A new method for indirect estimation of neonatal, infant, and child mortality trends using summary birth histories

Short title: Age-specific indirect mortality estimation

Authors

Roy Burstein *
Haidong Wang
Robert C Reiner, Jr.
Simon Iain Hay *

*Joint Corresponding Authors

Author Affiliations

Roy Burstein
Institute for Health Metrics and Evaluation
University of Washington Department of Global Health
Email: royburst@uw.edu

Haidong Wang
Institute for Health Metrics and Evaluation
University of Washington Department of Global Health

Robert C Reiner, Jr.
Institute for Health Metrics and Evaluation
University of Washington

Simon Iain Hay
Institute for Health Metrics and Evaluation
University of Washington
Email: sihay@uw.edu

Abstract

Background

The addition of neonatal mortality targets in the Sustainable Development goals highlights the increased need for age-specific quantification of mortality trends, detail which is not provided by summary birth histories. Several methods exist to indirectly estimate trends in under-5 mortality from summary birth histories, 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 summary birth history data available from household surveys and censuses.

Methods and Findings

We analyzed 243 Demographic and Health Surveys (DHS) from 76 countries, which collected both complete and summary birth histories 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 data to train a discrete hazards generalized additive model that was able 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 weights to potential births of mothers from summary birth history data. Age-specific estimates were evaluated using cross-validation, using an external database of an additional 243 non-DHS census and survey data sources, and overall under-5 mortality was compared 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 variance 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 Study, and significantly better than the standard indirect (Brass), especially in the five years preceding a survey. External validation using census and survey data found close agreement with concurrent direct estimates of mortality in the neonatal and infant age groups.

Conclusions

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 summary birth history data for estimation of trends in neonatal 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.

Author Summary

Why Was This Study Done?

- Recently, and particularly in light of the Sustainable Development Goals, there has been increased interest in rigorous measurement of neonatal and infant mortality trends at both national and subnational levels.
- Summary birth histories are a widely available data source for child mortality but have only been validated to produce estimates of under-5 mortality and are not widely used to make estimates of mortality for other age groups.
- A method was needed that would enable summary birth histories to be utilized in the estimation of age-specific (such as neonatal and infant) mortality trends.

What Did the Researchers Do and Find?

- We extracted data from 243 Demographic and Health Surveys (DHS) which contained complete birth histories – data from which age-specific trends could be directly estimated and from which summary birth histories could be produced.
- We trained a discrete time survival model that could be used to predict mortality from summary birth histories and used both cross-validation as well as a systematic external validation on an additional database of 243 summary birth history only census and survey sources, in order to assess how well the model could reproduce age-specific mortality trends.
- We found that in both out of sample DHS data and on a large database of external data, model predictions of age-specific mortality closely matched validation data, while also performing nearly as well as the current best model used for overall under-5 mortality.

What Do These Findings Mean?

- This method can help improve the empirical evidence base upon which global, national, and subnational neonatal, infant, and other age-specific child mortality estimates are made, including by making new use of publicly available summary birth history data on over 150 million children already extracted and analyzed for this paper.

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 due to 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 mortality in addition to under-5 mortality.[7]

Household survey and census-based child mortality questionnaires are available either as complete birth histories (CBH), also sometimes known as full birth histories, or summary birth histories (SBH). CBH are preferred over SBH 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 to directly by age group. In contrast, SBH surveys only ask each mother how many children she has birthed (CEB), how many of her children have died to date (CD), her age, and sometimes about the time since first birth and/or marriage. Nevertheless, SBH are widely available in many censuses and other sample surveys, due in part to 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 SBH.[8–11] However, such methods have yet to be specifically adapted for wider application to estimate

age-specific mortality among under-5s from SBH; subsequently, past assessments of neonatal and infant mortality have been informed by comparably less data, especially outside of VR settings.

Indirect trends in child mortality from SBH 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 Supplementary Information section 1. In brief, the standard indirect method uses simulated coefficients applied to the ratio of CD to CEB, aggregated at different maternal age (or time since first birth) 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 children died to mothers in each year preceding the survey. MAP distributions are produced by maternal age, CEB, and region. The period-specific aggregations of expected children died and 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 where both SBH and CBH are available, and thus the opportunity to train and test new methods.

In this paper we describe and test a novel method for indirect estimation of age-specific mortality using SBH, 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 complete and summary birth histories 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 Supplementary Information table 1.

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 livebirths 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. CBH 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 one month at

death, in months if the child was under two years old at death, and in years if the child was two or older at death.

We further analyzed an additional 243 censuses and household surveys from 93 low and middle income countries (LMICs), in order to demonstrate how the method can be applied in datasets where 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 Supplementary Information, along with their GHDx record identification number, where links to data distributors are provided.

Statistical Model

Our goal was to develop a model which 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 SBH, since we

ultimately wanted to use this model to estimate mortality trends in datasets where only SBH 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 which describe the risk of event over exposure time. Special care is taken for data that are right-censored, where event status is unobserved after a certain period. In 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 $\boldsymbol{X}$. 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 $Y_{i,a}$ is included for each row and set to 1 if the child died in that age bin. A no-covariate baseline hazard could then be determined by fitting the following logistic regression model:

208 $$Y_{i,a} \sim Bernoulli(q_a)$$

209 $$logit(q_a) = \sum_{a=1}^{A} I_{i,a}\beta_a$$

210 Note that fixed effects are estimated for each age bin without an intercept term, so that each β_a is in
211 reference to zero, and thus each e^{β_a} are interpretable as the probability of mortality in age group a,
212 conditional on survival to age group a, or (q_a). In the discrete case we thus refer to $q_a$as the probability
213 of death within age bin a, though 'mortality rate' is often used interchangeably. This basic model can be
214 extended to include individual-level covariates, random effects to account for hierarchical data,
215 transformations or smoothing splines on covariates to improve prediction, and interactions with age-bin
216 dummies in order to allow for non-proportional effects of covariates.

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

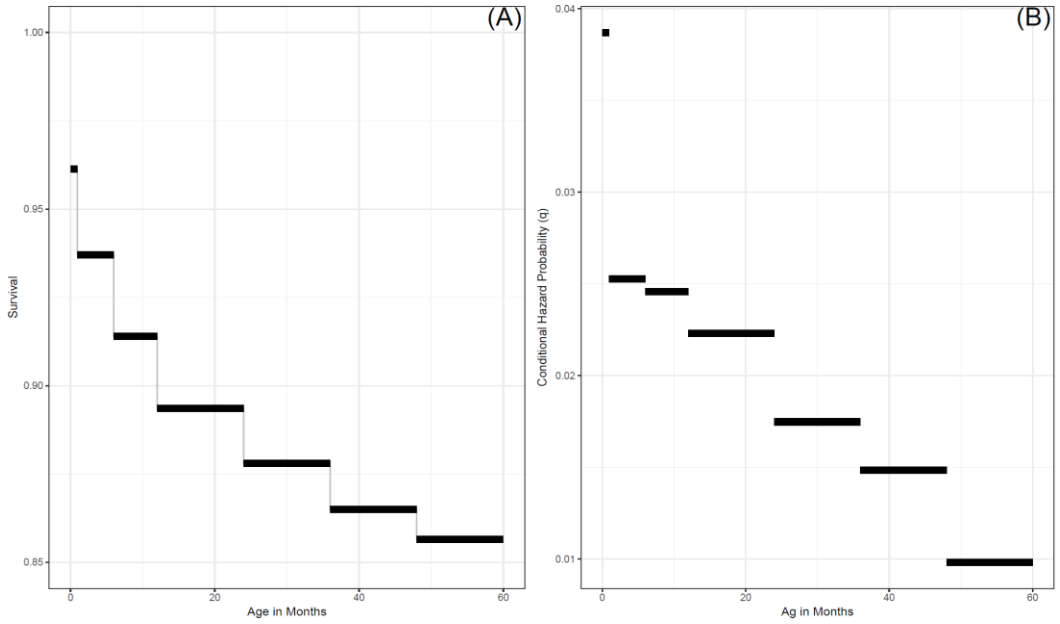


Figure 1 *Illustrating the estimated pooled baseline discrete hazard and survival functions from the 2011 Burundi DHS dataset, fit using the seven age bins* $a \in (1 = NN, 2 = PNN1, 3 = PNN2, 4 = 1yr, 5 = 2yr, 6 = 3yr, 7 = 4yr)$*. Note that we are estimating discrete hazards, and thus hazards (shown 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* $\widehat{S_a} = \prod_{\alpha=1}^{a}(1 - q_\alpha)$

We trained the model on the pooled complete birth history database with the purpose of making predictions in situations where only SBH 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 *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 child-bearing 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 which vary over the mother's life.

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

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

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

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

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

Where $g_*(\cdot)$ represent thin plate regression spline smooths, with $g_1(\cdot)$ having separate smooths for each age bin a. 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[21] index 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, allowing for prediction in countries without training data.

The variables in the second smooth represent child and mother level covariates. $\frac{CD_m}{CEB_m}$ is the ratio of children died to children ever 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 woman's fertility experiences.

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

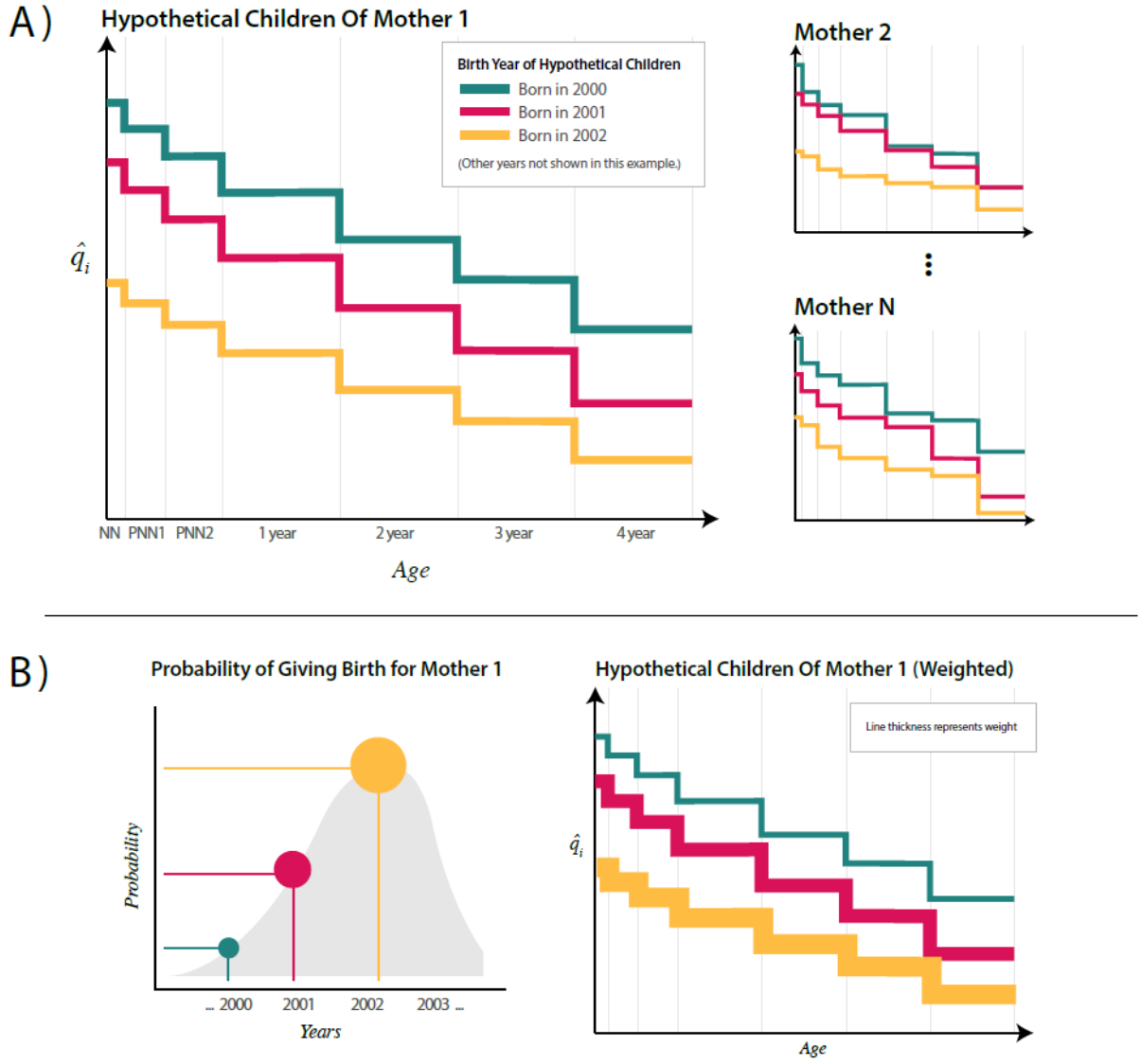
All covariates were centered and scaled by their standard deviations for model fitting. Models were fit separately by the same regions used by Rajaratnam and colleagues[9]. Uncertainty in predictions was ascertained by taking 1000 multivariate normal draws from the variance-covariance matrix of fitted model parameters, including fitted random effects values.[22] In cases where 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 1000 independent normal draws. Models were fit using restricted maximum likelihood with the `bam` command from the `mgcv` 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 summary birth history 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

287 which indicated the likelihood that each hypothetical child actually existed. This process is illustrated in

288 Figure 2.


A)
Hypothetical Children Of Mother 1
Birth Year of Hypothetical Children
Born in 2000
Born in 2001
Born in 2002
(Other years not shown in this example.)
Mother 2
Mother N
$\hat{q}_i$
NN PNN1 PNN2 1 year 2 year 3 year 4 year
Age
B)
Probability of Giving Birth for Mother 1
Probability
... 2000 2001 2002 2003 ...
Years
Hypothetical Children Of Mother 1 (Weighted)
Line thickness represents weight
$\hat{q}_i$
Age
C)
Weighted mean trend in $\hat{q}(PNN2)$
$\hat{q}_i$
2000 2001 2002 2003 2004
All Hypothetical Children Born in 2000
2001 2002 2003 2004 2005
All Hypothetical Children Born in 2001
2002 2003 2004 2005 2006
All Hypothetical Children Born in 2002

289

Figure 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 three 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 twelve years old. B) Probability of birth distributions are applied to each hypothetical birth from each mother. These are derived from the empirical map distributions from Rajaratnam et al. 2010, where a different probability is available by woman's age, CEB, region of residence, and year prior to the survey. These probabilities are multiplied by mother 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 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 are 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.

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. POB distributions are compiled from empirical distributions which describe, for each year preceding a survey, the probability of birth based on mothers' age, 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 a, for child born in year yr from mother m as:

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

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 $\hat{q}_{m,a,yr}$ 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 all quantities are calculated by repeating the aggregation procedure 1000 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 two 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. In our second approach we applied the method to national-representative non-DHS surveys which only collected SBH 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 and using their SBH variables to produce indirect age-specific time trends. We then used direct estimates from the CBH 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 survey.. We fit the model, used the fitted parameters to make indirect estimates from SBH, and compared to direct estimates from CBH. 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 5 metrics to assess out of sample predictive performance: (1) *Mean Error (ME)* to capture systematic bias. A mean error of zero indicates a perfectly unbiased estimate. ME is an absolute 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 percentage of total variance in the directly tabulated hazards explained by the modelled estimates. Each of the metrics were assessed for each age bin, as well as for $5q0$.

Single-year age-specific direct tabulation of CBH have relatively small sample sizes and can produce somewhat noisy estimates of a 'truth' for comparison. Since we are interested in modelling the actual underlying trend and not the noisy observed values, very good predictive performance in this case could actually signal over-fitting. Furthermore, mean-based metrics are sensitive to large outliers in errors, which could emerge spuriously where validation data are noisy. In other words, validation data with a larger sample size is 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) of 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 summary birth history data is 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 five-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 Global Burden of Disease 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 maternal age cohorts (MAC), and one based on time since first birth (TSFB), see [11].

We also used this cross-validation framework to compare our model to several other specifications. These results are presented in the Supplementary Information Section III.A.

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. 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 where only SBH was collected), we compared estimates from these data to directly estimated mortality, where concurrent CBH data were available. We used the same set of metrics described for the cross-validation assessment, and again smoothed CBH estimates.

Data and Code Availability

All datasets used for this analysis are listed in the Supplementary Information. Source availability information is available from the Global Health Data Exchange (GHDx, http://ghdx.healthdata.org/) for each source used.

All code for the analyses and figures 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 which 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	$\bar{q}_a$	***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 DHS database, for the 15 years prior to the survey being taken.

Figure 3 plots the agreement between age-specific mortality rates from the validation data compared to out of sample estimates. We also see 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 due to 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 leading to more stable estimates.

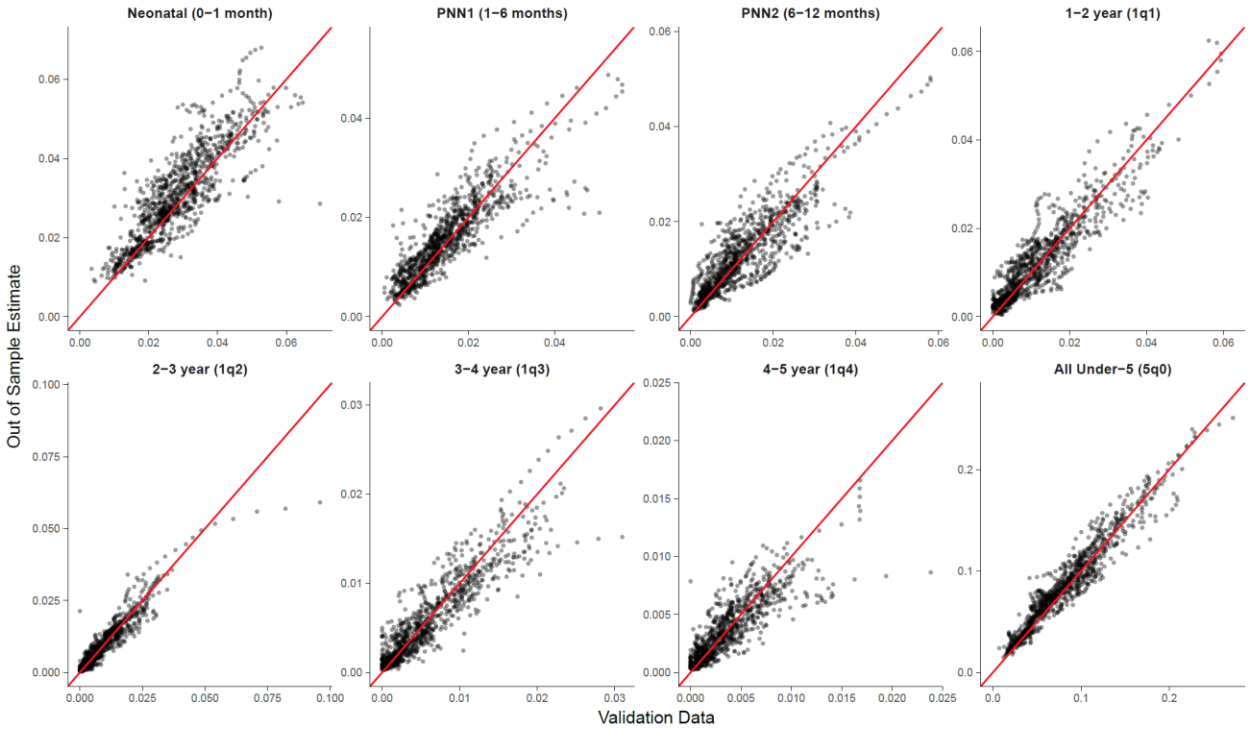
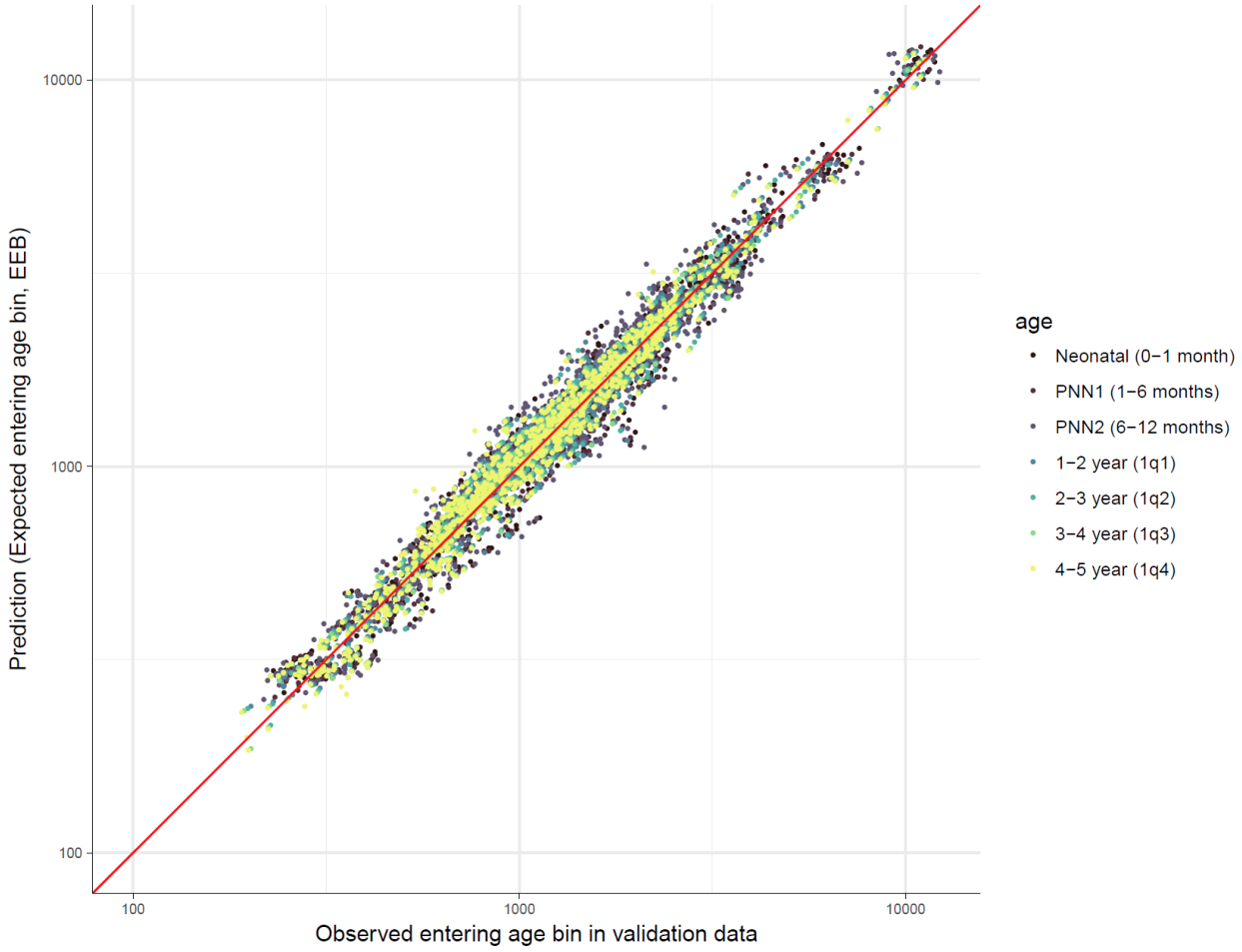


Figure 3: Out of sample predictions of mortality probability compared against loess-smoothed validation data. Each point represents a country-age mortality estimate ($q_{a,yr}$) for each held-out survey from the DHS database. Red line indicates unity.

Figure 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.



443

444 *Figure 4: Comparison of EEB, expected number of children entering each bin, and observed children entering each age bin from*

445 *CBH validation data. Each point represents a survey-year-age bin, both axes are on log 10 scale.* $R^2 = 0.97$.

446

447 At the subnational level, model performance was somewhat weaker. There was a similar pattern in

448 direction of bias across the ages, though bias remained minimal overall. There was more variability in

449 the errors, with MAPE ranging from 20.0% in the neonatal group to 38.0% in 4 year olds. Percent of

450 variance explained was also somewhat lower than at the national level. R^2 of the subnational $5q0$

451 estimates was 0.91. Some of this difference was likely due to smaller sample sizes in the subnational

452 data compared to the national validation. Our validation data, which were based on direct estimates

453 from CBH, represent realizations of the underlying probability, and thus the empirical probability from

454 the validation is measured with noise. Despite aggregating to 5 year bins, the average number of
455 children born in each 5-year aggregated subnational observation was 520, compared to 1769 in each
456 annual national observation, and 4148 (27%) of each survey-administrative area-age bin observation
457 had no observed deaths. Supplementary Information section III.B replicates our national level figures
458 and tables for the first administrative subnational level.

459 Figure 5 shows the out of sample estimated trends in age-specific mortality rates estimated using the
460 2013 Nigeria DHS and compared to the directly estimated validation data. In the Supplementary
461 Information we provide similar plots for each country with extended discussion on those results. Overall,
462 the model was able to reproduce trends in the validation data in Nigeria and in most other countries.
463 Performance was suboptimal in cases where test and train data differed significantly (i.e. in Benin), and
464 where trends were unique to a given country (i.e. Lesotho), see Supplementary Information attachment
465 for figures.

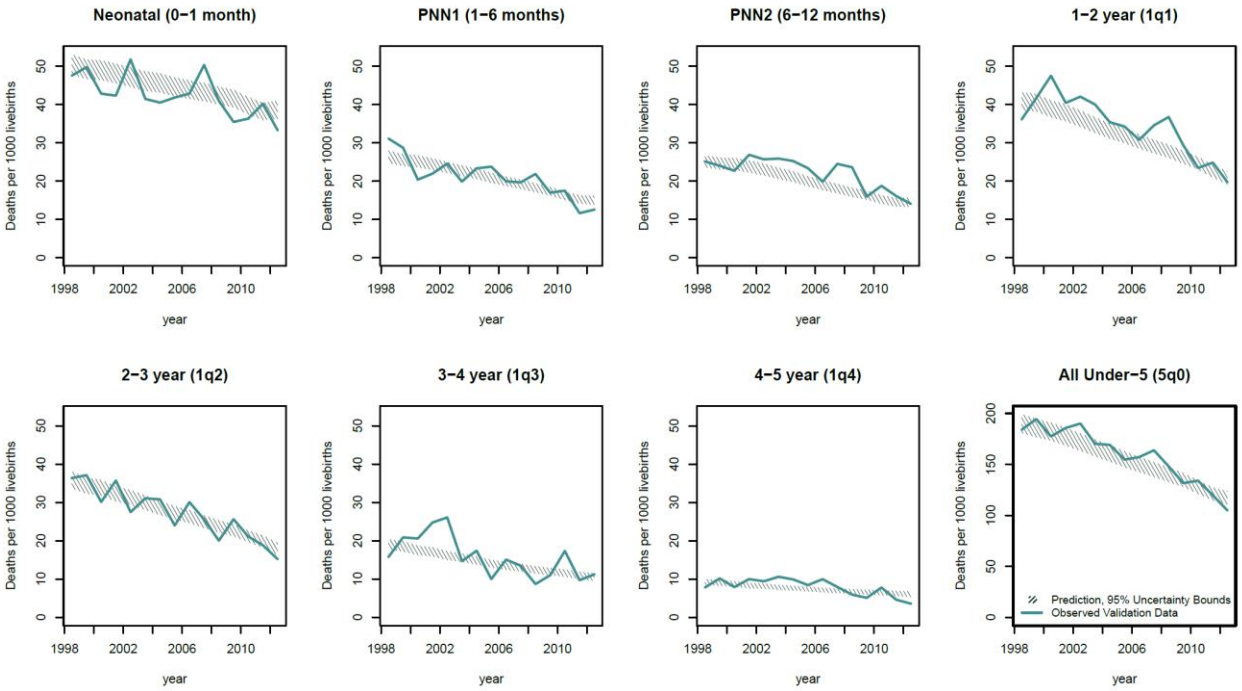


Figure 5: Trends in mortality for each age bin from the 2013 Nigeria DHS. Thick blue lines are validation data, hatched lines are the 95% uncertainty bounds on the out of sample predictions. Sampling variation is evident in the blue line through year on year spikes. The target of prediction was the overall 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 Supplemental Information.

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. Figure 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 five-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. Supplementary Information figure 3 shows trends for each survey in the testing data. We note that for certain surveys with no GBD-combined estimates, such as 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.

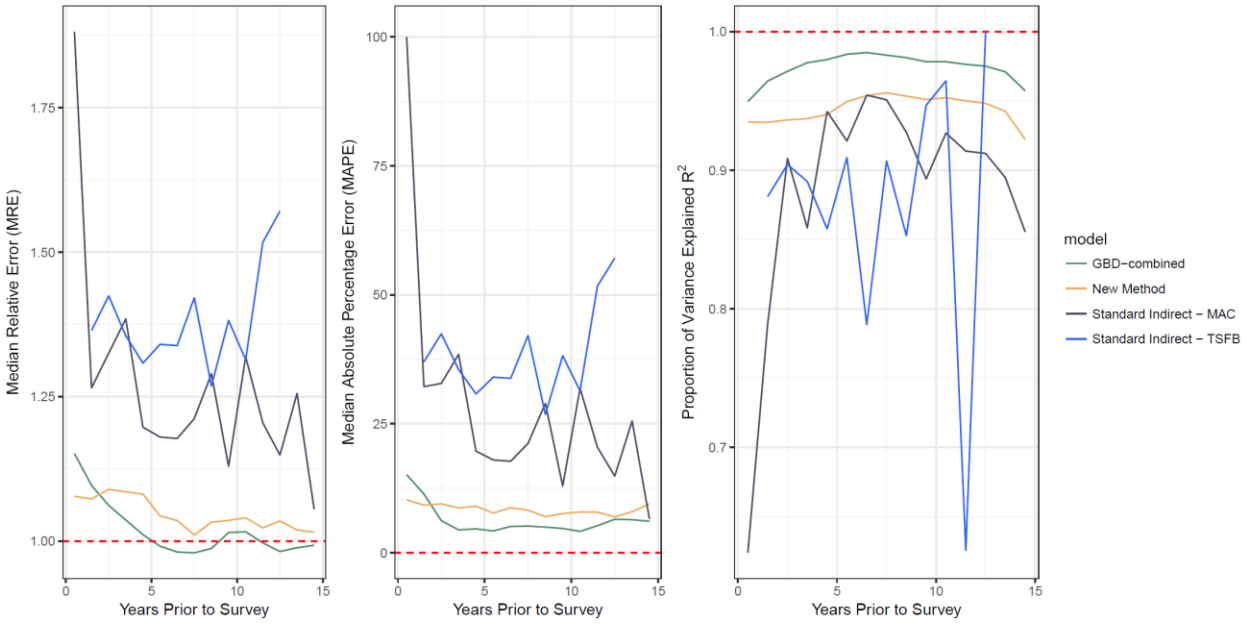


Figure 6 Comparing predictive validity metrics across different methods for indirect estimation of 5q0. Both the GBD-combined and new methods greatly out-perform the standard indirect methods, particularly in the most recent five years.

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 (see supplementary information tables 1 and 2 for a full source lists). 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 where only SBH data were available. For comparison, we also identified 10,655 country-year-age pairs where two concurrent CBH direct estimates were available.

Full trend plots for each country with available data are in the Supplemental Information Figure 6. In the majority of cases, trends from SBH-only data closely match contemporaneous trends from CBH data. There are several SBH-only surveys that exhibit overall source-level bias relative to concurrent trends.

Table 2 summarizes our findings for these paired comparisons for neonatal, 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, 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), see Supplementary Table 4.

Age-bin	$\overline{q}_{a,cbh}$	$\overline{q}_{a,sbh}$	***ME***	***SDE***	***MRE***	***MAPE***	R^2	R^2 source corrected
Neonatal	0.030	0.033	0.0026	0.011	1.06	17.7%	0.52	0.96
Infant	0.061	0.064	0.0030	0.019	1.05	16.4%	0.64	0.96
Under-5	0.093	0.096	0.0029	0.029	1.04	16.4%	0.74	0.97

Table 2 Summary results for the external validation comparisons across 16,527 country-year data pairs where a CBH and SBH estimate were both available.

Figure 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.

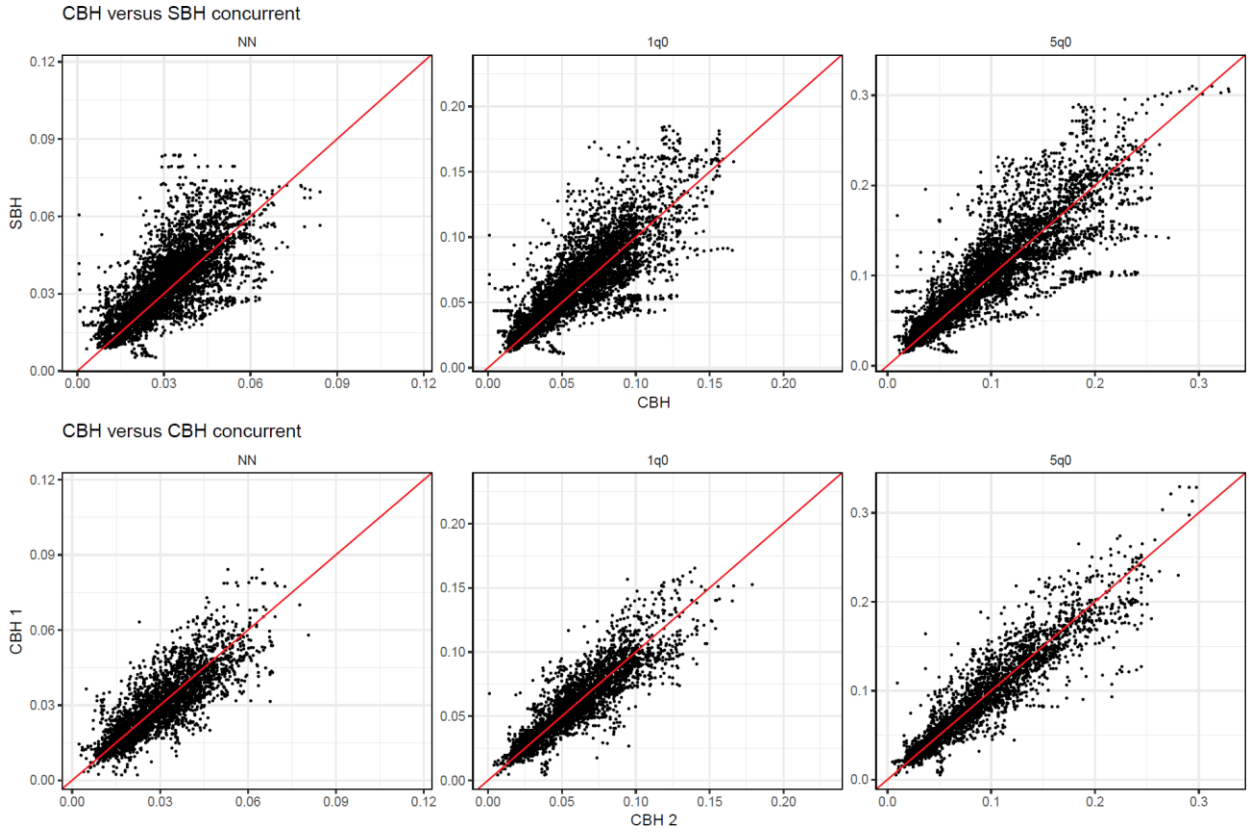


Figure 7. Each country-year concurrent estimate for neonatal, infant, and under-5 mortality. The top row compared concurrent estimates from the SBH-only data with CBH direct estimates. The bottom row shows the same comparison from concurrent CBH estimates, theoretically representing a baseline level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

Discussion

Our new method for indirect estimation produces age-specific mortality trends consistent with those produced using complete birth history data in most cases at the country and first administrative unit level, as well as producing $5q0$ 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 summary birth history data and found considerable agreement where comparisons could be made to contemporaneous estimates from complete birth histories. This new method greatly expands the potential utility of summary birth history 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 complete birth histories as time-to-event data, we were able to directly model the quantity of interest, the conditional probability of death q at various ages from birth until age 5. Leveraging existing data from millions of complete birth histories, we inferred hazard functions that vary across countries, surveys, mothers, and their individual children using only covariates that were available in SBH. 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. Maternal age cohort 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 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 summary birth history data used.[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. Modelling groups such as IGME and GBD regularly exclude data sources due to quality concerns and the data synthesizing models used by such groups 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 neonatal mortality to 12 deaths per 1000 livebirths.[7] Until now, estimates of neonatal mortality have depended mostly on CBH data, or VR where 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 census for monitoring progress toward the neonatal mortality SDG.

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 quantify 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, where 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 children ever born. Our preliminary analyses indicate this is generally true (see Supplementary Information section III.E). Future research should focus on modelling 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, rather than national level covariates, as well as implementing models which account for spatial autocorrelation in residuals. Fifth, by relying on concurrent SBH and CBH

estimates as 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 summary birth history data for estimation of trends in neonatal 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.

Acknowledgements

Roy Burstein and Simon I Hay were funded by Bill and Melinda Gates Foundation (Grant # OPP1093011).

Additionally, we thank Bob Thompson for project coordination, and Sarah Ray, Chloe Shields, Nancy Fullman, Laura Dwyer-Lindgren, Aaron Osgood-Zimmerman, Drew Dolgert, Emmanuela Gakidou, Jon Wakefield, Katie Wilson, Nicole Weaver, and Molly Nixon for useful comments on the methods and drafts of this manuscript. We also thank Nat Henry and Michael Collison for help with data extraction.

References

1. Murray CJL. Shifting to Sustainable Development Goals — Implications for Global Health. N Engl J Med. 2015;373: 1390–1393. doi:10.1056/NEJMp1510082

2. Mathers C, Boerma T. Mortality Measurement Matters: Improving Data Collection and Estimation Methods for Child and Adult Mortality. PLOS Med. 2010;7: e1000265. doi:10.1371/journal.pmed.1000265

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. doi:10.1186/1478-7954-12-14

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. doi:10.1016/S0140-6736(16)31575-6

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. doi:10.1016/S0140-6736(15)00120-8

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. doi:10.1016/S0140-6736(17)31758-0

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: https://sustainabledevelopment.un.org/post2015/transformingourworld

8. Manual X [Internet]. [cited 5 Feb 2017]. Available: 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. doi:10.1371/journal.pmed.1000253

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. doi:10.1371/journal.pone.0137713

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: http://demographicestimation.iussp.org/

12. Brass W. Methods for Estimating Fertility and Mortality from Limited and Defective Data. North Carolina: 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. doi:10.1080/00324728.1975.10410187

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

15. Preston S, Heuveline P, Guillot M. Demography: Measuring and Modeling Population Processes. 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. doi:10.1371/journal.pone.0176366

17. Tolles J, Lewis RJ. Time-to-Event Analysis. JAMA. 2016;315: 1046–1047. doi:10.1001/jama.2016.1825

18. Box-Steffensmeier JM, Jones BS. Event History Modeling: A Guide for Social Scientists. 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. doi:10.1002/sim.1203

20. Allison PD. Discrete-Time Methods for the Analysis of Event Histories. Sociol Methodol. 1982;13: 61–98. doi:10.2307/270718

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. doi:10.1016/S0140-6736(17)31833-0

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. doi:10.2307/2669316

23. Wood S. mgcv: Mixed GAM Computation Vehicle with Automatic Smoothness Estimation [Internet]. 2017. Available: https://cran.r-project.org/web/packages/mgcv/index.html

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: http://arxiv.org/abs/1711.00800

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. doi:10.1016/S2214-109X(16)30212-1

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: http://www.dhsprogram.com/publications/publication-SAR13-Spatial-Analysis-Reports.cfm

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. doi:10.1002/psp.484

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. doi:10.1371/journal.pmed.1001298

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. doi:10.1371/journal.pmed.1001296