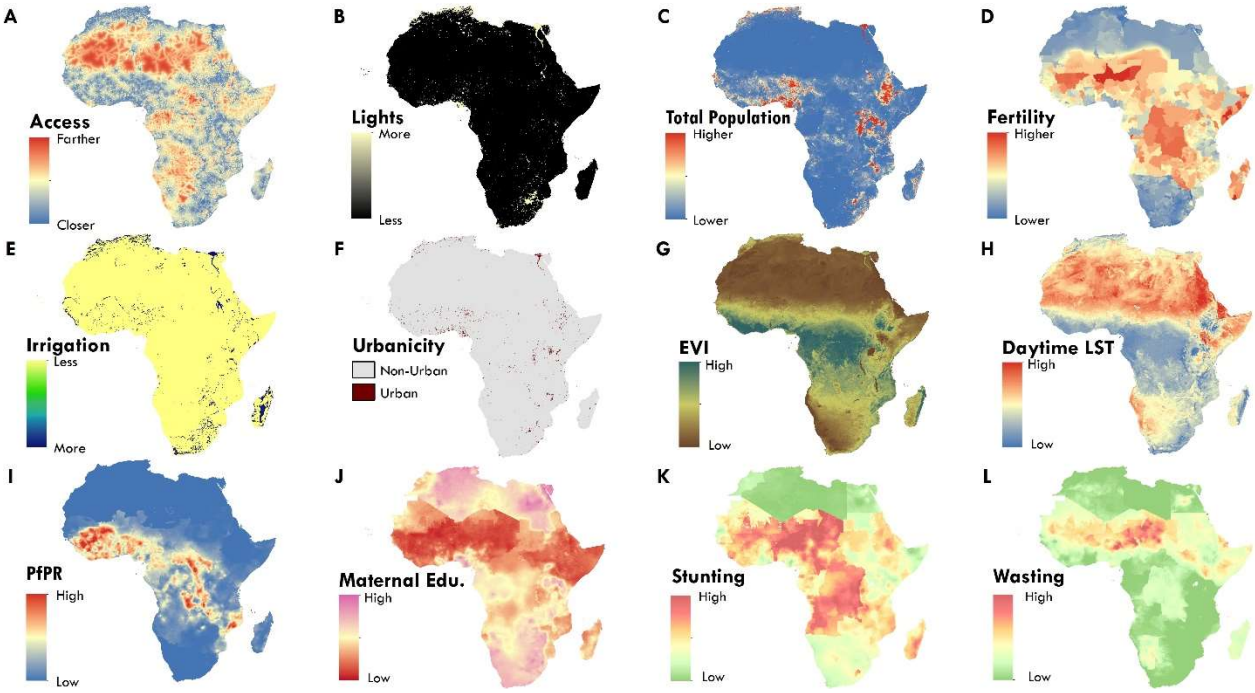
10. Covariates and covariate transformation

10.1 Covariates

Supplementary Table 5. Description of covariate layers used and their sources. For years, if multiple years were provided (included in parentheses), we took a synoptic mean.

Plot	Covariate	Years	Source
A	Travel time (in seconds) to nearest population centre (of at least 50,000 population)	2000	Uchida & Nelson, 2008: Agglomeration Index: Towards a New Measure of Urban Concentration http://forobs.jrc.ec.europa.eu/products/gam/index.php
B	Night-time light intensity	2000 (1997-2002); 2005 (2003-2007); 2010 (2008-2012); 2015 (2013)	Image and data processing by NOAA's National Geophysical Data Center. DMSP data collected by US Air Force Weather Agency https://www.ngdc.noaa.gov/eog/dmsp/downloadV4composites.html
C	Population	2000, 2005, 2010, 2015	Christopher T. Lloyd, Alessandro Sorichetta & Andrew J. Tatem. High resolution global gridded data for use in population studies. Scientific Data 4, Article number: 170001 (2017) doi:10.1038/sdata.2017.1 http://www.worldpop.org.uk/data/get_data/
D	Fertility proxy index (ratio of under-5 children to women of reproductive age [15-49 years])	2000, 2005, 2010, 2015	Christopher T. Lloyd, Alessandro Sorichetta & Andrew J. Tatem. High resolution global gridded data for use in population studies. Scientific Data 4, Article number: 170001 (2017) doi:10.1038/sdata.2017.1 http://www.worldpop.org.uk/data/get_data/
E	Irrigation (area irrigated per pixel)	2000	Siebert at al 2005: Development and validation of the global map of irrigation areas https://www.uni-frankfurt.de/45218039/Global_Irrigation_Map
F	Urban-rural dichotomous (as measured by Boolean classification)	2000 (1999-2002); 2005 (2003-2008); 2010 (2008-2012); 2015 (2013-2014)	Pesaresi et al 2016: Operating procedure for the production of the Global Human Settlement Layer from Landsat data of the epochs 1975, 1990, 2000, and 2014 http://ghsl.jrc.ec.europa.eu/data.php
G	Enhanced vegetation index (0-1 scale)	2000 (2000-2002); 2005 (2003-2007); 2010 (2008-2012); 2015 (2013-2015)	Huete et al 1999: MODIS vegetation index Algorithm theoretical basis document https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table/mod13a1
H	Average daytime temperature in C°	2000 (2000-2002); 2005 (2003-2007); 2010 (2008-2012); 2015 (2013-2015)	Wan 1999: MODIS Land-Surface Temperature Algorithm Theoretical Basis Document https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table/mod11a2
I	*Plasmodium falciparum* parasite rate (*Pf*PR)	(1998-2002)->2000 (1998-2000); 2005 (2003-2007); 2010 (2008-2012); 2015 (2013-2016)	Bhatt et al 2015: The effect of malaria control on Plasmodium falciparum in Africa between 2000 and 2015 http://www.map.ox.ac.uk/
J	Average years of education for women aged 15-49 years	(2000-2002)->2000 (2000-2002); 2005 (2003-2007); 2010 (2008-2012); 2015 (2013-2015)	Currently unpublished, correspondence to Simon Hay.
K	Prevalence of moderate and severe stunting for children under 5 (as measured by -2SD height for age Z-scores)	(2000-2002)->2000 (2000-2002); 2005 (2003-2007); 2010 (2008-2012); 2015 (2013-2015)	Currently unpublished, correspondence to Simon Hay.
L	Prevalence of moderate and severe wasting for children under 5 (as measured by -2SD weight for height Z-scores)	(2000-2002)->2000 (2000-2002); 2005 (2003-2007); 2010 (2008-2012); 2015 (2013-2015)	Currently unpublished, correspondence to Simon Hay.

Supplementary Figure 16: Images of spatial covariates used (2015 values shown here). EVI = enhanced vegetation index; LST = land surface temperature; PfPR = Plasmodium falciparum parasite rate.



10.2 Covariate selection and transformation with stacking

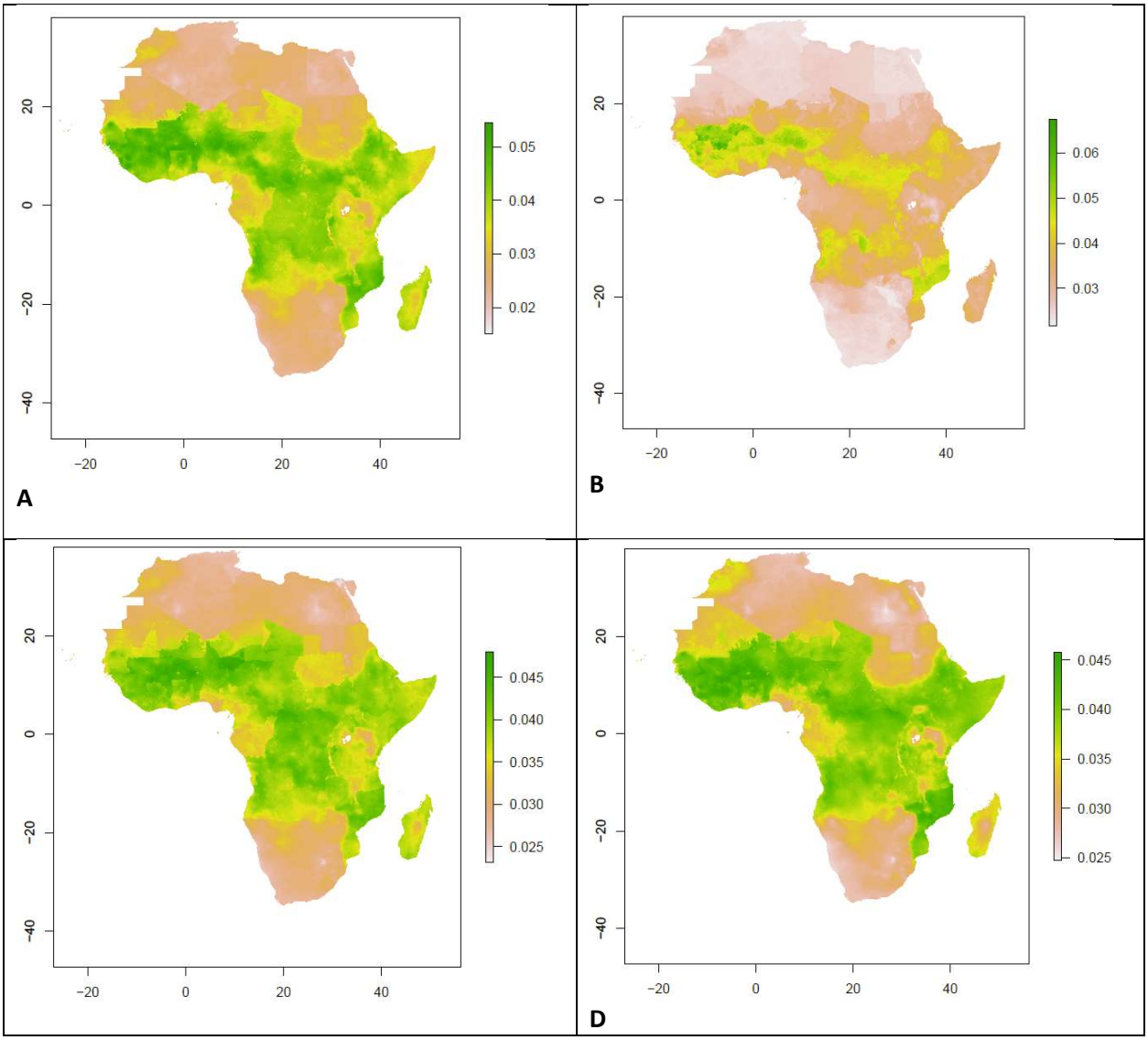
Stacked generalization/regression, or stacking, is a method of combining (ie, ensembling) multiple predictive models in order to enhance predictive validity relative to a single approach or model[9,10] and has been shown to be effective for geostatistical exercises.[11] In short, a suite of child models are fit using different modelling approaches and then combined, or stacked, by using a secondary learner, or stacker. Our implementation of stacking largely follows the approach described in Bhatt and colleagues.[11]

Our stacking hierarchy features two levels: a collection of 'child' models and a stacker model. Although there are possible other configurations, this direct, two-stage approach was taken because of its simplicity, computational tractability, and previous success.[11] The child models used for this study include a generalized additive model, a gradient boost machine (also known as boosted regression trees), and penalized regressions (ridge, lasso, and elastic net). All analyses were performed in R,[12] using the mgcv, dismo, and glmnet packages.

Each child model was fit 5+1 times, where five is the number of folds for out-of-sample cross validation, plus one full data model. Data are divided into five folds and for each fold a given model is fit while leaving out one fold of data and training on the remaining data. The left-out fold subsequently serves as out-of-sample test dataset for prediction. Once the out-of-sample cross-validation process is complete for every fold the child model is then trained on the entire dataset. By the end of this process, each observation has an out-of-sample cross-validated prediction and full-fit prediction for each child model. Both the cross-validated predictions and the full-fit predictions are preserved for the subsequent stacking.

The final stacker model (the geostatistical model described in the next section) is fit using the cross-validated predictions of the child models as covariates. By fitting on the cross-validated predictions we incorporate the generalization capabilities of each child model and otherwise reduce overfitting. If correlations in predictions from a pair of child models exceeded 0.99 we dropped one of the models. In all cases, elastic net and lasso regressions met this threshold of correlation and thus we dropped elastic net. Model correlations are reported in supplementary table 6. Final predictions from the geostatistical model were produced using the full-fit predictions from the constituent child models. Supplementary figure 12 illustrates child model full-fit predictions for each child model for age bin 1 in 2000. Child models were fit separately for each age bin and time period.

Supplementary Figure 17. Plots illustrating child model estimates of mortality for one age bin and year (age bin 1 in 2000). Each of these estimated surfaces is then used as a covariate in the geostatistical model. A. Generalized Additive Model (GAM); B. Boosted Regression Trees (BRT); C. Ridge regression; and D. Lasso regression.



C	

*Supplementary Table 6. Pairwise correlations between stacking model fits. Elastic Net was dropped due to high correlation with lasso. * = correlation exceeding 0.99.*

Model 1	Model 2	Age Bin 1	Age Bin 2	Age Bin 3	Age Bin 4
GAM	**BRT**	0.928	0.947	0.947	0.939
GAM	**Ridge**	0.850	0.956	0.961	0.954
GAM	**Elastic Net**	0.831	0.941	0.945	0.935
GAM	**Lasso**	0.824	0.940	0.940	0.933
BRT	**Ridge**	0.820	0.934	0.931	0.922
BRT	**Elastic Net**	0.828	0.938	0.938	0.937
BRT	**Lasso**	0.822	0.938	0.938	0.938
Ridge	**Elastic Net**	0.953	0.978	0.986	0.979
Ridge	**Lasso**	0.953	0.974	0.982	0.977
Elastic Net	**Lasso**	0.991*	0.997*	0.997*	0.995*

11. Geostatistical model

11.1 Model Geographies

In total, we ran four geostatistical models, one for each age bin.

11.2 Model description

The underlying statistical model, fitted separately for each of the four age bins was a spatially and temporally explicit hierarchical generalised linear regression model for binomial data, using the logit link function:

$$N_i^+ \sim Binomial(p_i, N_i)$$

$$logit(p_i) = \alpha + \boldsymbol{X}_i\boldsymbol{\beta} + \epsilon_{GP\,i}$$

$$\epsilon_{GP} \sim GP\left(0, \boldsymbol{K}_{space} \otimes \boldsymbol{K}_{time}\right)$$

$$\boldsymbol{K}_{space} = \left(\tau 2^{\nu-1}\Gamma(\nu)\right)^{-1}(\kappa\boldsymbol{D})^{\nu}K_{\nu}(\kappa\boldsymbol{D})$$

$$\boldsymbol{K}_{time_{k,l}} = \rho^{|t_k - t_l|}$$

where N_i are the number of monthly exposures recorded in survey cluster/period *i*, N_i^+ are the number of deaths resulting from all monthly exposures, p_i is the estimated cluster-level 5-year period mortality probability of death, modelled as a logit-linear function of the global intercept α, cluster-level covariate values $\boldsymbol{X}_i$ and vector of regression coefficients $\boldsymbol{\beta}$. One fixed effect was included for each of the predictions from the cross-validated 'child' stacking models. Spatiotemporally correlated residuals ϵ_{GP} are drawn from a three-dimensional, zero-mean Gaussian process (GP) with covariance matrix constructed as the Kronecker product of a spatial covariance matrix $\boldsymbol{K}_{space}$ and temporal covariance

matrix $\boldsymbol{K}_{time}$. $\boldsymbol{K}_{space}$ was defined by a stationary Matérn covariance function over the Euclidean distance matrix $\boldsymbol{D}$ between all survey cluster locations, with spatial decay parameter κ, spatial smoothness/complexity parameter ν, precision parameter τ, modified Bessel function of the second kind K_ν, and Gamma function $\Gamma(.)$. Since this Matérn covariance function is the stationary solution of a stochastic partial differential equation (SPDE), it enables the use of efficient statistical machinery for modelling with SPDEs, as described below. We set the complexity parameter ν to be fixed at 2. $\boldsymbol{K}_{time}$ was defined by the covariance function corresponding to the discrete-time autoregressive stochastic process of the first order (AR1). The AR1 process is typically defined over a nominal random variable: $x_t = \rho\, x_{t-1} + N(0, \sigma^2)$ where t indexes time (in our case the period) and ρ (which is constrained such that $|\rho| < 1$) and σ^2 are parameters. When convolving the space and time correlation structures with this definition of the AR1 process, the spatial variance $1/\tau$ and the temporal variance σ^2 would be non-identifiable. We therefore omit σ^2 from our definition above, and represent overall space-time variance via the parameter $1/\tau$.

11.3 Priors

The following minimally informative priors were specified over parameters in all four age bin models: $\alpha \sim N(0, 1000)$, $\beta \sim N(0, 1000)$, $\log((1+\rho)/(1-\rho)) \sim N(0, 0.15)$. INLA sets an uncorrelated multivariate normal prior on log-transformations of κ and τ, and by default it determines priors based on the characteristics of the finite elements mesh (described in the section below). We use the default minimally informative priors that INLA suggests which, in our setting, yielded $\theta_1 = \log(\tau) \sim N(0.378, 10)$ and $\theta_2 = \log(\kappa) \sim N(-1.64, 10)$.

11.4 Model fitting

Models were fitted by integrated nested Laplace approximations (INLA) and a stochastic partial differential equation (SPDE) representation of the Gaussian-Markov random field (GMRF) approximation to the GP model,[14] using the INLA R package.[15] The INLA-SPDE approach makes use of the close correspondence between a GMRF defined on a sufficiently dense lattice and a GP, the efficient numerical routines enabled by representing GMRFs as SPDEs, and efficient inference over the parameters of these models using the INLA method. These approximations enable us to carry out full Bayesian inference over the model for a very large dataset, where other inference methods (such as MCMC) would be computationally prohibitive. While the result is an approximation to the model posterior, this approach has been shown to have extremely high accuracy when compared with MCMC in both theoretical and real-world mapping problems.[16] We defined the GMRF on a lattice constructed by constrained, refined Delaunay triangulation within a convex hull no closer than five decimal degrees from the coastline of Africa (including Madagascar). Over land this lattice was constrained to have edge length no greater than 0.35 decimal degrees and over sea no greater than five decimal degrees. This was the densest possible lattice we were able to use over the study area at this time before running into computational issues. These fitted models were then used to generate 1,000 posterior samples each of mapped monthly mortality probability estimates for each of the four age groups by random sampling from the numerical approximation to the joint posterior density of the model parameters. These estimates were combined to estimate the pixel-level (marginal) predictive posterior mean neonatal and under-5 mortality probabilities and prediction uncertainty intervals (0.025% and 0.975%).

11.5 Model Results

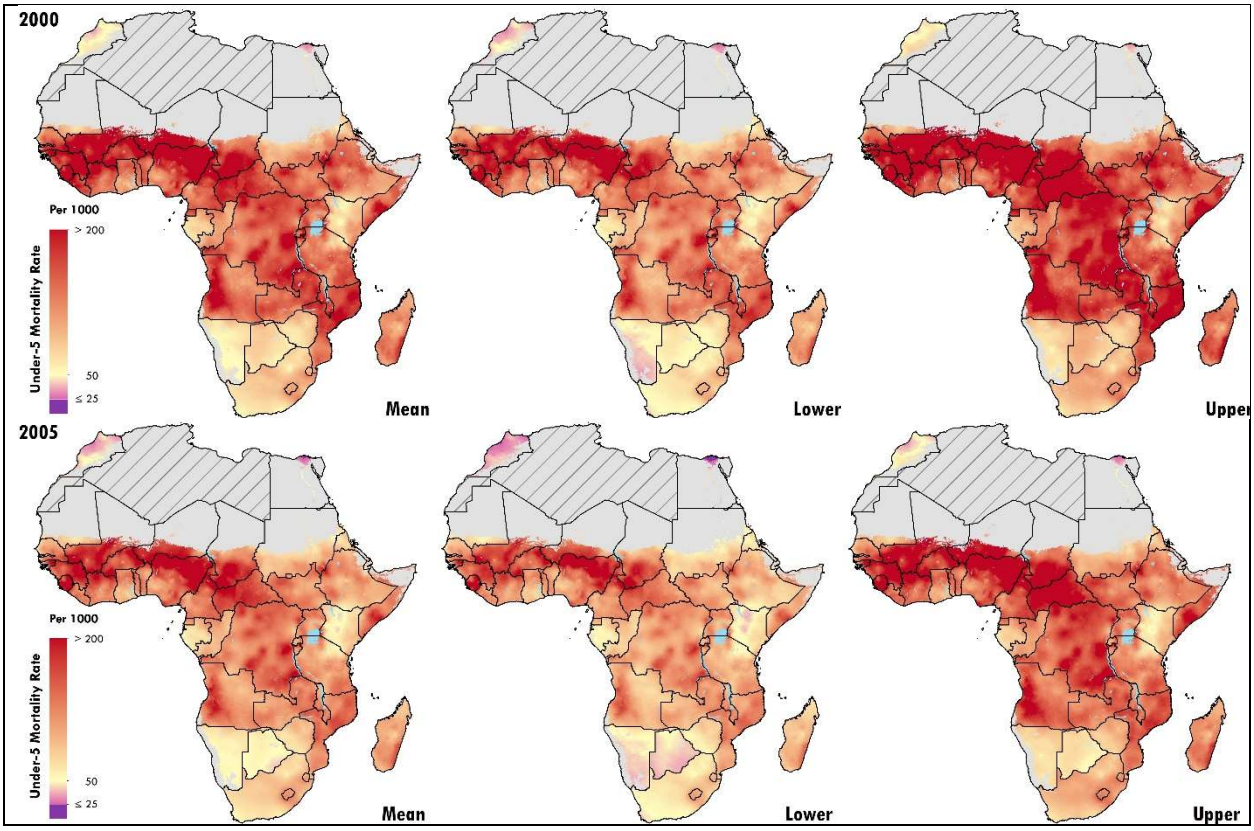
Geostatistical model results are presented in the Supplementary Table 7. Spatial Matérn covariance parameters κ and τ have been transformed (as have their lower and upper 95% uncertainty intervals) to be more interpretable. Range represents the distance in decimal degrees at which point approximately 90% of correlation has decayed, and is taken to be $\sqrt{8}/\kappa$. Nominal variance can be interpreted as the variance at each data point, in logit space, it is calculated as: $4\pi\kappa^2\tau^2$**.** The auto-regressive correlation coefficient for time, ρ, has not been transformed. Fixed effects include an intercept and the cross-validated predictions from the four included 'child' stacking models.

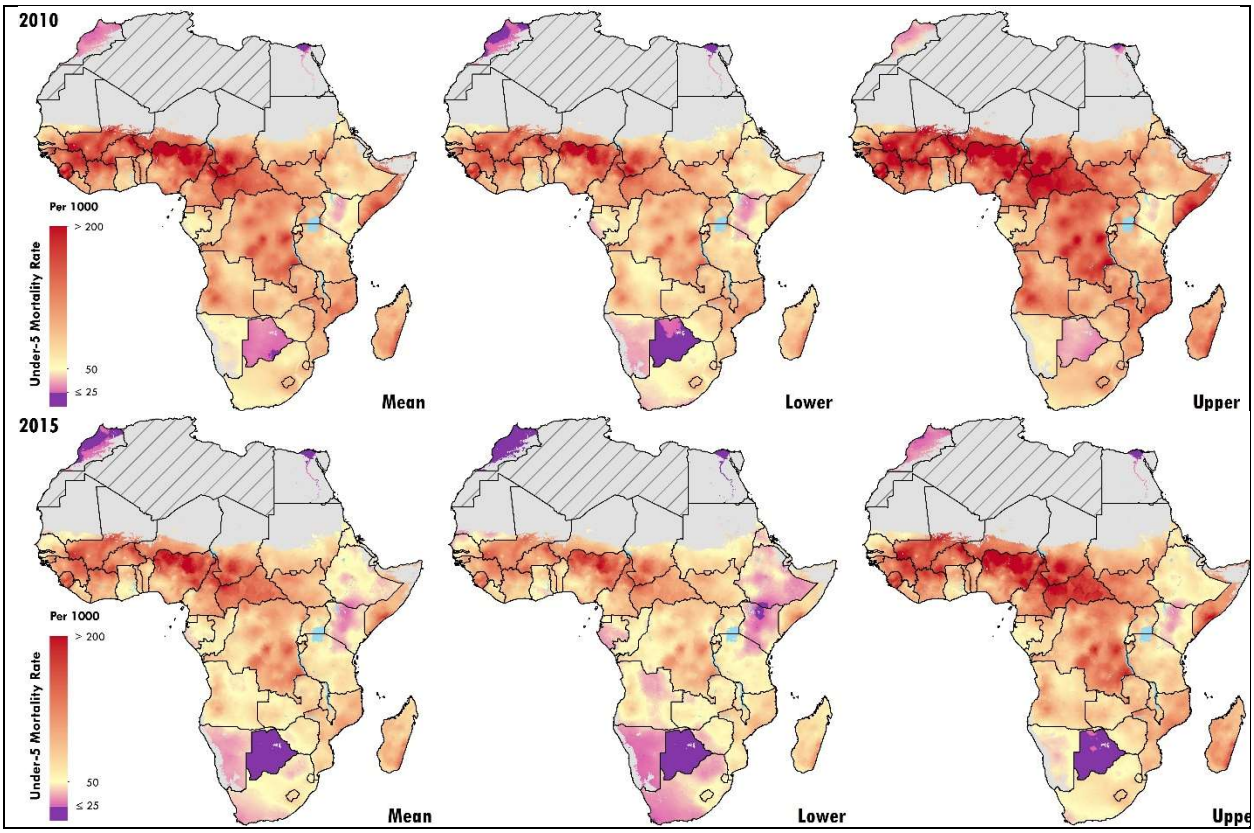
Supplementary Table 7: Model fits for each age-specific model. BRT = boosted regression trees; GAM = generalized additive model; GP = Gaussian process

	0 – 1 month			1 – 11 months			12 – 35 months			36 – 59 months		
	median	**2.5%**	**97.5%**	**median**	**2.5%**	**97.5%**	**median**	**2.5%**	**97.5%**	**median**	**2.5%**	**97.5%**
intercept	-0.291	-0.669	0.074	-0.235	-0.556	0.076	-0.931	-1.254	-0.622	-1.200	-1.669	-0.752
child model: lasso	0.521	0.229	0.812	0.399	0.210	0.589	0.291	0.132	0.449	0.089	-0.104	0.281
child model: ridge	-0.077	-0.356	0.203	0.038	-0.149	0.224	0.119	-0.060	0.297	0.085	-0.120	0.290
child model: BRT	0.360	0.238	0.481	0.396	0.312	0.478	0.345	0.280	0.409	0.424	0.349	0.500
child model: GAM	0.132	-0.020	0.283	0.138	0.015	0.259	0.128	0.041	0.214	0.255	0.151	0.358
GP: Range (Decimal Degrees)	4.914	4.097	5.869	3.058	2.687	3.477	3.356	2.976	3.885	6.307	5.084	7.767
GP: Nominal Variance	0.058	0.046	0.073	0.099	0.085	0.118	0.160	0.137	0.188	0.146	0.109	0.194
GP: AR1 correlation coefficient ρ	0.940	0.911	0.961	0.941	0.923	0.957	0.956	0.942	0.967	0.971	0.952	0.983

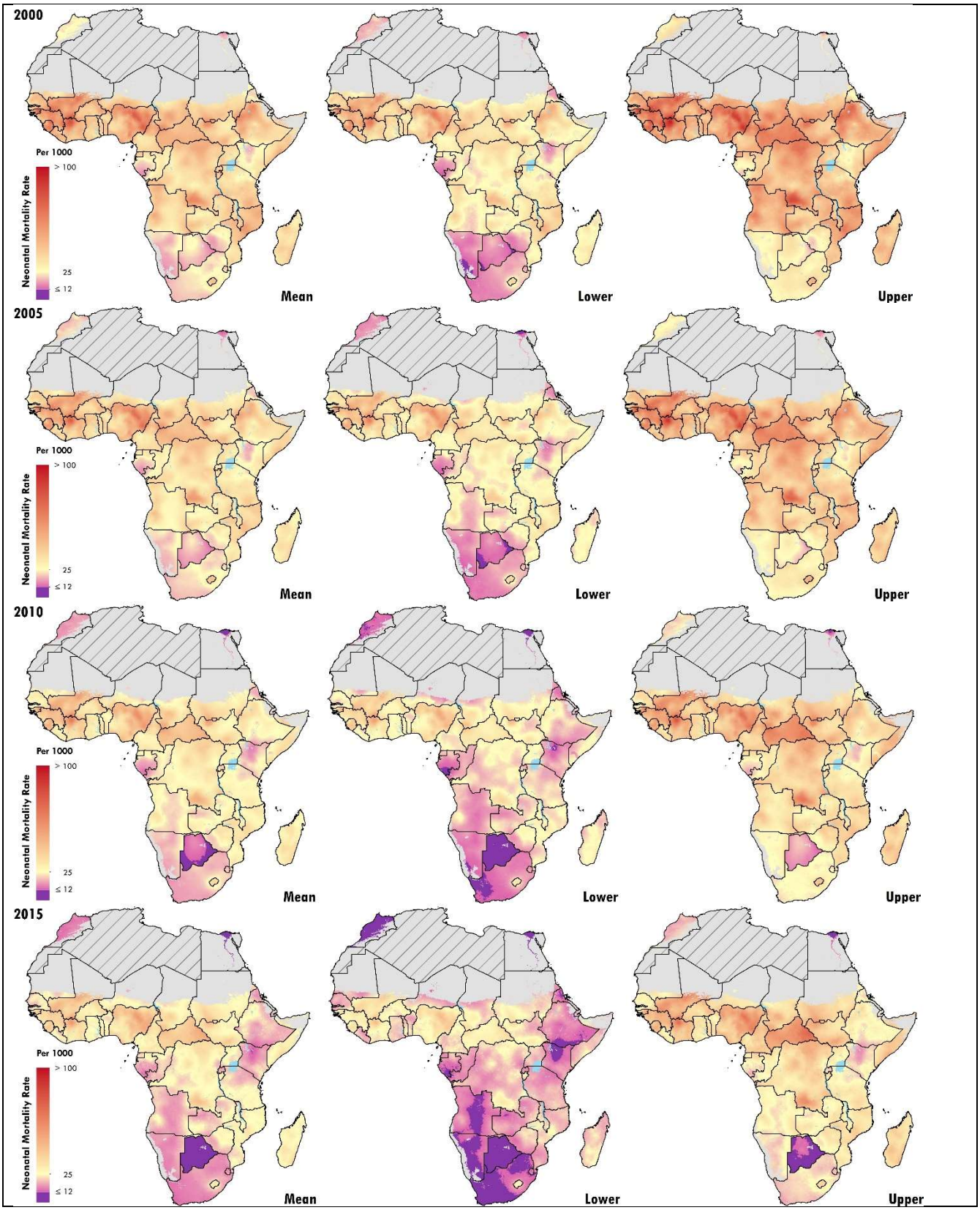
The figures below show model prediction surfaces for neonatal and under-5 mortality for the four study periods, as well as surfaces representing the lower (2.5%) and upper (97.5%) uncertainty bounds for each pixel.

Supplementary Figure 18. Mean, lower, and upper uncertainty interval predictions for under-5 mortality (5q0) in 2000, 2005, 2010, and 2015. Uncertainty intervals were taken as the 2.5% and 97.5% quantiles of 1,000 predictive draws for each pixel.





Supplementary Figure 19. Mean, lower, and upper uncertainty interval predictions for neonatal mortality in 2000, 2005, 2010, and 2015. Uncertainty intervals were taken as the 2.5% and 97.5% quantiles of 1,000 predictive draws for each pixel.


2000
2005
2010
2015
Per 1000
> 100
25
≤ 12
Neonatal Mortality Rate
Mean
Lower
Upper

Supplementary Table 8 shows the posterior expectation of the deviance $D(\theta) = -2Log(p(y|\theta))$ for each model. In addition to the full models (described above), we ran a 'null' model, which only included an intercept and the spatiotemporal random effect. The relative differences of these deviances show how much additional information, at minimum, was gained by including covariates. Across all models the percent difference in deviance between the full and null model ranged from 0.99% to 1.32%.

Supplementary Table 8. Posterior expectation of the deviance for each model, both full and null, and their relative differences.

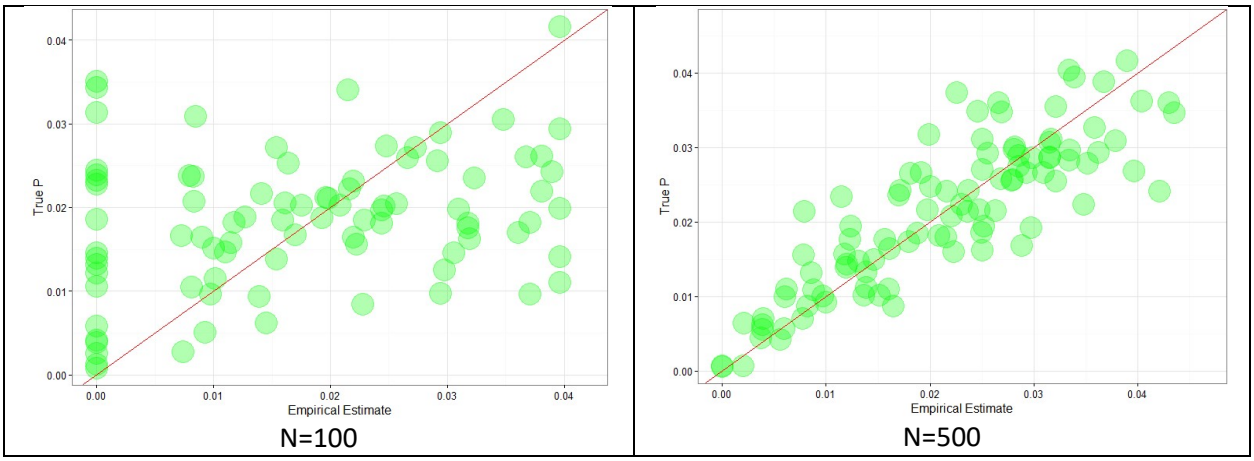
Age bin	Deviance		Percent difference
	Full model	Null model	
0 – 1 month	296222.6	300147.3	1.32%
1 – 11 months	339985.2	344045.8	1.19%
12 – 35 month	322265.1	325793.8	1.09%
36 – 59 months	242887.9	245299.5	0.99%

11.6 Model validation

In order to understand how well our model predicts mortality in locations where there are no data, we wish to estimate metrics of out-of-sample predictive validity. These are typically mean error (ME, to indicate level of bias), root-mean-squared error (RMSE, to indicate total variation in errors), correlation, and 95% coverage of the predictive intervals.

Model validation with these particular data presents a unique challenge. For relatively rare binomial data with small sample sizes, there is too much variance in the raw data to accurately assess model fit. In Supplementary Figure 15, we illustrate this using simulated data. Here we show that the empirical estimate taken as the ratio of events to observations in simulated binomial data with small *p* and relatively small *N* cannot reliably retrieve the probability (true p) they were simulated from. This means that even if we modelled mortality perfectly, we may not know it because many of our sample sizes are so small relative to the mortality proportion we are trying to estimate. The plot on the left shows this experiment done with N=100 (plus or minus some noise) and the plot on the right shows the same done with N=500 (plus or minus some noise). Adding sample size decreases variance in the empirical estimates, allowing for validation to be done. Many of our clusters are small, with median CBH clusters at 15, 150, 298, and 268 months for age bins 1 through 4, respectively. Furthermore, the monthly mortality probabilities which we estimate, particularly for age bins 3 and 4 are typically well below 0.001.

Supplementary Figure 20. Illustration of sample size impact on the ability to retrieve true probability from empirical binomial estimates.



11.6.1 Spatial aggregation

We chose to aggregate spatially proximal data points within administrative 1 and 2 areas to stabilize estimates of model predictive validity. This offers several benefits: 1) points close to each other should be more similar, and thus if we hold out points randomly in space we will likely inflate metrics of predictive validity because we would have predicted them well based on neighboring points; and 2) by aggregating points and thus creating areas of data sparseness in the model fitting, we more closely reflect true data missingness patterns over un-sampled administrative areas. At the same time it is important to note that choice of aggregation level can be somewhat arbitrary and model fit will generally improve with larger sample sizes at larger areas of aggregation, and different administrative levels will have varying sample sizes.

To predict out of sample, stacking and geostatistical models are run five times for each age bin, each time holding out data assigned to each respective fold. All data points were assigned identified membership within a given administrative unit area. Each administrative area was then randomly assigned a fold. This was done for both level 1 and 2 administrative areas. Once each model had run, we calculated ME, RMSE, correlation, and 95% coverage for the data we held out. We aggregated predictions and data estimates at each administrative unit by taking weighted aggregates based on sample size adjusted for integration and SBH-adjustment weights. RMSE, ME, and correlation were calculated by comparing mean aggregated estimates and predictions across administrative units, weighted on the aggregated sample size at each unit. Coverage was calculated by simulating predictions of child deaths from posterior draws of probability of death and observed sample sizes at each cluster location, thus representing our ability to reproduce hold out data within the specified level of certainty.

11.6.2 Metrics of predictive validity

The tables below show how each model performed based on the metrics of bias, total variance, and coverage. Tables are shown for first administrative unit-level predictions (Supplementary Table 9), and second administrative unit-level predictions (Supplementary Table 10). The number of data points and

exposure months in a given holdout area could vary greatly, and validation metrics will be sensitive to this.

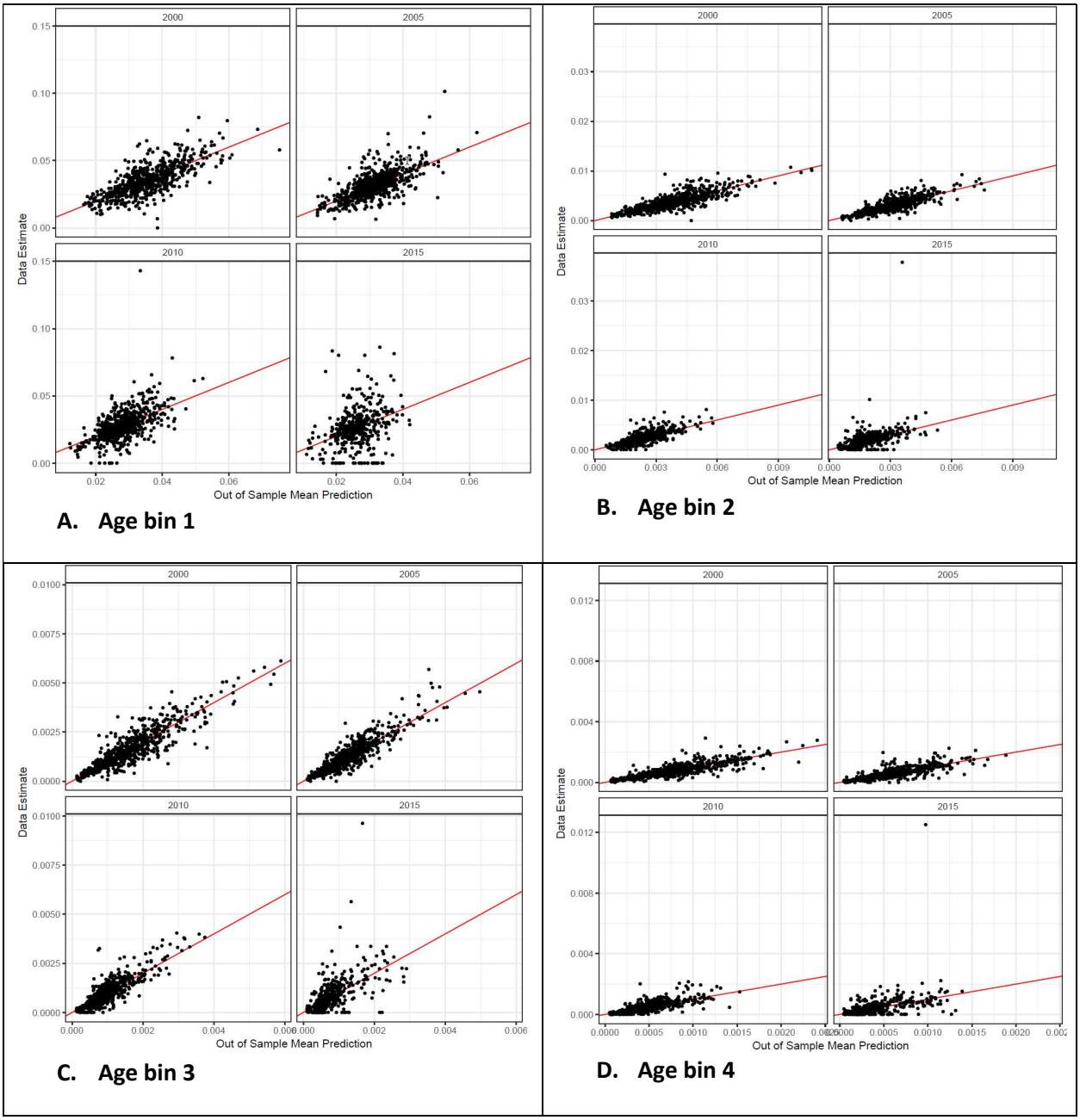
Supplementary Table 9. Out-of-sample predictive validity for first administrative-level holdout predictions

Age Bin	Avg estimated monthly probability	Mean Exposure months	Mean Error	RMSE	Correlation	95% Coverage
0-1 month	0.0328	783	-0.0003	0.0058	0.8111	95.0%
1-11 months	0.0031	8134	0.0000	0.0006	0.9043	95.3%
12-35 month	0.0013	16640	0.0000	0.0003	0.9432	92.7%
36-59 months	0.0006	17180	0.0000	0.0002	0.8989	93.8%

Supplementary Table 10. Out-of-sample predictive validity for second administrative-level holdout predictions

Age Bin	Avg estimated monthly probability	Mean Exposure months	Mean Error	RMSE	Correlation	95% Coverage
0-1 month	0.0327	104	-0.0002	0.0107	0.6236	95.1%
1-11 months	0.0031	1110	0.0000	0.0011	0.8017	95.5%
12-35 month	0.0013	2342	0.0000	0.0005	0.8670	93.5%
36-59 months	0.0006	2446	0.0000	0.0003	0.7766	94.4%

Supplementary Figure 21. Data estimates for aggregated administrative 1 level holdouts versus mean out of sample predictions for the same locations. Estimates are taken at each point where there was data in a holdout area and then aggregated weighted by data sample size. Red lines indicate equivalence.



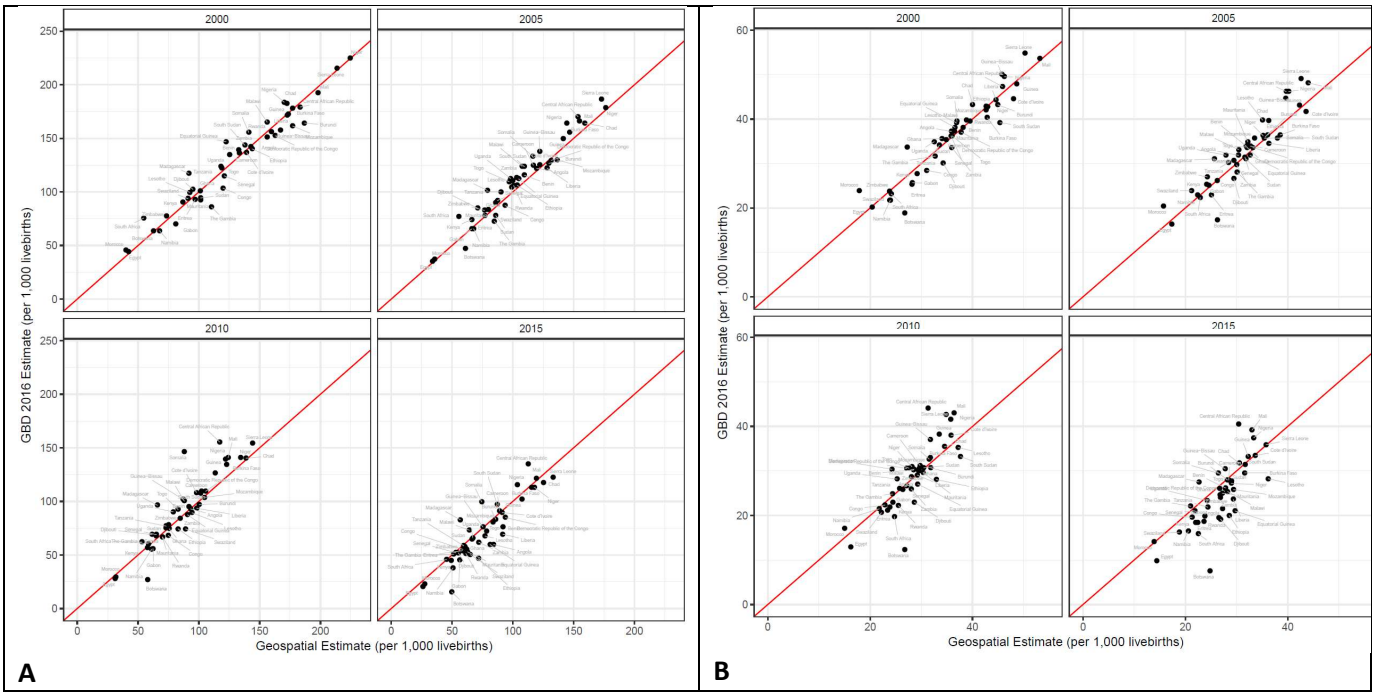
Supplementary Figure 22. Data estimates for aggregated administrative 1 level holdouts versus mean out of sample predictions for the same locations. Estimates are taken at each point where there was data in a holdout area and then aggregated weighted by data sample size. Red lines indicate equivalence.

Predictive agreement in the most recent time period was more difficult to assess because all survey data from this time period (2012-2017) included only data with partial period coverage, and thus smaller sample sizes than the previous time periods. As more data become available for future iterations of this work, we expect these predictions to improve.

12. Calibration to national estimates

As described in the main methods, we calibrated our mapped mortality rate estimates to match at the national level (or first administrative level where available) with mortality estimates from GBD 2016.[3] GBD estimates aggregate a wider pool of mortality data that are only available at a national level, including vital registration data. Supplementary figure 17 shows differences in population-weighted national-level estimates of mortality rates from the uncalibrated maps produced by our geostatistical modelling framework and GBD 2016 estimates for all countries and all four time periods. For each period, the average GBD estimate across the 5-years was used, with exception of the more recent period, for which GBD only makes estimates up to 2016 and thus 2017 was excluded. Maps were scaled to GBD by multiplying each pixel-draw by the ratio of national level GBD estimate to national level geostatistical estimate. National level geospatial estimates generally agreed well with GBD estimates as the median under-5 ratio was 1.01 and neonatal was 1.00 with 95% range of 0.81 and 1.22 and 0.76 and 1.19 respectively. There were some extreme outliers with Botswana in 2015 on the low end (0.33 for child and 0.32 for neonatal), and Madagascar 2015 (1.48 for child) and Central African Republic 2010 (1.40 for neonatal) at the high end.

Supplementary Figure 23. Comparison of nationally aggregated population-weighted geospatial estimates, and GBD estimates for the same years. A) Under-5 mortality estimates. B) Neonatal mortality estimates. Red lines indicate unity. The median ratio between GBD and this study's estimates for under-5 mortality was 1.01 [inter-quartile range (IQR): 0.95, 1.14] and for neonatal mortality 1.00 [IQR 0.93, 1.05]. Red lines indicate equivalence.

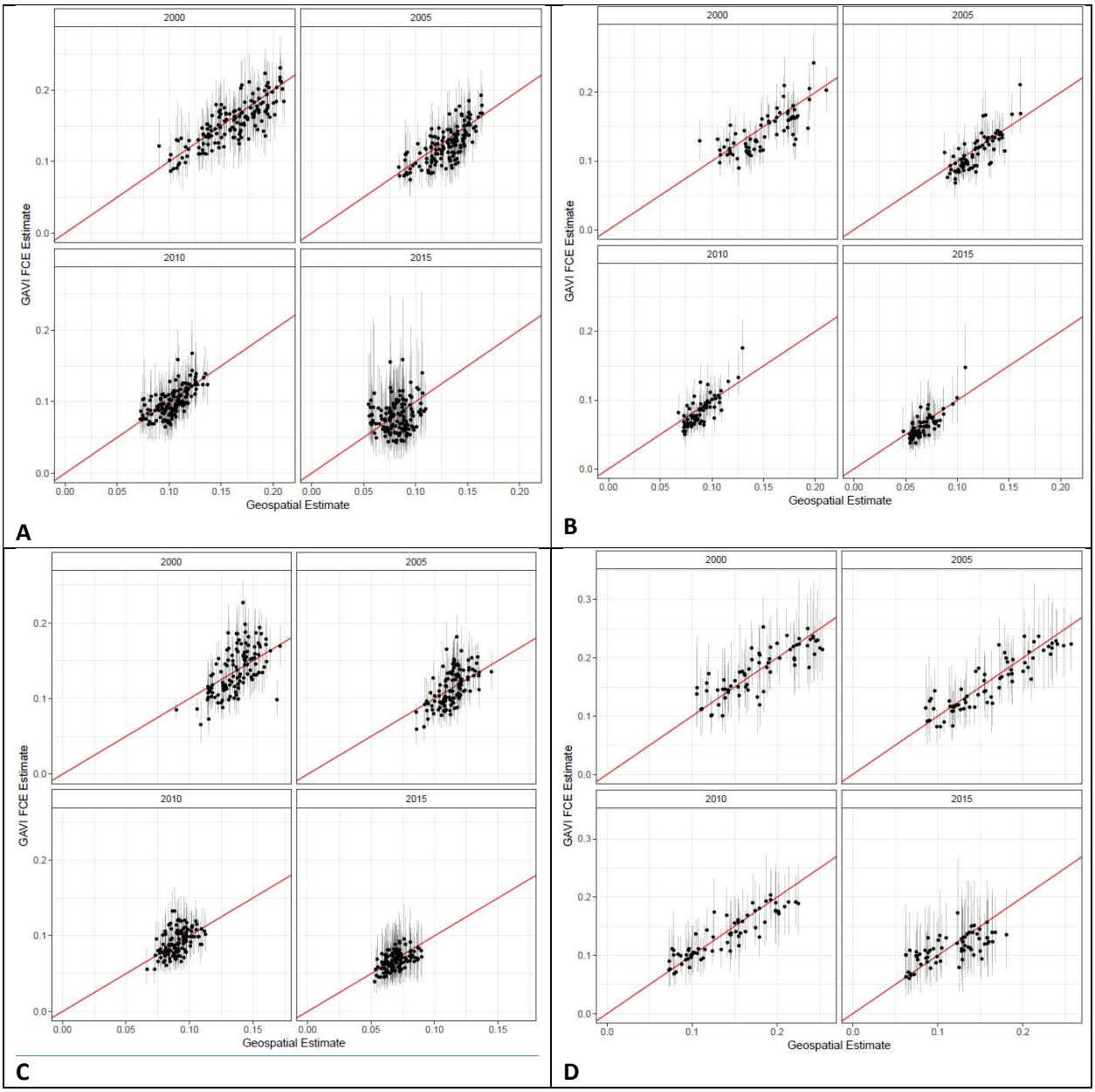


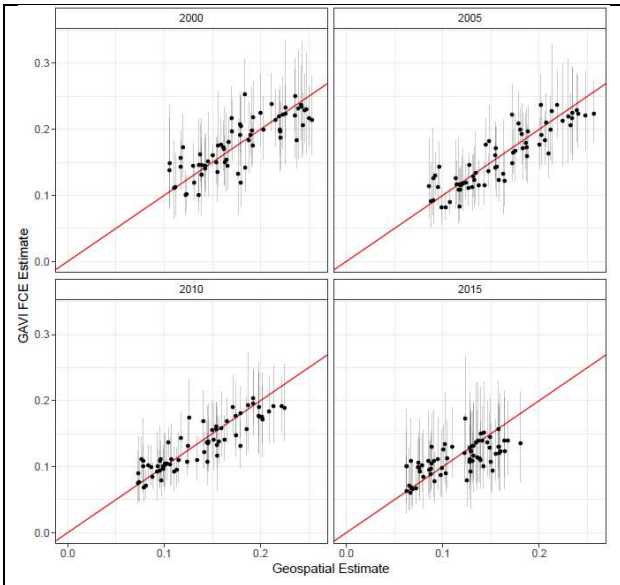
13. Verification and comparison against other subnational child mortality models

13.1 Gavi Full Country Evaluation small area estimates

As part of the Gavi Full Country Evaluation (Gavi FCE),[18] evaluation teams produced under-5 mortality estimates for Mozambique, Zambia, Uganda, Chad, and Cameroon. Supplementary Figure 18 (A, B, C, D, and E) shows agreement between aggregated population-weighted estimates from this study with the district level small area estimates of that one.

Supplementary Figure 24 Comparison of district- and county-level aggregated population-weighted geospatial estimates, and Gavi-FCE estimates of under-5 mortality for the same years. A) Mozambique. B) Zambia, C) Uganda, D) Chad, and E) Cameroon. Red lines indicate equivalence.


2000
2005
2010
2015
GAVI FCE Estimate
Geospatial Estimate
A
B
C
D

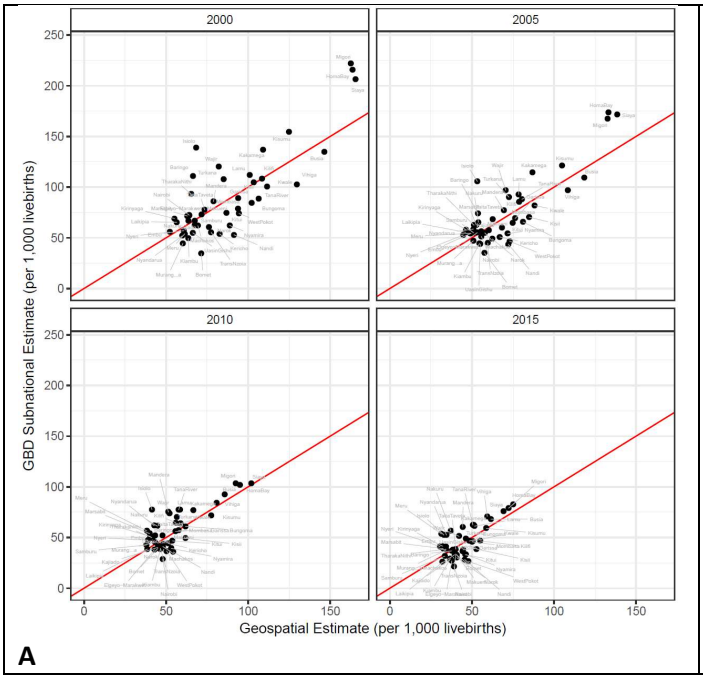


E

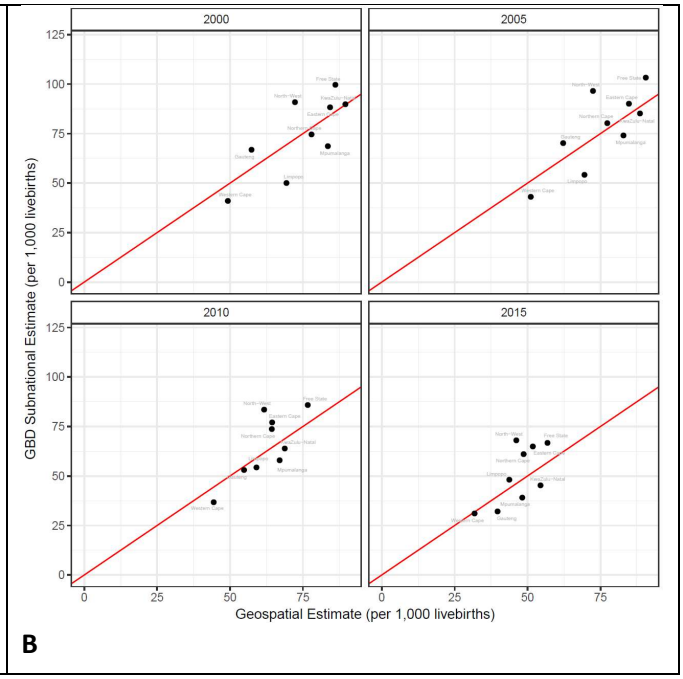
13.2 GBD 2015 subnational estimates in Kenya and South Africa

GBD 2016 included subnational estimates for Kenyan counties and South African regions. Comparisons are shown in Supplementary Figure 19 (A and B) below.

Supplementary Figure 25. Comparison of subnational aggregated population-weighted geospatial estimates, and GBD estimates of under-5 mortality for the same years. A) Kenyan Counties. B) South African Regions. Red lines indicate equivalence.



A



B

14. Source code

Source code is available through http://ghdx.healthdata.org/

References

1 Maternal, Newborn & Child Health Data - Countdown to 2030. http://countdown2030.org/ (accessed Feb 5, 2017).

2 Central Statistical Agency (Ethiopia), Government of Ethiopia, United Nations Population Fund (UNFPA), United Nations Development Programme (UNDP). Ethiopia Population and Housing Census 2007. Addis Ababa, Ethiopia: Central Statistical Agency (Ethiopia).

3 GBD 2016 Mortality Collaborators. Global, regional, and national under-5 mortality, adult mortality, age-specific mortality, and life expectancy, 1970-2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet* Under review.

4 Manual X. http://www.un.org/esa/population/publications/Manual_X/Manual_X.htm (accessed July 20, 2016).

5 Trussell TJ. A re-estimation of the multiplying factors for the Brass technique for determining childhood survivorship rates. *Popul Stud* 1975; **29**: 97–107.

6 Methods for estimating fertility and mortality from limited and defective data; 1975. http://unesdoc.unesco.org/Ulis/cgi-bin/ulis.pl?catno=19743&gp=0&lin=1 (accessed Feb 5, 2017).

7 Rajaratnam JK, Tran LN, Lopez AD, Murray CJL. Measuring Under-Five Mortality: Validation of New Low-Cost Methods. *PLOS Med* 2010; **7**: e1000253.

8 Lloyd CT, Sorichetta A, Tatem AJ. High resolution global gridded data for use in population studies. *Sci Data* 2017; **4**: 170001.

9 Wolpert DH. Stacked generalization. *Neural Netw* 1992; **5**: 241–59.

10 Breiman L. Stacked regressions. *Mach Learn* 1996; **24**: 49–64.

11 Bhatt S, Cameron E, Flaxman SR, Weiss DJ, Smith DL, Gething PW. Improved prediction accuracy for disease risk mapping using Gaussian Process stacked generalisation. *ArXiv161203278 Stat* 2016; published online Dec 10. http://arxiv.org/abs/1612.03278 (accessed Feb 5, 2017).

12 R: The R Project for Statistical Computing. https://www.r-project.org/ (accessed Feb 5, 2017).

13 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 Stat Methodol* 2009; **71**: 319–92.

14 Lindgren F, Rue H, Lindström J. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *J R Stat Soc Ser B Stat Methodol* 2011; **73**: 423–98.

15 The R-INLA project. http://www.r-inla.org/ (accessed July 27, 2016).

16 Schrödle B, Held L, Riebler A, Danuser J. Using integrated nested Laplace approximations for the evaluation of veterinary surveillance data from Switzerland: a case-study. *J R Stat Soc Ser C Appl Stat* 2011; **60**: 261–79.

17 Blangiardo M, Cameletti M. Spatial and Spatio-temporal Bayesian Models with R - INLA. John Wiley & Sons, 2015.

18 Gavi Full Country Evaluations Team. Gavi Full Country Evaluations: 2015 Dissemination Report. Seattle, WA: IHME, 2015.

Chapter 2: Development and validation of a new method for indirect estimation of neonatal, infant, and child mortality trends using summary birth histories

(Reprinted from https://doi.org/10.1371/journal.pmed.1002687)

Abstract

The addition of neonatal (NN) mortality targets in the Sustainable Development Goals highlights the increased need for age-specific quantification of mortality trends, detail that is not provided by summary birth histories (SBHs). Several methods exist to indirectly estimate trends in under-5 mortality from SBHs; however, efforts to monitor mortality trends in important age groups such as the first month and first year of life have yet to utilize the vast amount of SBH data available from household surveys and censuses.

We analyzed 243 Demographic and Health Surveys (DHS) from 76 countries, which collected both complete and SBHs from 8.5 million children from 2.3 million mothers to develop a new empirically based method to indirectly estimate time trends in age-specific mortality. We used complete birth history (CBH) data to train a discrete hazards generalized additive model in order to predict individual hazard functions for children based on individual-, mother-, and country-year-level covariates. Individual-level predictions were aggregated over time by assigning probability weights to potential birth years from mothers from SBH data. Age-specific estimates were evaluated in three ways: using cross-validation, using an external database of an additional 243 non-DHS census and survey data sources, and comparing overall under-5 mortality to existing indirect methods.

Our model was able to closely approximate trends in age-specific child mortality. Depending on age, the model was able to explain between 80% and 95% of the variation in the validation data. Bias was close to zero in every age, with median relative errors spanning from 0.96 to 1.09. For trends in all under-5s, performance was comparable to the methods used for the Global Burden of Disease (GBD) study and significantly better than the standard indirect (Brass) method, especially in the 5 years preceding a survey. For the 15 years preceding surveys, the new method and GBD methods could explain more than 95% of the variation in the validation data for under-5s, whereas the standard indirect variants tested could only explain up to 88%. External validation using census and survey data found close agreement with concurrent direct estimates of mortality in the NN and infant age groups. As a predictive method based on empirical data, one limitation is that potential issues in these training data could be reflected in the resulting application of the method out of sample.

This new method for estimating child mortality produces results that are comparable to current best methods for indirect estimation of under-5 mortality while additionally producing age-specific estimates. Use of such methods allows researchers to utilize a massive amount of SBH data for estimation of trends in NN and infant mortality. Systematic application of these methods could further improve the evidence base for monitoring of trends and inequalities in age-specific child mortality.

Introduction

Monitoring levels and trends of child mortality is a key component to understanding progress in child survival and for targeting additional policy and financial assistance to accelerate gains [1]. A complete, prospective, and continuous registration of births and deaths is the preferred source of information on child mortality [2], but in countries where child mortality is highest, deaths often go unrecorded because of poor or nonexistent vital registration (VR) systems [3]. In the absence of quality VR data, trends in under-5 mortality are typically estimated using retrospectively collected household sample survey and census data that ask mothers about births and deaths of their children [4,5].

Age-specific under-5 mortality varies widely both by and within country [4,6], and thus, it is critical to estimate levels and trends by age group with as much data as possible. The implications have high national and global relevance, particularly as the UN Sustainable Development Goals explicitly emphasized neonatal (NN) mortality in addition to under-5 mortality [7].

Household survey- and census-based child mortality questionnaires are available as either complete birth histories (CBHs), also sometimes known as full birth histories, or summary birth histories (SBHs). CBHs are preferred over SBHs because they capture detailed vital event histories on each child born to the surveyed mothers. Information on dates of birth and ages at death can thus be tabulated directly by age group and for specific years. In contrast, SBH surveys only ask each mother how many children she has birthed (children ever born [CEB]), how many of her children have died to date (children died [CD]), her age, and sometimes information about the time since first birth (TSFB) and/or marriage. Nevertheless, SBHs are widely available in many censuses and other sample surveys, in part because of the relative simplicity of collecting them. To utilize this vast source of data, several methods have been developed to indirectly estimate trends in under-5 mortality ($5q0$) from SBHs [8–11]. However, such methods have yet to be specifically adapted for wider application to estimate age-specific mortality among under-5s from SBHs; subsequently, past assessments of NN and infant mortality have been informed by comparably fewer data, especially outside of VR settings.

The Demographic and Health Surveys (DHS) have been widely collected in low- and middle-income countries (LMICs) since 1984 and provide a major source of CBH data. DHS surveys also collect SBH information, and the linked SBH–CBH serves as the basis for the method we describe in this paper. Other large survey families, such as the Multiple Indicator Cluster Surveys (MICS), collect SBHs in some settings and CBHs in others. Censuses are the largest global source of SBH data, representing data on many millions more children than are available in CBH surveys and often offering high sample sizes within small spatial areas. The DHS program takes steps to ensure high-quality and consistent data collection, including probing to ensure that CBH and SBH tabulations are aligned. Censuses have much more variation in collection methods and quality. For example, SBH modules are sometimes collected not from mothers but from household informants who are often male heads of household. Indirect trends in child mortality from SBHs are currently estimated using either the standard indirect method [8,11–15], a version of which is used by the UN Inter-Agency Groups for Child Mortality Estimation (IGME), or the combination of two methods outlined by Rajaratnam and colleagues [9], used in the Global Burden of Disease (GBD) study. For detailed review on these methods, see S1 Text. In brief, the standard indirect method uses simulated coefficients applied to the ratio of CD to CEB, aggregated at different maternal age (or TSFB) cohorts to estimate mortality rates and locate them in time. The GBD methods use pooled DHS survey data to inform two types of indirect estimation models, which are then combined to produce final estimates. The maternal age cohort (MAC)-based method is fundamentally

similar to the standard indirect method. The maternal age period (MAP) method uses empirical distributions, tabulated from DHS CBH data, describing the proportion of children born as well as the proportion of CD to mothers of specific ages in each year preceding the survey. Separate MAP distributions are produced by maternal age, CEB, and region. The period-specific aggregations of expected CD and children born derived from these distributions are used to locate mortality risk in time in SBH data.

Other methods such as cohort change and birth history imputation have been proposed [10,16], but in general, the development of new methods for indirect estimation of age-specific mortality has been understudied. Furthermore, none of the major existing methods have explored the use of predictive covariates measured at the individual mother or child level. The continued investment in collection of DHS surveys over the past 30 years has provided a massive dataset in which both SBHs and CBHs are available and thus the opportunity to empirically train and test new methods.

In this paper, we describe and test a novel method for indirect estimation of age-specific mortality using SBHs, based on a discrete hazards survival analysis model. This approach differs from existing popular indirect methods in two main respects: it produces a cohesive set of age-specific trend estimates without reliance on model life tables, thus allowing for the flexibility to estimate mortality rates for younger age groups such as neonates, and it is fit and predicted at the individual level, utilizing time-varying individual covariates.

Methods

Data

We analyzed 243 DHS (https://dhsprogram.com/) surveys from 76 countries, collecting CBHs and SBHs on 8,504,688 children from 2,346,538 mothers. We included DHS surveys and related Macro Malaria Indicator Surveys conducted since 1988 and available by October 2017. A full listing of the surveys used with summary information can be found in the S1 Table.

Birth history data in DHS surveys are recorded as follows: women are asked a series of questions about how many sons and daughters they have given live births to, including how many live with them now and how many have died. Certain probing questions are included to get more accurate responses. These data are aggregated to CEB and CD, forming the SBH component of the data. CBHs are also collected for each child born to the mother. Month and year of birth are recorded, as is age if the child is still alive. If the child reporting on had died, age at death is recorded in days if the child was under 1 month at death, in months if the child was under 2 years at death, and in years if the child was 2 years or older at death.

We further analyzed an additional 243 censuses and household surveys from 93 LMICs in order to demonstrate how the method can be applied in datasets in which only SBH was collected and to validate our results against concurrent CBH data. Of the SBH-only sources used, 71 were census, 81 were UNICEF Multiple Indicator Cluster Surveys (MICS), and the rest were from other household survey families such as Living Standards Measurement Surveys and other country-specific household surveys. The DHS datasets, as well as an additional 99 other CBH data sources, were used for comparison.

To identify data sources, we searched the Global Health Data Exchange (GHDx, http://ghdx.healthdata.org/) for national census and survey data in LMICs with the following key words: complete birth history, summary birth history, child mortality, and infant mortality. This was

further supplemented by bespoke searches on national statistics agency websites. We used only data sources for which individual-level data were available. A full listing of these data sources can be found in the S1 Table and S2 Table, along with their GHDx record identification number, where links to data distributors are provided.

Statistical model

Our goal was to develop a model that could be used to predict age-specific trends in mortality using SBH data only. We used CBH data to train the model, since it allowed us to identify mortality risk in time and age. For independent variables, we only used attributes that were available from SBHs, since we ultimately wanted to use this model to estimate mortality trends in datasets in which only SBHs are available.

We treated data from CBH as time to event, or survival data [17,18]. The goal of survival modeling is to estimate the underlying hazard or survival functions that describe the risk of event over exposure time. Special care is taken for data that are right censored, for which event status is unobserved after a certain period. In the context of child mortality data, the "event" of interest is a death, "exposure time" is age since birth, and right censoring occurs when a child is reported alive.

Most survival models can be expressed in the general form $h(age|\beta \boldsymbol{X}) = h_{0,age}e^{\beta \boldsymbol{X}}$,
where h_0 represents the baseline hazard function over age, which is shifted by weighted effects of covariates $\mathbf{X}$. The baseline hazard function can be fit either parametrically to a variety of smooth functions defined either by probability distributions or as flexible splines [19] or discretely, either using arbitrary age bins or in data-defined age bins, as in the widely used Cox proportional hazards model. Covariates will generally shift the hazard function and as such have a proportional effect across ages. This proportionality can be relaxed using age-varying covariates.

For this analysis, we adopted a discrete time survival analysis (DTSA) [20] approach to modeling the baseline hazard function. In a DTSA model, age is split into discrete bins, which conforms well to the discrete nature of age reporting in CBH data. The baseline hazard function is flexibly parameterized using fixed effects dummies, I, for chosen discrete age bins. This is achieved by reshaping input data such that every row in the new dataset is associated with each age bin, a, entered into by each child, i, in the data. Censored age bins for any child are not included in the reshaped data. An indicator variable $\mathrm{Y_{i,a}}$ is included for each row and recorded as 1 if the child died in that age bin and 0 if they survived it. A no-covariate baseline hazard could then be determined by fitting the following logistic regression model:

$$Y_{i,a} \sim Bernoulli(q_a)$$

$$logit(q_a) = \sum_{a=1}^{A} I_{i,a}\beta_a$$

Note that fixed effects are estimated for each age bin without an intercept term, so that each β_a is in reference to zero, and thus each e^{β_a} are interpretable as the probability of mortality in age group a, conditional on survival to age group a, or (q_a). In the discrete case we thus refer to $q_a$as the probability

of death within age bin a, though 'mortality rate' is commonly used interchangeably to describe the same quantity. This basic model can be extended to include individual-level covariates, random effects to account for hierarchical data, transformations or smoothing splines on covariates to improve prediction, and interactions with age-bin dummies in order to allow for non-proportional effects of covariates.

For this application, we used the following seven age bins for this analysis: live birth to 29 days (NN), 30 days to 5 months inclusive (post-neonatal 1 [PNN1]), 6 to 11 months inclusive (post-neonatal 2 [PNN2]), 12 to 23 months inclusive (1yr), 24 to 35 months inclusive (2yr), 36 to 47 months inclusive (3yr), and 48 to 59 months inclusive (4yr). These bins were chosen to align with the way in which age information is collected in DHS such that each age bin would have identifiable data on children entering and dying within it. We separated the first year of life further into three age bins because there is a high and quickly changing mortality hazard during this period. The NN period during the first month of life is further split because it is often of separate public health interest because of the unique epidemiology of causes of death during this period. Fig 1 shows this simple baseline hazard function fit to the 2011 Burundi DHS dataset for illustration purposes.

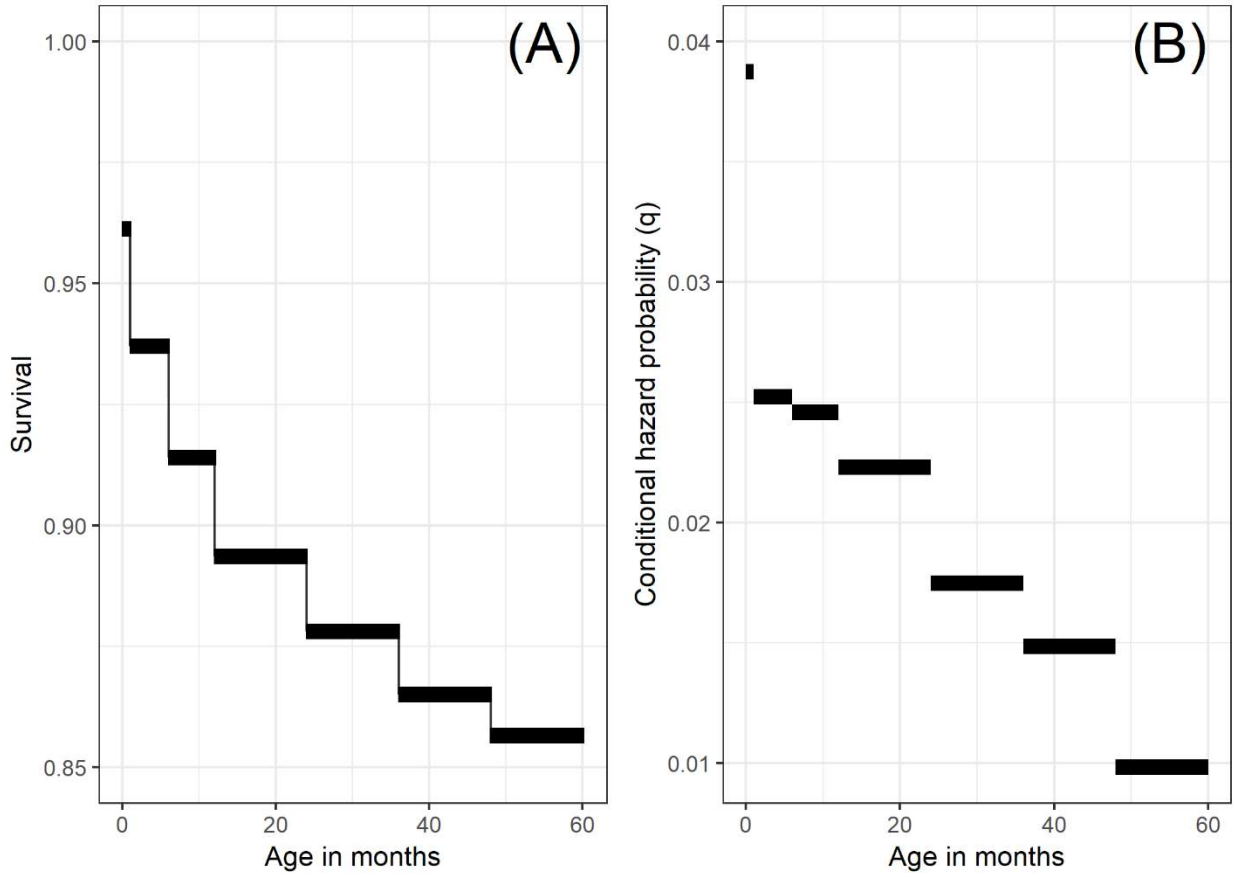


Fig 1. *Illustrating the estimated pooled baseline discrete hazard and survival functions from the 2011 Burundi DHS dataset, fit using the seven age bins* $a \in (1 = NN,\ 2 = PNN1, 3 = PNN2, 4 = 1yr,\ 5 = 2yr,\ 6 = 3yr,\ 7 = 4yr)$. *Note that we are estimating discrete hazards, and thus, hazards (shown in panel B) are interpreted as a conditional probability rather than a conditional rate. The survival function (shown in panel A), showing estimated survival at the end of each age bin, is calculated directly from estimated hazards as* $\widehat{S_a} = \prod_{\alpha=1}^{a}(1 - q_\alpha)$. *DHS, Demographic and Health Surveys; NN, neonatal; PNN1, post-neonatal 1; PNN2, post-neonatal 2.*

We trained the model on the pooled CBH database with the purpose of making predictions in situations for which only SBHs are available, as in census data. As such, we were limited to using covariates from the training data, which were also available in SBH-only datasets. Certain covariates, such as year of birth and mother's age at birth, were found to be highly predictive of mortality but could not be ascertained directly from SBH data. In order to account for them, we approach predicting from the perspective of a hypothetical child. Specifically, for any given woman in the target SBH data, we wished to predict hazard functions for all hypothetical children she could have had over the course of her childbearing years. For example, if a 30-year-old woman was observed in a dataset collected in 2010, we could predict a separate hazard function for a potential child born to her each year going back until she was 12 in 1992. Hazard functions for these hypothetical children could be differentiated by covariate values that vary over the mother's life.

We specified the following generalized additive DTSA model for the conditional probability of death for every age bin a of each child i to each mother m:

$$Y_{m,i,a} \sim Bernoulli(q_{m,i,a})$$

$$logit(q_{m,i,a}) = \sum_{\alpha=1}^{7}[I_{i,\alpha}\beta_{\alpha}] + \sum_{\alpha=1}^{7}[g_{1,\alpha}(yr_i, SDI_{c,yr,i})I_{i,\alpha}] + g_2\left(\frac{CD_m}{CEB_m}, CEB_{m,yr}, MothAge_{m,yr}\right) + \nu_{svy} + \eta_{country,a}$$

$$\nu \sim Normal(0, \sigma_{\nu}^2)$$

$$\eta \sim Normal(0, \sigma_{\eta}^2)$$

where $g_*(\cdot)$ represents thin plate regression spline smooths, with $g_1(\cdot)$ having separate smooths for each age bin a, operationalized by the age bin dummy variable I. $\mathrm{yr_i}$ represents the year of birth for child i. This is directly observed in the training data but for prediction is assigned for each hypothetical child. $SDI_{c,yr,i}$ represents the Socio-demographic Index (SDI) [21] for the country *c* that child *i* was born in at their year of birth yr. SDI is a composite average, expressed on a scale of 0 to 1, of income per capita, average educational attainment, and fertility rates and has been found to be a strong predictor of child mortality [21]. The interaction of SDI and year of birth allows the secular trend in mortality for each age bin to vary by the level of development in each country, thus allowing for prediction in countries without training data.

The variables in the second smooth represent child- and mother-level covariates. $\frac{CD_m}{CE_m}$ is the ratio of children died to children born to each mother m at the time of the survey. $MothAge_{m,i,yr}$ is the mother's age at the year of birth. This is observed in the training data and assigned for prediction of hypothetical children in the same way as yr. Finally, $CEB_{m,yr}$ is the number of children born to the mother at the time of birth yr of child i. This is directly observed in the training data. For prediction, we use empirical probability of birth distributions [9] to impute this value for each hypothetical child. Much in the same way that the standard indirect method interacts $\frac{CD}{CEB}$ with fertility ratios, this interaction is

included to address the fact that the relationship between $\frac{CD}{CEB}$ and q is mediated by the fertility experiences of the women reporting $\frac{CD}{CEB}$ [15]. This differs from previous approaches, which used aggregate levels of fertility, and instead depends on individual women's fertility experiences.

Finally, ν and η are independent normal random intercepts for each survey and each age bin within country.

All covariates were centered and scaled by their standard deviations for model fitting. Models were fit separately by the same regions used by Rajaratnam and colleagues [9]. Uncertainty in predictions was ascertained by taking 1,000 multivariate normal draws from the variance–covariance matrix of fitted model parameters, including fitted random effects values [22]. In cases for which prediction data had random effects levels not used in the training data (for a new survey or a new country), estimated variances $\widehat{\sigma_\nu^2}$ and $\widehat{\sigma_\eta^2}$ were used to simulate 1,000 independent normal draws. Models were fit using restricted maximum likelihood with the bam command from the mgcv package in the R Statistical Computing Language Version 3.4.3 [23,24].

Conversion to trends

Using the model described above, we estimated age-specific mortality hazards for individual hypothetical children to mothers responding to SBH questionnaires. These hazard functions of hypothetical children must then be converted into trends in age-specific mortality. To do so, we aggregated estimates of mortality among hypothetical children born in the period using weights that indicated the likelihood that each hypothetical child actually existed. This process is illustrated in Fig 2.

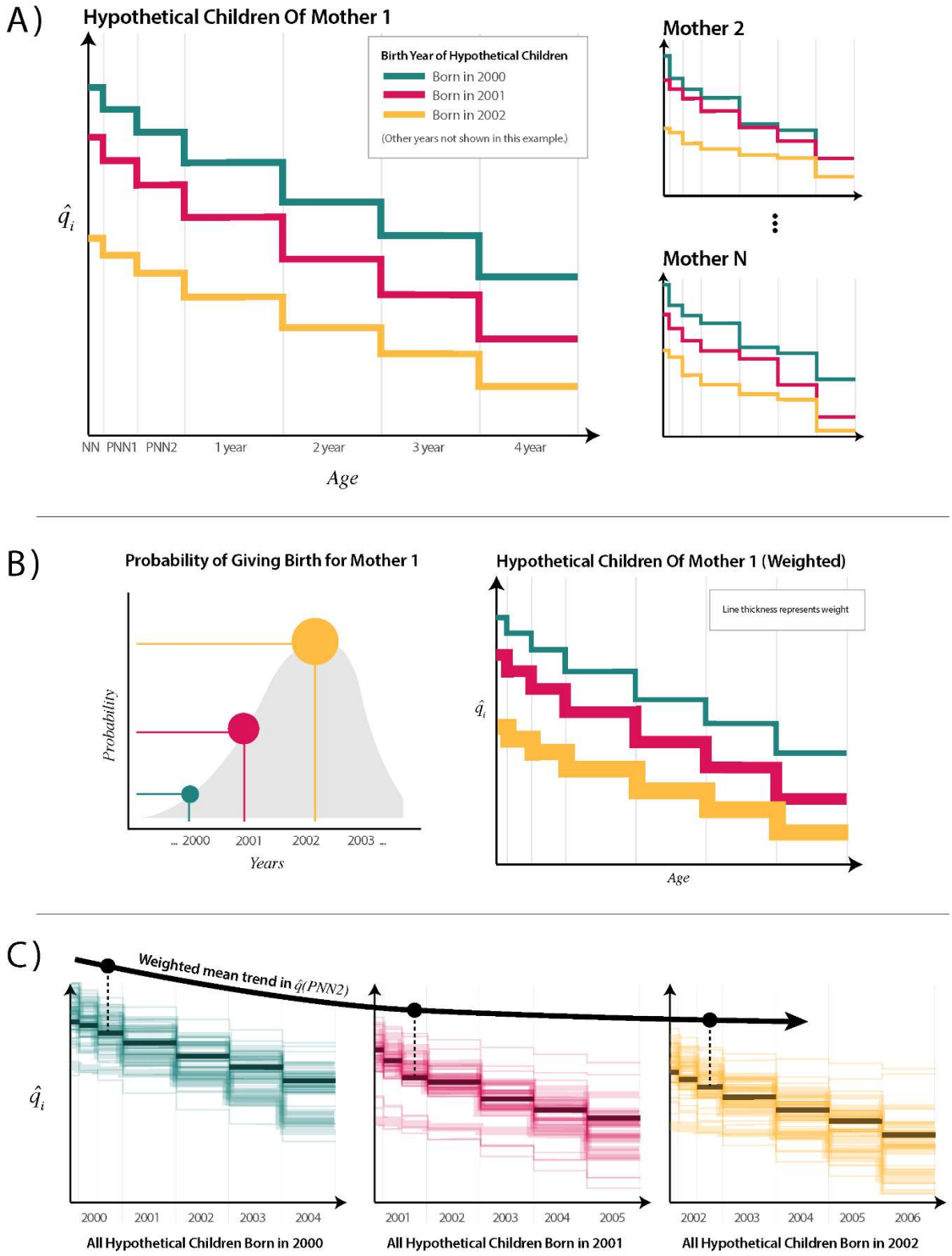


Fig 2. Illustration of procedure to convert discrete hazard functions for hypothetical children to population-level age-specific trends.

(A) Discrete hazard functions are estimated for each hypothetical child from each mother in the target SBH dataset. Here, we color all children born in the same year with the same color. Only 3 years are shown for simplicity in this example. In real data, the years of birth of hypothetical children would vary by mother based on her age, such that there would be one hypothetical child for each year going back in time from the survey until the mother was 12 years old. (B) Probability of birth distributions is applied to each hypothetical birth from each mother. These are derived from the empirical map distributions from Rajaratnam and colleagues [9], in which a different probability is available by woman's age, CEB, region of residence, and year prior to the survey. These probabilities are multiplied by each mother's CEB and carried through to subsequent age bins to estimate the expected number of children entering each age bin (EEB) using estimated survival probabilities. As such, line thicknesses get slightly smaller with each subsequent age bin. The EEB value for each hypothetical child's age bin represents the number of children entering that age bin that the hypothetical child represents for their given mother. (C) All hypothetical children to mothers are grouped by year of birth. The estimated mortality probabilities for each age bin from all hypothetical children born in the same year are pooled, and EEB is used to calculate a weighted mean. Trends are drawn across $\hat{q}_a$ for each year, indicated here by a trend in the third age bin. This aggregation procedure can be done for any grouping of women to make estimates for a survey cluster, a district, or a whole country. CEB, children ever born; EEB, expected entering bin; NN, neonatal; PNN1, post-neonatal 1; PNN2, post-neonatal 2; SBH, summary birth history.

From the model, we obtained estimates of $\hat{q}_{m,a,yr}$: the probability of death in age bin *a* for a hypothetical child born in yr to mother m. To obtain estimates of $\hat{q}_{m,yr}$: age-bin and period-specific hazards representative of the population surveyed, we weighted each child based on their probability of birth. Each hypothetical child was assigned a probability of birth ($POB_{m,yr}$) using the birth distributions used for the GBD-MAP method. Probability of birth distributions are compiled from empirical distributions that describe, for each year preceding a survey, the probability of birth based on mothers' age and CEB and by region. Distributions were matched based on geographical region, mothers' age, CEB, and yr to each hypothetical child.

We then assigned a weight to each age bin of each hypothetical child. We defined the expected number of children entering each age bin (expected entering bin [EEB]) *a* for child born in year yr from mother m as the following:

$$EEB_{m,a,yr} = POB_{m,yr} * CEB_m * \hat{S}_{m,a,yr}$$

where $\hat{S}_{m,a,yr}$ is the estimated survival until age bin a, and EEB is the estimate of the number of children entering each age bin for the hypothetical child born to mother m in year yr, given each mother's overall fertility and the estimated mortality experiences of her children over time.

We aggregated our estimates across $\hat{q}_{a,yr}$ by taking a weighted mean such that

$$\hat{q}_{a,yr} = \frac{\sum_{m=1}^{M} \hat{q}_{m,a,yr} EEB_{m,a,yr}}{\sum_{m=1}^{M} EEB_{m,a,yr}} = \frac{Expected\ Deaths_{a,yr}}{Expected\ Children\ Entering_{a,yr}}$$

The benefit of predicting at the individual level is that weighted means can be aggregated for any population desired. Also, this procedure conveniently provides not only estimates of $\hat{q}_{a,yr}$ and expected children entering each bin but also the numbers of expected deaths. For nationally representative estimates, survey weights can also be included into this procedure by multiplying weights into the summands. Finally, age bins can be combined as independent conditional probabilities to produce trends in wider age bins that may be of interest, such as $1\hat{q}0$ or $5\hat{q}0$.

Uncertainty in aggregate estimates of all quantities are calculated by repeating the aggregation procedure 1,000 times based on the predictive draws of $\hat{q}_{m,a,yr}$. We report the 2.5% and 97.5% quantiles.

Validation and verification

We developed three approaches to model validation. We first used cross-validation on the DHS data in order to assess how well age-specific mortality trends estimated from our method could reproduce those directly estimated from CBH data. We then compared our indirect $5q0$ estimates to those produced using existing methods. Finally, we applied the method to nationally representative non-DHS surveys that only collected SBHs and compared those results to contemporaneous direct estimates.

We developed the first model validation framework to assess out-of-sample predictive validity, holding out entire DHS surveys from the database, fitting the predictive model in their absence, and using their SBH variables to produce indirect age-specific time trends. We then used direct estimates from the CBHs of these held surveys to reproduce age-specific trends to serve as a basis for validation.

For each country in the DHS database, we held out the most recent DHS. We fit the model, used the fitted parameters to make indirect estimates from SBHs, and compared to direct estimates from CBHs. This was repeated for each country. Using the most recent survey represents a particularly difficult test because doing so requires several years of out-of-sample projection from the time since the penultimate survey in that country.

Our aim was to minimize the bias and magnitude of errors (the difference between estimates and validation data). We used the following five metrics to assess out-of-sample predictive performance: (1) Mean error (ME) to capture systematic bias. An ME of zero indicates a perfectly unbiased estimate. ME is an absolute (as opposed to relative) metric and thus cannot be compared across age bins. (2) Standard deviation of the errors (SDE) to capture how much variation there is in out-of-sample errors across countries and years. The smaller the SDE, the more precise the errors are. Again, SDE is an absolute metric. (3) Median relative error (MRE) to capture relative bias. MRE is simply the ratio of estimate to validation data, and as such, an MRE close to one indicates no bias. MRE allows us to compare bias on a relative scale across age bins. (4) Median absolute percentage error (MAPE) to capture the relative scale of errors. This is calculated as the ratio of the absolute error to the direct validation estimate multiplied by 100. The MAPE represents overall relative accuracy of the estimates, with a value close to zero indicating high accuracy. (5) The coefficient of determination (R^2) represents the proportion of total variance in the directly tabulated hazards explained by the modeled estimates. Each of the metrics were assessed for each age bin as well as for $5q0$.

Single-year age-specific direct tabulations of CBHs have relatively small sample sizes and can produce somewhat noisy estimates of a "truth" for comparison. Since we are interested in modeling the actual underlying trend and not the noisy observed values, very good predictive performance in this case could actually signal overfitting. Furthermore, mean-based metrics are sensitive to large outliers in errors, which could emerge spuriously when validation data are noisy. In other words, validation data with a larger sample size are expected to produce a precise approximation of the true underlying mortality hazard. We dealt with this in two ways. First, following Rajaratnam and colleagues [9], we smoothed the noisy validation trends using loess (with α set to 0.85). Second, we weighted all of our metrics of

predictive performance by the sample sizes (number entering each age bin) as tabulated from the raw validation data.

We also used this same validation approach to evaluate our estimates of the numbers of children expected to enter each age bin, or EEB. This is done by comparing EEB with the direct tabulations of numbers of children entering each bin, in each year, from the validation data.

With increasing interest in subnational child mortality estimation [6,25–28], it is also critical to assess the validity of these results at subnational levels of aggregation. Most SBH data are geographically identifiable to the first administrative level—typically referred to as states or regions in most countries [6]. We aggregated to the first administrative unit, defined using the Global Administrative Unit Layers (GAUL) shape file made available by the FAO (http://www.fao.org/geonetwork/srv/en/metadata.show?id=12691). In order to obtain large enough sample sizes for stable comparison in the validation, we also aggregated data into 5-year bins preceding each survey. As such, each administrative area only supplied three estimates, and thus, we did not smooth them.

We also compared how well the proposed method estimated trends in $5q0$ relative to existing methods, since a well-behaving method for age-specific trends should also be able to accurately reproduce trends in $5q0$. We thus compared out-of-sample trends in $5q0$ estimated from our test data to those produced by the GBD methods, as well as the standard indirect method. GBD-combined indirect estimates for each available survey were taken from the GBD mortality database (Available online: https://vizhub.healthdata.org/mortality/) and were produced by combining MAP and MAC estimates. For the standard indirect method, we matched model life tables to countries as used by IGME. We included two variants of the standard indirect method: one based on MACs and one based on TSFB (see [11]).

We also used this cross-validation framework to compare our model to several other model specifications. These results are presented in the S1 Fig.

Finally, in order to better establish external validity of this method, we also sought to understand its performance on non-DHS data. By nature of joint data collection, CBH and SBH data from DHS are presumed to be highly consistent. For this reason, SBHs from DHS could be different enough from data for which only SBHs are collected—for example, in censuses—that a cross-validation based on DHS alone may not be sufficient evidence that method performs well for these sources. Thus, for a more practical perspective on the performance of this method in settings where it is intended to be used (i.e., in data for which only SBHs were collected), we compared estimates from these data to directly estimated mortality for which concurrent CBH data were available. First, we applied our method to 243 nationally representative SBH surveys and censuses in order to estimate trends in NN, infant, and under-5 mortality. We then directly estimated these same trends from CBH surveys in the same countries, as described above for the DHS cross-validation, which served as a basis for comparison. We then identified CBH–SBH estimate pairs from the same country-years. Looking at the differences in these pairs of estimates, we used the same set of predictive validity metrics described for the cross-validation assessment in order to assess how similar the age-specific indirect estimates were to contemporaneous direct estimates. Furthermore, in order to understand how our model performed across different data source types, we stratified this comparison across MICS, censuses, and other survey sources.

Data and code availability

All datasets used for this analysis are listed in the Supporting information (S1 Table for the DHS and S2 Table for additional surveys and censuses used for external validation). Each source in the tables is supplied with an ID number associated with a record in the GHDx (http://ghdx.healthdata.org/). Each GHDx record links to data providers for each survey.

All code for the analyses in this manuscript is available at https://github.com/royburst/sbh_agespecific_indirect_paper_code. In the near future, we plan to release a package for R that allows users to apply this indirect method to any SBH dataset.

Results

Table 1 shows summary statistics from our cross-validation. The table shows the mean estimates of age-specific mortalities, q_a, across all countries and age bins, along with aggregated out-of-sample predictive validity metrics for estimates of q_a. We find little bias across all ages, as indicated by very small MEs and MREs close to one. The bias that does exist tends to be slightly over in the younger age bins and slightly under in the older age bins. We also see relatively small SDE across all age bins, indicating that, on average, there is a not large variation in out-of-sample errors across countries and years. Relative variance in errors, measured by MAPE, increases as q_a decreases as a function of age.

Age-bin	$\overline{q}_a$	*ME*	*SDE*	*MRE*	*MAPE*	R^2
NN	0.031	0.0022	0.005	1.05	9.5%	0.82
PNN1	0.015	0.0010	0.004	1.09	15.9%	0.80
PNN2	0.013	0.0004	0.004	1.08	17.3%	0.82
1yr	0.013	0.0002	0.004	1.02	18.2%	0.88
2yr	0.009	-0.0001	0.002	1.00	16.2%	0.93
3yr	0.006	-0.0002	0.002	0.96	20.7%	0.88
4yr	0.003	-0.0001	0.001	0.97	20.6%	0.81
5q0	0.083	0.0033	0.010	1.05	8.8%	0.95

Table 1: *Overall out of sample predictive validity metrics for each age bin and mean direct estimates of q_a across all country years in the Demographic and Health Surveys database, for the 15 years prior to the survey being taken. Abbreviations: NN = livebirth to 29 days (Neonatal); PNN1 = 30 days to 5 months inclusive (Post Neonatal 1); PNN2 = 6 to 11 months inclusive; 1yr = 12 to 23 months inclusive; 2yr = 24 to 35 months inclusive; 3yr = 36 to 47 months inclusive; 4yr = 48 to 59 months inclusive; ($\overline{q_a}$= average estimated mortality probability; ME = mean error; SDE = standard deviation of the errors; MRE = median relative error; MAPE = median absolute percentage error; R^2= coefficient of determination.*

Fig 3 plots the agreement between age-specific mortality rates from the validation data compared to out-of-sample estimates. We also see a relatively high proportion of variance explained as measure by R^2, with all age bins above 0.80. Predictive validity metrics for the combined $5q0$ age bin perform better than for the smaller age bins, as the model can explain 95% of the variance in input data. This is likely because of several reasons: errors are averaged over when collapsing across ages; relative metrics are less sensitive with a larger overall q_a; age bins with larger relative errors tend to have lower hazards, which contribute less overall mortality and thus impact metrics in the combined group less; and larger sample sizes lead to more stable estimates.

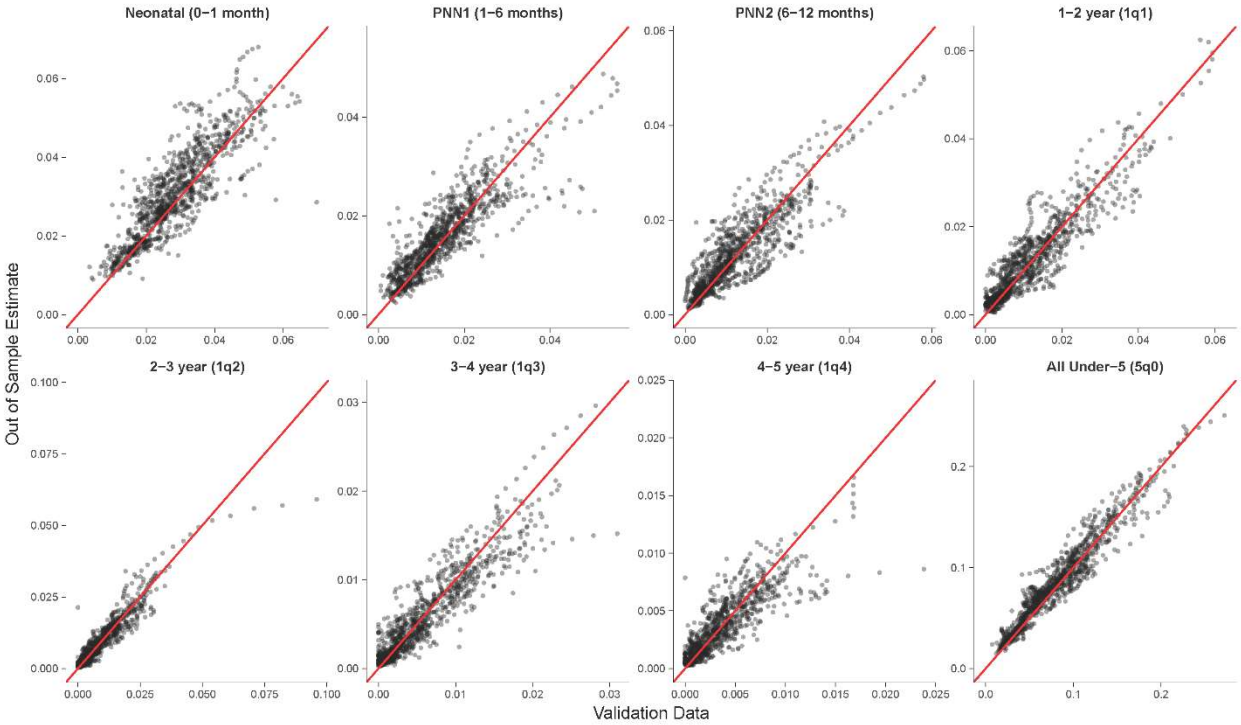


Fig 3. *Out-of-sample predictions of mortality probability compared against loess-smoothed validation data. Each point represents a country–age mortality estimate ($q_{a,yr}$) for each held-out survey from the DHS database. Red line indicates unity. DHS, Demographic and Health Surveys; loess, local regression; PNN1, post-neonatal 1; PNN2, post-neonatal 2.*

Fig 4 compares EEB with the observed number of children entering each age bin from the validation data. There was high agreement across age groups, with MRE ranging from 1.015 to 1.032 and MAPE ranging from 6.8% to 11.0%, indicating small errors and potentially a very slight upward bias in the EEB estimates. There is no clear difference in EEB performance across age bins. Overall, R^2 was 0.97. This indicates that empirical probability of birth distributions can be reliably used to approximate sample sizes for indirect estimates. This also adds support to the favorable validation results shown above, as EEB weights are an important component of aggregating trends to the national level and because empirical probability of birth distributions are used to impute CEB at birth for prediction.

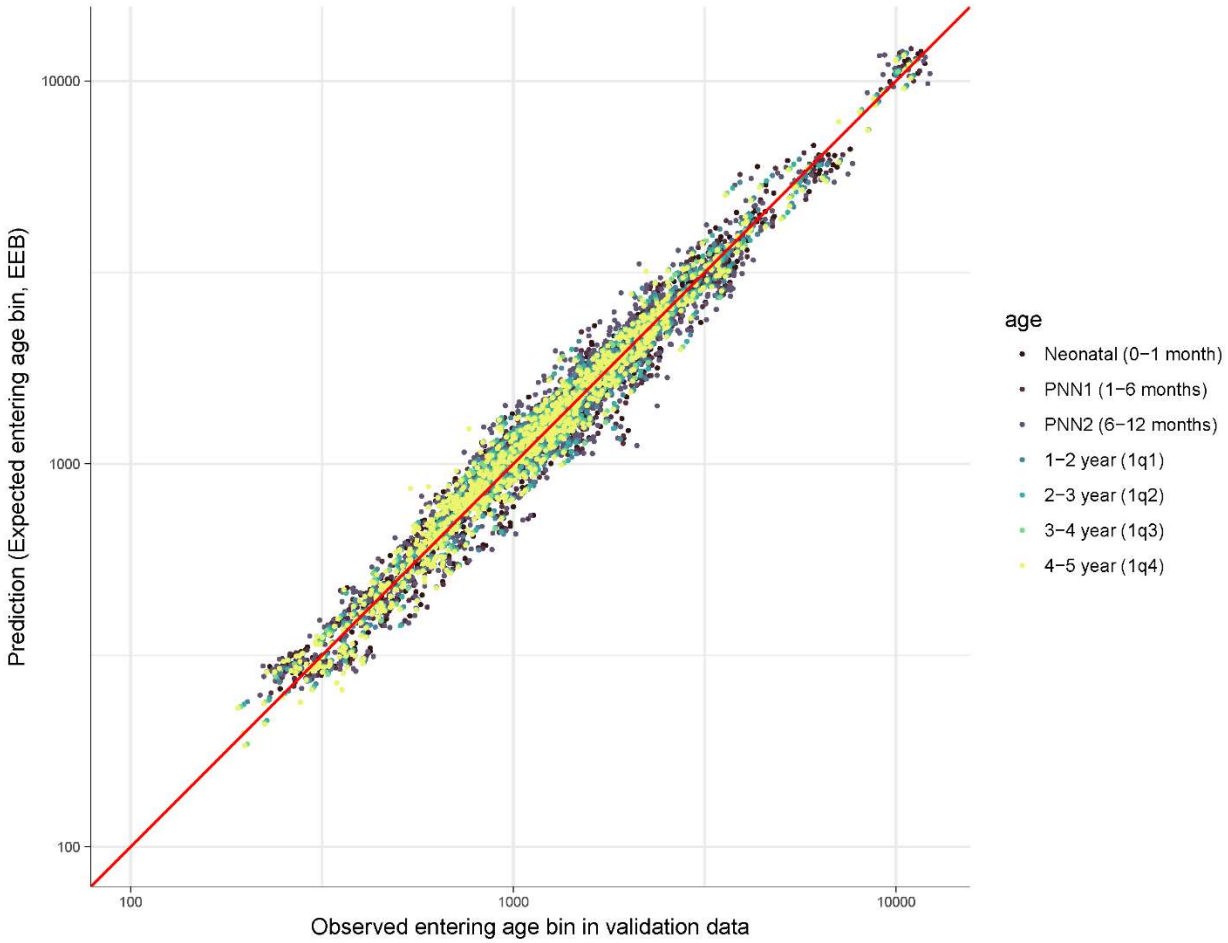


Fig 4. *Comparison of EEB and observed children entering each age bin from CBH validation data.*
Each point represents a survey year age bin; both axes are on log 10 scale. R^2 *= 0.97. CBH, complete birth history; EEB, expected entering bin; PNN1, post-neonatal 1; PNN2, post-neonatal 2.*

At the subnational level, model performance was somewhat weaker. There was a similar pattern in direction of bias across the ages, though bias remained minimal overall. There was more variability in the errors, with MAPE ranging from 20.0% in the NN group to 38.0% in 4-year-olds. The percentage of variance explained was also somewhat lower than at the national level. R^2 of the subnational 5q0 estimates was 0.91. Some of this difference was likely due to smaller sample sizes in the subnational data compared to the national validation. Our validation data, which were based on direct estimates from CBH, represent realizations of the underlying probability, and thus, the empirical probability from the validation is measured with noise. Despite aggregating to 5-year bins, the average number of children born in each 5-year aggregated subnational observation was 520, compared to 1,769 in each annual national observation, and 4,148 (27%) of each survey-administrative area-age bin observation had no observed deaths. S2 Fig and S3 Table replicate Fig 3 and Table 1, respectively, at the first administrative subnational level.

Fig 5 shows the out-of-sample estimated trends in age-specific mortality rates estimated using the 2013 Nigeria DHS and compared to the directly estimated validation data. In the S4 Fig, we provide similar plots for each country with extended discussion on those results. Overall, the model was able to reproduce trends in the validation data in Nigeria and in most other countries. Performance was suboptimal in cases for which test and train data differed significantly (for example, in Benin) and when trends were unique to a given country (for example, in Lesotho).

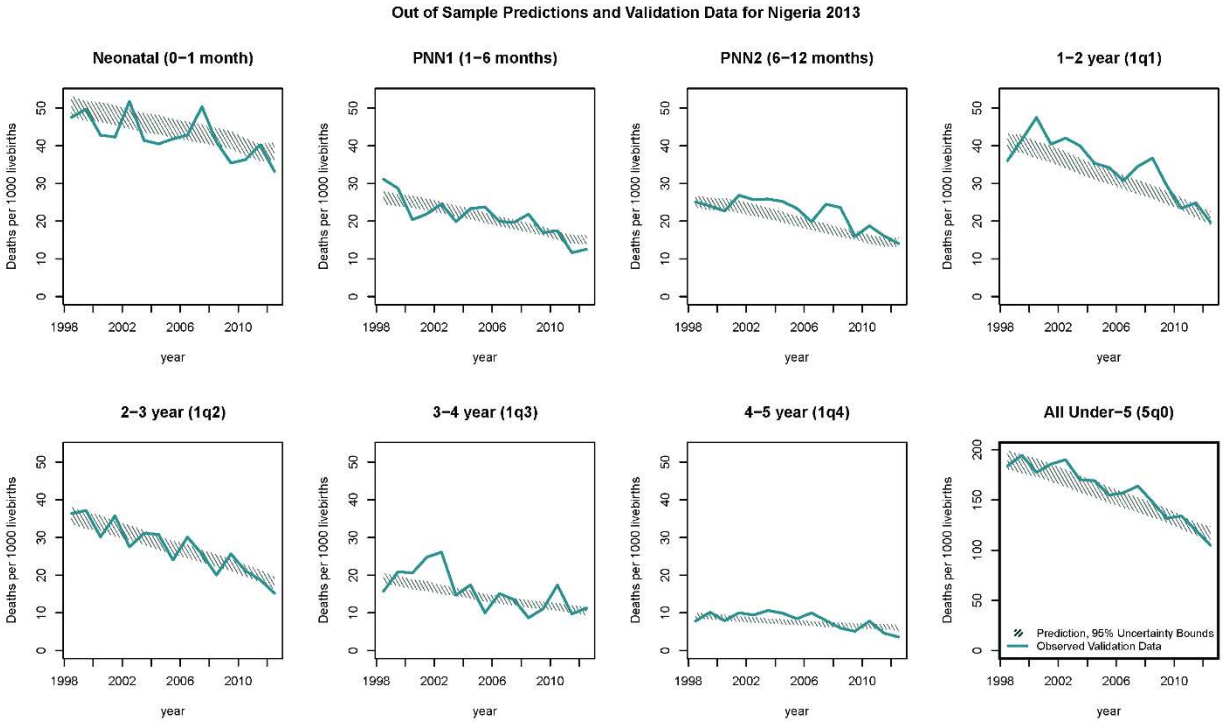


Fig 5. *Trends in mortality for each age bin from the 2013 Nigeria DHS. Thick blue lines are validation data; hatched lines are the 95% uncertainty bounds on the out-of-sample predictions. Sampling variation is evident in the blue line through year-on-year spikes. The target of prediction was the overall time trend, leading to a smoother prediction. Axis scales are fixed except for* $5q0$*, which is the combination of the mortality rates from the seven age bins. Similar plots for each country in the validation data are available in in the S4 Fig. DHS, Demographic and Health Surveys; PNN1, post-neonatal 1; PNN2, post-neonatal 2.*

We compared predictive validity in our out-of-sample estimates to indirect estimates of trends in $5q0$ made from the same SBH holdout data, using the GBD-combined method and the standard indirect method. Fig 6 compares predictive validity metrics for the three methods over the 15 years preceding the survey. Confirming results from Rajaratnam and colleagues [9], we find unstable estimates from the standard indirect method in the most recent 5 years preceding surveys. Near overlap in the MRE and MAPE over time indicates that the new method and the GBD-combined methods generally produce similarly performing results. Overall, we estimated a MAPE of 8.8% for the new method, 6.1% for the GBD-combined method, and 25.6% and 36.5% for the MAC and TSFB variants of the standard indirect method, respectively. The new method and GBD-combined methods each had an MRE of 1.05 and 1.02, whereas the MAC and TSFB variants of the standard indirect method each had an MRE of 1.25 and 1.36. Finally, the R^2 for the new method was 0.95, whereas it was 0.98 for the GBD-combined method and 0.80 for the MAC variant and 0.88 for the TSFB variance of the standard indirect method. If we excluded

the most recent 5 years, MRE and MAP remained largely the same, but the R^2 for the MAC and TSFB variants rose to 0.92 and 0.88, respectively. S3 Fig shows trends for each survey in the testing data. We note that for certain surveys with no GBD-combined estimates, such as the Malawi DHS 2016, we are able to produce accurate trends using the new method. It is possible that the non-GBD methods would have performed even better by comparison if these were included. Furthermore, several of the surveys in the testing set were used to train the GBD-combined models while remaining out of sample for the new method and the standard indirect method, potentially giving the GBD-combined method a slight advantage in this comparison.

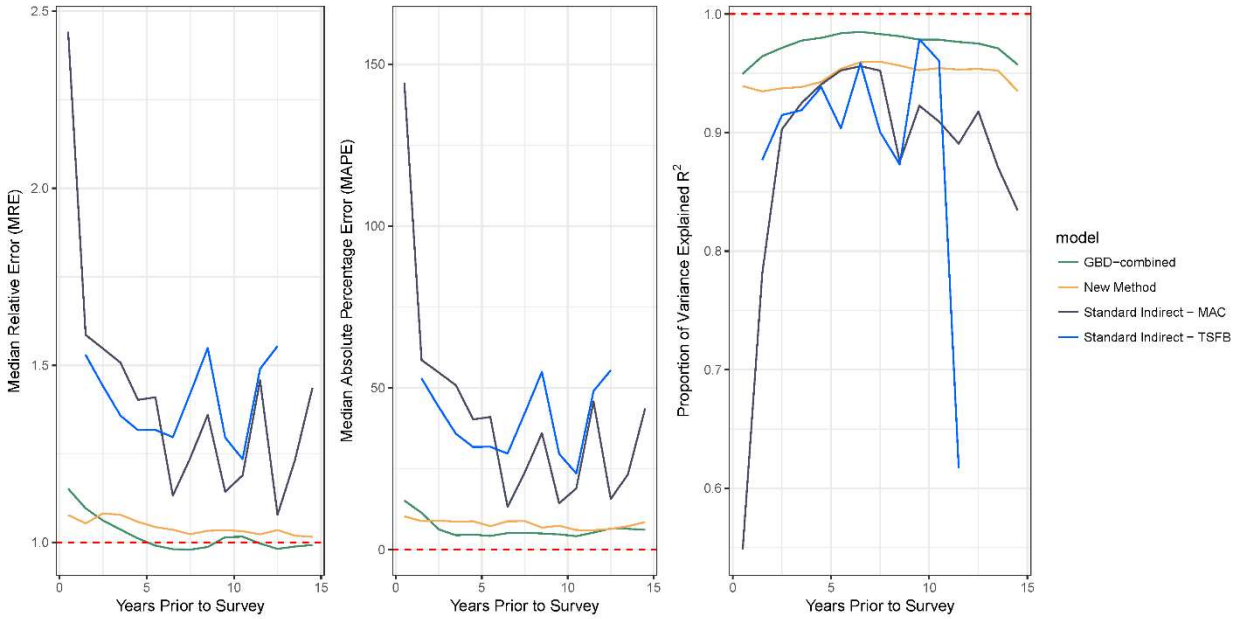


Fig 6. Comparing predictive validity metrics across different methods for indirect estimation of 5q0. Both the GBD-combined and new methods greatly outperform the standard indirect methods, particularly in the most recent 5 years. Note that whereas the standard indirect methods produce estimates in only a few years for any given survey, the location of these times vary based on the population in each survey as determined by the observed fertilities in each source, and thus, we are able to make annual estimates going back 15 years. GBD, Global Burden of Disease; MAC, maternal age cohort; TSFB, time since first birth.

For external validation, we identified 243 censuses and surveys from 93 countries, in which only SBH was collected. As a basis for comparison, we identified 316 CBH datasets. Applying our method, we estimated trends from each SBH-only data source and identified 16,527 estimate pairs for which we had contemporaneous SBH- and CBH-derived estimates in a single country-year. Estimates for any year after 1990 and within 15 years of the survey data were kept. We further identified 2,694 country-year–age pairs of data from 524 unique country-years for which only SBH data were available. For comparison, we also identified 10,655 country-year–age pairs for which two concurrent CBH direct estimates were available.

Full trend plots for each country with available data are in S6 Fig. In the majority of cases, trends from SBH-only data closely match contemporaneous trends from CBH data. There were several SBH-only surveys that exhibit overall source-level bias relative to concurrent trends.

Table 2 summarizes our findings for these paired comparisons for NN, infant (under-1), and under-5 mortality. We find close agreement across validation metrics. Overall variance was slightly higher and unadjusted R^2 was slightly lower than in the cross-validation assessment. Much of this additional variance could be explained by survey; by simply controlling for data source using survey-level fixed effects, we find large improvements in R^2, with each age bin around 0.96. We further found these results to be robust across SBH data type (census, MICS, and other surveys).

Age-bin	SBH SOURCE	$\bar{q}_{a,cbh}$	$\bar{q}_{a,sbh}$	*ME*	*SDE*	*MRE*	*MAPE*	R^2	R^2corr
Neonatal	**ALL**	0.030	0.033	0.0026	0.010	1.06	17.7%	0.52	0.96
Infant	**ALL**	0.061	0.064	0.0030	0.018	1.05	16.4%	0.64	0.96
Under-5	**ALL**	0.093	0.096	0.0029	0.028	1.04	16.4%	0.74	0.97
Neonatal	**MICS**	0.038	0.041	0.0030	0.009	1.09	17.9%	0.5	0.96
Infant	**MICS**	0.075	0.080	0.0047	0.019	1.05	16.1%	0.67	0.97
Under-5	**MICS**	0.123	0.129	0.0055	0.029	1.05	16.3%	0.79	0.98
Neonatal	**CENSUS**	0.030	0.032	0.0019	0.010	1.02	18.4%	0.55	0.96
Infant	**CENSUS**	0.062	0.063	0.0013	0.019	1.00	17.2%	0.68	0.97
Under-5	**CENSUS**	0.097	0.096	-0.0005	0.026	0.98	17.0%	0.79	0.98
Neonatal	**OTHER**	0.029	0.032	0.0029	0.010	1.07	17.5%	0.48	0.96
Infant	**OTHER**	0.056	0.059	0.0034	0.017	1.07	15.8%	0.57	0.95
Under-5	**OTHER**	0.083	0.087	0.0037	0.028	1.06	16.2%	0.62	0.96

Table 2. *Summary results for the external validation comparisons across 16,527 country-year data pairs for which a CBH and SBH estimate were both available. We also show a sub-analysis by data source, indicating good robustness of results across source type. $\bar{q}$, average estimated mortality probability; ME, mean error; SDE, standard deviation of the errors; MRE, median relative error; MAPE, median absolute percentage error; R^2, coefficient of determination.*

Fig 7 shows a scatterplot of each country-year concurrent estimate. We also plot the same comparison for country-year pairs for which two direct CBH estimates are available. The comparison of CBH to CBH estimates represents a theoretical baseline difference we would expect to see in concurrent estimates. The similarity between the two sets of scatterplots highlights that much, though not all, of the variation we see between indirect and direct also exists between direct estimates and would be expected even given the best available survey data from which direct estimates could be made.

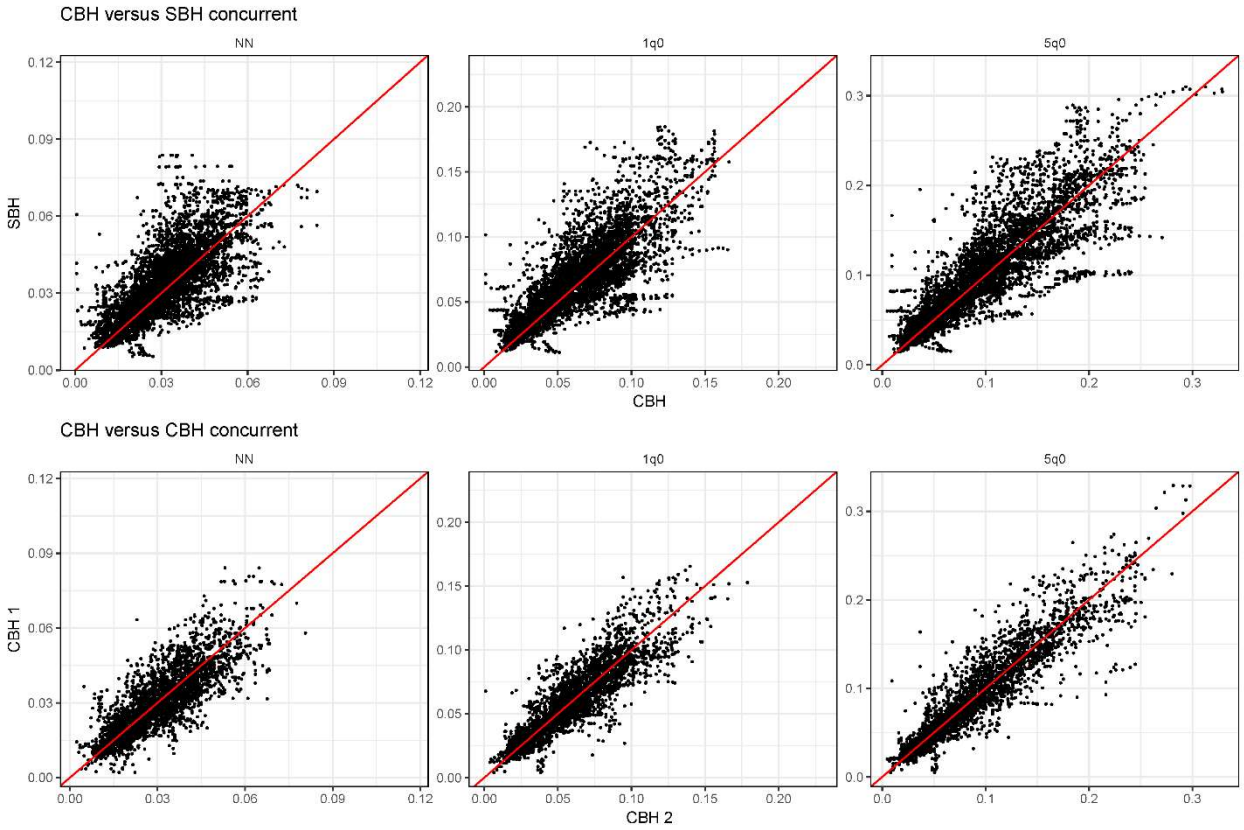


Fig 7. *Each country-year concurrent estimate for NN, infant, and under-5 mortality. The top row compared concurrent estimates from the SBH-only data with CBH direct estimates. The bottom row shows the same comparison from concurrent CBH estimates, theoretically representing a baseline level of variance we would expect in concurrent estimates. Comparing red lines indicate unity. CBH, complete birth history; NN, neonatal; SBH, summary birth history.*

Discussion

Our new method for indirect estimation produces age-specific mortality trends consistent with those produced using CBH data in most cases at the country- and first administrative unit–level, as well as producing $_5q_0$ estimates that improve on the standard indirect method and are closely comparable in performance with the current best-performing method [9]. We applied the method to external SBH data and found considerable agreement when comparisons could be made to contemporaneous estimates from CBHs. This new method greatly expands the potential utility of SBH data and fills a critical gap in the literature on indirect methods, extending indirect mortality estimation toward specific age bins of interest.

There are two main methodological innovations introduced by this new approach: using hierarchical survival analysis to model individual-level hazard functions and developing a hybrid approach to locating mortality risk in time. By viewing CBHs as time-to-event data, we were able to directly model the quantity of interest, the conditional probability of death $_nq_x$ at various ages from birth until age 5. Leveraging existing data from millions of CBHs, we inferred hazard functions that vary across countries, surveys, mothers, and their individual children using only covariates that were available in SBHs. These hazard functions, built up from flexibly chosen discrete age bins, then allowed us to produce indirect age-specific estimates for children born at various times. Since these estimates are made at the

individual level, they could then be aggregated to any population. Furthermore, accompanying model uncertainty is included in all predictions.

All indirect methods must rely on some approximation in order to locate mortality risk in time, since SBH does not provide explicit information on time of birth or death. MAC methods such as GBD-MAC and the standard indirect approach rely on observed fertility patterns to locate the mean time of risk for each maternal age group. They typically assume unchanging fertility and furthermore ignore recent mortality experiences to children from older mothers. The GBD-MAP method relies on empirical distributions of births and deaths to distribute risks in terms of years prior to survey. This allows older mothers to contribute information from more recent births but also runs the danger of overgeneralizing trends to the level at which data were pooled. Our new method utilizes several sources of information in order to locate mortality risk and to overcome some of the limitations with previous methods. First, secular trends over time are incorporated in the model but are allowed to vary by country-SDI to avoid overgeneralization and allow for prediction in countries not in the training data. Second, individual-level time-varying covariates allow us to predict hazard functions for hypothetical children born throughout different times in each mother's life so that all potential children, including recent births to older women, are incorporated. In order to aggregate trends, we use weights derived for the GBD-MAP method, which put more weight on hypothetical children that were more likely to have existed.

In applying our method to a variety of SBH-only data sources, we found that performance varied across specific sources, and validation metrics in the external data were slightly worse than in the DHS cross-validation assessment. It could be argued that the utility of any indirect method will depend on the quality of SBH data collected [15]. Though to the contrary, indirect methods such as ours, which have been validated externally as well as against high-quality DHS data, can also serve as a tool to assess the quality of these data sources. For example, by comparing trends from multiple available sources within a country, as we show in supplementary Fig 5, certain sources stand out as problematic. See, for example, the upward trend apparent in the Ghana 2000 census or the downward shift seen in the 2008 Cambodia census. These two censuses suffer from different data quality issues, and a future research should focus on understanding the topology of potential issues in SBH data. Modeling groups such as IGME and GBD regularly exclude data sources because of quality concerns that arise in vetting. Furthermore, we found that much of the variation in the external validation could be explained using source-specific effects. In practice, the data-synthesizing models used by groups such as IGME and GBD can account for source-level biases using fixed or random effects.

As global child mortality has declined rapidly in recent years, it has become clear that improvements have not been equal across all ages in early childhood [21]. The Sustainable Development Goals now have an explicit target of reducing NN mortality to 12 deaths per 1,000 live births [7]. Until now, estimates of NN mortality have depended mostly on CBH data or VR, when it is available. If no data are available, estimates are completely modeled based on external information. Until complete and reliable VR data are available from all countries, SBH data should be considered an “inexpensive” alternative to costlier CBH surveys. As we have demonstrated through extensive and systematic external validation, this new method now opens the possibility of leveraging a huge amount of SBH data available from surveys and censuses for monitoring progress toward the NN mortality Sustainable Development Goal.

Limitations and directions for future research

These results should be interpreted within the context of several limitations. First, despite being widely seen as high quality, and thus the basis for many child mortality estimates, DHS CBH data can suffer from certain issues such as selection biases [29] and misplacement of births [30]. By serving as the empirical basis upon which our model was trained, potential issues in these data could be reflected in the resulting application of it. Future research should focus on quantifying such issues and adjusting empirically based indirect methods to accommodate them. Second, the method presented here relies on formalizing existing relationships between covariates in the data to drive predictions. As such, when these relationships do not hold, predictions can suffer. Given the lack of period-based information in any one given SBH survey, it is expected that indirect estimates will poorly capture rapid changes in mortality [9]. This is partially mitigated in our approach by incorporating individual-level covariates, in which case mortality experiences from younger mothers will be more heavily weighted in recent periods. Third, by using GBD-MAP probability of birth distributions, we assume that fertility experiences are relatively stable over time among women in the same region, age, and number of CEB. Our preliminary analyses indicate this is generally true (see S7 Fig). Future research should focus on modeling these distributions at the individual level as well, potentially jointly fit within one model. Fourth, subnational predictions could likely be improved in the future by using subnational-level covariates rather than national-level covariates like SDI, as well as by implementing models that account for spatial autocorrelation in residuals. Fifth, by relying on concurrent SBH and CBH estimates as the basis for external validation, we could not ascertain the performance of this method in locations where only SBH exists, and thus, our sample may be somewhat biased toward higher-quality data. Finally, we validated the new model on one specific set of age bins, chosen to align with data collection and the typically used age breakdowns in previous research on child mortality. Future research can further validate other age bins and consider further distinguishing trends by sex.

Conclusions

This new method introduces a novel approach to indirect estimation of child mortality. It produces results comparable to current best methods for indirect estimation of under-5 mortality while additionally producing age-specific estimates at both national and subnational levels, supplying researchers a tool with which to utilize a massive amount of SBH data for estimation of trends in NN and infant mortality at various geographic levels. Systematic application of these methods could further improve the evidence base for monitoring of trends and inequalities in age-specific child mortality.

References

1. Murray CJL. Shifting to Sustainable Development Goals—Implications for Global Health. N Engl J Med. 2015;373: 1390–1393. pmid:26376045
2. Mathers C, Boerma T. Mortality Measurement Matters: Improving Data Collection and Estimation Methods for Child and Adult Mortality. PLoS Med. 2010;7: e1000265. pmid:20405053
3. Phillips DE, Lozano R, Naghavi M, Atkinson C, Gonzalez-Medina D, Mikkelsen L, et al. A composite metric for assessing data on mortality and causes of death: the vital statistics performance index. Popul Health Metr. 2014;12: 14. pmid:24982595
4. Wang H, and others. Global, regional, national, and selected subnational levels of stillbirths, neonatal, infant, and under-5 mortality, 1980–2015: a systematic analysis for the Global Burden of Disease Study 2015. The Lancet. 2016;388: 1725–1774. pmid:27733285

5. You D, Hug L, Ejdemyr S, Idele P, Hogan D, Mathers C, et al. Global, regional, and national levels and trends in under-5 mortality between 1990 and 2015, with scenario-based projections to 2030: a systematic analysis by the UN Inter-agency Group for Child Mortality Estimation. The Lancet. 2015;386: 2275–2286.
6. Golding N, Burstein R, Longbottom J, Browne AJ, Fullman N, Osgood-Zimmerman A, et al. Mapping under-5 and neonatal mortality in Africa, 2000–15: a baseline analysis for the Sustainable Development Goals. The Lancet. 2017;390: 2171–2182. pmid:28958464
7. United Nations. Transforming our world: the 2030 Agenda for Sustainable Development.:. Sustainable Development Knowledge Platform [Internet]. New York, NY: United Nations (UN); 2015. Available from: https://sustainabledevelopment.un.org/post2015/transformingourworld. [cited 2018 Jan 10].
8. Manual X [Internet]. [cited 5 Feb 2017]. Available from: http://www.un.org/esa/population/publications/Manual_X/Manual_X.htm
9. Rajaratnam JK, Tran LN, Lopez AD, Murray CJL. Measuring Under-Five Mortality: Validation of New Low-Cost Methods. PLoS Med. 2010;7: e1000253. pmid:20405055
10. Hill K, Brady E, Zimmerman L, Montana L, Silva R, Amouzou A. Monitoring Change in Child Mortality through Household Surveys. PLoS ONE. 2015;10: e0137713. pmid:26605920
11. Moultrie T, Dorrington R, Hill A, Hill K, Timaeus I, Zaba B. Tools for Demographic Estimation [Internet]. International Union for the Scientific Study of Population, Paris. Available from: http://demographicestimation.iussp.org/. [cited 2018 May 5].
12. Brass W. Methods for Estimating Fertility and Mortality from Limited and Defective Data. Chapel Hill, NC: Laboratories for Population Statistics, Univ. of NC; 1975.
13. Trussell TJ. A re-estimation of the multiplying factors for the Brass technique for determining childhood survivorship rates. Popul Stud. 1975;29: 97–107. pmid:22091807
14. Feeney G. Estimating Infant Mortality Trends from Child Survivorship Data. Popul Stud. 1980;34: 109–128.
15. Preston S, Heuveline P, Guillot M. Demography: Measuring and Modeling Population Processes. Malden, MA: Wiley; 2000.
16. Brady E, Hill K. Testing survey-based methods for rapid monitoring of child mortality, with implications for summary birth history data. PLoS ONE. 2017;12: e0176366. pmid:28441434
17. Tolles J, Lewis RJ. Time-to-Event Analysis. JAMA. 2016;315: 1046–1047. pmid:26954413
18. Box-Steffensmeier JM, Jones BS. Event History Modeling: A Guide for Social Scientists. Cambridge, UK: Cambridge University Press; 2004.
19. Royston P, Parmar MKB. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. Stat Med. 2002;21: 2175–2197. pmid:12210632
20. Allison PD. Discrete-Time Methods for the Analysis of Event Histories. Sociol Methodol. 1982;13: 61–98.
21. Wang H, Abajobir AA, Abate KH, Abbafati C, Abbas KM, Abd-Allah F, et al. Global, regional, and national under-5 mortality, adult mortality, age-specific mortality, and life expectancy, 1970–2016: a systematic analysis for the Global Burden of Disease Study 2016. The Lancet. 2017;390: 1084–1150. pmid:28919115
22. King G, Tomz M, Wittenberg J. Making the Most of Statistical Analyses: Improving Interpretation and Presentation. Am J Polit Sci. 2000;44: 347–361.
23. Wood S. mgcv: Mixed GAM Computation Vehicle with Automatic Smoothness Estimation [Internet]. 2017. Available from: https://cran.r-project.org/web/packages/mgcv/index.html. [cited 2017 Sep 15].
24. Wood SN. Generalized Additive Models: An Introduction with R. 1 edition. Boca Raton, FL: Chapman and Hall/CRC; 2006.
25. Wakefield J, Fuglstad G-A, Riebler A, Godwin J, Wilson K, Clark SJ. Estimating Under Five Mortality in Space and Time in a Developing World Context. ArXiv171100800 Stat. 2017; Available from: http://arxiv.org/abs/1711.00800. [cited 2018 Jan 10].
26. Burke M, Heft-Neal S, Bendavid E. Sources of variation in under-5 mortality across sub-Saharan Africa: a spatial analysis. Lancet Glob Health. 2016;4: e936–e945. pmid:27793587
27. Pezzulo C, Bird T, Utazi EC, Sorichetta A, Tatem AJ, Yourkavitch J, et al. Geospatial modeling of child mortality across 27 countries in Sub-Saharan Africa. 2016; Available from: http://www.dhsprogram.com/publications/publication-SAR13-Spatial-Analysis-Reports.cfm. [cited 2018 Jan 10].

28. Storeygard A, Balk D, Levy M, Deane G. The Global Distribution of Infant Mortality: A subnational spatial view. Popul Space Place. 2008;14: 209–229. pmid:22962545
29. Walker N, Hill K, Zhao F. Child Mortality Estimation: Methods Used to Adjust for Bias due to AIDS in Estimating Trends in Under-Five Mortality. PLoS Med. 2012;9: e1001298. pmid:22952437
30. Silva R. Child Mortality Estimation: Consistency of Under-Five Mortality Rate Estimates Using Full Birth Histories and Summary Birth Histories. PLoS Med. 2012;9: e1001296. pmid:22952436

Supporting Information for Chapter 2

S1 Table:

Input data sources for development and cross-validation of the method. This table lists each of the surveys used for training and testing the model. All were either Demographic and Health Surveys or related Malaria Indicator Surveys. Raw sample sizes for number women and number of children are also given. More information on each survey, including download links, can be found by searching the GHDx ID at http://ghdx.healthdata.org/. The most recent survey for each country was used for validation, and marked with an "X" in the table.

Abbreviations: GHDx = Global Health Data Exchange.

Model Region	Country	GHDx ID	Survey	Year	# Mothers	# Children	Test Data
North Africa / Middle East	Afghanistan	157018	MACRO_DHS	2016	26,598	125,715	X
Asia	Albania	18834	MACRO_DHS	2009	4,817	12,766	X
Sub-Saharan Africa, West/Central	Angola	56169	MACRO_MIS	2011	6,517	22,925	X
Asia	Armenia	31750	MACRO_DHS	2010	3,780	8,424	X
Asia	Armenia	18854	MACRO_DHS	2005	4,276	10,297	
Asia	Armenia	18843	MACRO_DHS	2000	4,372	11,286	
Asia	Bangladesh	157021	MACRO_DHS	2014	16,079	43,772	X
Asia	Bangladesh	55956	MACRO_DHS	2012	16,014	45,834	
Asia	Bangladesh	18913	MACRO_DHS	2007	9,849	30,527	
Asia	Bangladesh	18902	MACRO_DHS	2004	10,138	33,597	
Asia	Bangladesh	26826	MACRO_DHS	2000	9,353	31,906	
Asia	Bangladesh	18878	MACRO_DHS	1997	8,081	29,344	
Asia	Bangladesh	18889	MACRO_DHS	1994	8,541	32,581	
Sub-Saharan Africa, West/Central	Benin	79839	MACRO_DHS	2012	12,522	47,152	X
Sub-Saharan Africa, West/Central	Benin	18959	MACRO_DHS	2006	13,814	57,232	
Sub-Saharan Africa, West/Central	Benin	18950	MACRO_DHS	2001	4,612	19,398	
Sub-Saharan Africa, West/Central	Benin	18938	MACRO_DHS	1996	1,804	8,158	
Latin America and the Caribbean	Bolivia	19016	MACRO_DHS	2008	11,720	40,355	X
Latin America and the Caribbean	Bolivia	18971	MACRO_DHS	1998	7,634	29,473	
Latin America and the Caribbean	Bolivia	18990	MACRO_DHS	1994	6,053	24,174	
Latin America and the Caribbean	Bolivia	18979	MACRO_DHS	1989	5,542	22,338	
Sub-Saharan Africa, South/East	Botswana	19019	MACRO_DHS	1988	3,279	10,670	X
Latin America and the Caribbean	Brazil	19046	MACRO_DHS	1996	8,390	25,513	X
Sub-Saharan Africa, West/Central	Burkina Faso	19133	MACRO_DHS	2011	13,247	56,178	X
Sub-Saharan Africa, West/Central	Burkina Faso	19088	MACRO_DHS	2003	9,474	41,520	
Sub-Saharan Africa, West/Central	Burkina Faso	19076	MACRO_DHS	1999	4,916	22,145	
Sub-Saharan Africa, West/Central	Burkina Faso	19064	MACRO_DHS	1993	4,778	20,655	
Sub-Saharan Africa, South/East	Burundi	30431	MACRO_DHS	2011	5,954	24,520	X

Asia	Cambodia	157024	MACRO_DHS	2014	11,723	33,290	X
Asia	Cambodia	30379	MACRO_DHS	2011	11,856	37,511	
Asia	Cambodia	19167	MACRO_DHS	2006	10,791	40,457	
Asia	Cambodia	19156	MACRO_DHS	2000	9,930	40,990	
Sub-Saharan Africa, West/Central	Cameroon	19274	MACRO_DHS	2011	11,023	42,312	X
Sub-Saharan Africa, West/Central	Cameroon	19211	MACRO_DHS	2004	7,557	29,455	
Sub-Saharan Africa, West/Central	Cameroon	19198	MACRO_DHS	1998	3,847	15,187	
Sub-Saharan Africa, West/Central	Cameroon	19188	MACRO_DHS	1991	2,839	11,612	
Sub-Saharan Africa, West/Central	Central African Republic	19292	MACRO_DHS	1995	4,388	16,936	X
Sub-Saharan Africa, West/Central	Chad	157025	MACRO_DHS	2015	14,156	68,989	X
Sub-Saharan Africa, West/Central	Chad	19315	MACRO_DHS	2004	4,643	21,448	
Sub-Saharan Africa, West/Central	Chad	19305	MACRO_DHS	1997	5,865	25,739	
Latin America and the Caribbean	Colombia	218566	MACRO_DHS	2016	25,433	62,580	X
Latin America and the Caribbean	Colombia	19324	MACRO_DHS	2005	26,536	71,254	
Latin America and the Caribbean	Colombia	19359	MACRO_DHS	2000	7,830	21,267	
Latin America and the Caribbean	Colombia	19350	MACRO_DHS	1995	7,500	21,830	
Latin America and the Caribbean	Colombia	19341	MACRO_DHS	1990	5,365	15,964	
Sub-Saharan Africa, South/East	Comoros	76850	MACRO_DHS	2013	2,934	11,497	X
Sub-Saharan Africa, South/East	Comoros	19370	MACRO_DHS	1996	1,695	7,913	
Sub-Saharan Africa, West/Central	Congo	56151	MACRO_DHS	2012	8,787	31,948	X
Sub-Saharan Africa, West/Central	Congo	19391	MACRO_DHS	2005	5,152	16,687	
Sub-Saharan Africa, West/Central	Cote d'Ivoire	18533	MACRO_DHS	2012	7,498	28,211	X
Sub-Saharan Africa, West/Central	Cote d'Ivoire	18531	MACRO_DHS	1999	2,048	7,575	
Sub-Saharan Africa, West/Central	Cote d'Ivoire	18519	MACRO_DHS	1994	6,108	24,870	
Sub-Saharan Africa, West/Central	Democratic Republic of the Congo	76878	MACRO_DHS	2014	14,182	59,276	X
Sub-Saharan Africa, West/Central	Democratic Republic of the Congo	19381	MACRO_DHS	2007	7,148	29,548	
Latin America and the Caribbean	Dominican Republic	77819	MACRO_DHS	2013	6,687	18,167	X
Latin America and the Caribbean	Dominican Republic	19456	MACRO_DHS	2007	19,541	58,037	
Latin America and the Caribbean	Dominican Republic	19444	MACRO_DHS	2002	17,032	53,667	
Latin America and the Caribbean	Dominican Republic	19431	MACRO_DHS	1999	901	2,871	
Latin America and the Caribbean	Dominican Republic	19421	MACRO_DHS	1996	5,942	19,784	
Latin America and the Caribbean	Dominican Republic	19410	MACRO_DHS	1991	4,864	17,163	
North Africa / Middle East	Egypt	154897	MACRO_DHS	2014	19,770	59,266	X
North Africa / Middle East	Egypt	26842	MACRO_DHS	2008	14,778	48,619	
North Africa / Middle East	Egypt	19521	MACRO_DHS	2005	17,552	61,455	
North Africa / Middle East	Egypt	19529	MACRO_DHS	2003	8,275	30,298	
North Africa / Middle East	Egypt	19511	MACRO_DHS	2000	14,164	54,780	
North Africa / Middle East	Egypt	19493	MACRO_DHS	1996	13,329	56,390	
North Africa / Middle East	Egypt	19482	MACRO_DHS	1993	8,983	38,076	

North Africa / Middle East	Egypt	19472	MACRO_DHS	1989	8,091	35,519	
Sub-Saharan Africa, South/East	Eritrea	19539	MACRO_DHS	2002	6,009	24,370	X
Sub-Saharan Africa, South/East	Ethiopia	218568	MACRO_DHS	2016	10,274	41,392	X
Sub-Saharan Africa, South/East	Ethiopia	21301	MACRO_DHS	2011	10,896	45,540	
Sub-Saharan Africa, South/East	Ethiopia	19557	MACRO_DHS	2005	9,339	39,881	
Sub-Saharan Africa, South/East	Ethiopia	19571	MACRO_DHS	2000	10,143	44,174	
Sub-Saharan Africa, West/Central	Gabon	76706	MACRO_DHS	2012	6,383	23,109	X
Sub-Saharan Africa, West/Central	Gabon	19579	MACRO_DHS	2001	4,499	16,878	
Sub-Saharan Africa, West/Central	Ghana	157027	MACRO_DHS	2014	6,511	23,118	X
Sub-Saharan Africa, West/Central	Ghana	21188	MACRO_DHS	2008	3,299	11,888	
Sub-Saharan Africa, West/Central	Ghana	19627	MACRO_DHS	2003	3,992	15,086	
Sub-Saharan Africa, West/Central	Ghana	19614	MACRO_DHS	1999	3,499	13,188	
Sub-Saharan Africa, West/Central	Ghana	19604	MACRO_DHS	1994	3,501	13,280	
Sub-Saharan Africa, West/Central	Ghana	19587	MACRO_DHS	1988	3,453	14,216	
Latin America and the Caribbean	Guatemala	157031	MACRO_DHS	2015	17,178	55,398	X
Latin America and the Caribbean	Guatemala	19656	MACRO_DHS	1999	4,350	18,581	
Latin America and the Caribbean	Guatemala	19637	MACRO_DHS	1995	8,794	38,753	
Sub-Saharan Africa, West/Central	Guinea	69761	MACRO_DHS	2012	6,950	27,683	X
Sub-Saharan Africa, West/Central	Guinea	19683	MACRO_DHS	2005	6,259	27,115	
Sub-Saharan Africa, West/Central	Guinea	19670	MACRO_DHS	1999	5,413	22,943	
Latin America and the Caribbean	Guyana	21348	MACRO_DHS	2009	3,484	10,929	X
Latin America and the Caribbean	Haiti	65118	MACRO_DHS	2012	8,671	29,013	X
Latin America and the Caribbean	Haiti	19720	MACRO_DHS	2006	6,547	24,830	
Latin America and the Caribbean	Haiti	19708	MACRO_DHS	2000	6,459	26,437	
Latin America and the Caribbean	Haiti	19695	MACRO_DHS	1995	3,288	12,547	
Latin America and the Caribbean	Honduras	95440	MACRO_DHS	2012	15,854	49,263	X
Latin America and the Caribbean	Honduras	19728	MACRO_DHS	2006	13,991	50,093	
Asia	India	19963	MACRO_DHS	2006	84,609	256,782	X
Asia	India	19950	MACRO_DHS	2000	80,872	268,879	
Asia	India	19787	MACRO_DHS	1993	79,322	275,143	
Asia	Indonesia	76705	MACRO_DHS	2012	32,129	83,650	X
Asia	Indonesia	20021	MACRO_DHS	2007	30,420	84,726	
Asia	Indonesia	20011	MACRO_DHS	2003	27,317	79,791	
Asia	Indonesia	19999	MACRO_DHS	1997	26,562	86,276	
Asia	Indonesia	19990	MACRO_DHS	1994	26,045	90,326	
Asia	Indonesia	19979	MACRO_DHS	1991	21,065	74,329	
North Africa / Middle East	Jordan	77517	MACRO_DHS	2012	10,304	42,275	X
North Africa / Middle East	Jordan	21206	MACRO_DHS	2009	9,124	38,199	
North Africa / Middle East	Jordan	20083	MACRO_DHS	2007	9,916	43,460	
North Africa / Middle East	Jordan	20073	MACRO_DHS	2002	5,494	25,296	
North Africa / Middle East	Jordan	20060	MACRO_DHS	1997	5,038	24,243	
North Africa / Middle East	Jordan	20051	MACRO_DHS	1990	5,853	32,812	
Asia	Kazakhstan	20103	MACRO_DHS	1999	3,364	8,106	X
Asia	Kazakhstan	20092	MACRO_DHS	1995	2,649	6,866	
Sub-Saharan Africa, South/East	Kenya	157057	MACRO_DHS	2014	23,245	83,591	X

Sub-Saharan Africa, South/East	Kenya	21365	MACRO_DHS	2009	6,102	22,534	
Sub-Saharan Africa, South/East	Kenya	20145	MACRO_DHS	2003	5,865	22,074	
Sub-Saharan Africa, South/East	Kenya	20132	MACRO_DHS	1998	5,717	23,351	
Sub-Saharan Africa, South/East	Kenya	20120	MACRO_DHS	1993	5,415	23,899	
Sub-Saharan Africa, South/East	Kenya	20109	MACRO_DHS	1989	5,507	25,173	
Asia	Kyrgyzstan	77518	MACRO_DHS	2012	5,601	16,180	X
Asia	Kyrgyzstan	20154	MACRO_DHS	1997	2,776	8,781	
Sub-Saharan Africa, South/East	Lesotho	157058	MACRO_DHS	2014	4,540	11,710	X
Sub-Saharan Africa, South/East	Lesotho	21382	MACRO_DHS	2010	5,191	14,429	
Sub-Saharan Africa, South/East	Lesotho	20167	MACRO_DHS	2005	4,832	14,708	
Sub-Saharan Africa, West/Central	Liberia	77385	MACRO_DHS	2013	7,559	30,804	X
Sub-Saharan Africa, West/Central	Liberia	34279	MACRO_MIS	2009	3,643	14,872	
Sub-Saharan Africa, West/Central	Liberia	20191	MACRO_DHS	2007	5,701	22,123	
Sub-Saharan Africa, South/East	Madagascar	21409	MACRO_DHS	2009	12,970	48,464	X
Sub-Saharan Africa, South/East	Madagascar	20223	MACRO_DHS	2004	5,845	20,799	
Sub-Saharan Africa, South/East	Madagascar	20212	MACRO_DHS	1997	5,233	21,654	
Sub-Saharan Africa, South/East	Madagascar	20202	MACRO_DHS	1992	4,369	18,931	
Sub-Saharan Africa, South/East	Malawi	218581	MACRO_DHS	2016	18,988	68,074	X
Sub-Saharan Africa, South/East	Malawi	21393	MACRO_DHS	2010	18,041	72,301	
Sub-Saharan Africa, South/East	Malawi	20263	MACRO_DHS	2005	9,298	35,883	
Sub-Saharan Africa, South/East	Malawi	20252	MACRO_DHS	2000	10,337	40,421	
Sub-Saharan Africa, South/East	Malawi	20235	MACRO_DHS	1992	3,718	16,330	
Asia	Maldives	21311	MACRO_DHS	2009	6,106	20,136	X
Sub-Saharan Africa, West/Central	Mali	77388	MACRO_DHS	2013	8,480	33,803	X
Sub-Saharan Africa, West/Central	Mali	20274	MACRO_DHS	2006	11,566	52,140	
Sub-Saharan Africa, West/Central	Mali	20315	MACRO_DHS	2001	10,379	48,407	
Sub-Saharan Africa, West/Central	Mali	20301	MACRO_DHS	1996	7,935	37,921	
Sub-Saharan Africa, West/Central	Mauritania	20322	MACRO_DHS	2001	4,584	19,202	X
North Africa / Middle East	Morocco	20361	MACRO_DHS	2004	8,660	32,494	X
North Africa / Middle East	Morocco	20371	MACRO_DHS	1992	4,986	22,657	
Sub-Saharan Africa, South/East	Mozambique	55975	MACRO_DHS	2011	10,624	37,984	X
Sub-Saharan Africa, South/East	Mozambique	20394	MACRO_DHS	2004	9,732	37,443	
Sub-Saharan Africa, South/East	Mozambique	20382	MACRO_DHS	1997	6,798	25,752	
Asia	Myanmar	157061	MACRO_DHS	2016	7,796	22,989	X
Sub-Saharan Africa, South/East	Namibia	150382	MACRO_DHS	2013	6,453	18,090	X
Sub-Saharan Africa, South/East	Namibia	20428	MACRO_DHS	2007	6,636	19,522	
Sub-Saharan Africa, South/East	Namibia	20417	MACRO_DHS	2000	4,780	14,946	
Sub-Saharan Africa, South/East	Namibia	20404	MACRO_DHS	1992	3,710	13,372	
Asia	Nepal	21240	MACRO_DHS	2011	8,800	26,615	X
Asia	Nepal	20462	MACRO_DHS	2006	7,791	26,394	
Asia	Nepal	20450	MACRO_DHS	2001	7,772	28,955	
Asia	Nepal	20437	MACRO_DHS	1996	7,479	29,156	
Latin America and the Caribbean	Nicaragua	126952	NIC/DHS_ENDESA	2012	11,295	31,244	X

Latin America and the Caribbean	Nicaragua	20487	MACRO_DHS	2001	9,275	34,157	
Latin America and the Caribbean	Nicaragua	20478	MACRO_DHS	1998	9,696	36,820	
Sub-Saharan Africa, West/Central	Niger	74393	MACRO_DHS	2012	9,209	44,183	X
Sub-Saharan Africa, West/Central	Niger	20499	MACRO_DHS	2006	7,205	34,378	
Sub-Saharan Africa, West/Central	Niger	20537	MACRO_DHS	1998	5,921	28,888	
Sub-Saharan Africa, West/Central	Niger	20518	MACRO_DHS	1992	5,068	23,841	
Sub-Saharan Africa, West/Central	Nigeria	77390	MACRO_DHS	2013	27,451	119,386	X
Sub-Saharan Africa, West/Central	Nigeria	30991	MACRO_MIS	2010	4,632	19,644	
Sub-Saharan Africa, West/Central	Nigeria	21433	MACRO_DHS	2008	23,751	104,808	
Sub-Saharan Africa, West/Central	Nigeria	20567	MACRO_DHS	2003	5,111	23,038	
Sub-Saharan Africa, West/Central	Nigeria	20552	MACRO_DHS	1990	6,477	28,123	
Asia	Pakistan	77521	MACRO_DHS	2013	11,965	50,238	X
Asia	Pakistan	20595	MACRO_DHS	2007	8,798	39,049	
Asia	Pakistan	20584	MACRO_DHS	1991	5,905	27,369	
Latin America and the Caribbean	Peru	209930	MACRO_DHS	2014	17,488	47,633	X
Latin America and the Caribbean	Peru	270471	MACRO_DHS	2012	16,620	47,261	
Latin America and the Caribbean	Peru	270470	MACRO_DHS	2011	15,636	46,194	
Latin America and the Caribbean	Peru	270469	MACRO_DHS	2010	15,886	46,780	
Latin America and the Caribbean	Peru	270404	MACRO_DHS	2009	16,887	50,084	
Latin America and the Caribbean	Peru	275090	MACRO_DHS	2008	28,613	89,220	
Latin America and the Caribbean	Peru	20649	MACRO_DHS	2000	18,931	65,453	
Latin America and the Caribbean	Peru	20638	MACRO_DHS	1996	19,835	72,390	
Latin America and the Caribbean	Peru	20626	MACRO_DHS	1992	10,244	38,783	
Asia	Philippines	142943	MACRO_DHS	2013	10,125	31,680	X
Asia	Philippines	21421	MACRO_DHS	2008	8,639	28,518	
Asia	Philippines	20699	MACRO_DHS	2003	8,750	30,443	
Asia	Philippines	20683	MACRO_DHS	1998	8,662	32,626	
Asia	Philippines	20674	MACRO_DHS	1993	9,197	35,863	
Sub-Saharan Africa, South/East	Rwanda	157063	MACRO_DHS	2015	8,736	30,058	X
Sub-Saharan Africa, South/East	Rwanda	56040	MACRO_DHS	2011	8,501	32,639	
Sub-Saharan Africa, South/East	Rwanda	21222	MACRO_DHS	2008	4,686	18,421	
Sub-Saharan Africa, South/East	Rwanda	20740	MACRO_DHS	2005	7,045	30,072	
Sub-Saharan Africa, South/East	Rwanda	20722	MACRO_DHS	2000	6,539	27,602	
Sub-Saharan Africa, South/East	Rwanda	20711	MACRO_DHS	1992	4,292	19,440	
Sub-Saharan Africa, West/Central	Sao Tome and Principe	26866	MACRO_DHS	2009	2,014	7,620	X
Sub-Saharan Africa, West/Central	Senegal	286772	MACRO_DHS	2016	5,836	22,740	X
Sub-Saharan Africa, West/Central	Senegal	218592	MACRO_DHS	2015	5,906	23,250	
Sub-Saharan Africa, West/Central	Senegal	191270	MACRO_DHS	2014	5,733	22,365	
Sub-Saharan Africa, West/Central	Senegal	111432	MACRO_DHS	2013	5,650	22,563	
Sub-Saharan Africa, West/Central	Senegal	56063	MACRO_DHS	2011	10,652	42,510	
Sub-Saharan Africa, West/Central	Senegal	11540	MACRO_MIS	2009	13,134	53,608	
Sub-Saharan Africa, West/Central	Senegal	26855	MACRO_DHS	2005	9,593	39,895	
Sub-Saharan Africa, West/Central	Senegal	20780	MACRO_DHS	1997	6,097	27,448	
Sub-Saharan Africa, West/Central	Senegal	20767	MACRO_DHS	1993	4,534	20,815	

Sub-Saharan Africa, West/Central	Sierra Leone	131467	MACRO_DHS	2013	12,352	47,392	X
Sub-Saharan Africa, West/Central	Sierra Leone	21258	MACRO_DHS	2008	5,876	21,136	
Sub-Saharan Africa, South/East	South Africa	20796	MACRO_DHS	1998	8,223	22,934	X
North Africa / Middle East	Sudan	20813	MACRO_DHS	1990	5,277	25,805	X
Sub-Saharan Africa, South/East	Swaziland	20829	MACRO_DHS	2007	3,488	11,410	X
Asia	Tajikistan	74460	MACRO_DHS	2012	6,172	19,938	X
Sub-Saharan Africa, South/East	Tanzania	218593	MACRO_DHS	2016	9,721	37,169	X
Sub-Saharan Africa, South/East	Tanzania	21331	MACRO_DHS	2010	7,326	29,777	
Sub-Saharan Africa, South/East	Tanzania	20875	MACRO_DHS	2005	7,573	30,557	
Sub-Saharan Africa, South/East	Tanzania	20865	MACRO_DHS	1999	2,935	11,952	
Sub-Saharan Africa, South/East	Tanzania	20852	MACRO_DHS	1996	6,083	24,890	
Sub-Saharan Africa, South/East	Tanzania	20841	MACRO_DHS	1992	6,913	29,143	
Sub-Saharan Africa, West/Central	The Gambia	77384	MACRO_DHS	2013	6,845	26,601	X
Asia	Timor-Leste	21274	MACRO_DHS	2010	7,969	35,998	X
Sub-Saharan Africa, West/Central	Togo	77515	MACRO_DHS	2014	6,944	26,264	X
Sub-Saharan Africa, West/Central	Togo	20909	MACRO_DHS	1998	6,289	26,269	
Sub-Saharan Africa, West/Central	Togo	20896	MACRO_DHS	1988	2,536	10,782	
North Africa / Middle East	Tunisia	20926	MACRO_DHS	1988	3,856	16,463	X
North Africa / Middle East	Turkey	20954	MACRO_DHS	2004	7,360	22,443	X
North Africa / Middle East	Turkey	20947	MACRO_DHS	1998	5,578	17,791	
North Africa / Middle East	Turkey	20936	MACRO_DHS	1993	5,923	19,762	
Sub-Saharan Africa, South/East	Uganda	56021	MACRO_DHS	2011	6,393	28,609	X
Sub-Saharan Africa, South/East	Uganda	13109	MACRO_MIS	2010	3,143	13,863	
Sub-Saharan Africa, South/East	Uganda	21014	MACRO_DHS	2006	6,417	30,090	
Sub-Saharan Africa, South/East	Uganda	20993	MACRO_DHS	2001	5,501	23,410	
Sub-Saharan Africa, South/East	Uganda	20976	MACRO_DHS	1995	5,465	22,752	
Sub-Saharan Africa, South/East	Uganda	20964	MACRO_DHS	1989	3,563	16,074	
Asia	Ukraine	21024	MACRO_DHS	2007	4,811	8,007	X
Asia	Uzbekistan	21033	MACRO_DHS	1996	3,018	9,650	X
Asia	Vietnam	21058	MACRO_DHS	2002	5,390	14,383	X
Asia	Vietnam	21049	MACRO_DHS	1997	5,352	15,517	
North Africa / Middle East	Yemen	112500	MACRO_DHS	2013	14,688	64,602	X
North Africa / Middle East	Yemen	21068	MACRO_DHS	1992	5,059	27,081	
Sub-Saharan Africa, South/East	Zambia	77516	MACRO_DHS	2014	12,421	49,207	X
Sub-Saharan Africa, South/East	Zambia	21117	MACRO_DHS	2007	5,410	21,366	
Sub-Saharan Africa, South/East	Zambia	21102	MACRO_DHS	2002	5,831	23,805	
Sub-Saharan Africa, South/East	Zambia	21090	MACRO_DHS	1997	5,996	24,799	
Sub-Saharan Africa, South/East	Zambia	21079	MACRO_DHS	1992	5,186	22,122	
Sub-Saharan Africa, South/East	Zimbabwe	157066	MACRO_DHS	2015	7,253	20,791	X
Sub-Saharan Africa, South/East	Zimbabwe	55992	MACRO_DHS	2011	6,725	19,279	
Sub-Saharan Africa, South/East	Zimbabwe	21163	MACRO_DHS	2006	6,281	19,489	
Sub-Saharan Africa, South/East	Zimbabwe	21151	MACRO_DHS	1999	4,207	14,184	
Sub-Saharan Africa, South/East	Zimbabwe	21139	MACRO_DHS	1994	4,388	16,777	
Sub-Saharan Africa, South/East	Zimbabwe	21126	MACRO_DHS	1989	3,005	12,405	

S2 Table:

External validation data sources. An additional 342 surveys were used for external validation. 243 of these were SBH-only surveys, the rest were additional CBH sources used for comparison but not used in training the model.

Abbreviations: CBH = complete birth history; SBH = summary birth history.

Country	GHDx nid	Data Type	Children	Citation
Afghanistan	56099	cbh	113,806	Central Statistics Organization (Afghanistan), ICF Macro, Indian Institute of Health Management Research (IIHMR), Ministry of Public Health (Afghanistan), World Health Organization Regional Office for the Eastern Mediterranean (EMRO-WHO). Afghanistan Special Demographic and Health Survey 2010. Fairfax, United States: ICF International.
Afghanistan	56830	sbh	64,660	Central Statistics Organization (Afghanistan), United Nations Children's Fund (UNICEF). Afghanistan Multiple Indicator Cluster Survey 2010-2011. New York, United States: United Nations Children's Fund (UNICEF), 2013.
Albania	595	sbh	8,661	National Institute of Statistics (Albania), United Nations Children's Fund (UNICEF). Albania Multiple Indicator Cluster Survey 2000. New York, United States: United Nations Children's Fund (UNICEF).
Albania	608	sbh	7,531	National Institute of Statistics (Albania), United Nations Children's Fund (UNICEF). Albania Multiple Indicator Cluster Survey 2005. New York, United States: United Nations Children's Fund (UNICEF).
Albania	56862	sbh	29,794	European Union (EU), Government of Albania, Institute of Statistics (Albania). Albania Census 2011.
Algeria	627	cbh	29,406	National Office of Statistics (Algeria), Ministry of Health, Population and Hospital Reform (Algeria), League of Arab States. Algeria Family Health Survey 2002-2003.
Algeria	210614	cbh	58,390	Ministry of Health and Population (Algeria), United Nations Children's Fund (UNICEF). Algeria Multiple Indicator Cluster Survey 2012-2013. New York, United States: United Nations Children's Fund (UNICEF), 2016.
Angola	687	sbh	23,911	National Institute of Statistics (Angola), United Nations Children's Fund (UNICEF). Angola Multiple Indicator Cluster Survey 2001. New York, United States: United Nations Children's Fund (UNICEF).
Angola	672	cbh	2,931	COSEP-Consulting Ltd., ConsaÃºde Ltd., Macro International, Inc, Ministry of Health (Angola). Angola Malaria Indicator Survey 2006-2007. Fairfax, United States: ICF International.
Angola	30394	sbh	41,890	National Institute of Statistics (Angola), Oxford Policy Management, United Nations Children's Fund (UNICEF). Angola Integrated Inquiry into People's Well-Being 2008-2009.
Armenia	811	sbh	143,754	National Statistical Service of the Republic of Armenia, Minnesota Population Center. Armenia Population and Housing Census 2001 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. Minneapolis: University of Minnesota.
Azerbaijan	881	sbh	12,900	State Statistics Committee of Azerbaijan, United Nations Children's Fund (UNICEF). Azerbaijan Multiple Indicator Cluster Survey 2000. New York, United States: United Nations Children's Fund (UNICEF).
Azerbaijan	18865	cbh	13,565	Macro International, Inc, State Statistical Committee of Azerbaijan. Azerbaijan Demographic and Health Survey 2006. Fairfax, United States: ICF International.

Bangladesh	18920	cbh	319,622	Associates for Community and Population Research (ACPR), International Centre for Diarrhoeal Disease Research, Bangladesh (ICDDR,B), Johns Hopkins University (JHU), Mitra and Associates, National Institute of Population Research and Training (NIPORT), ORC Macro. Bangladesh Special Demographic and Health Survey 2001. Fairfax, United States: ICF International.
Bangladesh	126906	sbh	705,724	Bangladesh Bureau of Statistics. Bangladesh Multiple Indicator Cluster Survey 2009. Dhaka, Bangladesh: Bangladesh Bureau of Statistics.
Bangladesh	151086	sbh	112,041	Bangladesh Bureau of Statistics (BBS), Government of Bangladesh, Ministry of Planning (Bangladesh), United Nations Children's Fund (UNICEF). Bangladesh Multiple Indicator Cluster Survey 2012-2013. New York, United States: United Nations Children's Fund (UNICEF), 2015.
Belize	1089	sbh	4,206	Statistical Institute of Belize, United Nations Children's Fund (UNICEF). Belize Multiple Indicator Cluster Survey 2006. New York, United States: United Nations Children's Fund (UNICEF).
Belize	76699	sbh	8,888	Statistical Institute of Belize, United Nations Children's Fund (UNICEF). Belize Multiple Indicator Cluster Survey 2011. New York, United States: United Nations Children's Fund (UNICEF), 2013.
Benin	206075	cbh	45,183	National Institute of Statistics and Economic Analysis (INSAE) (Benin), United Nations Children's Fund (UNICEF). Benin Multiple Indicator Cluster Survey 2014. New York, United States: United Nations Children's Fund (UNICEF), 2017.
Bhutan	40028	sbh	31,697	National Statistics Bureau (Bhutan), United Nations Children's Fund (UNICEF), United Nations Population Fund (UNFPA). Bhutan Multiple Indicator Cluster Survey 2010. New York, United States: United Nations Children's Fund (UNICEF).
Bolivia	1245	sbh	12,777	National Institute of Statistics (Bolivia), World Bank (WB), Inter-American Development Bank (IDB), United Nations Economic Commission for Latin America and the Caribbean (CEPAL). Bolivia Household Survey 2000. La Paz, Bolivia: National Institute of Statistics (Bolivia).
Bolivia	1289	sbh	11,779	Population Development and Environment (PODEMA), National Directorate of Epidemiology (Bolivia), United Nations Children's Fund (UNICEF). Bolivia Multiple Indicator Cluster Survey 2000. New York, United States: United Nations Children's Fund (UNICEF).
Bolivia	1362	sbh	485,628	National Institute of Statistics (Bolivia), Minnesota Population Center. Bolivia National Census of Population and Housing 2001 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. Minneapolis: University of Minnesota.
Bolivia	19001	cbh	45,116	Macro International, Inc, Ministry of Health and Sports (Bolivia), National Institute of Statistics (Bolivia). Bolivia Demographic and Health Survey 2003-2004. Fairfax, United States: ICF International.
Botswana	1400	cbh	18,073	Central Statistics Office (Botswana). Botswana Family Health Survey 1996. Gaborone, Botswana: Central Statistics Office (Botswana).
Botswana	1404	sbh	12,745	Central Statistics Office (Botswana), United Nations Children's Fund (UNICEF). Botswana Multiple Indicator Cluster Survey 2000. New York, United States: United Nations Children's Fund (UNICEF), 2015.
Botswana	294205	sbh	92,210	Central Statistics Office (Botswana), Minnesota Population Center. Botswana Population and Housing Census 2001 from the Integrated Public Use Microdata Series, International. Minneapolis: University of Minnesota, 2017.
Botswana	21970	sbh	22,722	Central Statistics Office (Botswana). Botswana Demographic Survey 2006. Gaborone, Botswana: Central Statistics Office (Botswana).
Botswana	22125	cbh	10,858	Statistics Botswana, Botswana Familiy Health Survey 2007-2008
Botswana	294235	sbh	95,455	Central Statistics Office (Botswana), Minnesota Population Center. Botswana Population and Housing Census 2011 from the Integrated Public Use Microdata Series, International. Minneapolis: University of Minnesota, 2017.
Brazil	19035	cbh	15,363	Brazilian Society for Family Welfare (BEMFAM), Macro International, Inc. Brazil Demographic and Health Survey 1991. Fairfax, United States: ICF International.

Brazil	38230	sbh	4,830,714	Brazilian Institute of Geography and Statistics (IBGE), Minnesota Population Center. Brazil General Census 2000 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. Minneapolis: University of Minnesota.
Brazil	1488	sbh	177,981	Brazilian Institute of Geography and Statistics (IBGE). Brazil National Household Sample Survey 2002. Rio de Janeiro, Brazil: Brazilian Institute of Geography and Statistics (IBGE).
Brazil	1489	sbh	174,600	Brazilian Institute of Geography and Statistics (IBGE). Brazil National Household Sample Survey 2003. Rio de Janeiro, Brazil: Brazilian Institute of Geography and Statistics (IBGE).
Brazil	1490	sbh	180,887	Brazilian Institute of Geography and Statistics (IBGE). Brazil National Household Sample Survey 2004. Rio de Janeiro, Brazil: Brazilian Institute of Geography and Statistics (IBGE).
Brazil	80311	sbh	182,709	Brazilian Institute of Geography and Statistics (IBGE). Brazil National Household Sample Survey 2005. Rio de Janeiro, Brazil: Brazilian Institute of Geography and Statistics (IBGE).
Brazil	93528	sbh	179,995	Brazilian Institute of Geography and Statistics (IBGE). Brazil National Household Sample Survey 2006. Rio de Janeiro, Brazil: Brazilian Institute of Geography and Statistics (IBGE).
Brazil	93490	sbh	174,714	Brazilian Institute of Geography and Statistics (IBGE). Brazil National Household Sample Survey 2007. Rio de Janeiro, Brazil: Brazilian Institute of Geography and Statistics (IBGE).
Brazil	93487	sbh	166,527	Brazilian Institute of Geography and Statistics (IBGE). Brazil National Household Sample Survey 2008. Rio de Janeiro, Brazil: Brazilian Institute of Geography and Statistics (IBGE).
Brazil	93522	sbh	166,438	Brazilian Institute of Geography and Statistics (IBGE). Brazil National Household Sample Survey 2009.
Brazil	105322	sbh	3,915,750	Minnesota Population Center, Brazilian Institute of Geography and Statistics. Brazil Demographic Census 2010 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. Minneapolis: University of Minnesota, 2013.
Brazil	106724	sbh	145,313	Brazilian Institute of Geography and Statistics (IBGE). Brazil National Household Sample Survey 2011. Rio de Janeiro, Brazil: Brazilian Institute of Geography and Statistics (IBGE), 2012.
Brazil	156581	sbh	144,682	Brazilian Institute of Geography and Statistics (IBGE). Brazil National Household Sample Survey 2012. Rio de Janeiro, Brazil: Brazilian Institute of Geography and Statistics (IBGE).
Brazil	156583	sbh	144,033	Brazilian Institute of Geography and Statistics (IBGE). Brazil National Household Sample Survey 2013. Rio de Janeiro, Brazil: Brazilian Institute of Geography and Statistics (IBGE).
Brazil	238441	sbh	141,577	Brazilian Institute of Geography and Statistics (IBGE). Brazil National Household Sample Survey 2014. Rio de Janeiro, Brazil: Brazilian Institute of Geography and Statistics (IBGE).
Brazil	281548	sbh	135,941	Brazilian Institute of Geography and Statistics (IBGE). Brazil National Household Sample Survey 2015. Rio de Janeiro, Brazil: Brazilian Institute of Geography and Statistics (IBGE).
Burkina Faso	1927	sbh	25,541	National Institute of Statistics and Demography (Burkina Faso), United Nations Children's Fund (UNICEF). Burkina Faso Multiple Indicator Cluster Survey 2006. New York, United States: United Nations Children's Fund (UNICEF).
Burkina Faso	105403	sbh	946,618	Minnesota Population Center, National Institute of Statistics and Demography (Burkina Faso). Burkina Faso Population and Housing Census 2006 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. Minneapolis: University of Minnesota, 2013.
Burkina Faso	188785	sbh	27,986	ICF International, National Institute of Statistics and Demography (Burkina Faso), National Program for the Fight Against Malaria (PNLP) (Burkina Faso). Burkina Faso Malaria Indicator Survey 2014. Fairfax, United States: ICF International, 2015.
Burundi	1970	sbh	20,848	Burundi Institute of Statistics and Economic Studies. Burundi Priority Survey 1998-1999.
Burundi	1994	sbh	13,864	Burundi Institute of Statistics and Economic Studies, United Nations Children's Fund (UNICEF). Burundi Multiple Indicator Cluster Survey 2000. New York, United States: United Nations Children's Fund (UNICEF).

Burundi	1981	sbh	25,768	United Nations Children's Fund (UNICEF), Burundi Institute of Statistics and Economic Studies, United Nations Population Fund (UNFPA). Burundi Multiple Indicator Cluster Survey 2005. New York, United States: United Nations Children's Fund (UNICEF).
Burundi	108080	sbh	13,647	Burundi Institute of Statistics and Economic Studies, ICF Macro, Ministry of Public Health and the Fight Against AIDS (Burundi), National Institute of Public Health (Burundi). Burundi Malaria Indicator Survey 2012-2013. Fairfax, United States: ICF International, 2013.
Cambodia	19170	cbh	18,969	Macro International, Inc, Ministry of Health (Cambodia), National Institute of Public Health (Cambodia). Cambodia Special Demographic and Health Survey 1998. Fairfax, United States: ICF International.
Cambodia	35322	sbh	708,771	National Institute of Statistics (Cambodia), Minnesota Population Center. Cambodia General Population Census 1998 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. Minneapolis: University of Minnesota.
Cambodia	2002	sbh	58,962	National Institute of Statistics (Cambodia). Cambodia Intercensal Population Survey 2004.
Cambodia	35329	sbh	663,330	National Institute of Statistics (Cambodia), Minnesota Population Center. Cambodia General Population Census 2008 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. Minneapolis: University of Minnesota, 2011.
Cameroon	2053	sbh	14,980	Department of Statistics and Accounting, Ministery of the Economy and Finance (Cameroon) and United Nations Children's Fund (UNICEF). Cameroon Multiple Indicator Cluster Survey 2000. New York, United States: United Nations Children's Fund (UNICEF).
Cameroon	105800	sbh	1,046,493	Minnesota Population Center, National Institute of Statistics (Cameroon), Central Bureau of the Census and Population Studies (Cameroon). Cameroon Population and Housing Census 2005 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. Minneapolis: University of Minnesota, 2013
Cameroon	244455	cbh	26,201	Ministry of Public Health (Cameroon), National Institute of Statistics (Cameroon), United Nations Children's Fund (UNICEF). Cameroon Multiple Indicator Cluster Survey 2014. New York, United States: United Nations Children's Fund (UNICEF), 2017.
Cape Verde	27511	cbh	16,772	Cape Verde National Statistics Institute (INE) Division of Reproductive Health-Centers for Disease Control and Prevention (CDC). (2000) Cape Verde Reproductive Health Survey 1998. Atlanta, United States: Centers for Disease Control and Prevention (CDC).
Central African Republic	2209	sbh	59,424	Division of Statistics and Economic Studies (Central African Republic), Ministry of Economy, Planning and International Cooperation (Central African Republic), United Nations Children's Fund (UNICEF). Central African Republic Multiple Indicator Cluster Survey 2000. New York, United States: United Nations Children's Fund (UNICEF).
Central African Republic	2223	sbh	35,974	United Nations Children's Fund (UNICEF). Central African Republic Multiple Indicator Cluster Survey 2006. New York, United States: United Nations Children's Fund (UNICEF).
Central African Republic	82832	sbh	37,331	Central African Institute of Statistics, Economic and Social Studies (ICASEES) (Central African Republic), ICF International. Central African Republic Multiple Indicator Cluster Survey 2010-2011. Fairfax, United States: ICF International, 2013.
Chad	2244	sbh	21,938	United Nations Children's Fund (UNICEF), Census Bureau (Chad), National Institute of Statistical, Economic and Demographic Studies (Chad). Chad Multiple Indicator Cluster Survey 2000. New York, United States: United Nations Children's Fund (UNICEF).
Chad	76701	sbh	61,128	Ministry of Planning, Economy, and International Cooperation (Chad), National Institute of Statistical, Economic and Demographic Studies (Chad), United Nations Children's Fund (UNICEF). Chad Multiple Indicator Cluster Survey 2010. New York, United States: United Nations Children's Fund (UNICEF), 2014.
Chile	2311	sbh	623,140	National Institute of Statistics (INE) (Chile), Minnesota Population Center. Chile General Population and Housing Census 2002 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. Minneapolis: University of Minnesota.

China	294255	sbh	4,260,950	China Statistics Press, Minnesota Population Center. China National Population Census 2000 from the Integrated Public Use Microdata Series, International. Minneapolis: University of Minnesota, 2017.
Colombia	3029	sbh	1,771,037	National Administrative Department of Statistics (DANE) (Colombia), Minnesota Population Center. Colombia General Census 2005-2006 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. Minneapolis: University of Minnesota.
Colombia	21281	cbh	55,239	ICF Macro, Profamilia. Colombia Demographic and Health Survey 2009-2010. Calverton, United States: ICF Macro, 2011.
Comoros	3114	sbh	16,560	United Nations Development Programme (UNDP), United Nations Children's Fund (UNICEF). Comoros Multiple Indicator Cluster Survey 2000. New York, United States: United Nations Children's Fund (UNICEF).
Congo	3133	sbh	18,574	ICF Macro, National Center for Statistics and Economic Studies (Congo, Rep.). Congo AIDS Indicator Survey 2009. Calverton, United States: ICF Macro.
Costa Rica	3243	sbh	189,555	National Institute of Statistics and Censuses (INEC) (Costa Rica), Minnesota Population Center. Costa Rica National Population and Housing Census 2000 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. Minneapolis: University of Minnesota.
Costa Rica	227111	sbh	174,898	Minnesota Population Center, Costa Rica National Institute of Statistics and Census. Costa Rica Census 2011 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. Minneapolis: University of Minnesota, 2015.
Cuba	60935	sbh	13,154	Ministry of Public Health (Cuba), United Nations Children's Fund (UNICEF). Cuba Multiple Indicator Cluster Survey 2010-2011. New York, United States: United Nations Children's Fund (UNICEF).
Cuba	169975	sbh	12,553	Ministry of Public Health (Cuba), National Office of Statistics (Cuba), United Nations Children's Fund (UNICEF). Cuba Multiple Indicator Cluster Survey 2014. New York, United States: United Nations Children's Fund (UNICEF), 2005.
Democratic Republic of the Congo	3161	sbh	39,356	Ministry of Planning and Reconstruction (Congo, DR), United Nations Children's Fund (UNICEF). Congo, DR Multiple Indicator Cluster Survey 2001. New York, United States: United Nations Children's Fund (UNICEF).
Democratic Republic of the Congo	26998	sbh	40,648	National Statistical Institute (Congo, DR), Ministry of Planning (Congo, DR), United Nations Children's Fund (UNICEF). Congo, DR Multiple Indicator Cluster Survey 2010. New York, United States: United Nations Children's Fund (UNICEF).
Djibouti	3392	cbh	10,574	Department of Statistics and Demographic Studies (Djibouti), League of Arab States, Ministry of Health (Djibouti), Pan Arab Project for Family Health (PAPFAM). Djibouti Family Health Survey 2002.
Djibouti	3404	sbh	10,479	Ministry of Economy, Finance, and Planning in charge of Privatization (Djibouti), Ministry of Health (Djibouti), United Nations Children's Fund (UNICEF). Djibouti Multiple Indicator Cluster Survey 2006. New York, United States: United Nations Children's Fund (UNICEF).
Dominican Republic	27069	sbh	8,619	Center for Social and Demographic Studies (CESDEM), United Nations Children's Fund (UNICEF). Dominican Republic Multiple Indicator Cluster Survey 2000. New York, United States: United Nations Children's Fund (UNICEF).
Dominican Republic	151296	sbh	415,435	National Statistics Office (Dominican Republic), Minnesota Population Center. Dominican Republic Census 2002 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. Minneapolis: University of Minnesota.
Dominican Republic	3455	cbh	18,903	National Statistics Office (Dominican Republic), United Nations Children's Fund (UNICEF). Dominican Republic National Multipurpose Household Survey 2006. Santo Domingo, Dominican Republic: National Statistics Office (Dominican Republic).
Dominican Republic	151304	sbh	466,537	National Statistics Office (Dominican Republic), Minnesota Population Center. Dominican Republic Census 2010 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. Minneapolis: University of Minnesota.

Dominican Republic	200697	cbh	58,946	National Statistics Office (Dominican Republic), United Nations Children's Fund (UNICEF). Dominican Republic Multiple Indicator Cluster Survey 2014. New York, United States: United Nations Children's Fund (UNICEF), 2016.
Ecuador	27615	cbh	33,765	Center for Studies of Population and Social Development (CEPAR), Division of Reproductive Health-Centers for Disease Control and Prevention (CDC). (1995) Ecuador Reproductive Health Survey 1994. Atlanta, United States: Centers for Disease Control and Prevention (CDC).
Ecuador	27621	cbh	34,046	Center for Studies of Population and Social Development (CEPAR) (Ecuador), Division of Reproductive Health-Centers for Disease Control and Prevention (CDC). Ecuador Reproductive Health Survey 1999. Atlanta, United States: Centers for Disease Control and Prevention (CDC), 2001.
Ecuador	3549	sbh	613,208	National Institute of Statistics and Censuses (Ecuador), Minnesota Population Center. Ecuador Population and Housing Census 2001 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. Minneapolis: University of Minnesota.
Ecuador	27630	cbh	24,696	Center for Studies of Population and Social Development (CEPAR) (Ecuador) and Division of Reproductive Health-Centers for Disease Control and Prevention (CDC). (2005) Ecuador Reproductive Health Survey 2004. Quito, Ecuador: CEPAR.
Ecuador	105801	sbh	673,109	Minnesota Population Center, National Institute of Statistics and Census (INEC) (Ecuador). Ecuador Population and Housing Census 2010 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. Minneapolis: University of Minnesota, 2013.
Egypt	157026	sbh	16,984	El-Zanaty and Associates, ICF International, Ministry of Health and Population (Egypt), National Population Council (Egypt). Egypt Special Demographic and Health Survey 2015. Fairfax, United States: ICF International, 2015.
El Salvador	27590	cbh	32,149	El Salvador Demographic Association (ADS), Division of Reproductive Health-Centers for Disease Control and Prevention (CDC). El Salvador Reproductive Health Survey 1998. Atlanta, United States: Centers for Disease Control and Prevention (CDC).
El Salvador	27599	cbh	24,442	AsociaciÃ³n DemogrÃ¡fica SalvadoreÃ±a (ADS), Division of Reproductive Health-Centers for Disease Control and Prevention (CDC). (2004) El Salvador Reproductive Health Survey 2002-2003. San Salvador, El Salvador: ADS.
El Salvador	56476	sbh	270,154	Minnesota Population Center, General Administration of Statistics and Censuses (El Salvador), Ministry of Economy (El Salvador). El Salvador Population and Housing Census 2007 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. Minneapolis: University of Minnesota, 2012.
El Salvador	27606	cbh	25,228	AsociaciÃ³n DemogrÃ¡fica SalvadoreÃ±a (ADS), Division of Reproductive Health-Centers for Disease Control and Prevention (CDC). (2009) El Salvador Reproductive Health Survey 2008. San Salvador, El Salvador: ADS.
El Salvador	200636	cbh	24,689	General Administration of Statistics and Censuses (El Salvador), Ministry of Health (El Salvador), United Nations Children's Fund (UNICEF). El Salvador Multiple Indicator Cluster Survey 2014. New York, United States: United Nations Children's Fund (UNICEF), 2017.
Equitoria Guinea	3655	sbh	12,919	Ministry of Planning, Economic Development and Public Investment (Equatorial Guinea), United Nations Children's Fund (UNICEF). Equatorial Guinea Multiple Indicator Cluster Survey 2000. New York, United States: United Nations Children's Fund (UNICEF).
Eritrea	19546	cbh	14,268	Macro International, Inc, National Statistics Office (Eritrea). Eritrea Demographic and Health Survey 1995-1996. Calverton, United States: Macro International, Inc.
Ethiopia	227133	sbh	816,117	Minnesota Population Center, Ethiopia Central Statistical Agency. Ethiopia Population and Housing Census 2007 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. Minneapolis: University of Minnesota, 2015.

Fiji	105854	sbh	41,450	Minnesota Population Center, Bureau of Statistics (Fiji). Fiji Population and Housing Census 2007 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. Minneapolis: University of Minnesota, 2013
Georgia	3970	sbh	13,793	National Center for Disease Control (Georgia), State Department of Statistics of Georgia, United Nations Children's Fund (UNICEF). Georgia Multiple Indicator Cluster Survey 2005. New York, United States: United Nations Children's Fund (UNICEF).
Georgia	27494	cbh	8,284	Georgia Center for Disease Control (NCDC), Georgian Ministry of Labor Health and Social Affairs (MOLHSA), Division of Reproductive Health, Centers for Disease Control and Prevention (CDC). Georgia Reproductive Health Survey 2005. Atlanta, United States: Centers for Disease Control and Prevention (CDC).
Georgia	95336	cbh	7,773	Division of Reproductive Health, Centers for Disease Control and Prevention (CDC), Georgia Ministry of Labor, Health and Social Affairs, National Center for Disease Control and Public Health (Georgia), National Statistics Office of Georgia (GeoStat). Georgia Reproductive Health Survey 2010-2011.
Ghana	38508	sbh	1,134,199	Ghana Statistical Service, Minnesota Population Center. Ghana Population and Housing Census 2000 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. Minneapolis: University of Minnesota.
Ghana	4679	sbh	22,681	Ghana Statistical Service. Ghana Living Standards Measurement Survey 2005-2006. Accra, Ghana: Ghana Statistical Service.
Ghana	4694	sbh	15,689	Ministry of Health (MOH) (Ghana), Ghana Statistical Service and United Nations Children's Fund (UNICEF). Ghana Multiple Indicator Cluster Survey 2006. New York, United States: United Nations Children's Fund (UNICEF).
Ghana	21173	cbh	25,710	Ghana Health Service, Ghana Statistical Service, Macro International, Inc. Ghana Special Demographic and Health Survey 2007-2008. Calverton, United States: Macro International, Inc, 2010.
Ghana	160576	cbh	34,751	Ghana Statistical Service, Ministry of Health (Ghana), United Nations Children's Fund (UNICEF). Ghana District Multiple Indicator Cluster Survey 2007-2008.
Ghana	151306	sbh	1,275,941	Ghana Statistical Service, Minnesota Population Center. Ghana Census 2010 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. Minneapolis: University of Minnesota.
Ghana	56241	sbh	2,016	Institute of Statistical, Social and Economic Research, University of Ghana, United Nations Children's Fund (UNICEF). Ghana - Accra Multiple Indicator Cluster Survey 2010-2011. New York, United States: United Nations Children's Fund (UNICEF), 2014.
Ghana	63993	cbh	31,145	Centers for Disease Control and Prevention (CDC), Ghana Statistical Service, Government of Japan, ICF Macro, Ministry of Health (Ghana), Navrongo Health Research Centre, USAID, United Nations Children's Fund (UNICEF), United Nations Population Fund (UNFPA). Ghana Multiple Indicator Cluster Survey 2011. New York, United States: United Nations Children's Fund (UNICEF), 2013.
Ghana	286788	sbh	13,050	Ghana Health Service, Ghana Statistical Service, ICF International, National Malaria Control Program (Ghana), National Public Health and Reference Laboratory (NHPRL)(Ghana). Ghana Malaria Indicator Survey 2016. Fairfax, United States: ICF International, 2017.
Guatemala	45718	sbh	23,378	Inter-American Development Bank (IDB), National Statistics Institute (Guatemala), World Bank. Guatemala Living Standards Measurement Survey 2000. Washington DC, United States: World Bank.
Guatemala	27563	cbh	28,731	Guatemala Ministry of Health and Social Assistance, University of Valle, Division of Reproductive Health-Centers for Disease Control and Prevention (CDC). (2003) Guatemala Reproductive Health Survey 2002. Atlanta, United States: Centers for Disease Control and Prevention (CDC).
Guatemala	4779	cbh	45,017	Guatemala Ministry of Health and Social Assistance, University of Valle and Division of Reproductive Health-Centers for Disease Control and Prevention (CDC). Guatemala Reproductive Health Survey 2008-2009. Atlanta, United States: Centers for Disease Control and Prevention (CDC).

Guinea-Bissau	4808	sbh	23,844	Secretary State of Planning, National Institute of Statistics and Census (INEC), United Nations Children's Fund (UNICEF). Guinea-Bissau Multiple Indicator Cluster Survey 2000. New York, United States: United Nations Children's Fund (UNICEF).
Guinea-Bissau	4818	sbh	24,016	United Nations Children's Fund (UNICEF), Government of Guinea-Bissau. Guinea-Bissau Multiple Indicator Cluster Survey 2006. New York, United States: United Nations Children's Fund (UNICEF).
Guinea-Bissau	174049	cbh	27,607	National Statistics Institute (Guinea-Bissau), United Nations Children's Fund (UNICEF). Guinea-Bissau Multiple Cluster Indicator Survey 2014. New York, United States: United Nations Children's Fund (UNICEF), 2016.
Guyana	4905	cbh	2,887	Bureau of Statistics (Guyana), World Bank. Guyana Living Standards Measurement Survey 1992-1993.
Guyana	4916	sbh	11,691	Bureau of Statistics (Guyana), United Nations Children's Fund (UNICEF). Guyana Multiple Indicator Cluster Survey 2000. New York, United States: United Nations Children's Fund (UNICEF)
Guyana	4837	cbh	4,923	Central Bureau of Statistics (Ghana), Guyana Responsible Parenthood Association (GRPA), Ministry of Health (Guyana), ORC Macro, Pan American Health Organization (PAHO). Guyana AIDS Indicator Survey 2005. Calverton, United States: ORC Macro.
Guyana	4926	sbh	12,157	United Nations Children's Fund (UNICEF), Bureau of Statistics (Guyana). Guyana Multiple Indicator Cluster Survey 2006. New York, United States: United Nations Children's Fund (UNICEF).
Guyana	200598	cbh	11,161	Bureau of Statistics (Guyana), Ministry of Health (Guyana), United Nations Children's Fund (UNICEF). Guyana Multiple Indicator Cluster Survey 2014. New York, United States: United Nations Children's Fund (UNICEF), 2016.
Haiti	106473	sbh	439,200	Minnesota Population Center, Haitian Institute of Statistics and Informatics. Haiti Population and Housing Census 2003 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. Minneapolis: University of Minnesota, 2013
Haiti	26680	cbh	4,711	Global Fund to Fight Aids Tuberculosis and Malaria (GFATM). Haiti Global Fund Household Survey 2008.
Honduras	27551	cbh	23,535	Honduras Family Planning Association (ASHONPLAFA), Ministry of Health (Honduras), and Division of Reproductive Health-Centers for Disease Control and Prevention (CDC). Honduras Reproductive Health Survey 2001. Tegucigalpa, Honduras: Honduras Family Planning Association (ASHONPLAFA).
Honduras	5009	sbh	21,774	National Institute of Statistics (Honduras). Honduras Survey of Living Conditions 2004. Tegucigalpa, Honduras: National Institute of Statistics (Honduras).
India	23183	sbh	757,856	International Institute for Population Sciences (India). India District Level Household Survey 1998-1999. Mumbai, India: International Institute for Population Sciences (India).
India	23219	sbh	1,391,228	International Institute for Population Sciences (India). India District Level Household Survey 2002-2005. Mumbai, India: International Institute for Population Sciences (India).
India	23258	sbh	1,818,042	International Institute for Population Sciences (India). India District Level Household Survey 2007-2008. Mumbai, India: International Institute for Population Sciences (India), 2010.
India	165390	sbh	259,032	International Institute for Population Sciences (India). India District Level Household Survey 2012-2014. New Delhi, India: Ministry of Health and Family Welfare (India).
Indonesia	6535	cbh	568,692	Central Bureau of Statistics (Indonesia). Indonesia Intercensal Population Survey 1995.
Indonesia	6767	sbh	466,726	Central Bureau of Statistics (Indonesia), Ministry of Health (Indonesia), World Bank. Indonesia National Socioeconomic Survey 1998.
Indonesia	6790	sbh	438,949	Central Bureau of Statistics (Indonesia), Ministry of Health (Indonesia), World Bank. Indonesia National Socioeconomic Survey 1999.
Indonesia	6811	sbh	384,103	Central Bureau of Statistics (Indonesia), Ministry of Health (Indonesia), World Bank. Indonesia National Socioeconomic Survey 2000.

Indonesia	56554	sbh	9,961,117	Minnesota Population Center, Statistics Indonesia. Indonesia Population Census 2000 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. Minneapolis: University of Minnesota, 2012.
Indonesia	6842	sbh	458,846	Statistics Indonesia, Indonesia National Socioeconomic Survey 2001
Indonesia	43510	sbh	432,996	Statistics Indonesia. Indonesia National Socioeconomic Survey 2002.
Indonesia	6874	sbh	449,842	Statistics Indonesia. Indonesia National Socioeconomic Survey 2003.
Indonesia	6904	sbh	501,844	Statistics Indonesia. Indonesia National Socioeconomic Survey 2004.
Indonesia	5376	sbh	510,692	Statistics Indonesia. Indonesia National Socioeconomic Survey 2005.
Indonesia	5401	sbh	545,163	Statistics Indonesia. Indonesia National Socioeconomic Survey 2006.
Indonesia	6970	sbh	582,333	Statistics Indonesia. Indonesia National Socioeconomic Survey 2007.
Indonesia	43526	sbh	131,645	Statistics Indonesia. Indonesia National Socioeconomic Survey 2008.
Indonesia	43552	sbh	129,963	Statistics Indonesia. Indonesia National Socioeconomic Survey 2009.
Indonesia	30235	sbh	125,345	Statistics Indonesia. Indonesia National Socioeconomic Survey 2010.
Indonesia	56558	sbh	10,489,232	Minnesota Population Center, Statistics Indonesia. Indonesia Population Census 2010 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. Minneapolis: University of Minnesota, 2012.
Indonesia	91740	sbh	10,485,859	Statistics Indonesia. Indonesia Population Census 2010.
Indonesia	85265	sbh	130,728	Statistics Indonesia. Indonesia National Socioeconomic Survey 2011.
Indonesia	150884	sbh	534,256	Central Bureau of Statistics (Indonesia). Indonesia National Socioeconomic Survey 2012. Jakarta, Indonesia: Central Bureau of Statistics (Indonesia).
Indonesia	151184	sbh	518,178	Statistics Indonesia. Indonesia National Socioeconomic Survey 2013. Jakarta, Indonesia: Statistics Indonesia.
Iran	39396	sbh	628,318	Statistical Centre of Iran, Minnesota Population Center. Iran General Census of Population and Housing 2006 from the Integrated Public Use Microdata Series, International: Version 6.1 [Machine-readable database]. Minneapolis: University of Minnesota, 2011.
Iraq	7028	cbh	62,359	United Nations Children's Fund (UNICEF), Central Organization for Statistics and Information Technology (Iraq), Kurdistan Regional Statistics Office. Iraq Multiple Indicator Cluster Survey 2006. New York, United States: United Nations Children's Fund (UNICEF).
Iraq	23429	sbh	34,284	Ministry of Health (Iraq), Central Organization for Statistics and Information Technology (Iraq), Kurdistan Regional Statistics Office, World Health Organization (WHO), Ministry of Health (Kurdistan). Iraq Family Health Survey 2006-2007.
Iraq	76707	cbh	136,878	Central Organization for Statistics and Information Technology (Iraq), Kurdistan Regional Statistics Office, Ministry of Health (Iraq), United Nations Children's Fund (UNICEF). Iraq Multiple Indicator Cluster Survey 2011. New York, United States: United Nations Children's Fund (UNICEF), 2013.
Jamaica	7140	sbh	8,532	United Nations Children's Fund (UNICEF). Jamaica Multiple Indicator Cluster Survey 2000. New York, United States: United Nations Children's Fund (UNICEF).
Jamaica	39450	sbh	93,521	Statistical Institute of Jamaica (STATIN), Minnesota Population Center. Jamaica Population Census 2001 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. Minneapolis: University of Minnesota, 2011.
Jamaica	7149	sbh	7,206	Statistical Institute of Jamaica (STATIN) and United Nations Children's Fund (UNICEF). Jamaica Multiple Indicator Cluster Survey 2005. New York, United States: United Nations Children's Fund (UNICEF).
Kazakhstan	7340	sbh	23,329	Agency of the Republic of Kazakhstan on Statistics and United Nations Children's Fund (UNICEF). Kazakhstan Multiple Indicator Cluster Survey 2006. New York, United States: United Nations Children's Fund (UNICEF).

Kazakhstan	76702	sbh	22,048	Agency of the Republic of Kazakhstan on Statistics, United Nations Children's Fund (UNICEF). Kazakhstan Multiple Indicator Cluster Survey 2010-2011. New York, United States: United Nations Children's Fund (UNICEF), 2013.
Kazakhstan	260403	sbh	21,482	Ministry of National Economy (Kazakhstan), United Nations Children's Fund (UNICEF). Kazakhstan Multiple Indicator Cluster Survey 2015. New York, United States: United Nations Children's Fund (UNICEF), 2017.
Kenya	39481	sbh	964,933	Central Bureau of Statistics (CBS) (Kenya), Minnesota Population Center. Kenya Population and Housing Census 1999 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. Minneapolis: University of Minnesota.
Kenya	153943	sbh	976,094	Central Bureau of Statistics (Kenya), UK Department for International Development (DFID), United Nations Development Programme (UNDP), United Nations Population Fund (UNFPA), United States Agency for International Development (USAID). Kenya Population and Housing Census 1999.
Kenya	7387	sbh	31,278	Central Bureau of Statistics (Kenya), United Nations Children's Fund (UNICEF). Kenya Multiple Indicator Cluster Survey 2000. New York, United States: United Nations Children's Fund (UNICEF).
Kenya	7375	sbh	37,583	Central Bureau of Statistics (Kenya), UK Department for International Development (DFID), United States Agency for International Development (USAID), European Union (EU), Danish International Development Agency (DANIDA), World Bank (WB), United Nations Development Programme (UNDP). Kenya Integrated Household Budget Survey 2005-2006. Nairobi, Kenya: Central Bureau of Statistics (Kenya).
Kenya	57990	sbh	14,769	Centers for Disease Control and Prevention (CDC), KEMRI Wellcome Trust Research Programme (KWTRP), Kenya National Bureau of Statistics, Ministry of Public Health and Sanitation (Kenya), National Coordinating Agency for Population and Development (Kenya), Population Services International (PSI). Kenya Malaria Indicator Survey 2007.
Kenya	133219	sbh	26,155	Centers for Disease Control and Prevention (CDC), Kenya Medical Research Institute (KEMRI), Kenya National Bureau of Statistics, Ministry of Public Health and Sanitation (Kenya), National AIDS Control Council (Kenya), National AIDS and STI Control Program (Kenya), National Coordinating Agency for Population and Development (Kenya), National Public Health Laboratory Services, Ministry of Public Health and Sanitation (Kenya), USAID. Kenya AIDS Indicator Survey 2007. Nairobi, Kenya: Kenya National Bureau of Statistics.
Kenya	7427	sbh	2,390,112	Kenya National Bureau of Statistics, USAID, United Nations Population Fund (UNFPA), United States Census Bureau. Kenya Population and Housing Census 2009.
Kenya	106512	sbh	2,388,668	Minnesota Population Center, Kenya National Bureau of Statistics. Kenya Population Census 2009 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. Minneapolis: University of Minnesota, 2013.
Kenya	218579	sbh	14,087	ICF International, Kenya National Bureau of Statistics, National Malaria Control Program (NMCP) (Kenya). Kenya Malaria Indicator Survey 2015. Fairfax, United States: ICF International, 2015.
Kyrgyzstan	39466	sbh	256,968	National Statistical Committee of the Kyrgyz Republic, Minnesota Population Center. Kyrgyzstan National Population Census 1999 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. Minneapolis: University of Minnesota.
Kyrgyzstan	7540	sbh	12,820	United Nations Children's Fund (UNICEF), National Statistical Committee of the Kyrgyz Republic. Kyrgyzstan Multiple Indicator Cluster Survey 2005-2006. New York, United States: United Nations Children's Fund (UNICEF).
Kyrgyzstan	106520	sbh	280,046	Minnesota Population Center, National Statistical Committee of the Kyrgyz Republic. Kyrgyzstan Population and Housing Census 2009 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. Minneapolis: University of Minnesota, 2013.

Kyrgyzstan	162283	cbh	14,527	National Statistical Committee of the Kyrgyz Republic, United Nations Children's Fund (UNICEF). Kyrgyzstan Multiple Indicator Cluster Survey 2014. New York, United States: United Nations Children's Fund (UNICEF), 2015.
Laos	103973	cbh	56,802	Ministry of Education and Sports (Laos), Ministry of Health (Laos), Ministry of Planning and Investment (Laos). Laos Multiple Indicator Cluster Survey 2011-2012. New York, United States: United Nations Children's Fund (UNICEF), 2013.
Lebanon	44861	cbh	9,867	Central Administration of Statistics (Lebanon), League of Arab States, Ministry of Social Affairs (Lebanon), Pan Arab Project for Family Health (PAPFAM). Lebanon Family Health Survey 2004.
Lesotho	7721	sbh	15,482	Bureau of Statistics (Lesotho), United Nations Children's Fund (UNICEF). Lesotho Multiple Indicator Cluster Survey 2000. New York, United States: United Nations Children's Fund (UNICEF).
Liberia	151310	sbh	209,401	Liberia Institute for Statistics and Geo-information Services (LISGIS), Minnesota Population Center. Liberia Census 2008 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. Minneapolis: University of Minnesota.
Liberia	56828	sbh	13,917	ICF International, Liberia Institute for Statistics and Geo-information Services (LISGIS), National Malaria Control Program (Liberia). Liberia Malaria Indicator Survey 2011. Fairfax, United States: ICF International, 2012.
Liberia	286768	sbh	13,869	ICF International, Liberia Institute for Statistics and Geo-information Services (LISGIS), National Malaria Control Program (Liberia). Liberia Malaria Indicator Survey 2016. Fairfax, United States: ICF International, 2017.
Libya	107340	cbh	49,554	League of Arab States, National Center for Disease Control (Libya), Pan Arab Project for Family Health (PAPFAM). Libya Family Health Survey 2007.
Madagascar	27020	sbh	21,615	National Institute of Statistics (Madagascar), United Nations Children's Fund (UNICEF). Madagascar Multiple Indicator Cluster Survey 2000. New York, United States: United Nations Children's Fund (UNICEF).
Madagascar	69806	sbh	23,462	ICF International, National Institute of Statistics (Madagascar), National Program for the Fight Against Malaria (PNLP) (Madagascar), Pasteur Institute of Madagascar (IPM). Madagascar Malaria Indicator Survey 2011. Fairfax, United States: ICF International.
Madagascar	125594	cbh	9,956	National Institute of Statistics (Madagascar), United Nations Children's Fund (UNICEF). Madagascar - South Multiple Indicator Cluster Survey 2012. New York, United States: United Nations Children's Fund (UNICEF), 2015.
Madagascar	111438	sbh	22,075	ICF International, National Institute of Statistics (Madagascar), National Program for the Fight Against Malaria (PNLP) (Madagascar), Pasteur Institute of Madagascar (IPM). Madagascar Malaria Indicator Survey 2013. Fairfax, United States: ICF International, 2013.
Madagascar	218580	sbh	27,819	ICF International, Ministry of Public Health (Madagascar), National Institute of Statistics (Madagascar), National Program for the Fight Against Malaria (PNLP) (Madagascar), Pasteur Institute of Madagascar (IPM). Madagascar Malaria Indicator Survey 2016. Fairfax, United States: ICF International, 2017.
Malawi	40179	sbh	729,974	National Statistical Office (Malawi), Minnesota Population Center. Malawi Population and Housing Census 1998 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. Minneapolis: University of Minnesota, 2011.
Malawi	7919	cbh	78,960	United Nations Children's Fund (UNICEF), National Statistics Office (Malawi). Malawi Multiple Indicator Cluster Survey 2006. New York, United States: United Nations Children's Fund (UNICEF).
Malawi	40186	sbh	875,423	National Statistical Office (Malawi), Minnesota Population Center. Malawi Population and Housing Census 2008 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. Minneapolis: University of Minnesota, 2011.
Malawi	77387	sbh	8,026	ICF International, National Malaria Control Program (Malawi). Malawi Malaria Indicator Survey 2012. Fairfax, United States: ICF International, 2013.

Malawi	157059	sbh	8,126	ICF International, Ministry of Health (Malawi), National Malaria Control Program (Malawi), National Statistical Office of Malawi. Malawi Malaria Indicator Survey 2014. Fairfax, United States: ICF International, 2015.
Malawi	161662	cbh	72,579	National Statistical Office of Malawi, United Nations Children's Fund (UNICEF). Malawi Multiple Indicator Cluster Survey 2013-2014. New York, United States: United Nations Children's Fund (UNICEF), 2015.
Mali	40235	sbh	663,345	Central Census Bureau (Mali), Minnesota Population Center. Mali General Population and Housing Census 1998 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. Minneapolis: University of Minnesota.
Mali	151311	sbh	840,735	Central Census Bureau (Mali), Minnesota Population Center. Mali Census 2009 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. Minneapolis: University of Minnesota.
Mali	218587	sbh	28,960	ICF International, INFO-STAT (Mali), Ministry of Health and Public Hygiene (Mali), National Institute of Public Health Research (INRSP) (Mali), National Institute of Statistics (INSTAT) (Mali). Mali Malaria Indicator Survey 2015. Fairfax, United States: ICF International, 2016.
Mali	248224	cbh	55,820	Ministry of Health (Mali), Ministry of Planning (Mali), National Institute of Statistics (INSTAT) (Mali), United Nations Children's Fund (UNICEF). Mali Multiple Indicator Cluster Survey 2015. New York, United States: United Nations Children's Fund (UNICEF), 2017.
Mauritania	26871	cbh	13,246	Macro International, Inc, Ministry of Health and Social Affairs (Mauritania), National Office of Statistics (Mauritania). Mauritania Special Demographic and Health Survey 2003-2004. Calverton, United States: Macro International, Inc.
Mauritania	8115	sbh	35,683	National Office of Statistics (Mauritania), United Nations Children's Fund (UNICEF). Mauritania Multiple Indicator Cluster Survey 2007. New York, United States: United Nations Children's Fund (UNICEF).
Mauritania	152783	cbh	33,648	National Office of Statistics (Mauritania), United Nations Children's Fund (UNICEF). Mauritania Multiple Indicator Cluster Survey 2011. New York, United States: United Nations Children's Fund (UNICEF), 2015.
Mexico	43771	sbh	5,645,050	National Institute of Statistics, Geography, and Informatics (Mexico), Minnesota Population Center. Mexico Population and Housing Census 2000 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. Minneapolis: University of Minnesota.
Mexico	43776	sbh	5,076,280	National Institute of Statistics, Geography, and Informatics (Mexico), Minnesota Population Center. Mexico Population and Housing Census 2005 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. Minneapolis: University of Minnesota.
Mexico	23982	cbh	78,420	National Institute of Statistics, Geography, and Informatics (Mexico), National Institute for Public Health (Mexico), National Population Council (Mexico). Mexico National Survey of Demographic Dynamics 2006. Aguascalientes, Mexico: National Institute of Statistics, Geography, and Informatics (Mexico).
Mexico	24006	sbh	158,141	National Institute of Statistics and Geography (INEGI) (Mexico). Mexico National Survey of Demographic Dynamics 2009.
Mexico	56480	sbh	6,073,966	Minnesota Population Center, National Institute of Statistics and Geography (Mexico). Mexico Population and Housing Census 2010 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. Minneapolis: University of Minnesota, 2012.
Mexico	93321	sbh	50,526	National Institute of Statistics and Geography (INEGI) (Mexico). Mexico Household Income and Expenditure Survey 2010. Mexico City, Mexico: National Institute of Statistics and Geography (INEGI) (Mexico).
Mexico	240525	sbh	10,418,861	National Institute of Statistics and Geography (INEGI) (Mexico). Mexico Intercensal Survey 2015. Mexico City, Mexico: National Institute of Statistics and Geography (INEGI) (Mexico).
Mexico	294574	sbh	5,403,100	National Institute of Statistics, Geography and Informatics (INEGI), Minnesota Population Center. Mexico Intercensal Survey

				2015 from the Integrated Public Use Microdata Series, International. Minneapolis: University of Minnesota, 2017.
Moldova	20339	cbh	9,903	Macro International, Inc, National Scientific and Applied Center for Preventive Medicine (Moldova). Moldova Demographic and Health Survey 2005. Calverton, United States: Macro International, Inc.
Mongolia	43016	cbh	17,782	Ministry of Health and Social Welfare (Mongolia), National Statistical Office of Mongolia, United Nations Statistics Division (UNSD). Mongolia Reproductive Health Survey 1998.
Mongolia	8788	sbh	17,766	National Statistical Office of Mongolia. Mongolia Multiple Indicator Cluster Survey 2000. New York, United States: United Nations Children's Fund (UNICEF).
Mongolia	8777	sbh	15,378	National Statistical Office of Mongolia, United Nations Children's Fund (UNICEF). Mongolia Multiple Indicator Cluster Survey 2005. New York, United States: United Nations Children's Fund (UNICEF).
Mongolia	76704	sbh	18,061	National Statistical Office of Mongolia, United Nations Children's Fund (UNICEF). Mongolia Multiple Indicator Cluster Survey 2010. New York, United States: United Nations Children's Fund (UNICEF), 2013.
Morocco	20351	cbh	10,496	Macro International, Inc, Ministry of Public Health (Morocco). Morocco Special Demographic and Health Survey 1995. Calverton, United States: Macro International, Inc.
Morocco	56492	sbh	726,578	Minnesota Population Center, High Commission for Planning (Morocco). Morocco Population and Housing Census 2004 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. Minneapolis: University of Minnesota, 2012.
Morocco	126909	sbh	30,414	Ministry of Health (Morocco), Pan Arab Project for Family Health (PAPFAM), United Nations Children's Fund (UNICEF), United Nations Population Fund (UNFPA), World Health Organization (WHO). Morocco National Survey on Population and Family Health 2010-2011.
Mozambique	227143	sbh	1,359,295	Minnesota Population Center, Mozambique National Statistics Institute. Mozambique Census 2007 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. Minneapolis: University of Minnesota, 2015.
Mozambique	8906	sbh	29,263	ICF Macro, Ministry of Health (Mozambique), National Institute of Statistics (INE) (Mozambique). Mozambique AIDS Indicator Survey 2009. Calverton, United States: ICF Macro, 2010.
Mozambique	27031	cbh	42,215	United Nations Children's Fund (UNICEF), National Statistics Institute (Mozambique). Mozambique Multiple Indicator Cluster Survey 2008-2009. New York, United States: United Nations Children's Fund (UNICEF).
Myanmar	90696	cbh	60,796	Ministry of Health (Myanmar), Ministry of National Planning and Economic Development (Myanmar), United Nations Children's Fund (UNICEF). Myanmar Multiple Indicator Cluster Survey 2009-2010.
Nepal	162317	cbh	28,647	Central Bureau of Statistics (Nepal), United Nations Children's Fund (UNICEF). Nepal Multiple Indicator Cluster Survey 2014. New York, United States: United Nations Children's Fund (UNICEF), 2015.
Nicaragua	56520	sbh	295,576	Minnesota Population Center, National Institute of Statistics and Censuses (Nicaragua). Nicaragua Population and Housing Census 2005 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. Minneapolis: University of Minnesota, 2012.
Nicaragua	9270	cbh	34,055	Division of Reproductive Health, Centers for Disease Control and Prevention (CDC), National Institute for Development Information (Nicaragua). Nicaragua Reproductive Health Survey 2006-2007. Managua, Nicaragua: National Institute for Development Information (Nicaragua).
Niger	9439	sbh	21,570	Government of Niger, Macro International, Inc, United Nations Children's Fund (UNICEF). Niger Multiple Indicator Cluster Survey 2000. New York, United States: United Nations Children's Fund (UNICEF).

Nigeria	9506	sbh	55,461	National Bureau of Statistics (Nigeria), United Nations Children's Fund (UNICEF). Nigeria Multiple Indicator Cluster Survey 1999. Abuja, Nigeria: National Bureau of Statistics (Nigeria).
Nigeria	25006	sbh	52,723	Federal Office of Statistics (Nigeria). Nigeria Living Standards Survey 2003-2004.
Nigeria	9516	sbh	68,689	United Nations Children's Fund (UNICEF), National Bureau of Statistics (Nigeria). Nigeria Multiple Indicator Cluster Survey 2007. New York, United States: United Nations Children's Fund (UNICEF).
Nigeria	24915	sbh	55,508	Central Bank of Nigeria, National Bureau of Statistics (Nigeria), Nigerian Communications Commission (NCC). Nigeria General Household Survey 2008.
Nigeria	76703	sbh	100,531	National Bureau of Statistics (Nigeria), United Nations Children's Fund (UNICEF). Nigeria Multiple Indicator Cluster Survey 2011. New York, United States: United Nations Children's Fund (UNICEF), 2013.
Nigeria	218590	sbh	25,450	ICF International, National Bureau of Statistics (Nigeria), National Malaria Control Programme (Nigeria), National Population Commission of Nigeria. Nigeria Malaria Indicator Survey 2015. Fairfax, United States: ICF International, 2016.
Nigeria	218613	cbh	101,649	National Agency for the Control of AIDS (Nigeria), National Bureau of Statistics (Nigeria), National Primary Health Care Development Agency (NPHCDA) (Nigeria), United Nations Children's Fund (UNICEF). Nigeria Multiple Indicator Cluster Survey with National Immunization Coverage Survey Supplement 2016-2017. New York, United States: United Nations Children's Fund (UNICEF), 2018.
Pakistan	9919	cbh	24,151	Federal Bureau of Statistics (Pakistan) and World Bank. Pakistan Living Standards Measurement Survey 1991. Islamabad, Pakistan: Federal Bureau of Statistics (Pakistan).
Pakistan	9658	sbh	68,941	Federal Bureau of Statistics (Pakistan). Pakistan Integrated Household Survey 1998-1999. Islamabad, Pakistan: Federal Bureau of Statistics (Pakistan).
Pakistan	9720	sbh	68,782	Federal Bureau of Statistics (Pakistan). Pakistan Integrated Household Survey 2001-2002. Islamabad, Pakistan: Federal Bureau of Statistics (Pakistan).
Pakistan	24818	sbh	60,656	Federal Bureau of Statistics (Pakistan). Pakistan Social and Living Standards Measurement Survey 2005-2006. Islamabad, Pakistan: Federal Bureau of Statistics (Pakistan).
Pakistan	30634	sbh	56,277	Federal Bureau of Statistics (Pakistan). Pakistan Social and Living Standards Measurement Survey 2007-2008. Islamabad, Pakistan: Federal Bureau of Statistics (Pakistan).
Palestine	10001	cbh	26,074	Ministry of Health (Palestine), Palestinian Central Bureau of Statistics, United Nations Children's Fund (UNICEF), United Nations Population Fund (UNFPA). Palestine - West Bank and Gaza Strip Multiple Indicator Cluster Survey 2000. Ramallah, Palestine: Palestinian Central Bureau of Statistics.
Palestine	20596	cbh	22,478	Palestinian Central Bureau of Statistics. Palestine Demographic and Health Survey 2004.
Palestine	9999	cbh	51,635	League of Arab States, Palestinian Central Bureau of Statistics, United Nations Children's Fund (UNICEF). Palestine Family Health Survey 2006-2007.
Palestine	125591	cbh	55,823	Ministry of Health (Palestine), Palestinian Central Bureau of Statistics, United Nations Children's Fund (UNICEF), United Nations Population Fund (UNFPA). Palestine Multiple Indicator Cluster Survey 2010. New York, United States: United Nations Children's Fund (UNICEF), 2014.
Palestine	161590	cbh	31,817	Ministry of Health (Palestine), Palestinian Central Bureau of Statistics, United Nations Children's Fund (UNICEF). Palestine Multiple Indicator Cluster Survey 2014. New York, United States: United Nations Children's Fund (UNICEF), 2015.
Panama	40907	sbh	142,057	Department of Statistics and Census (Panama), Minnesota Population Center. Panama Population and Housing Census 2000 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. Minneapolis: University of Minnesota.

Panama	10224	sbh	13,639	Census and Statistics Directorate (Panama), Ministry of Economy and Finance (Panama), World Bank. Panama Living Standard Measurement Survey 2003. Washington DC, United States: World Bank.
Panama	106529	sbh	156,544	Minnesota Population Center, National Institute of Statistics and Census (Panama). Panama Population and Housing Census 2010 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. Minneapolis: University of Minnesota, 2013.
Papua New Guinea	58191	sbh	3,250,144	National Statistical Office (Papua New Guinea). Papua New Guinea Census 2000.
Paraguay	10326	cbh	8,262	Division of Reproductive Health-Centers for Disease Control and Prevention (CDC). Paraguay Contraceptive Prevalence Survey 1998. Atlanta, United States: Centers for Disease Control and Prevention (CDC).
Paraguay	10357	sbh	11,580	Department of Statistics, Surveys and Censuses (Paraguay). Paraguay Integrated Household Survey 1997-1998. AsunciÃ³n, Paraguay: Department of Statistics, Surveys and Censuses (Paraguay).
Paraguay	10350	sbh	20,082	Department of Statistics, Surveys and Censuses (Paraguay). Paraguay Integrated Household Survey 2000-2001. AsunciÃ³n, Paraguay: Department of Statistics, Surveys and Censuses (Paraguay).
Paraguay	227167	sbh	272,548	Minnesota Population Center, Paraguay Department of Statistics, Surveys and Censuses. Paraguay Population and Housing Census 2002 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. Minneapolis: University of Minnesota, 2015.
Paraguay	10370	cbh	14,799	Division of Reproductive Health-Centers for Disease Control and Prevention (CDC). (2005): Paraguay Reproductive Health Survey 2004. AsunciÃ³n, Paraguay, Paraguayan Center for Population Studies (CEPEP).
Paraguay	27525	cbh	11,367	Paraguay Center for Population Studies (CEPEP). Paraguay Reproductive Health Survey 2008. AsunciÃ³n, Paraguay: Paraguayan Center for Population Studies (CEPEP).
Paraguay	324470	cbh	14,355	General Directorate of Statistics, Surveys and Censuses (DGEEC)(Paraguay), Ministry of Public Health and Social Welfare (Paraguay), United Nations Children's Fund (UNICEF). Paraguay Multiple Indicator Cluster Survey 2016. New York, United States: United Nations Children's Fund (UNICEF), 2017.
Peru	10460	sbh	10,468	Instituto CuÃ¡nto. Peru National Living Standards Measurement Survey 2000. Lima, Peru: Instituto CuÃ¡nto.
Peru	41267	sbh	1,255,851	National Institute of Statistics and Informatics (INEI) (Peru), Minnesota Population Center. Peru National Population and Housing Census 2007 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. Minneapolis: University of Minnesota.
Peru	146860	cbh	44,725	ICF International, National Institute of Statistics and Informatics (INEI) (Peru). Peru Continuous Demographic and Health Survey 2013. Lima, Peru: National Institute of Statistics and Informatics (INEI) (Peru), 2014.
Peru	303663	cbh	74,559	National Institute of Statistics and Informatics (INEI) (Peru). Peru Demographic and Family Health Survey 2015. Lima, Peru: National Institute of Statistics and Informatics (INEI) (Peru), 2017.
Peru	303664	cbh	67,481	National Institute of Statistics and Informatics (INEI) (Peru). Peru Demographic and Family Health Survey 2016. Lima, Peru: National Institute of Statistics and Informatics (INEI) (Peru), 2017.
Philippines	135803	cbh	99,962	ICF International, National Statistics Office (Philippines). Philippines Demographic and Health Survey 2011.
Rwanda	42432	sbh	518,045	National Census Commission (Rwanda), Minnesota Population Center. Rwanda Population and Housing Census 2002 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. Minneapolis: University of Minnesota.
Rwanda	11324	sbh	21,025	National Institute of Statistics of Rwanda (NISR), Oxford Policy Management. Rwanda Integrated Living Conditions Survey 2005-2006. Kigali, Rwanda: National Institute of Statistics of Rwanda (NISR).

Rwanda	218773	sbh	547,998	National Institute of Statistics of Rwanda. Rwanda Population and Housing Census 2012. Kigali, Rwanda: National Institute of Statistics of Rwanda, 2015.
Rwanda	77391	sbh	11,726	ICF International, Ministry of Health (Rwanda). Rwanda Malaria Indicator Survey 2012-2013. Fairfax, United States: ICF International, 2014.
Sao Tome and Principe	27055	sbh	7,765	National Institute of Statistics (Sao Tome and Principe), United Nations Children's Fund (UNICEF). Sao Tome and Principe Multiple Indicator Cluster Survey 2000. New York, United States: United Nations Children's Fund (UNICEF).
Sao Tome and Principe	214640	cbh	7,492	Global Fund to Fight Aids Tuberculosis and Malaria (GFATM), ICF International, National Center for Endemic Diseases (CNE) (Sao Tome and Principe), National Institute of Statistics (Sao Tome and Principe), United Nations Children's Fund (UNICEF), United Nations Development Programme (UNDP). Sao Tome and Principe Multiple Indicator Cluster Survey 2014. New York, United States: United Nations Children's Fund (UNICEF), 2016.
Senegal	20786	cbh	51,506	Groupe SERDHA, Macro International, Inc, Ministry of Health and Prevention (Senegal). Senegal Demographic and Health Survey 1999-2000. Calverton, United States: Macro International, Inc.
Senegal	43142	sbh	564,269	Directorate of Forecasting and Statistics (Senegal), Minnesota Population Center. Senegal General Population and Housing Census 2002 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. Minneapolis: University of Minnesota.
Senegal	11516	sbh	18,520	Macro International, Inc, Research Center for Human Development (Senegal). Senegal Malaria Indicator Survey 2006. Calverton, United States: Macro International, Inc.
Sierra Leone	11639	sbh	17,032	Central Statistics Office (Sierra Leone), United Nations Children's Fund (UNICEF). Sierra Leone Multiple Indicator Cluster Survey 2000. New York, United States: United Nations Children's Fund (UNICEF).
Sierra Leone	11661	sbh	344,320	Statistics Sierra Leone and Minnesota Population Center. Sierra Leone Population and Housing Census 2004 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. Minneapolis: University of Minnesota, 2011.
Sierra Leone	11649	sbh	28,284	United Nations Children's Fund (UNICEF), Statistics Sierra Leone. Sierra Leone Multiple Indicator Cluster Survey 2005. New York, United States: United Nations Children's Fund (UNICEF).
Sierra Leone	76700	sbh	39,257	Statistics Sierra Leone, United Nations Children's Fund (UNICEF). Sierra Leone Multiple Indicator Cluster Survey 2010. New York, United States: United Nations Children's Fund (UNICEF).
Sierra Leone	286773	sbh	28,463	Catholic Relief Services (CRS), College of Medicine and Allied Health Sciences, University of Sierra Leone (COMAHS), ICF International, Roll Back Malaria Partnership, Statistics Sierra Leone. Sierra Leone Malaria Indicator Survey 2016. Fairfax, United States: ICF International, 2017.
Somalia	11774	cbh	20,034	Pan Arab Project for Family Health (PAPFAM), United Nations Children's Fund (UNICEF). Somalia Multiple Indicator Cluster Survey 2006. New York, United States: United Nations Children's Fund (UNICEF).
South Africa	105306	cbh	65,741	Central Statistical Service (South Africa). South Africa October Household Survey 1994.
South Africa	106684	cbh	62,706	Central Statistical Service (South Africa). South Africa October Household Survey 1995.
South Africa	106686	cbh	104,873	Central Statistical Service (South Africa). South Africa October Household Survey 1997.
South Africa	43152	sbh	1,799,625	Statistics South Africa, Minnesota Population Center. South Africa Census 2001 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. Minneapolis: University of Minnesota.
South Africa	43158	sbh	469,555	Statistics South Africa, Minnesota Population Center. South Africa Community Survey 2007 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. Minneapolis: University of Minnesota.

South Africa	280803	sbh	1,247,106	Statistics South Africa. South Africa Community Survey 2016. Pretoria, South Africa: Statistics South Africa, 2016.
South Sudan	106548	sbh	337,728	Minnesota Population Center, Southern Sudan Centre for Census, Statistics and Evaluation. Sudan - South Sudan Population and Housing Census 2008 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. Minneapolis: University of Minnesota, 2013
Sudan	24143	cbh	77,520	Ministry of Health (Southern Sudan), Federal Ministry of Health (Sudan), Southern Sudan Centre for Census, Statistics and Evaluation (SSCCSE), Central Bureau of Statistics (Sudan). Sudan Family Health Survey 2006.
Sudan	43167	sbh	2,810,742	National Population Census Council (Sudan), Central Bureau of Statistics (Sudan), Southern Sudan Centre for Census, Statistics and Evaluation (SSCCSE), Minnesota Population Center. Sudan Population and Housing Census 2008 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. Minneapolis: University of Minnesota, 2011.
Sudan	153643	cbh	47,092	Central Bureau of Statistics (Sudan), Ministry of Health (South Sudan). Sudan - North Multiple Indicator Cluster Survey 2010. New York, United States: United Nations Children's Fund (UNICEF), 2015.
Sudan	200617	cbh	52,245	Central Bureau of Statistics (Sudan), Federal Ministry of Health (Sudan), United Nations Children's Fund (UNICEF). Sudan Multiple Indicator Cluster Survey 2014. New York, United States: United Nations Children's Fund (UNICEF), 2016.
Suriname	12280	sbh	8,481	General Bureau of Statistics (Suriname), Pan American Health Organization (PAHO), United Nations Children's Fund (UNICEF), United Nations Development Programme (UNDP). Suriname Multiple Indicator Cluster Survey 1999-2000. New York, United States: United Nations Children's Fund (UNICEF).
Suriname	12289	sbh	10,503	General Statistical Office (Suriname), United Nations Children's Fund (UNICEF). Suriname Multiple Indicator Cluster Survey 2006. New York, United States: United Nations Children's Fund (UNICEF).
Swaziland	12320	sbh	13,812	Central Statistical Office (Swaziland), United Nations Children's Fund (UNICEF). Swaziland Multiple Indicator Cluster Survey 2000. New York, United States: United Nations Children's Fund (UNICEF).
Swaziland	30325	cbh	9,805	Central Statistical Office (Swaziland), United Nations Children's Fund (UNICEF). Swaziland Multiple Indicator Cluster Survey 2010. New York, United States: United Nations Children's Fund (UNICEF).
Swaziland	200707	cbh	9,830	Central Statistical Office (Swaziland), United Nations Children's Fund (UNICEF), United Nations Educational, Scientific and Cultural Organization (UNESCO), United Nations Population Fund (UNFPA). Swaziland Multiple Indicator Cluster Survey 2014. New York, United States: United Nations Children's Fund (UNICEF), 2016.
Syria	12399	sbh	55,015	United Nations Children's Fund (UNICEF), Central Bureau of Statistics (Syria), Ministry of Health (Syria), Pan Arab Project for Family Health (PAPFAM). Syria Multiple Indicator Cluster Survey 2006. New York, United States: United Nations Children's Fund (UNICEF).
Tajikistan	12455	cbh	8,527	National State Statistical Agency (Tajikistan), World Bank. Tajikistan Living Standards Measurement Survey 1999.
Tajikistan	12595	sbh	16,393	National State Statistical Agency (Tajikistan), United Nations Children's Fund (UNICEF). Tajikistan Multiple Indicator Cluster Survey 2000 . New York, United States: United Nations Children's Fund (UNICEF).
Tajikistan	12489	cbh	13,458	National State Statistical Agency (Tajikistan), World Bank. Tajikistan Living Standards Measurement Survey 2003.
Tajikistan	12608	sbh	23,127	United Nations Children's Fund (UNICEF), State Committee on Statistics of the Republic of Tajikistan. Tajikistan Multiple Indicator Cluster Survey 2005. New York, United States: United Nations Children's Fund (UNICEF).
Tanzania	43212	sbh	2,664,423	National Bureau of Statistics (Tanzania), Minnesota Population Center. Tanzania Population and Housing Census 2002 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. Minneapolis: University of Minnesota.

Tanzania	12630	sbh	33,014	National Bureau of Statistics (Tanzania), ORC Macro, Tanzania Commission for AIDS (TACAIDS). Tanzania AIDS Indicator Survey 2003-2004. Calverton, United States: ORC Macro.
Tanzania	12644	cbh	27,511	Macro International, Inc, National Bureau of Statistics (Tanzania), Office of Chief Government Statistician (OCGS-Zanzibar), Tanzania Commission for AIDS (TACAIDS), Zanzibar AIDS Commission (ZAC). Tanzania HIV/AIDS and Malaria Indicator Survey 2007-2008. Calverton, United States: Macro International, Inc.
Tanzania	77395	sbh	32,522	ICF International, National Bureau of Statistics (Tanzania), Office of Chief Government Statistician (OCGS-Zanzibar), Tanzania Commission for AIDS (TACAIDS), Zanzibar AIDS Commission (ZAC). Tanzania AIDS Indicator Survey 2011-2012. Fairfax, United States: ICF International, 2013.
Tanzania	294725	sbh	3,225,395	National Bureau of Statistics (Tanzania), Minnesota Population Center. Tanzania Population and Housing Census 2012 from the Integrated Public Use Microdata Series, International. Minneapolis: University of Minnesota, 2017.
Thailand	43231	sbh	213,031	National Statistical Office (Thailand), Minnesota Population Center. Thailand Population and Housing Census 2000 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. Minneapolis: University of Minnesota.
Thailand	12732	sbh	48,610	National Statistical Office (Thailand), United Nations Children's Fund (UNICEF). Thailand Multiple Indicator Cluster Survey 2005-2006. New York, United States: United Nations Children's Fund (UNICEF).
Thailand	148649	sbh	30,853	College of Population Studies, Chulalongkorn University (Thailand), Institute for Population and Social Research, Mahidol University (Thailand), International Health Policy Program (Thailand), Ministry of Education (Thailand), Ministry of Public Health (Thailand), Ministry of Social Development and Human Security (MSDHS) (Thailand), National Health Security Office (Thailand), National Statistical Office (Thailand), Thai Health Promotion Foundation, United Nations Children's Fund (UNICEF). Thailand Multiple Indicator Cluster Survey 2012. New York, United States: United Nations Children's Fund (UNICEF), 2016.
The Gambia	3922	sbh	16,875	Central Statistics Department (Gambia), United Nations Children's Fund (UNICEF). Gambia Multiple Indicator Cluster Survey 2000. New York, United States: United Nations Children's Fund (UNICEF).
The Gambia	3935	sbh	27,475	Gambia Bureau of Statistics (GBOS), United Nations Children's Fund (UNICEF). Gambia Multiple Indicator Cluster Survey 2005-2006. New York, United States: United Nations Children's Fund (UNICEF).
Timor-Leste	20888	cbh	17,889	ACIL Australia Pty Ltd., Australian National University, Ministry of Health (Timor-Leste), National Statistics Directorate (Timor-Leste), University of Newcastle (Australia). Timor-Leste Demographic and Health Survey 2003. Newcastle, Australia: University of Newcastle (Australia).
Togo	12896	sbh	17,832	Directorate General of Statistics and National Accounting (Togo), United Nations Children's Fund (UNICEF). Togo Multiple Indicator Cluster Survey 2006. New York, United States: United Nations Children's Fund (UNICEF).
Togo	40021	sbh	18,954	Directorate General of Statistics and National Accounting (Togo), United Nations Children's Fund (UNICEF). Togo Multiple Indicator Cluster Survey 2010. New York, United States: United Nations Children's Fund (UNICEF).
Trinidad and Tobago	12940	sbh	6,445	Central Statistical Office (Trinidad and Tobago), United Nations Children's Fund (UNICEF). Trinidad and Tobago Multiple Indicator Cluster Survey 2000. New York, United States: United Nations Children's Fund (UNICEF).
Trinidad and Tobago	12950	sbh	6,551	Central Statistical Office (Trinidad and Tobago) and United Nations Children's Fund (UNICEF). Trinidad and Tobago Multiple Indicator Cluster Survey 2006. New York, United States: United Nations Children's Fund (UNICEF).

Trinidad and Tobago	294807	sbh	35,604	Central Statistical Office (Trinidad and Tobago), Minnesota Population Center. Trinidad and Tobago Population and Housing Census 2011 from the Integrated Public Use Microdata Series, International [Machine-readable database]. Minneapolis: University of Minnesota, 2017.
Tunisia	76709	cbh	13,569	Ministry of Regional Development and Planning (Tunisia), National Institute of Statistics (Tunisia), United Nations Children's Fund (UNICEF). Tunisia Multiple Indicator Cluster Survey 2011-2012. New York, United States: United Nations Children's Fund (UNICEF), 2014.
Turkey	56509	sbh	1,638,302	Minnesota Population Center, State Institute of Statistics (Turkey). Turkey Population Census 2000 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. Minneapolis: University of Minnesota, 2012.
Turkey	32421	cbh	19,678	Institute of Population Studies, Hacettepe University, Ministry of Health (Turkey), State Planning Organization (Turkey), Turkish Statistical Institute. Turkey Demographic and Health Survey 2008. Ankara, Turkey: Institute of Population Studies, Hacettepe University.
Turkey	90439	sbh	55,121	Istanbul University, Ministry of Health (Turkey), Turkish Statistical Institute. Turkey Infant and Under-5 Mortality Survey 2011.
Turkmenistan	13064	sbh	12,070	Ministry of Health and Medical Industry (Turkmenistan), United Nations Children's Fund (UNICEF). Turkmenistan Multiple Indicator Cluster Survey 2006. New York, United States: United Nations Children's Fund (UNICEF), 2016.
Turkmenistan	264583	cbh	12,800	State Committee on Statistics of Turkmenistan, United Nations Children's Fund (UNICEF). Turkmenistan Multiple Indicator Cluster Survey 2015-2016. New York, United States: United Nations Children's Fund (UNICEF), 2017.
Uganda	43328	sbh	1,811,659	Uganda Bureau of Statistics, Minnesota Population Center. Uganda Population and Housing Census 2002 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. Minneapolis: University of Minnesota.
Uganda	81004	sbh	9,404	Uganda Bureau of Statistics. Uganda Living Standards Measurement Survey - Integrated Survey on Agriculture 2009-2010. Washington DC, United States: World Bank.
Uganda	55973	sbh	68,086	Centers for Disease Control and Prevention (CDC), ICF Macro, Ministry of Health (Uganda), Uganda Bureau of Statistics, Uganda Virus Research Institute. Uganda AIDS Indicator Survey 2011. Calverton, United States: ICF Macro.
Uganda	157065	sbh	17,128	ICF International, National Malaria Control Program, Ministry of Health (Uganda), Uganda Bureau of Statistics. Uganda Malaria Indicator Survey 2014-2015. Fairfax, United States: ICF International, 2015.
Ukraine	13218	cbh	8,144	Division of Reproductive Health-Centers for Disease Control and Prevention (CDC) and Kiev International Institute of Sociology. (2001) Ukraine Reproductive Health Survey 1999. Atlanta, United States: Centers for Disease Control and Prevention (CDC).
Ukraine	13197	sbh	7,872	United Nations Children's Fund (UNICEF). Ukraine Multiple Indicator Cluster Survey 2005. New York, United States: United Nations Children's Fund (UNICEF).
Ukraine	132739	sbh	10,715	StatInform Consulting, State Statistics Service (Ukraine), Ukrainian Center for Social Reforms (UCSR), United Nations Children's Fund (UNICEF). Ukraine Multiple Indicator Cluster Survey 2012. New York, United States: United Nations Children's Fund (UNICEF), 2014.
Uruguay	56577	sbh	95,208	Minnesota Population Center, National Institute of Statistics (Uruguay). Uruguay Extended National Household Survey 2006 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. Minneapolis: University of Minnesota, 2012.
Uruguay	151322	sbh	106,698	National Institute of Statistics (Uruguay), Minnesota Population Center. Uruguay Census 2011 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. Minneapolis: University of Minnesota.
Uzbekistan	13436	sbh	16,770	United Nations Children's Fund (UNICEF), Ministry of Macroeconomics and Statistics (Uzbekistan). Uzbekistan Multiple Indicator Cluster Survey 2000. New York, United States: United Nations Children's Fund (UNICEF).

Uzbekistan	21039	cbh	11,607	Analytical and Information Center of the Ministry of Health of Uzbekistan, Macro International, Inc, Ministry of Macroeconomics and Statistics (Uzbekistan). Uzbekistan Special Demographic and Health Survey 2002. Calverton, United States: Macro International, Inc.
Uzbekistan	13445	sbh	26,751	United Nations Children's Fund (UNICEF), State Committee of the Republic of Uzbekistan on Statistics. Uzbekistan Multiple Indicator Cluster Survey 2006. New York, United States: United Nations Children's Fund (UNICEF).
Vanuatu	13465	sbh	6,316	Ministry of Health (Vanuatu), United Nations Children's Fund (UNICEF). Vanuatu Multiple Indicator Cluster Survey 2007-2008. New York, United States: United Nations Children's Fund (UNICEF).
Venezuela	43412	sbh	1,161,057	National Institute of Statistics (Venezuela), Minnesota Population Center. Venezuela Population and Housing Census 2002 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. Minneapolis: University of Minnesota.
Vietnam	43718	sbh	1,103,904	General Statistics Office (Viet Nam), Minnesota Population Center. Viet Nam Population and Housing Census 1999 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. Minneapolis: University of Minnesota.
Vietnam	13708	sbh	17,570	General Statistics Office (Viet Nam), United Nations Children's Fund (UNICEF). Vietnam Multiple Indicator Cluster Survey 2000. New York, United States: United Nations Children's Fund (UNICEF).
Vietnam	13544	sbh	20,964	General Statistics Office (Viet Nam), National Institute of Hygiene and Epidemiology (Viet Nam), ORC Macro. Vietnam AIDS Indicator Survey 2005. Calverton, United States: Macro International, Inc.
Vietnam	13719	sbh	16,447	General Statistics Office (Viet Nam), United Nations Children's Fund (UNICEF). Vietnam Multiple Indicator Cluster Survey 2006. New York, United States: United Nations Children's Fund (UNICEF).
Vietnam	43726	sbh	6,004,427	General Statistics Office (Viet Nam), Minnesota Population Center. Viet Nam Population and Housing Census 2009 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. Minneapolis: University of Minnesota.
Vietnam	57999	sbh	18,127	General Statistics Office (Viet Nam), United Nations Children's Fund (UNICEF). Vietnam Multiple Indicator Cluster Survey 2010-2011. New York, United States: United Nations Children's Fund (UNICEF).
Vietnam	152735	cbh	15,479	General Statistics Office (Viet Nam), United Nations Children's Fund (UNICEF). Vietnam Multiple Indicator Cluster Survey 2013-2014. New York, United States: United Nations Children's Fund (UNICEF), 2015.
Yemen	13795	sbh	54,378	Central Statistical Organization (Yemen), League of Arab States, Ministry of Public Health and Population (Yemen), Pan Arab Project for Family Health (PAPFAM). Yemen Family Health Survey 2003.
Yemen	13816	cbh	17,213	Ministry of Health (Yemen) and United Nations Children's Fund (UNICEF). Yemen Multiple Indicator Cluster Survey 2006. New York, United States: United Nations Children's Fund (UNICEF).
Zambia	151325	sbh	602,546	Central Statistical Office (Zambia), Minnesota Population Center. Zambia Census 2000 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. Minneapolis: University of Minnesota.
Zambia	151326	sbh	768,988	Central Statistical Office (Zambia), Minnesota Population Center. Zambia Census 2010 from the Integrated Public Use Microdata Series, International: [Machine-readable database]. Minneapolis: University of Minnesota.
Zimbabwe	35493	cbh	23,716	Central Statistical Office (Zimbabwe). Zimbabwe Multiple Indicator Monitoring Survey 2009. New York, United States: United Nations Children's Fund (UNICEF).
Zimbabwe	152720	cbh	32,285	United Nations Children's Fund (UNICEF), Zimbabwe National Statistics Agency. Zimbabwe Multiple Indicator Cluster Survey 2014. New York, United States: United Nations Children's Fund (UNICEF), 2015.

S3 Table:

Overall out of sample predictive validity metrics for each age bin and mean direct estimates of q_a across all first administrative unit years.

$\bar{q}$= average estimated mortality probability; ME = mean error; SDE = standard deviation of the errors; MRE = median relative error; MAPE = median absolute percentage error; R^2= coefficient of determination.

Age-bin	$\bar{q}_a$	*ME*	*SDE*	*MRE*	*MAPE*	R^2
NN	0.030	0.002	0.010	1.05	20.0	0.53
PNN1	0.017	0.000	0.007	1.05	26.5	0.62
PNN2	0.015	0.000	0.007	1.04	29.0	0.66
1yr	0.015	0.000	0.008	1.06	30.0	0.72
2yr	0.011	0.000	0.006	0.98	28.7	0.73
3yr	0.007	0.000	0.005	0.94	31.5	0.61
4yr	0.004	0.000	0.004	0.92	38.0	0.43
5q0	0.094	0.002	0.017	1.04	11.8	0.91

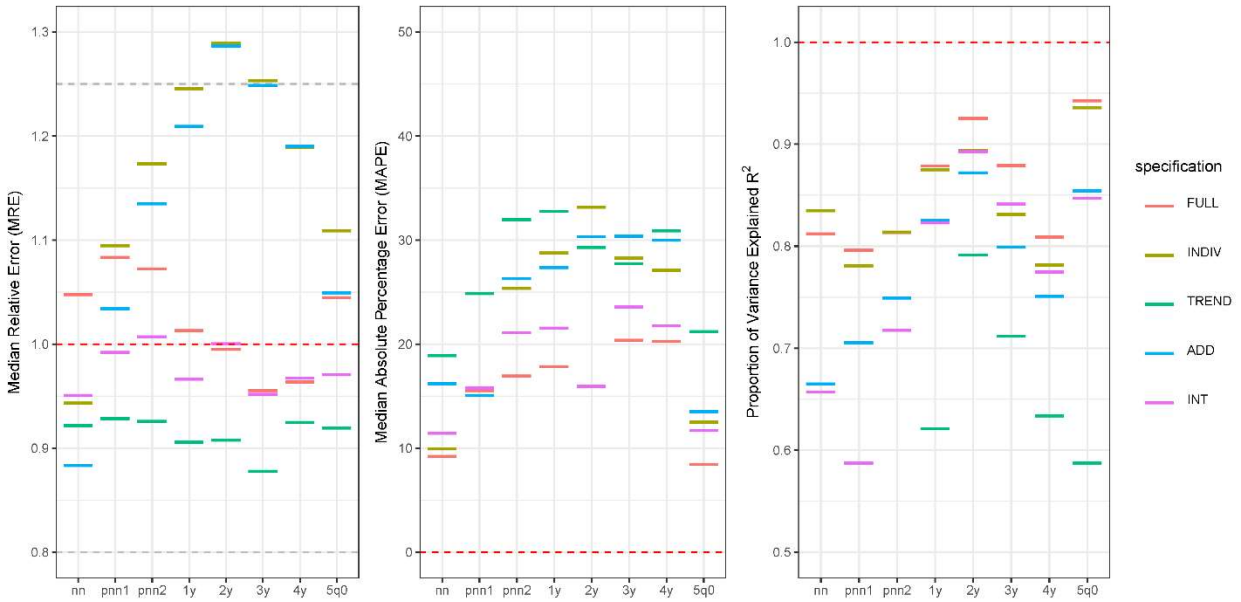
S1 Fig:

We used our validation framework to compare several model specifications. We tested and compared the following models: (FULL) The full model specification described in the methods section. (INT) All covariates from the full model specified with no smooths, variable interactions in smooths replaced with interaction terms. (ADD) All covariates from the full model, but specified with no smooths or interactions, all variables included additively and untransformed. (TREND) Only the year-SDI smooth included, with no individual covariates. (INDIV) Only the individual-level covariates used, with no year-SDI trend included.

This figure shows metrics of predictive validity for these different specifications. The full model does reliably best across all metrics, with only a few exceptions. Notably, full specification is outperformed in bias (MRE) for several age groups by the interaction only model. The full model generally outperforms all other models. These findings indicate that the predictive gains from using the more flexible (yet complex) GAM approach was appropriate here. The TREND specification does reliably worst across all metrics, indicating that there was value in including individual level time-varying covariates.

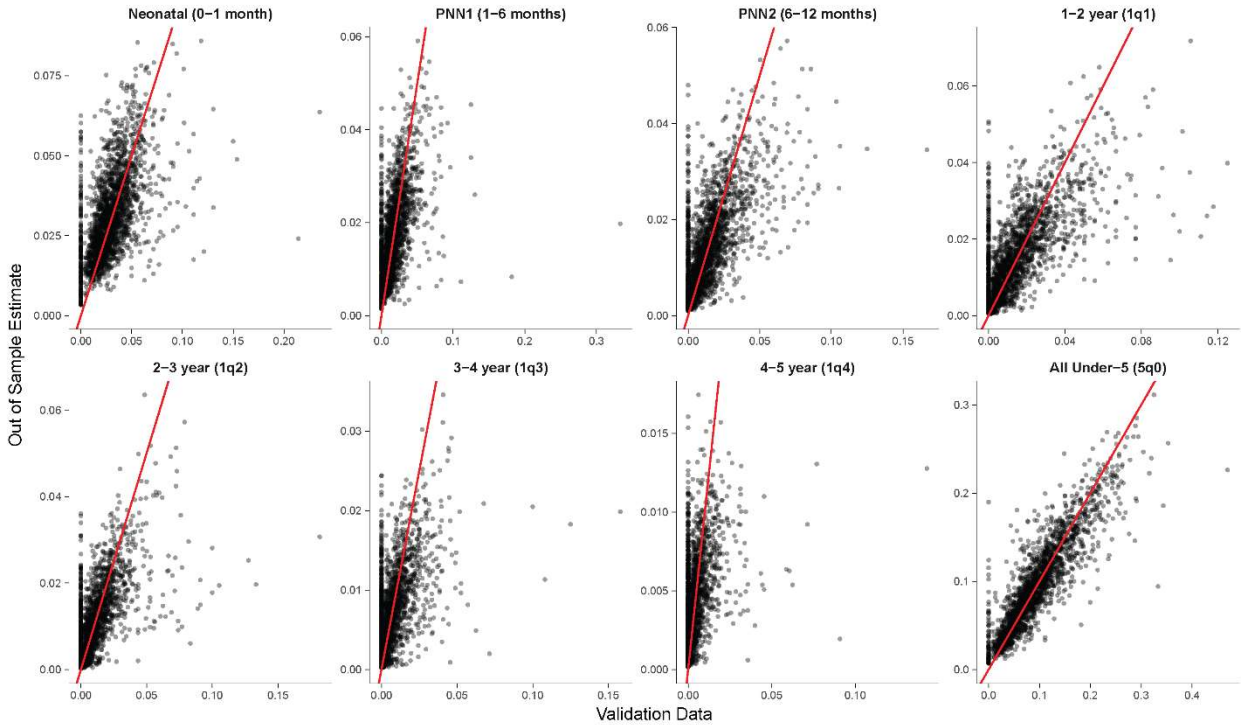
From left: MRE, MAPE, and R^2. Dashed red lines are the target values for perfect fit for each metric. Note that if metrics for certain specifications are missing, they are outside of the plotted range and can be considered very poorly performing.

Abbreviations: GAM = generalized additive model; MAPE = median absolute percentage error; MRE = median relative error



S2 Fig:

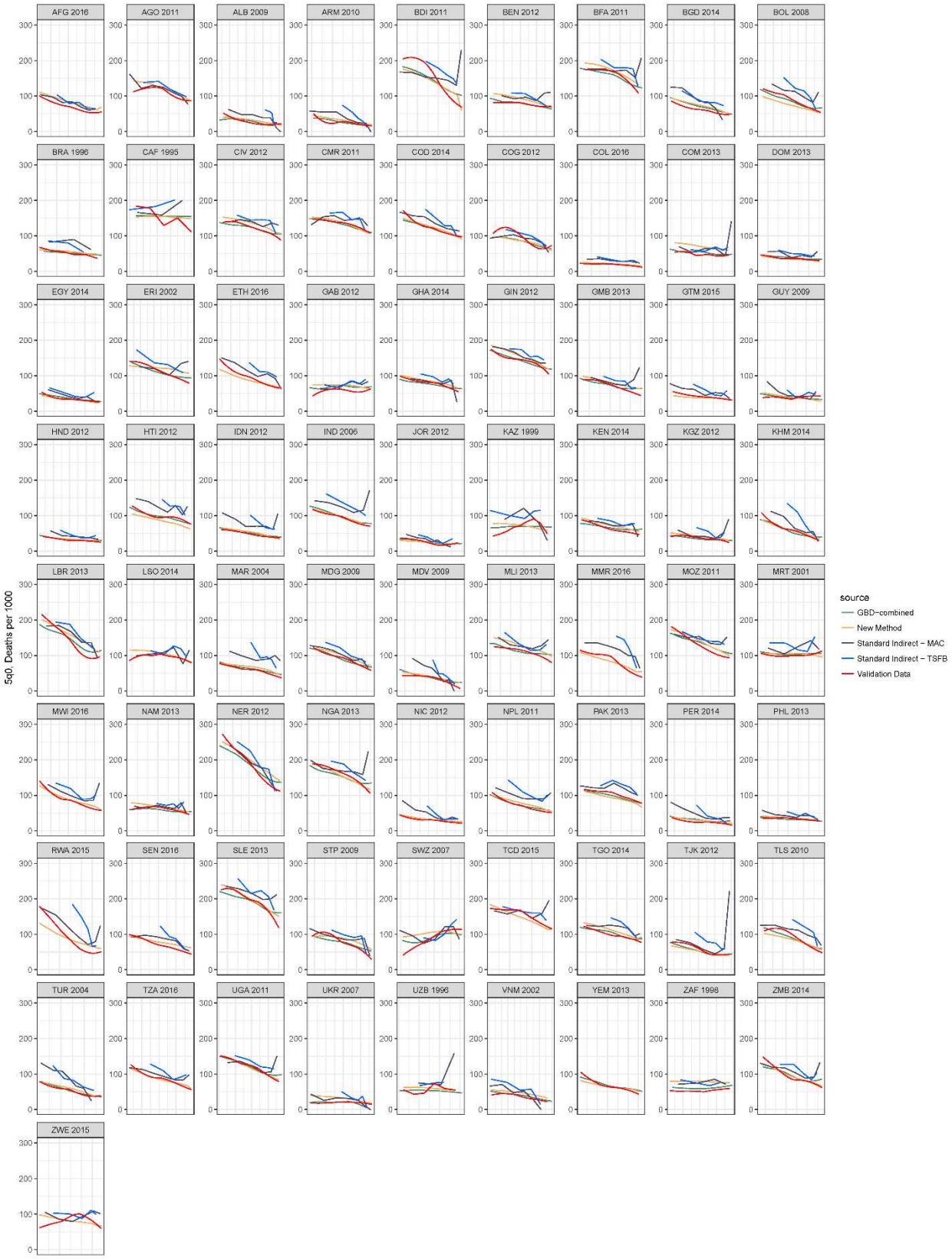
Out of sample predictions of mortality probability compared against loess-smoothed validation data at the first-administrative level. Each point represents an admin1-age mortality estimate ($q_{a,adm1,yr}$) for each held-out survey. Red line indicates unity.



S3 Fig:

This figure shows estimated out of sample $5q0$ trends for the 15 years preceding each survey in the test data, along with trends estimated via the GBD-combined and MAC and TSFB standard indirect methods. 60 of the surveys had GBD-combined estimates available from the GBD mortality database. Trends begin at the year of the survey and go back 15 years or until 1990. Red lines indicate validation data (smoothed with loess, $\alpha = 0.85$).

Abbreviations: GBD = Global Burden of Disease Study; MAC = maternal age cohort; TSFB = time since first birth.


AFG 2016
AGO 2011
ALB 2009
ARM 2010
BDI 2011
BEN 2012
BFA 2011
BGD 2014
BOL 2008
BRA 1996
CAF 1995
CIV 2012
CMR 2011
COD 2014
COG 2012
COL 2016
COM 2013
DOM 2013
EGY 2014
ERI 2002
ETH 2016
GAB 2012
GHA 2014
GIN 2012
GMB 2013
GTM 2015
GUY 2009
HND 2012
HTI 2012
IDN 2012
IND 2006
JOR 2012
KAZ 1999
KEN 2014
KGZ 2012
KHM 2014
LBR 2013
LSO 2014
MAR 2004
MDG 2009
MDV 2009
MLI 2013
MMR 2016
MOZ 2011
MRT 2001
MWI 2016
NAM 2013
NER 2012
NGA 2013
NIC 2012
NPL 2011
PAK 2013
PER 2014
PHL 2013
RWA 2015
SEN 2016
SLE 2013
STP 2009
SWZ 2007
TCD 2015
TGO 2014
TJK 2012
TLS 2010
TUR 2004
TZA 2016
UGA 2011
UKR 2007
UZB 1996
VNM 2002
YEM 2013
ZAF 1998
ZMB 2014
ZWE 2015
5q0, Deaths per 1000
0
100
200
300
source
GBD−combined
New Method
Standard Indirect − MAC
Standard Indirect − TSFB
Validation Data

S4 Fig:

The figures (collectively S4 Fig) show estimated out of sample age-specific mortality trends with uncertainty compared to validation data for each country. Surveys used for out of sample validation are labeled with an 'X' in Supplementary Table 1.

https://doi.org/10.6084/m9.figshare.7163300.v1

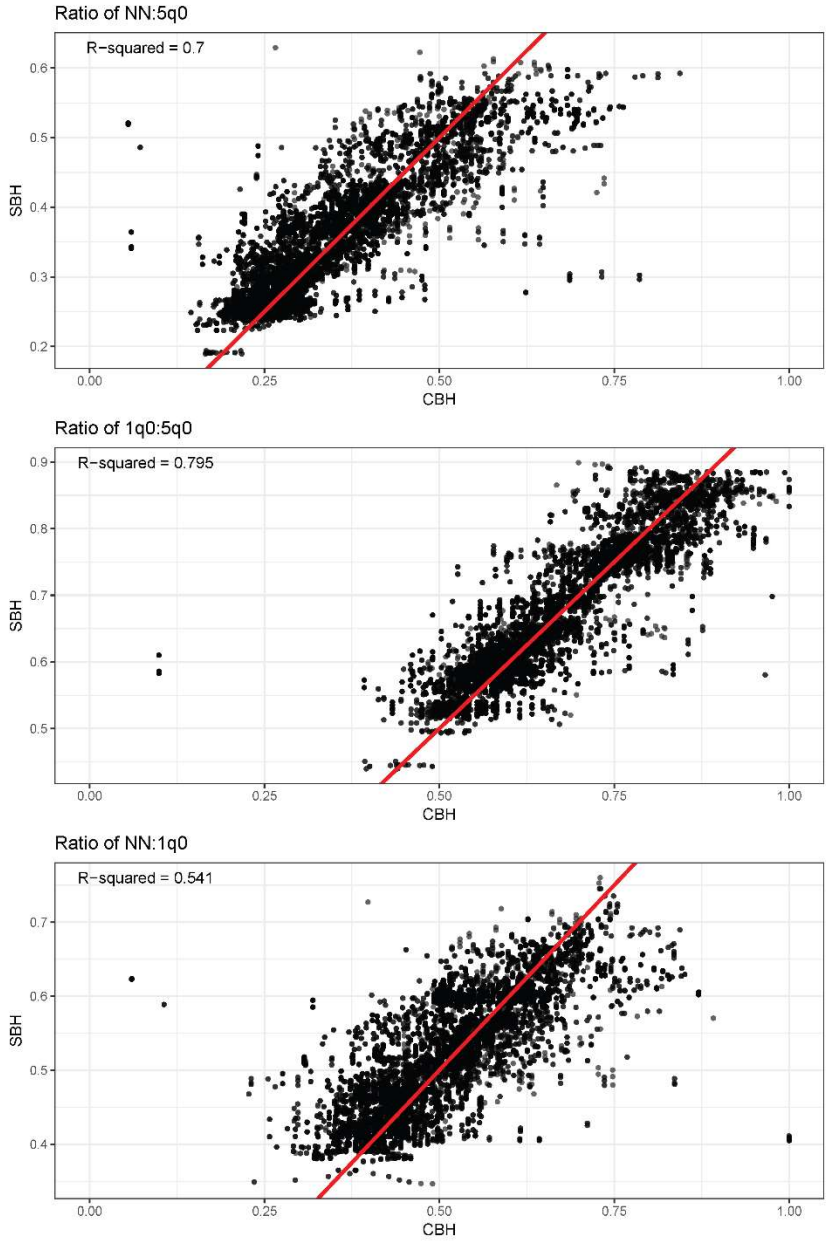
S5 Fig:

The figures (collectively S5 Fig) show trends for each country from the external validation. Trends estimated using the new indirect method for SBH-only data are in blue, direct estimates from CBH surveys are in red. Citations for censuses and surveys used for external validation additional to the training and testing DHS data are listed in S2 Table. Abbreviations: CBH = complete birth history; DHS = Demographic and Health Surveys; SBH = summary birth history.

https://doi.org/10.6084/m9.figshare.7163321.v1

S6 Fig:

For a test of consistency within the external validation, we looked at the overall ratios of mortality rates across age bins of interest. In this figure, we show scatterplots of the relationships between neonatal and under-5 mortality, infant and udner-5 mortality, and neonatal and infant mortality across country-year data pairs from contemporaneous SBH and CBH derived estimates. High overall agreement in these ratios adds further support to the predictive validity of this method. Each point is a contemporaneous country-year estimate. Red line indicates unity. Abbreviations: CBH = complete birth history; SBH = summary birth history.



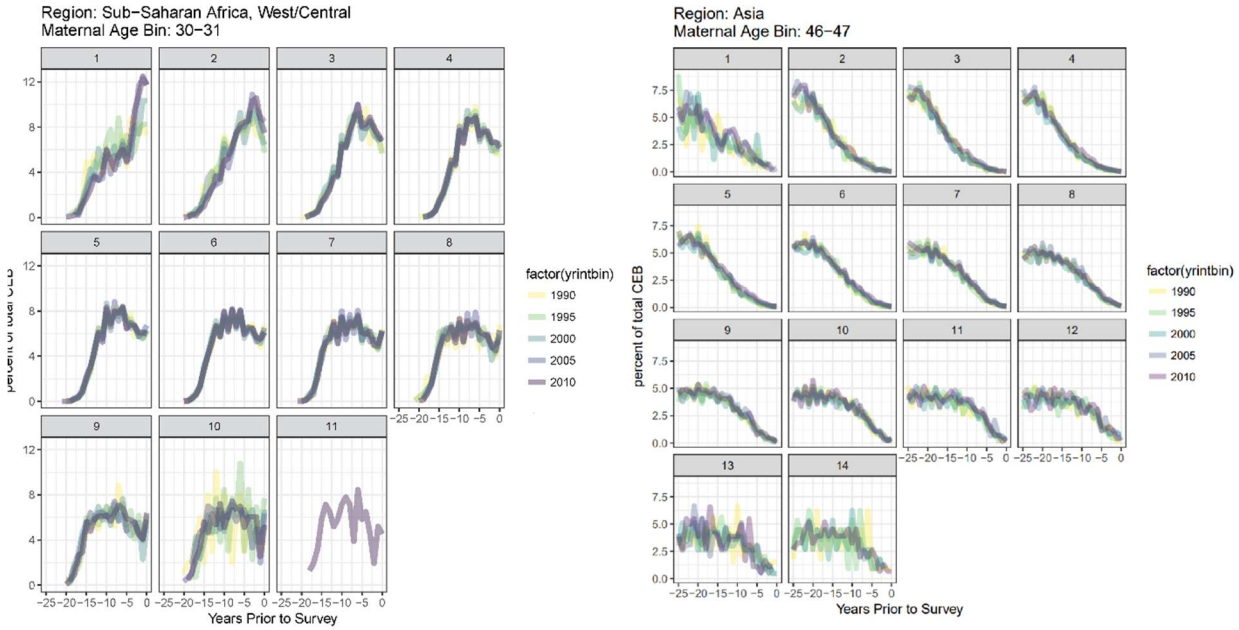
S7 Fig:

Using probability of birth distributions (POB) from the GBD maternal age period (MAP) method has the limitation of being stationary in time. In other words, birth histories are pooled for all mothers from a region with the same CEB, irrespective of year of birth. If distributions changed significantly over time, this could present concern for this analysis. Using the DHS database, we reproduced POB distributions by region, maternal age, CEB, and year of survey (five-year increments).

This figure shows two examples of empirical POB distributions made to explore changes in POB over time. Overall, we found very little change in POB distributions over time for most maternal age groups. The most change has occurred for low-fertility (1-2 CEB) older women, as the distributions seem to indicate there has been a trend toward delayed fertility in this group in recent years. Overall, this group

represents a very small proportion of the mortality exposure in the data. Noisy distributions in high-fertility women are due to low sample sizes. (a) Sub-Saharan Africa West/Central Region plots for women ages 30-31. (b) Asia region for women ages 46-47. A full set of distributions for all maternal ages and regions can be made available by the authors upon request.

Abbreviations: CEB = children ever born; DHS = Demographic and Health Surveys; GBD = Global Burden of Disease Study; POB = probability of birth.



S1 Text:

A brief review of previous approaches to indirect methods for child mortality estimation using summary birth history data:

Several popular methods have been developed to indirectly estimate $5q0$ and to locate this mortality risk in time. These methods are referenced in detail in the main manuscript and as such we provide a brief review of them here.

William Brass[1] laid the foundation for the first method to convert summary data on the total number of children born and died into a more conventional life-table measure, $xq0$, or the probability of death before reaching age x. Several refinements[2,3] have since been adopted since Brass's original work, which are widely used and well summarized in the UN's *Manual X: Indirect Techniques for Demographic Estimation*[4] and by Preston, Heuveline, and Guillot.[5] We refer to this as the standard indirect method.

Data requirements for the standard indirect method are the counts of number of children ever born (CEB) and the number of children died (CD) to mothers in 5-year age groups, beginning with 15-19 year

olds. The ratio of CD to CEB for each maternal age group ($\frac{CD_i}{CEB_i}$) is converted into a child-age (x) and period (t) specific mortality probability (q) using the following two equations:

$$xq0 = (a_i + b_i \frac{P_1}{P_2} + c_i \frac{P_2}{P_3}) \frac{CD_i}{CEB_i}$$

$$t(x) = d_i + e_i \frac{P_1}{P_2} + f_i \frac{P_2}{P_3}$$

where a_i through f_i are maternal age cohort specific coefficients estimated via simulation, and P_i are the average CEB for each 5-year maternal age group. Each equation is used for each maternal age cohort, and each maternal age cohort produces estimates of cumulative mortality since birth, $xq0$, for a child age x which is best identified for that age group of women. These estimates are then localized to a reference period $t(x)$, representing the year in which the estimate of $xq0$ most reliably refers. Estimates of mortality probabilities at specific ages $xq0$ are then converted to $5q0$ using model life tables. By locating risk from each maternal age cohort to one period in time, estimates of $5q0$ for the most recent periods come from mothers in the youngest age groups. These often do not represent overall mortality in the population because mortality to children of younger mothers is often higher.

The fertility ratios $\frac{P_1}{P_2}$ and $\frac{P_2}{P_3}$ used in the standard indirect method serve as an index for the earliness of fertility in the population. Fertility is critical to making accurate indirect estimates because it determines the amount of mortality risk (in time) children born to mothers of certain ages are exposed to. This is why the fertility ratios are multiplied by $\frac{CD_i}{CE_i}$ in the first equation. By using survey-specific fertility ratios, however, the standard indirect method explicitly assumes that this period measure of fertility represents the fertility experience of all maternal cohorts in the data.

In 2010, Rajaratnam and colleagues[6] proposed a suite of new methods for indirect estimation using summary birth histories. A major advancement brought by their approach was to fit empirical models directly from data rather than relying on simulations. The authors used data from 166 DHS surveys, where $5q0$ could be estimated directly from CBH, and where aggregate measures of CD_i and CEB_i were also available for the same mothers. The new methods could broadly be categorized on the basis of estimation - either by maternal age cohort (MAC) or maternal age period (MAP). Variants based on time since first birth, rather than maternal age, were also tested, as this information is also available in some censuses.

Following the standard indirect method, the MAC method is also based on aggregate SBH information from 5-year age cohorts of women surveyed to localize $5q0$ estimates in time for each age group. Similar to the standard indirect method, the MAC method utilizes two models, one to estimate q and another to estimate reference time. The equations differ from the standard indirect method in a number of key ways: $\frac{CD_i}{CE_i}$ is logit transformed, CEB_i is also included as a covariate, fertility ratios are included but not interacted with $\frac{CD_i}{CE_i}$, and country random intercepts are included. Models are fit directly on DHS data using logit($5q0$) as the response, thus removing the need to rely on model life

tables. By modeling cohort-specific $5q0$ directly, the MAC method resolves the young mother bias, but still relies on young mothers for more recent $5q0$ estimates and does not make use of information about more recent mortality of children from older mothers.

The maternal age period (MAP) method computes period-derived estimates based on frequency distributions from the same DHS database. Empirical distributions describing the proportion of children born as well as the proportion of children died to mothers in each year preceding the survey are tabulated from CBH data. Separate distributions are produced by maternal age and CEB for each of five global regions. These distributions are then applied to each mother on the basis of her region, age, and CEB in order to project back patterns of mortality risk and fertility likelihood, thus assigning each woman-year prior to the survey an expected number of births and deaths. These expected births and deaths are summed across the survey population by year and regressed on logit($5q0$) for that year. This generally yields coefficients near 1, indicating close agreement between the imputed annual $\frac{CD}{CEB}$ and observed logit($5q0$). The MAP method may over-generalize trends over time, since they are taken at the regional level. As such, individual and survey-level variation may not be well captured.

They then use loess smoothing to combine MAP, MAC, and their time since first birth variants to produce a combined result. Estimates resulting from this method are used for the Global Burden of Disease Study. We refer to this as the GBD-combined method. Using smoothed direct estimates from their DHS database as validation, they generally find large predictive improvements over the standard indirect approach, which has since been further verified by others.[7]

More recently Golding and Burstein, et al.[8] developed a method to approximate age-specific binomial samples for their mapping of child mortality in Africa. They again used a large training set of data where both CBH and SBH were available. Their method involved fitting two models, one to estimate a child-age and period-specific mortality probability, q, and another for the expected number of child months lived in the surveyed population for a specific child-age and period. Time lag and age indicator variables were used and interacted with logit($\frac{CD}{CEB}$) instead of fitting a separate reference time model in order to localize mortality risk in time, and variables for fertility ratios, year, and maternal age compositions were also included, along with country random effects. Reasonable predictive validity was achieved. Again, this method may also over generalize across time and space, since it relies on lag indicators and global coefficients for the other covariates. Furthermore, the model was fit and predicted on aggregate, rather than individual data.

Several other approaches to indirect estimation have also been described. For example, Kenneth Hill and colleagues [9,10] have proposed the cohort change and birth history imputation methods. The cohort change approach looks at change among true maternal cohorts between two surveys in order to estimate the $5q0$ for the inter-survey period. The birth history imputation approach involves imputing complete birth histories from summary birth histories, by using complete birth histories from a different survey in the same country and matching based on maternal characteristics. The birth history imputation approach has the benefit of producing age-period specific mortality estimates, though in their preliminary investigations, the authors report disappointing validation results.

References

1. Brass W. Methods for Estimating Fertility and Mortality from Limited and Defective Data. North Carolina: Laboratories for Population Statistics, Univ. of NC; 1975.

2. Trussell TJ. A re-estimation of the multiplying factors for the Brass technique for determining childhood survivorship rates. Popul Stud. 1975;29: 97–107. doi:10.1080/00324728.1975.10410187

3. Feeney G. Estimating Infant Mortality Trends from Child Survivorship Data. Popul Stud. 1980;34: 109–128. doi:10.2307/2173698

4. Manual X [Internet]. [cited 5 Feb 2017]. Available: http://www.un.org/esa/population/publications/Manual_X/Manual_X.htm

5. Preston S, Heuveline P, Guillot M. Demography: Measuring and Modeling Population Processes. Wiley; 2000.

6. Rajaratnam JK, Tran LN, Lopez AD, Murray CJL. Measuring Under-Five Mortality: Validation of New Low-Cost Methods. PLOS Med. 2010;7: e1000253. doi:10.1371/journal.pmed.1000253

7. Verhulst A. Child mortality estimation: An assessment of summary birth history methods using microsimulation. Demogr Res. 2016;34: 1075–1128. doi:10.4054/DemRes.2016.34.39

8. Golding N, Burstein R, Longbottom J, Browne AJ, Fullman N, Osgood-Zimmerman A, et al. Mapping under-5 and neonatal mortality in Africa, 2000–15: a baseline analysis for the Sustainable Development Goals. The Lancet. 2017;390: 2171–2182. doi:10.1016/S0140-6736(17)31758-0

9. Hill K, Brady E, Zimmerman L, Montana L, Silva R, Amouzou A. Monitoring Change in Child Mortality through Household Surveys. PLOS ONE. 2015;10: e0137713. doi:10.1371/journal.pone.0137713

10. Brady E, Hill K. Testing survey-based methods for rapid monitoring of child mortality, with implications for summary birth history data. PLOS ONE. 2017;12: e0176366. doi:10.1371/journal.pone.0176366

Chapter 3: Geographic accessibility and utilization of facility-based care in Zambia: a geostatistical analysis

Background

Understanding who accesses healthcare, why, and for what type of services is critical for planning health systems, which ultimately ought to be responsive and available to all potential users.[1,2] Patterns of utilization can be used to identify differences in treatment seeking among key populations or geographic areas. In certain cases, differences in utilization rates can result in diverging health trajectories, thus reinforcing health inequalities.[3] In practice, a better understanding of healthcare utilization can be used to more efficiently target allocation of resources to health facilities,[4] and to improve burden estimation via linked health facility information systems, where for example passive surveillance can miss a large proportion of cases, especially those occurring further away.[5,6]

Numerous studies have investigated the determinants of utilization. The healthcare utilization model developed by Andersen[7] posited that specific factors ultimately determine utilization; these include "predisposing" factors, or characteristics viewed as less mutable, such as race and age; "enabling" factors, or potential drivers and constraints to utilization, like insurance affiliation and travel time to health facilities; and need for healthcare. Empirical studies often aim to quantify the effects of various utilization determinants using multiple regression. In particular, assessments of geographic accessibility – the physical distance or travel time between a user and health services[8] – have consistently shown a negative relationship between greater distance or travel time and healthcare utilization. In many African countries, this relationship exists to some degree across a variety of healthcare services, including deliveries and antenatal care,[9–13] preventative services such as vaccinations,[14,15] HIV services,[16] and treatment for acute illnesses in children.[5,17] A number of studies have also shown that distance to care is a strongly associated with morbidity and mortality outcomes.[5,18–24]

Several studies have attempted to exploit the strong association observed between distance to health services and utilization in order to predict utilization rates across geographic areas. One notable example is a study by Alegana and colleagues,[25] wherein the authors combined household location and healthcare seeking behavior for febrile children from a Namibian Demographic and Health Survey (DHS) with a theoretical surface of travel to health facilities. These data were then used to fit curves representing the decay in utilization over travel time. Using the fitted distance decay[26] function and the theoretical surface of travel times, they predicted out a gridded surface representing the probability attendance in facilities by febrile children. Similar procedures were recently used to produce gridded surfaces of ANC attendance in the Lake endemic zones of Kenya.[4] Other studies used theoretical surfaces of travel times to describe geographic accessibility to essential surgery[27] and emergency care[28] across sub-Saharan Africa, as well as geographic accessibility to primary care in single countries such as Mozambique[29] and South Sudan.[30]

Efforts to quantify the relationship between distance and utilization and then predict utilization rates onto gridded geospatial surfaces have been subject to considerable data and methodological limitations. For instance, researchers have often resorted to using approximate measures of household and

population locations with respect to health facilities. There are several reasons for this. First, most countries do not have a fully geo-referenced census of health facilities.[31,32] Second, when georeferenced data are collected, GPS coordinates of households are often 'jittered' up to ten kilometers to support confidentiality; this is particularly common for publicly available data sets such as the DHS, [33] which is commonly used in this type of analysis. Despite knowledge of this jittering, extant studies have generally treated scrambled GPS coordinates as if the measurements are true and without introduced error.[4,13,25,34,35] Furthermore, geographic accessibility, defined as the distance or travel time to healthcare, is sometimes measured using straight-line distance,[15,21,34] which is likely an over-simplified model of human mobility as it ignores networked paths and roads and barriers to movement.

Studies that model utilization geographically are often univariate, viewing travel time as the only determinant and not accounting for other potentially important drivers or confounders.[25] Where interpolated geospatial covariates are incorporated, uncertainty in those interpolations are not incorporated into the predications.[4,35] Past analyses generally focus on one condition (eg, childhood malaria or suspected malaria), which results in discarding potentially useful data on other conditions.[4,25,35] Furthermore, most studies are limited to subnational areas and thus do not map utilization across an entire country. To date, no analyses have utilized a full probabilistic geostatistical model to account for correlated spatial residuals; instead, they have relied exclusively on one or more covariates. Finally, we are not aware of any studies that report uncertainty in their gridded predictions of utilization, making it impossible assess the utility of these predictions.

Since 1990, Zambia has made marked progress in increasing the coverage of many maternal and child health interventions and improving rates of childhood survival.[36–39] Nonetheless, national levels of under-5 mortality still exceeded 59 per 1,000 livebirths in 2016, wherein malaria and diarrhea accounted for 18% of child deaths and 10% of total national disease burden that year.[38] Health service delivery is largely dominated by the public system: about 90% of patients who seek care do so in government-run facilities.[40] Improving access to primary healthcare is a priority for the Ministry of Health, which as a stated mission "to provide equitable access to cost-effective, quality health services as close to the family as possible."[41] In 2011, the government abolished user fees for all services in primary healthcare facilities, and patients should expect free curative services for malaria and diarrhea at all levels of care. Despite this, potential barriers to healthcare utilization at facilities remain, including transportation cost and perceived poor quality.[40] Previous studies have found associations between utilization with a number of predisposing, enabling, and need factors, including distance to care,[42] educational attainment, and reported health status.[43]

In this study, we sought to overcome limitations in accessibility and utilization mapping using a triangulation of data sources from Zambia in order to answer three main questions: First, what is the status of geographic accessibility to health facilities in Zambia? Next, how does travel time to nearest health facility affect utilization rates? And finally, can we make a reasonable map of utilization rates using a geostatistical model?

Using a census of health facility locations, we first quantified geographic accessibility as travel time to nearest health facility across Zambia at a resolution of 1 x 1-kilometer (km) pixels. Then, based on household survey data with precise locations, we developed a geostatistical model, leveraging individual and household covariates and a flexible non-linear distance decay function, to simultaneously predict the probability of healthcare utilization for a number of different types of illness episodes for each 1 x 1-

km pixel in Zambia. We compared our mapped predictions to those made using a method from the literature. We used the results from our geostatistical model to better understand individual determinants of healthcare utilization, with a particular focus on the effect of geographic accessibility. Finally, we combined our predicted gridded surfaces with similarly resolved estimates of incident cases of diarrhea to approximate unmet need for corresponding healthcare services at the population level.

Data

Utilization

Utilization data came from the 2014 Zambia Household Health Expenditure and Utilization Survey (ZHHEUS), a nationally-representative two-stage household sample survey conducted by the Zambian Central Statistical Office, with support from the Ministry of Health and the Department of Economics at the University of Zambia. The ZHHEUS included 11,944 households and 59,514 individuals surveyed from January to March 2014.[44] Respondents or their proxies answered a number of demographic, economic, and health questions, including those focused on subjective health status and presence of acute or chronic illnesses. Specifically, respondents were asked whether they were ill within the past four weeks, and if so, what type of illness and symptoms they experienced and whether they sought care from a health provider. For this paper, we focus specifically on reported fever, malaria, and diarrhea, but indicators for 18 illnesses plus a free-response were recorded. Since we focus specifically on treatment in facilities, we coded responses claimed to have sought care at pharmacies, religious healers, herbalists, and from community health workers as not seeking care in a health facility. In total, these responses comprised less than 5% care sought by ill people.

We combine reported malaria and fever into one category for “febrile illness”. Given high endemicity in much of Zambia, we can assume that many of these reported malaria/fever episodes were indeed malaria.[45,46] Since the survey asked about a combination diseases and symptoms, it is difficult to assess the degree to which respondents could have known if they had truly had malaria. Henceforth, for brevity, we refer to this measure of fever and/or malaria as febrile illness in this manuscript.

To enable geo-positioning of households and utilization in relation to health facilities, we used exact GPS locations as recorded by ZHHEUS interviewers. All households in the data had GPS data.

Health facility locations

We triangulated multiple health facility data sources in an effort to approach data coverage of all health facilities. Our goal was to produce a complete georeferenced listing of health facilities in the country. First, we used the Ministry of Health’s 2012 *List of Health Facilities in Zambia*[47] as a baseline list (*n* = 1,956). Since this facility list lacked GPS coordinates, we merged this dataset with facility and GPS information provided by the 2005-2006 *Zambia Health Facility Census.*[48] We also supplemented these data with an unpublished list of geo-located facilities provided by the Zambian Central Statistical Office; facility coordinates from the 2011-2012 Access, Bottlenecks, Costs, and Equity (ABCE) facility survey;[49] and publically-available location data from www.zaplaces.com and www.facebook.com for any other facilities without GPS information. Last, if ZHHEUS respondents reported visiting facilities that were not

included in the master list, we appended these facilities to the master list and sought to link corresponding GPS information to those facilities.

This effort resulted in a final list of 2,020 facilities, 85% of which (n = 1,726) could be geo-located. Of the facilities missing GPS coordinates, 55% (n = 155) were marked with certainty as urban clinics. Urban areas have a higher density of facilities and so we expect their missingness to have minimal effects of estimates of travel time to nearest health facility.

Surface of theoretical travel times to health facilities

We constructed a 30 arc second (1x1 km at the equator) resolution gridded surface of theoretical travel times to the nearest health facility, wherein each pixel represented the amount of time it would take to travel from it to the nearest health facility. As a basis for a realistic and validated model of mobility, we used the global friction surface developed by Weiss and colleagues.[50] The friction surface generically quantifies the time it takes to travel across a pixel. Data on road networks were merged with geospatial layers representing land cover and elevation and then scaled by estimated travel times across these different surfaces. With the facility locations as inputs, and using an accumulated cost algorithm based on Dijkstra's algorithm, we calculated the travel time from each pixel in the friction surface to its nearest health facility. The result was a gridded surface of theoretical travel times to health facilities. Finally, we extracted the travel time values at each household location using recorded GPS readings from ZZHEUS.

Other geographic layers

We accessed several geographic data layers, representing incidence rates of diarrhea, under-5 and total population counts, and urban extents for each 1x1 km pixel in Zambia.

Gridded layers for incidence of diarrhea in children under-5 have been produced by the Institute for Health Metrics and Evaluation at the 5x5 km resolution.[51] We used candidate maps[52] representing 1000 draws from the posterior predictive distribution in order to propagate uncertainty in estimates of incidence rates. We disaggregated each pixel-draw to match the 1x1 km resolution of this analysis. Incidence rates were used in combination with gridded population data of under-5s to estimate cases per pixel. The WorldPop project provides population count estimates for children under 5.[53,54]

We also produced a map which classified ‘urban extents’ by matching urban classification in the ZHHEUS data. We extracted data from the Global Human Settlement Layer project[55] and log-transformed total population from WorldPop for each household in the ZHHEUS data and fit a logistic regression on whether the household was considered urban in the sample. The area under the receiver operating characteristic curve was 0.93, indicating a very good fit. A resulting surface representing the probability each pixel is urban was used for prediction from the geostatistical utilization model described in the next section.

Ethical approval

Ethical approval was granted by the University of Washington Human Subjects Division IRB (#51398: "Estimating Spatial Patterns in Health Care Utilization in Zambia").

Methods

Statistical model

We developed a statistical model representing the probability that an individual, given illness in the past four weeks, would seek care from a health facility. With this model, we had the dual goal of making inference on the effects of certain covariates on treatment seeking, as well using it to ultimately predict a surface representing utilization rates. We first subset the ZHHEUS data to the 13,150 (22% of total) respondents who reported having an illness in the past four weeks. The choice to seek treatment at a health facility was then encoded as a binary choice variable Y, with 0 representing not seeking care and 1 representing that the respondent did seek care. We then modeled the choice to seek in-facility treatment for individual i, living in household j, from location k as a Bernoulli random variable with probability p_{ijk}, where p_{ijk} represents the individual probability of seeking healthcare. The logit transformation of p_{ijk} was further modelled a linear combination of individual and household level covariates, and structured and unstructured random effects. The model can be written as:

$$
\begin{gathered}
Y_{ijk} \sim Bernoulli(p_{ijk}) \\
logit(p_{ijk}) = \alpha + f(tt_j) + ill\_F_i\beta_m + ill\_O_i\beta_d + \boldsymbol{X_i\beta_i} + \boldsymbol{X_{ij}\beta_j} + U_j + S_k \\
U_j \sim Normal(0, \sigma_j^2) \\
S_k \sim GP(0, \boldsymbol{K}) \\
f(tt_j) = \sum_{b=1}^{n} \beta_b X_b \\
\beta_b \sim N(\beta_{b-1}, \sigma_b^2)
\end{gathered}
$$

where $f(tt_j)$ is the sum of cubic spline basis functions on travel time to nearest health facility, which comprises the distance decay function. We placed spline knots at 0.02, 0.05, 0.15, 0.31, 0.76 hours, corresponding to the 16.7%, 33.3%, 50%, 66.7% 83.3% quantiles of the data, respectively. We placed additional knots at 1.5 and 2.5 hours in order to increase coverage in more remote locations. A boundary knot, beyond which the function would be forced to be linear, was placed at 4.6 hours, the most remote household observed in the data.

A number of covariates were selected in order to capture the individual and household level predisposing factors, enabling factors, and needs which could lead to treatment seeking. ill_F_i and ill_O_i are binary indicators for whether the respondent reported a febrile illness nor any illness other than diarrhea or a febrile illness, respectively. Reported diarrhea was used as the reference category. $\boldsymbol{X_i}$ are individual-level covariates including an indicator for whether or not the observation was for a child under the age of five, and an indicator for female sex. ZZHEUS did not ask about illness severity, so instead we used an indicator for whether or not the respondent claimed to have more than one illness in the past four weeks. $\boldsymbol{X_j}$ are household-level covariates. We used the log household expenditure per

capita as an indicator of household wealth. As a sensitivity analysis, we also tried an asset index, and found our results were robust to both specifications, and chose to use expenditure to keep our specification similar to a previous analysis of this dataset.[44] We included an indicator for whether or not the head of the household had ever been to school, and finally an indicator for whether the household was in an urban area.

Our model also included two random effects. U_j is a random effect for each household, meant to capture variation not explained by the covariates. We attempted to also include a cluster-level random effect, but found model identifiability issues if both a household and cluster level random effect were kept. S_k is a spatially correlated random effect term, modeled as a Gaussian process with covariance matrix $\boldsymbol{K}$, which was structured using the Matérn covariance function. Sample weights from the survey were used to weight the contribution of each observation in the log-likelihood.

First order random walk priors were used to penalize spline bases in order to impose smoothness.[56,57] Imposing this structure on spline smoothness allowed us to use a larger number of bases functions while avoiding overfitting. The standard deviation of the household random effect, σ_j^2, had a $gamma(1,1)$ prior. Priors for the hyper-parameters of the Matérn covariance function were $\log(\tau) \sim Normal(0,9)$ and $\log(\kappa) \sim Normal(0,9)$. All fixed effects coefficients which were not spline bases had zero mean normal priors with a standard deviation of 10.

The model was fit using Template Model Builder package[58] in R version 3.5.1.[59] Maximum a posteriori inference was implemented using a non-linear optimizer. Within this framework, the spatial effects were fitted using the stochastic partial differential equation (SPDE) approximation[60–62] on a finite elements mesh, with maximum edge lengths set to 0.05 degrees (see Supplementary Figure 1). Uncertainty in predictions was calculated using 10,000 multivariate normal draws from the fitted joint precision matrix across all model parameters.

We used draws of the fitted model parameters to construct gridded prediction surfaces wherein each pixel represented the individual probability of seeking healthcare for either febrile illness or diarrhea in children.[52] Individual level covariate values which were not used in mapping were integrated out using a Monte Carlo approach by drawing randomly with replacement from their observed joint distributions in the data.[63] Certain covariates were set to specific values for prediction. For example, to predict a map of utilization given childhood diarrhea, we set the under-5 indicators to one, and the febrile illness and other illness indicators to zeros. Furthermore, since we had a map of urbanicity with values for each location, those mapped values were used for prediction.

Model comparison

In order to compare our model with previous methods, we applied the method used by Alegana and colleagues[25] to the ZHHEUS data. In their paper, the authors used a theoretical travel time surface as a univariate predictor of utilization. The distance decay curve was assumed to take the following form:

$$Prob(Utilization) = \frac{C}{1 + e^{\frac{A - travel\ time}{B}}}$$

where A, B, and C are parameters to be fitted. Using data subsets for children under-5 with either febrile illness or diarrhea, we fitted two curves based on this method by minimizing the sum of squared errors in each sample. We compared predictions against those from our model using receiver operating characteristic (ROC) curves.

Results

Table 1 reports basic descriptive statistics for the respondents who reported an illness in past four weeks in the ZZHUES data. Of the 59,514 people surveyed, 13,150 (22%) reported an illness, and are included in Table 1. Of those reporting an illness 7,566 (58%) sought care from a health facility during that time. Of the 2164 children under 5 who were ill in the preceding four weeks, 1387 (64%) had a reported case of fever or malaria and 357 (17%) had a reported case of diarrhea. In total those with febrile illness and diarrhea collectively comprised 75% of all reported childhood illnesses. Of those children with febrile illness 925 (66%) sought care at a facility, while 224 (63%) of those with diarrhea sought care.

Variable Name	Mean	*N or (SD)*
Sought care from health facility	57.5%	*7566*
Travel time to nearest health facility	0.4	*(0.6)*
Illness: Diarrhea	7.6%	*1002*
Illness: Febrile	52.9%	*6962*
Ln(HH expenditure per capita)	4.1	*(1.6)*
Younger than 5	16.5%	*2164*
Urban residence	30.9%	*4063*
Female	54.3%	*7145*
>1 reported illness	35.4%	*4651*
Head of HH any schooling	88.9%	*11,695*
Unique households		*6811*

Table 11: Descriptive table of variables used in the regression model. Data was subset on those who had an illness (13,150). For binary variables, we report the percentage and number reporting. For continuous variables (ln(Household expenditure per capita) and Travel time to nearest facility), we report the mean and standard deviation in the sample. There were no missing values for any of the variables used.

Geographic accessibility

Figure 1A shows the 1 x 1-km gridded surface of travel times to Zambian facilities, overlaid with health facility locations. Travel times in Zambia ranged from 0 minutes (in grid cells with health facilities), up to 13.5 hours. As expected, travel time across Zambia was unevenly distributed; for instance, travel time was lowest along roads, since travel speeds were assumed to be much faster on roads. To better understand how travel times are distributed among the population, we extracted travel times at all ZHHEUS survey locations, giving us estimates of travel time to healthcare from each household in the survey data. We also overlaid the same travel time surface with matching pixel-level estimates of total population from WorldPop. Figure 1B shows the cumulative population distributions across travel time

as measured by both of these population data sources. We found that 97.7% of the weighted ZHHEUS sample lived within 2 hours of the health facility. This differed significantly from estimates produced using pixel-level population data from WorldPop, wherein 88.7% of the population lived within 2 hours of a facility and 97.7% lived within 4.5 hours from a health facility. Using the WorldPop surface, we calculated the average travel time to the nearest health facility throughout Zambia to be 45 minutes, while using the ZZHEUS sample, the average estimated time was 19 minutes.

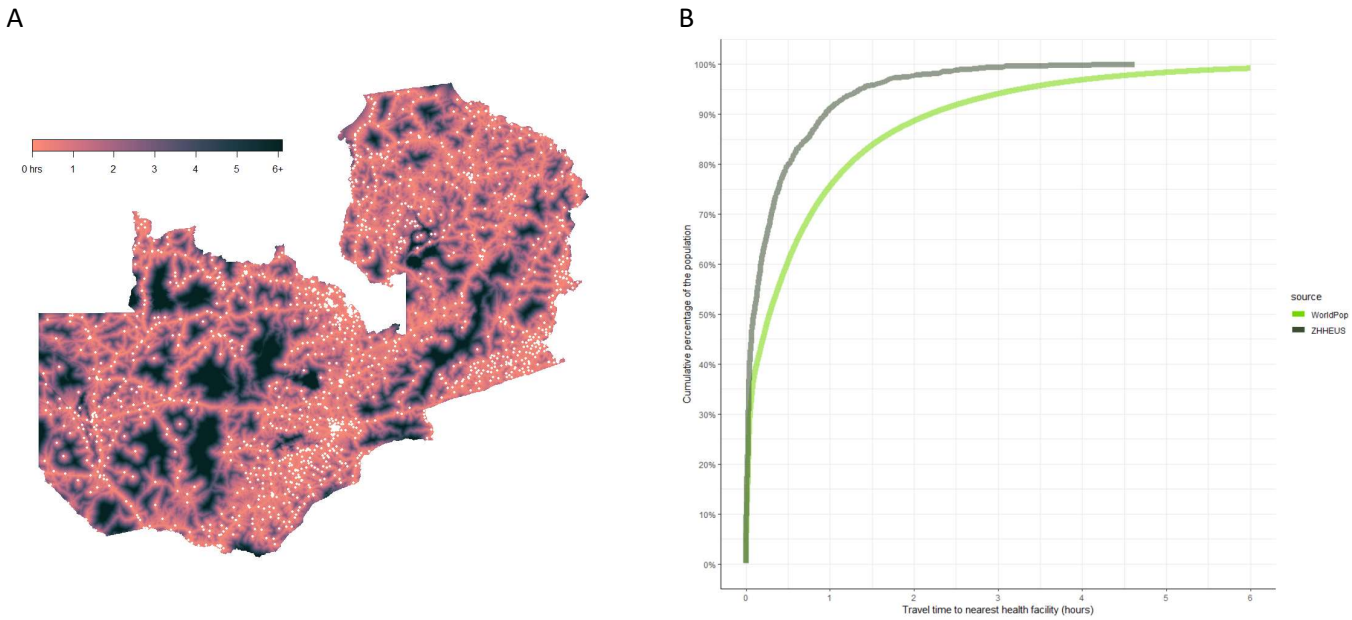


Figure 1. The map in (A) represents the gridded surface of theoretical travel times to the nearest health facilities, with all georeferenced health facilities indicated with white dots. (B) shows the cumulative population by distance to nearest health facility. The black line represents the cumulative weighted population from ZHHEUS, while the green line represents the cumulative weighted population from the WorldPop gridded population dataset.

Determinants of treatment seeking

Table 2 presents results from the geostatistical logistic regression. Results were largely in line to those reported by a previous analysis of this dataset,[44] with one key difference. All else equal, the odds of seeking care for febrile illness was 59% (95% uncertainty interval: 30%-95%) higher than that for diarrhea. The odds of seeking care were much higher for children under-5 (OR = 1.67, 1.48-1.89) relative to anyone older than 5, and for females (OR = 1.10, 1.01-1.21) relative to males. Having a household head with any education was associated with a 42% (22%-64%) higher odds of treatment seeking. Having multiple ailments also significantly increased the odds of seeking care. Furthermore, household expenditure had little measurable effect on treatment seeking (OR=1.01, 0.97-1.04). Interestingly, urban residence was significantly negatively associated with care seeking relative to rural residence (OR = 0.77, 0.63-0.93). In a sensitivity model run without travel time, the effect was positive. This indicates that urban residence could stand as a proxy for geographic accessibility, but that controlling for accessibility reveals a possible structural negative association. Variance inflation factor tests did not indicate that this was a spurious result due to multicollinearity (VIF < 2).

The fitted spatial hyper-parameters guiding the Matérn covariance function of the Gaussian process indicate that, after accounting for these covariates, some local spatial correlation remained; that is, 90% of the residual spatial correlation was diminished within 0.74 (0.50-1.07) decimal degrees. Supplementary figure 4 provides density plots of the priors and posteriors of the model hyper-parameters.

Fixed effects	***OR***	***95% UI***	
intercept	1.05	0.72	1.52
Travel Time		*< sum of bases functions>*	
Diarrhea		---	
Febrile Illness	1.59	1.30	1.95
Other Illness	1.23	1.01	1.51
ln(HH expenditure/person)	1.01	0.97	1.04
Older than 5		---	
5 or younger	1.67	1.47	1.89
Rural residence		---	
Urban residence	0.77	0.63	0.93
Male		---	
Female	1.10	1.01	1.21
Head of HH no schooling		---	
Head of HH some schooling	1.42	1.22	1.64
1 reported illness		---	
>1 reported illness	1.31	1.19	1.44
Hyper-parameters	***median***	***95% UI***	
SD of P-spline Random Walk	0.20	0.10	0.40
SD of HH Random Effect	0.45	0.35	0.57
Matern Range (Degrees)	0.74	0.50	1.07

Table 2. Fitted parameter estimates from the geospatial regression model. Fixed effects coefficients have been transformed into odds ratios. Covariates in italics are reference levels and thus do not have coefficient estimates.

Effect of travel time on utilization

Figure 2 shows the predicted marginal distance decay curves from the fitted smoothing splines representing the effects of travel time on utilization. We held other model covariates at average values such that the decay curves represented the expected response on treatment for the hypothetical average person in our data. We held illness indicators constant such that we could produce marginal smooths for diarrhea and febrile illness independently.

The probability of utilization declined quickly between 0 and 60 minutes from a health facility, and continued to decline less rapidly thereafter. This quick initial decline in utilization goes contrary to distance decay curves published in the literature, which do not typically drop off quickly.[21,25,35,64] Due to data sparsity of households located farther than 3 hours from a facility, uncertainty increased substantially beyond this threshold, to the extent that distance decay curves for febrile illness and diarrhea largely overlapped after 3 hours' distance.

The inset plot shows a close of the first five hours from the nearest health facility. Points indicate empirical estimates of utilization from the ZHHEUS data for febrile illness and diarrhea in red and blue, respectively across bins of travel time. This indicates that this quick drop-off is supported by the data. The empirical estimates also indicate that decay functions for the two conditions of interest do indeed appear to be proportional to each other, thus supporting the modelling choice of pooling data across conditions and using intercept shifts, rather than independent smooths, to represent different utilization rates across them.

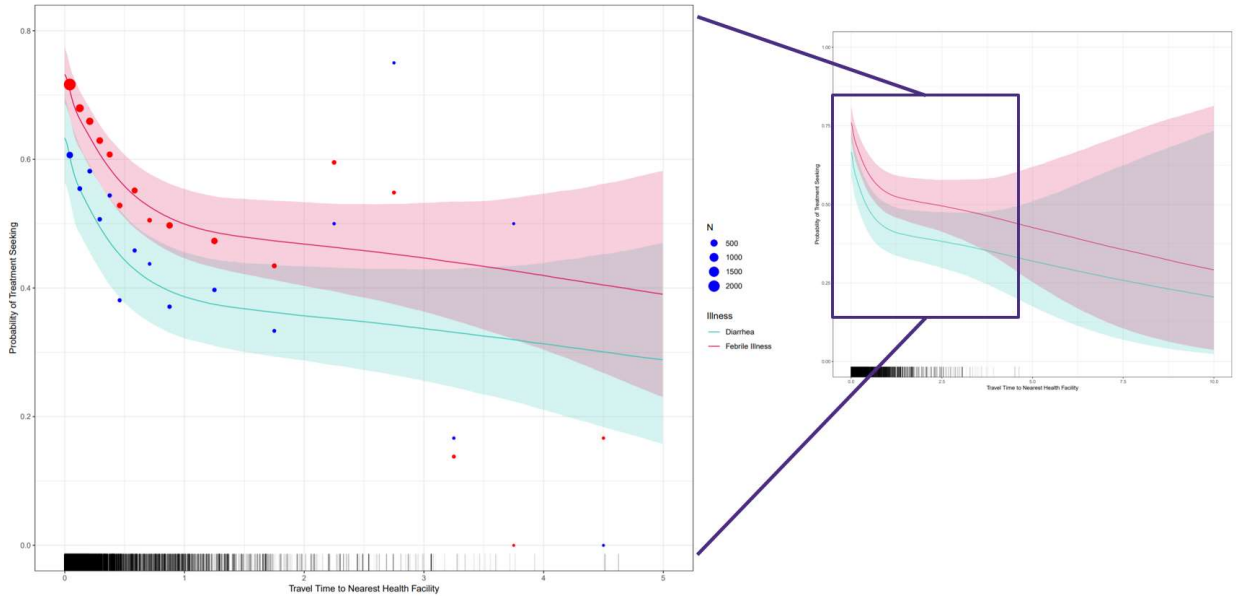


Figure 2. Marginal smooth curves representing the distance decay curves, or the effect of increasing travel time on probability of seeking care. Separate curves are drawn for diarrhea and febrile illness. Other covariates were set at their means. Shaded areas represent the 95% uncertainty interval. Black spikes across the x-axis represent travel time values of the input data.

Utilization mapping

Figure 3 shows the predicted utilization surfaces with uncertainty for febrile illness. The values of each grid cell represent the expected probability that a given individual residing within a grid cell will seek care. The map of utilization given diarrhea look very similar to febrile illness, except with lower utilization due to the intercept shift in the model (see Supplementary Figure 2). Higher utilization was generally found in places with lower travel times, as the influence of travel times is evident in this map. Though the uncertainty interval around prediction utilization rate was generally narrower in more populated areas, uncertainty remained fairly high across the country. Higher confidence in predictions in populated areas was due to the combined effects of the increased certainty in the fitted Gaussian process random effects where data were available and because remote and less populated areas

experienced high travel time values which were not available in the ZHHEUS data that the model was fit on. For instance, the average width of predicted grid cell uncertainty intervals across Zambia was 0.49, but this value narrowed to 0.19 when population was taken into account in the averaging.

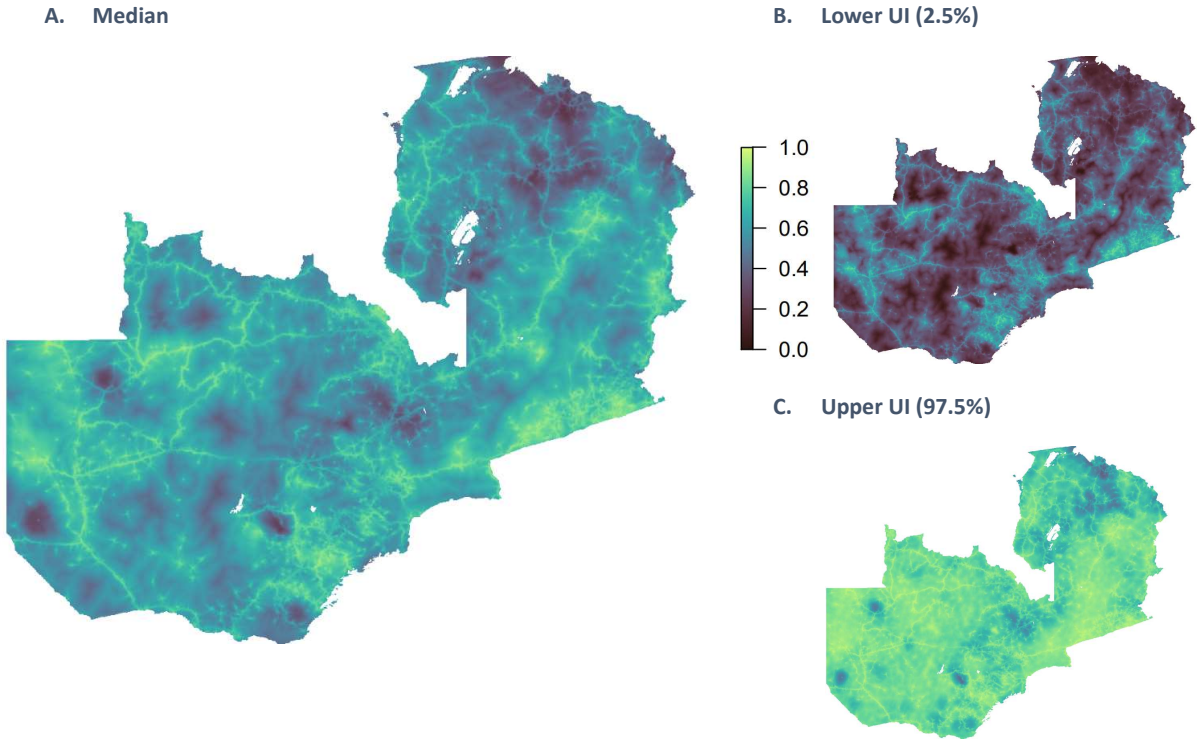


Figure 3. Maps of median prediction and 95% uncertainty interval for probability of utilization in children with febrile illnesses. For similar maps of diarrhea utilization, see Supplementary Figure 2.

Population metrics

To translate what these utilization measures could mean at the population level, we multiplied the gridded estimates of diarrhea utilization rates with incidence rates and with population counts for children under-5. This calculation produced gridded surfaces of numbers of diarrhea episodes that did not seek healthcare. At the national level, the facility-based utilization rate for diarrhea episodes in children under-5 was 51% (95%UI: 40%-60%). The estimated number of diarrhea cases seen at a health

facility was 4.1 million (3.2 million – 5.1 million), while the number not seen at a health facility was estimated at 3.3 million (2.5 million – 4.2 million). Table 3 shows these numbers aggregated to provincial level. While uncertainty is high, there is some discernable subnational variation in utilization rates. For example, in Northern province, for each diarrhea case seen at a health facility 1.6 [1.5-1.7] do not seek treatment at a facility. While in Eastern province, for each case that does not seek treatment at a health facility, 1.9 (1.8-2.1) do go to a health facility, indicating a three-fold difference in utilization between the two provinces.

	Cases Utilized		Cases Not Utilized	
Province	**Median**	**95% UI**	**Median**	**95% UI**
Central	361,958	[264,858 - 474,133]	359,057	[271,272 - 463,856]
Copperbelt	587,154	[443,751 - 763,061]	337,183	[240,064 - 468,176]
Eastern	701,524	[549,535 - 887,012]	367,031	[263,826 - 499,375]
Luapula	301,870	[218,785 - 402,398]	304,210	[224,970 - 403,214]
Lusaka	653,520	[490,128 - 855,856]	485,485	[351,624 - 656,475]
Muchinga	238,682	[175,600 - 313,668]	219,314	[166,119 - 284,354]
North-Western	203,877	[145,914 - 268,963]	180,947	[129,324 - 243,228]
Northern	273,801	[194,050 - 367,971]	429,902	[335,525 - 557,193]
Southern	509,452	[384,247 - 651,528]	374,497	[267,292 - 505,981]
Western	267,109	[199,217 - 347,657]	185,846	[133,275 - 251,485]

Table 12: Estimated number of diarrhea cases which did and did not utilize healthcare across the 10 Zambian provinces in 2015.

Comparison with prior method for utilization mapping

In our replication of the method used by Alegana and colleagues,[25] we estimated the parameters of the logistic curve A, B, and C to be 6.9, -1.8, and 0.8 and 5.9, -1.7, and 0.8 for febrile illnessand diarrhea, respectively. Similar to the modeled smoothing splines, these curves declined immediatedly before starting to flatten out (see Supplementary Figure 3). The distance decay curves modelled from ZHHEUS data fundamentally differed from most other published curves, which typically decay slowly at first before dropping off quickly. Figure 4 shows that comparison of ROC curves for febrile illness and diarrhea between our model and the method described in Alegana and colleagues. Area under the curve (AUC) for the median estimate from our model was 0.74 febrile illness and 0.73 for diarrhea, while AUCs using the approach by Alegana and colleagues were 0.57 and 0.60. ROC curves shown in figure 4 from our model were plotted for each draw from the posterior. Despite the uncertainty in our estimates, there were no draws from our model with lower AUCs than those produced using Alegana and colleagues' approach.

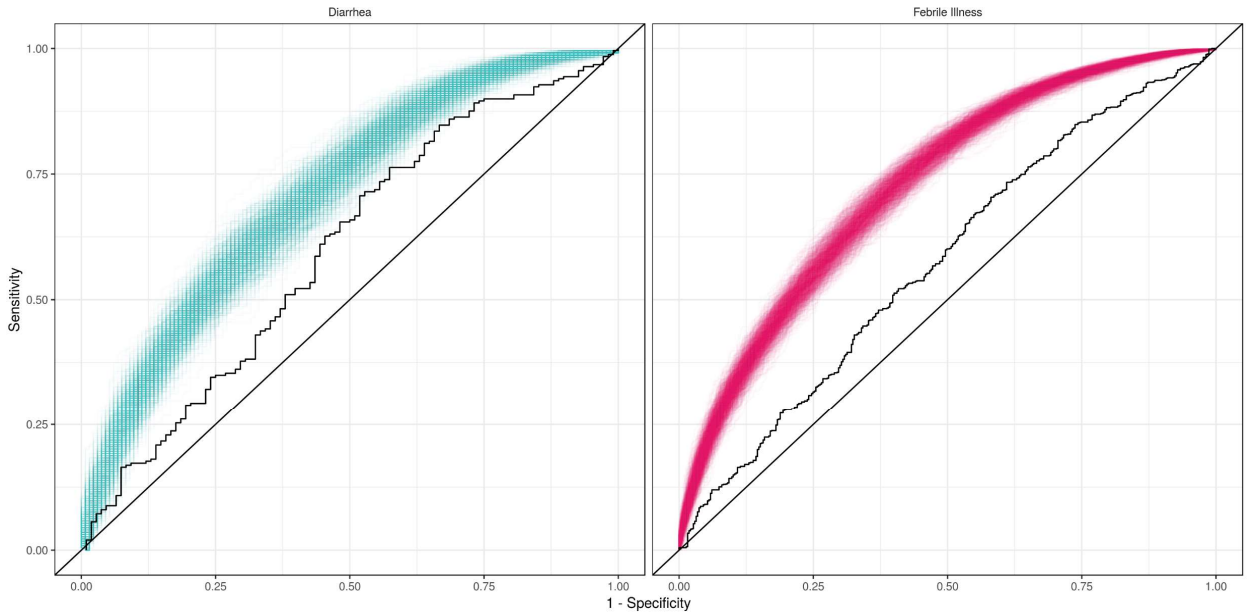


Figure 4. Receiver operating characteristic (ROC) curves for febrile illness and diarrhea. Colored curves are from 10,000 estimated draws from our model. Black curves are produced using the method from Alegana et al.

Discussion

In this analysis, we generated the most comprehensive geo-located dataset of Zambian health facilities, and used these data in combination with geospatially resolved layers of travel time and household information on utilization to generate a maps of healthcare access and utilization patterns in Zambia. Our results provide an update to the last national estimate of geographic accessibility conducted by Moise and colleagues,[65] which was based on a listing of 568 facilities in 2002 and estimated that 62% of the population lived within 15 km of a health facility.

We found that a large proportion of the Zambian population lives relatively close to facility care, but that small differences in travel time within the first hour from a health facility seem to impact utilization rates more than those same differences in distances for households that are further away from health facilities. Both findings go contrary to common perceptions, but have some support from a recent cross-country study by Karra and colleauges.[19] While a non-linear distance decay curve has been modelled in the past, the shape is typically concave – declining slowly at first then more rapidly – rather than the convex shape we observed. One possible explanation for this difference is the random 'jittering' of GPS coordinates used in previous studies. [25,35]

Other factors associated with increased utilization were largely consistent with previous studies in sub-Saharan Africa and included higher education, female sex, younger age. Interestingly though, we found that, when controlling for travel time, urban residence was negatively associated with utilization. This indicates that models that do not control for travel time and see a positive relationship with urban residence could be picking up a proxy effect of urban as representing geographic accessibility. Our

analysis indicated that, holding accessibility constant, there was a systematic tendency for less facility-based care in urban areas. More research is needed to better understand this phenomenon in the Zambian setting.

It is important to emphasize that this study focuses specifically at facility-based healthcare utilization. Our intention was to measure this phenomenon, not to make claims on whether care ought to be sought health facilities in all cases. In certain cases, people may seek effective care outside of health facilities. For example, diarrhea can be treated with for oral rehydration solution stocked at home form from previous healthcare visits, interactions with community health workers, and informal shops or privately owned pharmacies.

By combining information on travel times and household survey data, we offer the first-ever application of model-based geostatistics for healthcare utilization mapping. While the present analysis focuses on febrile illness and diarrhea in children as a use case, our method and underlying geostatistical framework could readily accommodate for estimation of utilization for other conditions, pending data collection and availability on a broader set of conditions. Accessibility to health facilities for each household were independently measured using a combination of precise GPS readings and the gridded surface of theoretical travel times, and thus did not rely on self-report. By estimating healthcare utilization in a joint probabilistic model, we could quantify uncertainty in our mapped estimates, an endeavor that has not been reported in previous work. Our study indicates that spatially resolved estimates of utilization were difficult to achieve with any practical degree of certainty, despite showing predictive performance improvements over previous models. These findings emphasize how the complex factors associated with the decision to seek care in a health facility go far beyond geographic accessibility or other geographic factors. As the science of utilization mapping evolves and improves, researchers should provide uncertainty estimates as to more transparently communicate the strength (or lack thereof) of evidence provided by these types of analyses.

This research also uncovered a disagreement between the implied population distribution from the representative sample of ZHHEUS and gridded population data from WorldPop. Depending on source population data, 75% to 92% of the country's residents live within an hour of a health facility. This is despite the fact that both sources utilize the 2010 census as either a basis for sampling or estimation. It is currently unclear the extent to which this difference is due to an under-sampled (and under-weighted) rural population in ZHHEUS or an over-allocation of rural population in the gridded WorldPop surface. This disaccord has practical implications for health planners in Zambia who aim to provide 'services as close to the family as possible,'[41] whereas based on these two sources, at a national level it is unknown whether 25% or 8% of the population which lives further than an hour from healthcare.

By further combining utilization rate estimates with spatially resolved estimates of disease incidence to estimate cases unseen by healthcare, we provide an important use case of utilization mapping. Results from this geospatial approach could in theory be used to give public health planners a better sense of where these unseen cases are likely to be found. Future research should focus on the application of utilization mapping as a possible approach to adjusting under-reporting of cases in administrative data sources such as DHIS2.[66]

Limitations and future research directions

The results of this study should be interpreted in light of several important limitations.

We needed to make simplifying assumptions due to a number of data limitations. First, our analysis was limited to an analytical resolution of 30 arc-seconds (approximately 1 x 1-km) due to the available resolution of the input friction surface,[50] implying that all travel within a pixel was treated as homogenous. Second, ZHHEUS provided limited data on remote households, and as a result the distance decay functions beyond 2 hours had high uncertainty. While this likely represents a small proportion of the population, it is also a particularly vulnerable one. Third, we assumed that self-reported illness episodes correctly captured illness experiences in the population. This could bias results, if, for example, respondents tended to report they had malaria only if they were diagnosed at a clinic, thus artificially increasing the utilization rates for febrile illness in our analysis. Furthermore, due to a lack of information about episode severity in the ZHHEUS questionnaire, we assumed a homogenous distribution of severity across the population. In an ideal model we would account for known effect modification in the distance decay function due to severity of illness.[5] Finally, utilization patterns were measured as part of a cross-sectional survey during what is usually considered the rainy season in Zambia, potentially affecting physical access due to flooding, and/or incurring atypical utilization rates as related to heightened risk for illnesses like malaria and diarrhea.

We assumed that distance to nearest health facility accurately operationalizes geographic accessibility in all cases. For some people, a nearby health facility may not be effectively accessible to due to poor perceived quality, wait times, or stigma. These are important potential gaps in access to healthcare which should be considered in future utilization mapping work.

Finally, spatial covariates could help improve accuracy and precision of predictions in future utilization, assuming that covariates of interest in fact vary spatially. In this analysis, we had the dual goal of inference and prediction. In lieu of making valid inference on the effect of covariates, it could be possible to produce improved geospatial predictions. For example, a spatial layer on maternal education[67] would likely be highly predictive of utilization, but itself was constructed from a model which relied heavily on travel times as a covariate, thus introducing strong multicollinearity into the model. Furthermore, error-in-covariate models should be considered when dealing with uncertain spatial covariates.

Conclusion

While at least three quarters of Zambians live within an hour of a health facility, we found that small differences in travel time to healthcare are independently associated with large declines in utilization rates within the first hour. The decision to seek care at a health facility is a complex process that is not easily reduced to geography. As such, a univariate model based solely on geographic accessibility is not sufficient for accurate prediction of utilization across a gridded surface. Improved prediction using a probabilistic model is possible but uncertainty remains high.

References

1. *WHO - The world health report 2000 - Health systems: improving performance*.

2. Bonfrer, I., van de Poel, E., Grimm, M. & Van Doorslaer, E. Does the distribution of healthcare utilization match needs in Africa? *Health Policy Plan.* **29**, 921–937 (2014).

3. McLaren, Z. M., Ardington, C. & Leibbrandt, M. Distance decay and persistent health care disparities in South Africa. *BMC Health Serv. Res.* **14**, 541 (2014).

4. Macharia, P. M., Odera, P. A., Snow, R. W. & Noor, A. M. Spatial models for the rational allocation of routinely distributed bed nets to public health facilities in Western Kenya. *Malar. J.* **16**, 367 (2017).

5. Moïsi, J. C. *et al.* Sensitivity of hospital-based surveillance for severe disease: a geographic information system analysis of access to care in Kilifi district, Kenya. *Bull. World Health Organ.* **89**, 102–111 (2011).

6. Poletti, P. *et al.* The hidden burden of measles in Ethiopia: how distance to hospital shapes the disease mortality rate. *BMC Med.* **16**, 177 (2018).

7. Andersen, R. M. Revisiting the Behavioral Model and Access to Medical Care: Does it Matter? *J. Health Soc. Behav.* **36**, 1–10 (1995).

8. Peters, D. H. *et al.* Poverty and Access to Health Care in Developing Countries. *Ann. N. Y. Acad. Sci.* **1136**, 161–171 (2008).

9. Tsegay, Y. *et al.* Determinants of antenatal and delivery care utilization in Tigray region, Ethiopia: a cross-sectional study. *Int. J. Equity Health* **12**, 30 (2013).

10. Luginaah, I. N. *et al.* Timing and utilization of antenatal care services in Liberia: Understanding the pre-Ebola epidemic context. *Soc. Sci. Med.* **160**, 75–86 (2016).

11. Masters, S. H. *et al.* Travel time to maternity care and its effect on utilization in rural Ghana: A multilevel analysis. *Soc. Sci. Med.* **93**, 147–154 (2013).

12. Yao, J., Murray, A. T. & Agadjanian, V. A geographical perspective on access to sexual and reproductive health care for women in rural Africa. *Soc. Sci. Med.* **96**, 60–68 (2013).

13. Gabrysch, S., Cousens, S., Cox, J. & Campbell, O. M. R. The Influence of Distance and Level of Care on Delivery Place in Rural Zambia: A Study of Linked National Data in a Geographic Information System. *PLOS Med* **8**, e1000394 (2011).

14. Otieno, N. A. *et al.* Demographic, socio-economic and geographic determinants of seasonal influenza vaccine uptake in rural western Kenya, 2011. *Vaccine* **32**, 6699–6704 (2014).

15. Shemwell, S. A. *et al.* Determinants of full vaccination status in children aged 12–23 months in Gurùé and Milange districts, Mozambique: results of a population-based cross-sectional survey. *Int. Health* **9**, 234–242 (2017).

16. Siedner, M. J. *et al.* GPS-measured distance to clinic, but not self-reported transportation factors, are associated with missed HIV clinic visits in rural Uganda. *AIDS Lond. Engl.* **27**, 1503–1508 (2013).

17. Feikin, D. R. *et al.* The impact of distance of residence from a peripheral health facility on pediatric health utilisation in rural western Kenya. *Trop. Med. Int. Health* **14**, 54–61 (2009).

18. Rutherford, M. E., Mulholland, K. & Hill, P. C. How access to health care relates to under-five mortality in sub-Saharan Africa: systematic review. *Trop. Med. Int. Health* **15**, 508–519 (2010).

19. Karra, M., Fink, G. & Canning, D. Facility distance and child mortality: a multi-country study of health facility access, service utilization, and child health outcomes. *Int. J. Epidemiol.* **46**, 817–826 (2017).

20. Schoeps, A., Gabrysch, S., Niamba, L., Sié, A. & Becher, H. The Effect of Distance to Health-Care Facilities on Childhood Mortality in Rural Burkina Faso. *Am. J. Epidemiol.* **173**, 492–498 (2011).

21. Målqvist, M., Sohel, N., Do, T. T., Eriksson, L. & Persson, L.-Å. Distance decay in delivery care utilisation associated with neonatal mortality. A case referent study in northern Vietnam. *BMC Public Health* **10**, 762 (2010).

22. Zaman, S. M. A. *et al.* The effect of distance on observed mortality, childhood pneumonia and vaccine efficacy in rural Gambia. *Epidemiol. Infect.* **142**, 2491–2500 (2014).

23. Manongi, R. *et al.* Inpatient child mortality by travel time to hospital in a rural area of Tanzania. *Trop. Med. Int. Health* **19**, 555–562 (2014).

24. Kadobera, D., Sartorius, B., Masanja, H., Mathew, A. & Waiswa, P. The effect of distance to formal health facility on childhood mortality in rural Tanzania, 2005–2007. *Glob. Health Action* **5**, 19099 (2012).

25. Alegana, V. A. *et al.* Spatial modelling of healthcare utilisation for treatment of fever in Namibia. *Int. J. Health Geogr.* **11**, 6 (2012).

26. Stock, R. Distance and the utilization of health facilities in rural Nigeria. *Soc. Sci. Med.* **17**, 563–570 (1983).

27. Juran, S. *et al.* Geospatial mapping of access to timely essential surgery in sub-Saharan Africa. *BMJ Glob. Health* **3**, e000875 (2018).

28. Ouma, P. O. *et al.* Access to emergency hospital care provided by the public sector in sub-Saharan Africa in 2015: a geocoded inventory and spatial analysis. *Lancet Glob. Health* **6**, e342–e350 (2018).

29. dos Anjos Luis, A. & Cabral, P. Geographic accessibility to primary healthcare centers in Mozambique. *Int. J. Equity Health* **15**, 173 (2016).

30. Macharia, P. M., Ouma, P. O., Gogo, E. G., Snow, R. W. & Noor, A. M. Spatial accessibility to basic public health services in South Sudan. *Geospatial Health* **12**, 510 (2017).

31. Noor, A. M., Alegana, V. A., Gething, P. W. & Snow, R. W. A spatial national health facility database for public health sector planning in Kenya in 2008. *Int. J. Health Geogr.* **8**, 13 (2009).

32. Noor, A. M., Gikandi, P. W., Hay, S. I., Muga, R. O. & Snow, R. W. Creating spatially defined databases for equitable health service planning in low-income countries: the example of Kenya. *Acta Trop.* **91**, 239–251 (2004).

33. The DHS Program - GPS Data Collection. Available at: https://dhsprogram.com/What-We-Do/GPS-Data-Collection.cfm. (Accessed: 17th September 2018)

34. Lohela, T. J., Campbell, O. M. R. & Gabrysch, S. Distance to Care, Facility Delivery and Early Neonatal Mortality in Malawi and Zambia. *PLOS ONE* **7**, e52110 (2012).

35. Ouma, P. O., Agutu, N. O., Snow, R. W. & Noor, A. M. Univariate and multivariate spatial models of health facility utilisation for childhood fevers in an area on the coast of Kenya. *Int. J. Health Geogr.* **16**, 34 (2017).

36. Colson, K. E. *et al.* Benchmarking health system performance across districts in Zambia: a systematic analysis of levels and trends in key maternal and child health interventions from 1990 to 2010. *BMC Med.* **13**, 69 (2015).

37. Ng, M. *et al.* Assessing the Contribution of Malaria Vector Control and Other Maternal and Child Health Interventions in Reducing All-Cause Under-Five Mortality in Zambia, 1990–2010. *Am. J. Trop. Med. Hyg.* **97**, 58–64 (2017).

38. Naghavi, M. *et al.* Global, regional, and national age-sex specific mortality for 264 causes of death, 1980–2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet* **390**, 1151–1210 (2017).

39. Golding, N. *et al.* Mapping under-5 and neonatal mortality in Africa, 2000–15: a baseline analysis for the Sustainable Development Goals. *The Lancet* **390**, 2171–2182 (2017).

40. Masiye, F., Kaonga, O. & Kirigia, J. M. Does User Fee Removal Policy Provide Financial Protection from Catastrophic Health Care Payments? Evidence from Zambia. *PLOS ONE* **11**, e0146508 (2016).

41. *Zambia National Health Strategic Plan 2017-2021*. (Zambia MOH).

42. Gabrysch, S., Cousens, S., Cox, J. & Campbell, O. M. R. The Influence of Distance and Level of Care on Delivery Place in Rural Zambia: A Study of Linked National Data in a Geographic Information System. *PLOS Med.* **8**, e1000394 (2011).

43. Zyaambo, C., Siziya, S. & Fylkesnes, K. Health status and socio-economic factors associated with health facility utilization in rural and urban areas in Zambia. *BMC Health Serv. Res.* **12**, 389 (2012).

44. Masiye, F. & Kaonga, O. Determinants of Healthcare Utilisation and Out-of-Pocket Payments in the Context of Free Public Primary Healthcare in Zambia. *Int. J. Health Policy Manag.* **5**, 693–703 (2016).

45. Bhatt, S. *et al.* The effect of malaria control on *Plasmodium falciparum* in Africa between 2000 and 2015. *Nature* **526**, 207–211 (2015).

46. Gething, P. W. *et al.* Mapping Plasmodium falciparum Mortality in Africa between 1990 and 2015. *N. Engl. J. Med.* **375**, 2435–2445 (2016).

47. MOH. *The 2012 List of Health Facilities in Zambia*. (Zambia Ministry of Health).

48. Zambia Health Facility Census 2005-2006 | GHDx. Available at: http://ghdx.healthdata.org/record/zambia-health-facility-census-2005-2006. (Accessed: 17th September 2018)

49. Access, Bottlenecks, Costs, and Equity (ABCE) Project | GHDx. Available at: http://ghdx.healthdata.org/series/access-bottlenecks-costs-and-equity-abce-project. (Accessed: 17th September 2018)

50. Weiss, D. J. *et al.* A global map of travel time to cities to assess inequalities in accessibility in 2015. *Nature* **553**, 333–336 (2018).

51. Reiner, R. C. *et al.* Variation in Childhood Diarrheal Morbidity and Mortality in Africa, 2000–2015. *N. Engl. J. Med.* **379**, 1128–1138 (2018).

52. 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).

53. Tatem, A. J. WorldPop, open data for spatial demography. *Scientific Data* (2017). doi:10.1038/sdata.2017.4

54. Pezzulo, C. *et al.* Sub-national mapping of population pyramids and dependency ratios in Africa and Asia. *Sci. Data* **4**, 170089 (2017).

55. Pesaresi, M. *et al.* A Global Human Settlement Layer From Optical HR/VHR RS Data: Concept and First Results. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **6**, 2102–2131 (2013).

56. Fahrmeir, L. & Lang, S. Bayesian inference for generalized additive mixed models based on Markov random field priors. *J. R. Stat. Soc. Ser. C Appl. Stat.* **50**, 201–220 (2001).

57. Lang, S. & Brezger, A. Bayesian P-Splines. *J. Comput. Graph. Stat.* **13**, 183–212 (2004).

58. Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H. & Bell, B. TMB: Automatic Differentiation and Laplace Approximation. *J. Stat. Softw.* **70**, (2016).

59. R: The R Project for Statistical Computing. Available at: https://www.r-project.org/. (Accessed: 5th February 2017)

60. Lindgren, F., Rue, H. & Lindström, J. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **73**, 423–498 (2011).

61. Bayesian Spatial Modelling with R-INLA | Lindgren | Journal of Statistical Software. doi:10.18637/jss.v063.i19

62. Blangiardo, M. & Cameletti, M. *Spatial and Spatio-temporal Bayesian Models with R - INLA*. (John Wiley & Sons, 2015).

63. Amoah, B., Giorgi, E., Heyes, D. J., van Burren, S. & Diggle, P. J. Geostatistical modelling of the association between malaria and child growth in Africa. *Int. J. Health Geogr.* **17**, 7 (2018).

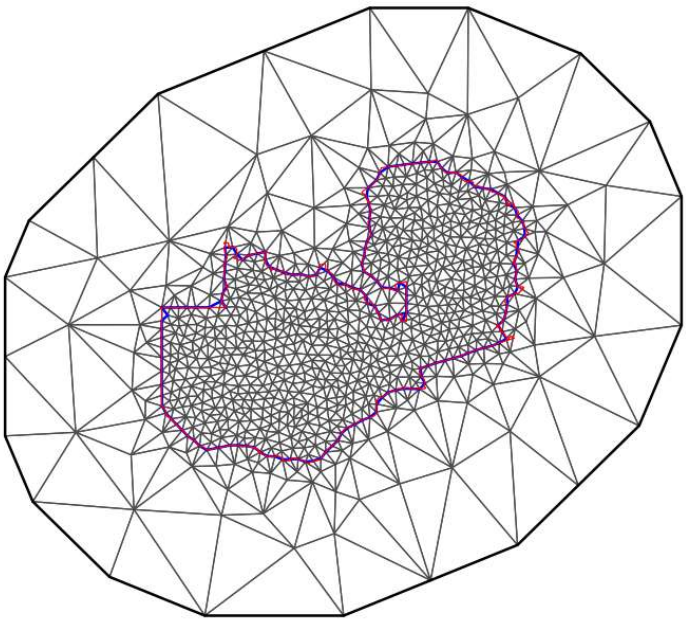
64. Tanser, F., Hosegood, V., Benzler, J. & Solarsh, G. New approaches to spatially analyse primary health care usage patterns in rural South Africa. *Trop. Med. Int. Health* **6**, 826–838 (2001).

65. Moise, I. K., Kalipeni, E. & Zulu, L. C. Analyzing Geographical Access to HIV Sentinel Clinics in Relation to Other Health Clinics in Zambia. *J. Map Geogr. Libr.* **7**, 254–281 (2011).

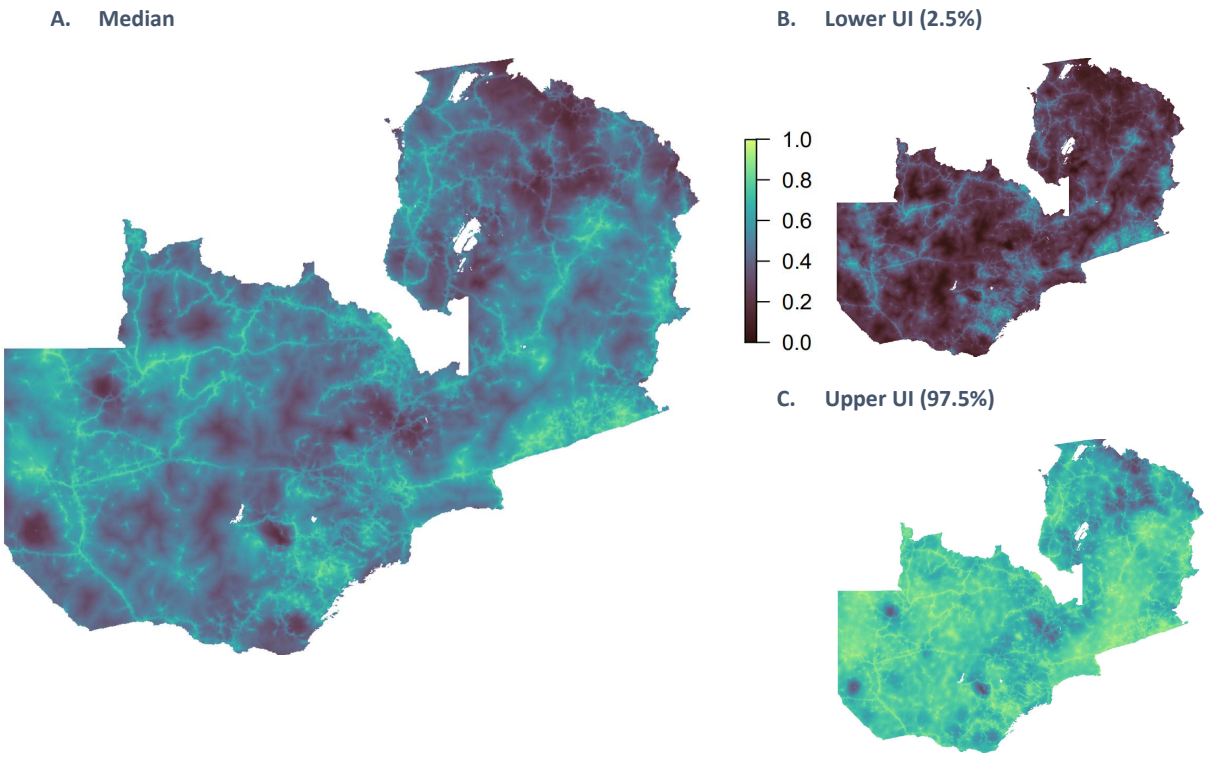
66. Okello, G. A. Producing Malaria Indicators Through District Health Information Software (DHIS2): Practices, Processes And Challenges In Kenya. (The Open University, 2017). doi:http://oro.open.ac.uk/52787/10/180111_Okello_GRADPROG.docx

67. Graetz, N. *et al.* Mapping local variation in educational attainment across Africa. *Nature* **555**, 48–53 (2018).

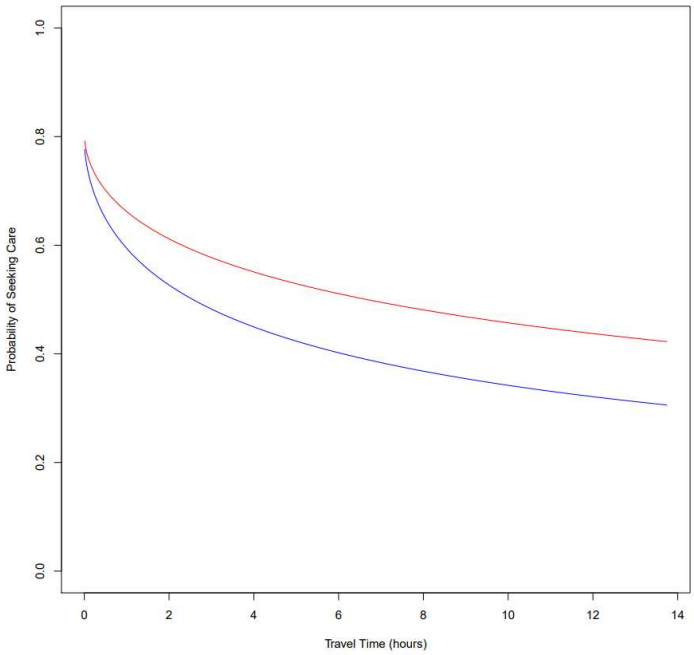
Supplementary Information for Chapter 3



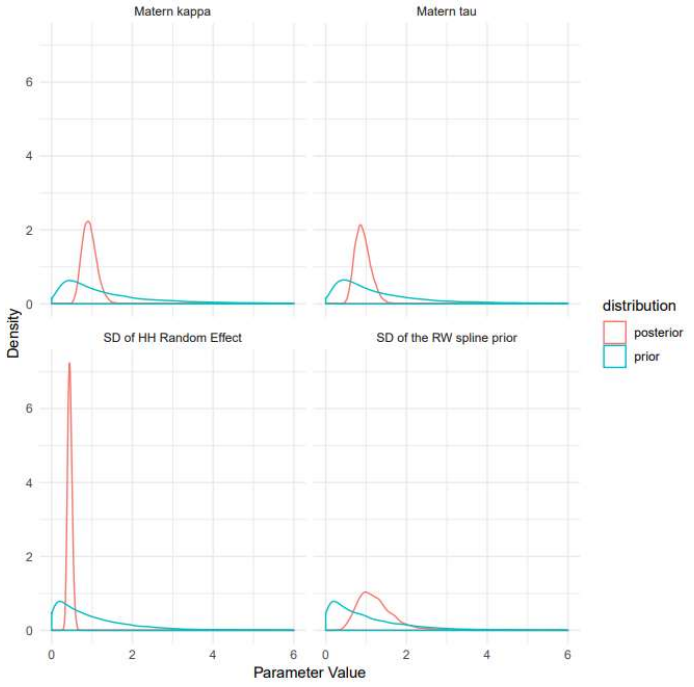
Supplementary Figure 26: Finite elements mesh used to make approximate inference on the Gaussian process random effects.



Supplementary Figure 27: Maps of median prediction and 95% uncertainty interval for probability of utilization in children under 5 with diarrhea.



Supplementary Figure 28: Distance decay curves for febrile illness (red) and diarrhea (blue) in children under 5 using the method in Alegana et al.



Supplementary Figure 29: Density functions of priors and posteriors of model hyper-parameters.