Global estimation of child mortality using a Bayesian B-spline bias-reduction model

Leontine Alkema and Jin Rou New

Department of Statistics and Applied Probability
and Saw Swee Hock School of Public Health
National University of Singapore
Singapore

e-mail: alkema@nus.edu.sg

September 9, 2013

Abstract: National estimates of the under-5 mortality rate (U5MR) are used to track progress in reducing child mortality and to evaluate countries' performance related to United Nations Millennium Development Goal 4, which calls for a reduction in the U5MR by two-thirds between 1990 and 2015. However, for the great majority of developing countries without well-functioning vital registration systems, estimating levels and trends in child mortality is challenging, not only because of limited data availability but also because of issues with data quality. Global U5MR estimates are often constructed without accounting for potential biases in data series, which may lead to inaccurate point estimates and/or credible intervals.
We describe a Bayesian penalized B-spline regression model for assessing levels and trends in the U5MR for all countries in the world, whereby biases in data series are estimated through the inclusion of a multilevel model to improve upon the limitations of current methods. B-spline smoothing parameters are also estimated through a multilevel model. Improved spline extrapolations are obtained through logarithmic pooling of the posterior predictive distribution of country-specific changes in spline coefficients with observed changes on the global level.
The proposed model is able to flexibly capture changes in U5MR over time, gives point estimates and credible intervals that reflect potential biases in data series and performs reasonably well in out-of-sample validation exercises. It has been accepted by the United Nations Inter-agency Group for Child Mortality Estimation to measure countries' progress in reducing U5MR, and to evaluate their performance with respect to Millennium Development Goal 4.

Keywords and phrases: Bayesian hierarchical model, Millennium Development Goal 4, Logarithmic pooling, Penalized B-spline regression model, Under-five mortality rate, United Nations Inter-agency Group for Child Mortality Estimation.

1. Introduction

The under-five mortality rate (U5MR) is a key barometer of the well-being of a country's children, and more broadly, an indicator of socioeconomic progress. U5MR is defined as the probability (expressed as a rate per 1,000 live births) that a child born in a given year will die before reaching the age of five if subject to current age-specific mortality rates (United Nations Inter-agency Group for Child Mortality Estimation 2012). National estimates of the U5MR are used to track progress in reducing child mortality and to evaluate countries' performance with respect to the United Nations' Millennium Development Goal 4 (MDG 4), which calls for a two-thirds reduction in the U5MR between 1990 and 2015 (United Nations Inter-agency Group for Child Mortality Estimation 2012), corresponding to an annual rate of reduction of 4.4%.

For the great majority of developing countries without well-functioning vital registration systems, estimating levels and trends in U5MR is challenging, not only because of limited data availability but also because of issues with data quality. Every year, the United Nations Inter-agency Group for Child Mortality Estimation (UN IGME, including the United Nations Children's Fund, the World Health Organization, the World Bank, and the United Nations Population Division) produces and publishes estimates of child mortality comparable across countries and years for 195 countries. In 2012, a Loess regression model was used to estimate the U5MR (United Nations Inter-agency Group for Child Mortality Estimation 2012). For each country, the default setting for its smoothness parameter α was determined by the type and availability of data in the country. A bootstrap method was used to assess the uncertainty in the U5MR estimates (Alkema and New 2012). A number of limitations with this approach were identified. The first limitation was that for a subset of countries, the fitted Loess curve was deemed to not fit the data well and post-hoc adjustments in the α value were necessary. The second limitation was that all observations were weighted equally to obtain point estimates; standard errors, potential data biases and indicators of data quality were not accounted for. The calibration of the resulting point estimates and uncertainty intervals left room for improvement.

Alternative methods for estimating child mortality for all countries have been developed by the Institute for Health Metrics and Evaluation (Rajaratnam et al. 2010; Wang et al. 2012), which uses Gaussian process regression modeling to obtain U5MR estimates. A model validation exercise to check model performance based on the 2010 version of the IHME approach also left room for improvement (Alkema et al. 2012), possibly explained by not fully accounting for potential data biases in this approach. To the best of our knowledge, the same exercise has not been repeated for the most recent iteration of the IHME model (Wang et al. 2012). We expect that issues with model calibration have not yet been fully addressed given that the data model has not been updated to incorporate the possibility of data biases.

In this paper, we propose an alternative U5MR estimation approach to improve upon the limitations and lack of calibration of existing methods. The approach is given by a Bayesian B-spline bias-reduction model, referred to as the B3 model. The UN IGME has decided to use the B3 model to assess countries' progress towards MDG 4 and B3 estimates are included in the upcoming 2013 versions of "A Promise Renewed Progress Report 2012" (United Nations Children's Fund, Division of Policy and Strategy 2012) and the "Child Mortality Report 2012" (United Nations Inter-agency Group for Child Mortality Estimation 2012).

The paper is organized as follows. Section 2 provides background information on child mortality estimation. In Section 3, we present the B3 model specification, followed by validation results and resulting U5MR estimates in Section 4. We end with a discussion of the model and scope for future research.

2. Background

U5MR data series are constructed from information from vital registration (VR) and sample vital registration (SVR) systems, surveys and censuses. U5MR data for selected countries are shown in Figures 1 and 2. The selected countries differ with respect to U5MR level and trend, as well as data availability and data quality.

In the Netherlands, data from the VR system capturing all births and deaths are available since 1950. Such data from well-functioning VR systems are the preferred data source for calculating U5MR. However, in 2013, 60 countries for which the UN IGME produces U5MR estimates did not have any data from VR systems. Among the 135 countries with VR or SVR systems, recording of birth and/or deaths is not necessarily complete; illustrations are given for Mexico and Moldova. In Mexico, VR data was deemed complete only since 2005. For Moldova, VR data are considered incomplete for all observation years.

For countries without (or with limited information from) well-functioning VR systems, complete or summary birth histories of women, collected in surveys and censuses, are often the main source of information on U5MR. A complete birth history lists all the live births a woman has had, including information on the date of birth of each child; whether the child is still alive, and if the child has died, the age at death. U5MR observations are calculated from such information through a synthetic cohort approach (Pedersen and Liu 2012), and referred to as direct estimates of U5MR. Many of these direct series are obtained from complete birth histories that were collected as part of the international household survey program Demographic and Health Surveys (DHS). Other direct series are obtained from data from survey programs similar to the DHS (here referred to as Other DHS as opposed to (Standard) DHS), as well as other national surveys (referred to as Others Direct). Examples of direct series are shown in Figures 1 and 2. Because of the retrospective nature of the data, direct series can extend for up to decades before the survey. For example, the DHS in Cambodia that was carried out in 2005–2006 provides data from 1979 to 2004.

As the name suggests, summary birth histories provide a summary of complete birth histories: they list the number of live births a woman has had and the number of children that have died. These summarized histories are more commonly collected than complete birth histories because of the simplicity of data collection. For summary birth histories, demographic models are used to calculate the U5MR from the recorded proportion of dead children for different time references (Brass 1964; United Nations 1983). Because of the dependency on models, these estimates based on summary birth histories are referred to as indirect estimates. Indirect series are most commonly obtained using information from censuses and surveys such as the Multiple Indicator Cluster Survey (MICS), an international survey program that collects summary birth histories in many developing countries. Examples of indirect series are shown in Figures 1 and 2. As discussed for direct data series, indirect series also provide data points for a long retrospective period. For example, the Cambodian census from 1998 provides indirect estimates from 1983 to 1994.

The availability of nationally-representative surveys and censuses carried out in developing countries varies greatly. For instance, a large number of data series are available from various sources in Pakistan, but only five data series are available for Papua New Guinea (PNG). Moreover, data series do not necessarily tell a similar story about levels and/or trends in U5MR. For example, in PNG, there are large differences between U5MR estimates from the various sources. In Pakistan, the DHS 2006–2007 survey suggests lower levels of U5MR than data from its registration system. The spread in data points for countries without data from well-functioning VR systems is not specific to the selected countries in Figures 1 and 2, but is observed in many developing countries as U5MR data are associated with a variety of data quality issues. Apart from sampling error, observations from non-VR sources may also be subject to bias and non-sampling error, e.g. because of recall biases when collecting birth histories. Specific data series may be entirely biased upwards or downwards, e.g. based on inaccuracies in the indirect estimation method that was used to translate the summary birth histories from a census or survey in U5MR observations.

Given issues with data quantity and quality, estimating the U5MR is challenging for many countries. A modeling approach needs to be flexible enough to capture short-term fluctuations in U5MR without being overly sensitive to erroneous data fluctuations.

3. Constructing U5MR estimates

We developed a modeling approach that combines a flexible curve fitting method with a comprehensive data model to account for data quality issues. Let u_i denote observed U5MR for observation i in country $c[i]$ and year $t[i]$ with $u_i = U_{c[i]}(t[i]) \cdot \varepsilon_i$, where $U_c(t)$ denotes the true U5MR in country c in year t and error multiplier $\varepsilon_i > 0$. On the natural log-scale, this corresponds to

$$y_i = f_{c[i]}(t[i]) + \delta_i, \tag{1}$$

where $y_i = \log(u_i)$, $f_c(t) = \log(U_c(t))$, and $\delta_i = \log(\varepsilon_i)$. To construct U5MR estimates, we use a Bayesian penalized spline regression model to represent $f_c(t)$, explained further in Section 3.1. The data and specification of error term δ_i (the data model) is discussed further in Section 3.2.

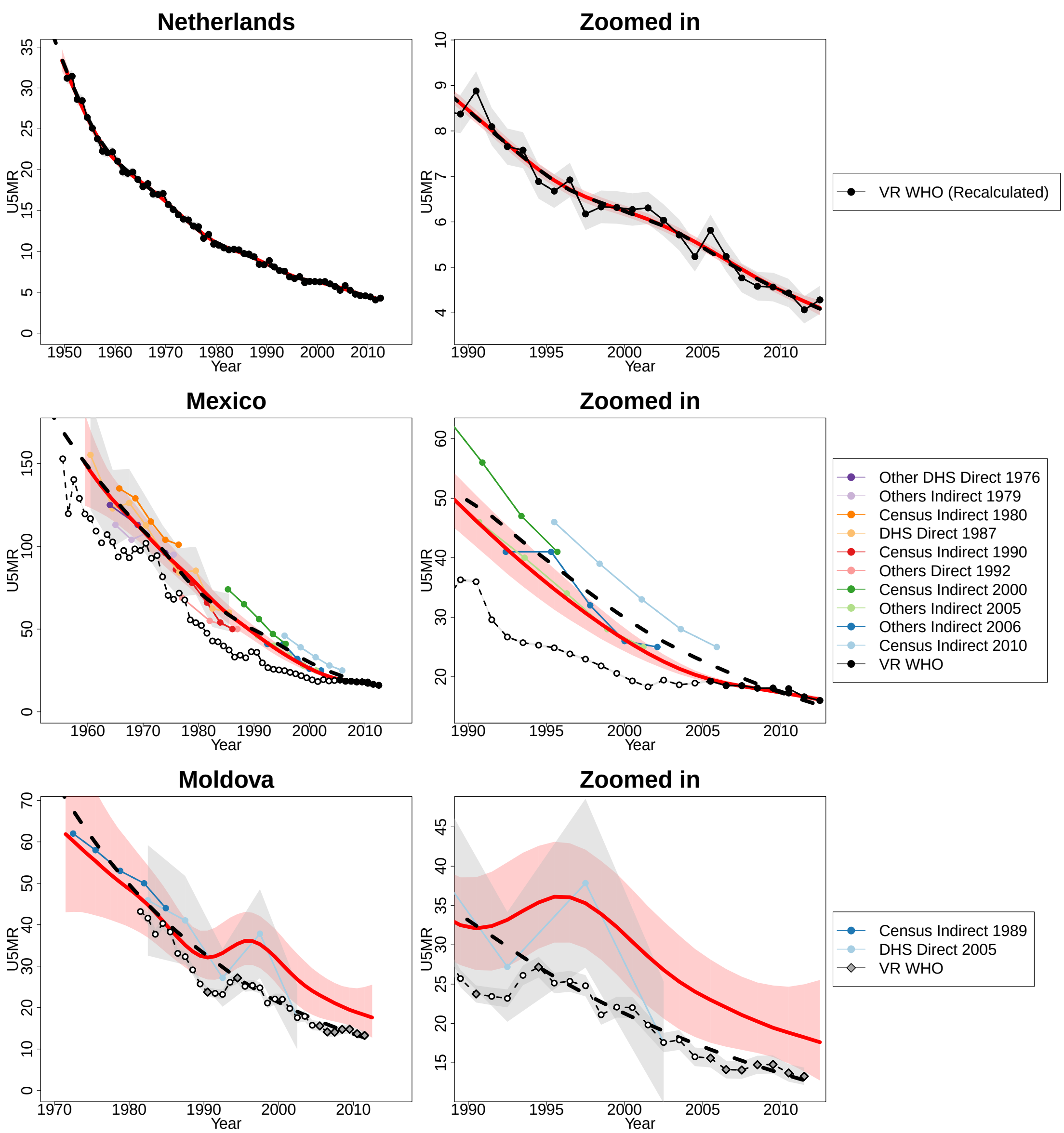

Fig 1: **U5MR data series and estimates for the Netherlands, Mexico and Moldova.** Connected dots represent data series from the same source, as explained in the legend. B3 estimates are illustrated by the solid red lines and 90% CIs are shown by the red shaded areas. The fitted Loess curve based on UN IGME 2012 methodology is illustrated with the black dashed line.

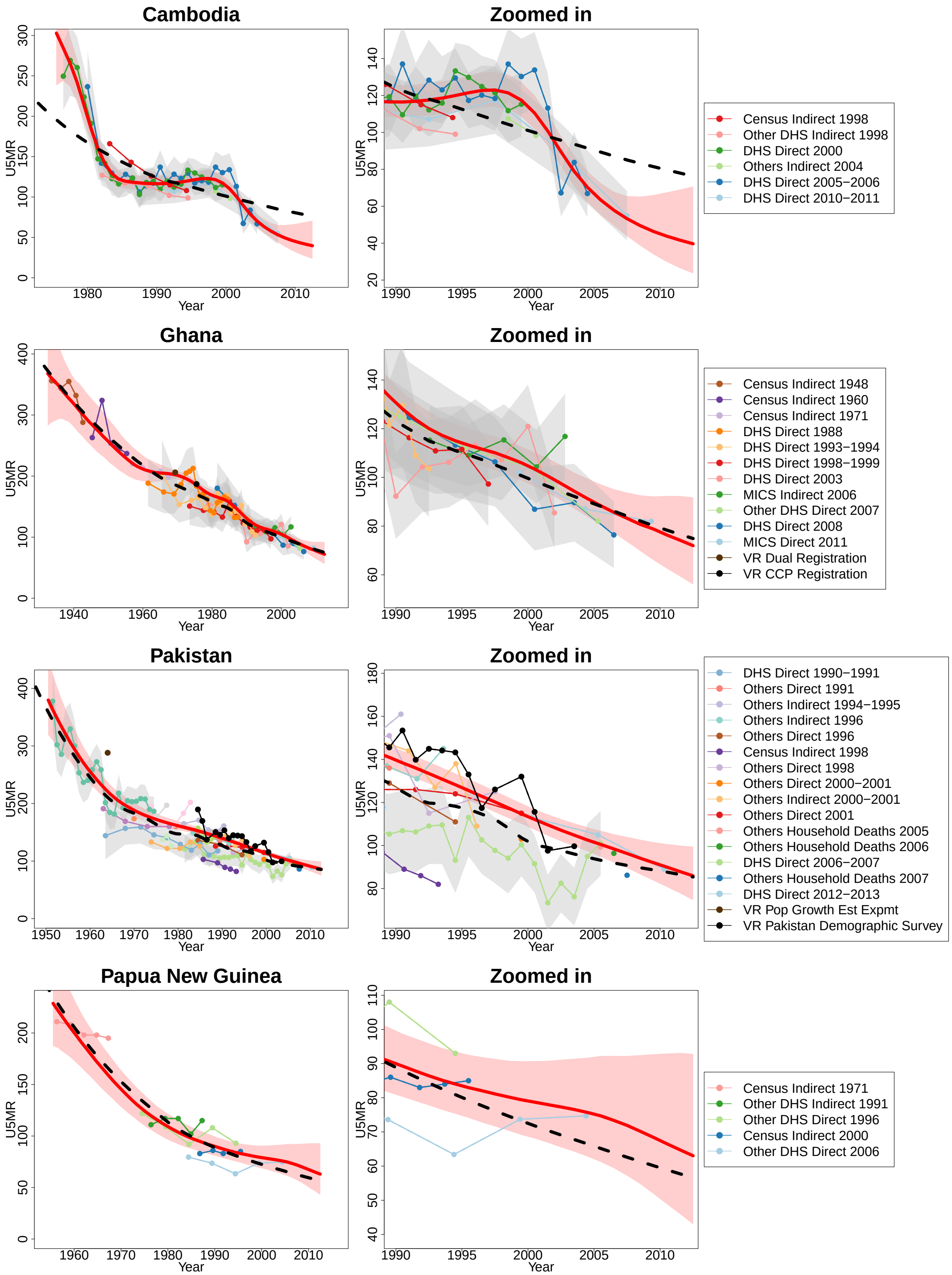

Fig 2: **U5MR data series and estimates for Cambodia, Ghana, Pakistan and Papua New Guinea.** Connected dots represent data series from the same source, as explained in the legend. B3 estimates are illustrated by the solid red lines and 90% CIs are shown by the red shaded areas. The fitted Loess curve based on UN IGME 2012 methodology is illustrated with the black dashed line.

Our analysis included all 195 countries for which the UN IGME publishes estimates except for the Democratic People's Republic of Korea (Korea DPR). This country was excluded because estimates cannot be constructed based on U5MR data alone (UN IGME uses estimates produced by the UN Population Division (United Nations, Department of Economic and Social Affairs, Population Division 2013)). For countries with high HIV prevalence, conflicts or natural disasters, we applied our proposed modeling approach with modifications based on the IGME 2012 estimation method, as explained in Section 3.3.

3.1. Bayesian penalized spline regression

The regression spline model for $f_c(t)$ from Eq.(1) is given by:

$$f_c(t) \quad = \quad \sum_{k=1}^{K_c} B_{c,k}(t)\alpha_{c,k}, \tag{2}$$

where $\alpha_{c,k}$ refers to spline coefficient k in country c and $B_{c,k}(t)$ the k-th spline, evaluated in year t, given by a third order B-spline (Eilers and Marx 1996; Eilers and Marx 2010).

B-splines are symmetric third-order polynomials that add up to unity in any given year, as illustrated in Figure 3 for Norway. Equally spaced knots were used such that the resulting splines are non-zero for a total of $4 \cdot I$ years, where I refers to the in-between-knots interval length. This distance was set using in-sample and out-of-sample validation; U5MR estimates were similar for interval lengths up to around 3 years. We chose $I = 2.5$ years such that each spline is non-zero for 10 years. In each country, the knots were placed such that the largest two splines $B_{c,K_c-2}(t)$ and $B_{c,K_c-1}(t)$ in the most recent observation year $t = t_{n_c}$ have equal height while $B_{c,K_c}(t)$ is close to zero (motivated in Section 3.6).

When fitting the spline model from Eq.(2) to the observations, second-order differences in adjacent spline coefficients ($\Delta^2\alpha_k = \alpha_k - 2\alpha_{k-1} + \alpha_{k-2}$) are penalized to guarantee smoothness of the resulting U5MR trajectory and the model was re-written to easily implement this penalization. Let $\boldsymbol{t}_c = (t_{c,1}, \ldots, t_{c,n_c})'$ refer to the vector of country-specific observation years with spline model estimate $f_c(\boldsymbol{t}_c) = (f(t_{c,1}), \ldots, f(t_{c,n_c}))'$. The spline model $f_c(\boldsymbol{t}_c) = \boldsymbol{B}_c(\boldsymbol{t}_c)\boldsymbol{\alpha}_c$, with $\boldsymbol{B}_c(\boldsymbol{t}_c) = (B_{c,1}(\boldsymbol{t}_c), \ldots, B_{c,K_c}(\boldsymbol{t}_c))$, $\boldsymbol{\alpha}_c = (\alpha_{c,1}, \ldots, \alpha_{c,K_c})'$, can be written as follows (Currie and Durban 2002; Eilers 1999; Eilers and Marx 2010):

$$\begin{aligned} f_c(\boldsymbol{t}_c) &= \boldsymbol{B}_c(\boldsymbol{t}_c)\boldsymbol{G}_{K_c}\boldsymbol{b}_c + \boldsymbol{Z}_c(\boldsymbol{t}_c)\boldsymbol{e}_c, \\ \boldsymbol{G}_{K_c} &= (\boldsymbol{1}_{K_c}\ \boldsymbol{g}_{K_c}),\ \text{where } \boldsymbol{g}_{K_c} = (1 - K_c/2, \ldots, K_c - K_c/2)', \\ \boldsymbol{Z}_c(\boldsymbol{t}_c) &= \boldsymbol{B}_c(\boldsymbol{t}_c)\boldsymbol{D}'_{K_c}(\boldsymbol{D}_{K_c}\boldsymbol{D}'_{K_c})^{-1}, \end{aligned} \tag{3}$$

where in difference matrix $\boldsymbol{D}_{K_c}$, $D_{K_c,i,i} = D_{K_c,i,i+2} = 1$, $D_{K_c,i,i+1} = -2$ and $D_{K_c,i,j} = 0$ otherwise.

The first part in Eq.(3), $\boldsymbol{B}_c(\boldsymbol{t}_c)\boldsymbol{G}_{K_c}\boldsymbol{b}_c$, describes the linear trend during the observation period, and the second part $\boldsymbol{Z}_c(\boldsymbol{t}_c)\boldsymbol{e}_c$ describes the fluctuations around the main trend. The unknown parameters are given by $\boldsymbol{b}_c = (b_{c,0}, b_{c,1})'$, and $\boldsymbol{e}_c = \boldsymbol{D}_{K_c}\boldsymbol{\alpha}_c$, where $\boldsymbol{e}_c = (e_{c,1}, \ldots, e_{c,Q_c})'$, with $Q_c = K_c - 2$ and $e_{c,q} = \Delta^2\alpha_{c,q+2}$ for $q = 1, \ldots, Q_c$. Second-order differences are penalized by imposing

$$e_{c,q} \quad \sim \quad N(0, \sigma_c^2),\ \text{for } q = 1, \ldots, Q_c, \tag{4}$$

where variance σ_c^2 determines the extent of smoothing; a smaller variance corresponds to smoother trajectories. In the limit when σ_c decreases to zero (as the penalty increases), a linear fit for log(U5MR) is obtained.

The model is fitted in the Bayesian framework. When estimating the spline coefficients, no information on levels or trends are exchanged across countries to avoid the situation where estimates for a country A with little information are pooled downwards because it is neighboring country B, where much progress has been made in reducing child mortality or vice versa. Information on spline coefficients is exchanged across countries only through a multilevel model for the variance of the differences in the spline coefficients, i.e. the variance of $e_{c,q} = \alpha_{c,q+2}$ - $2\alpha_{c,q+1} + \alpha_{c,q}$ is estimated hierarchically:

$$\log(\sigma_c) \quad \sim \quad N(\chi, \varphi_\sigma^2). \tag{5}$$

Spread out prior distributions are used for the $\boldsymbol{b}_c$'s and the hyperparameters for the hierarchical model for the σ_c's (see Appendix).

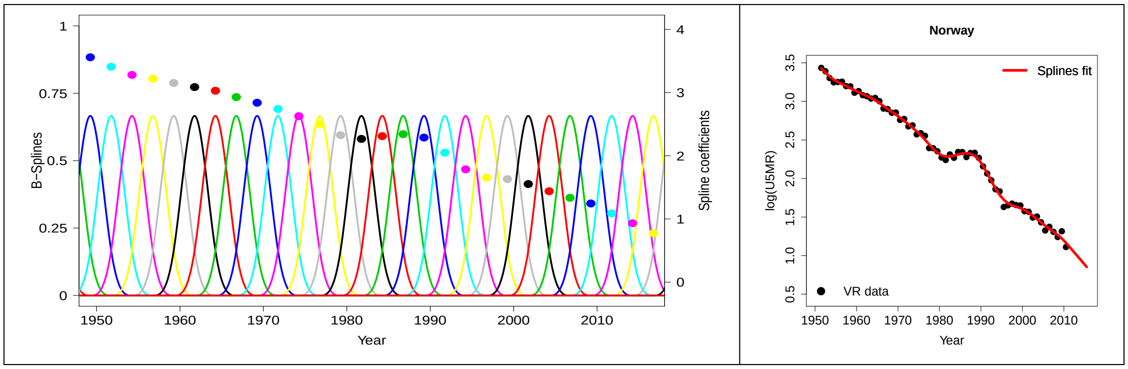

Fig 3: **Illustration of the B-spline regression model for Norway.** Left: B-splines and the estimated spline coefficients (dots, labeled on right axis). Right: Observed log(U5MR) plotted against time, together with a least-squares fit of the penalized spline model.

3.2. Database and data model

Under-five mortality data for all countries was taken from the UN IGME database. This database is publicly available on CME Info (http://www.child mortality.org). The 2013 version (used in this paper) is forthcoming.

Section 2 provided an introduction to U5MR data sources. A more detailed overview and explanation on data sources is given elsewhere (Hill et al. 2012). The break-down of the U5MR observations by main source types is given in Table 1. Based on potential differences in biases and non-sampling errors across data sources (explained further below), a distinction was made between series of observations from complete and summary birth histories (direct and indirect estimates), and observations based on different data sources and data collection methods (e.g. VR systems, records based on household deaths, and life tables obtained from reports).

	Number of data series	Number of observations
VR (including SVR)	110	2968
(Standard) DHS Direct (with reported sampling errors)	203	2902
(Standard) DHS Direct (without reported sampling errors)	15	56
Other DHS Direct (with reported sampling errors)	49	634
Other DHS Direct (without reported sampling errors)	25	107
MICS Indirect (with reported sampling errors)	55	248
MICS Indirect (without reported sampling errors)	20	80
Census Indirect	228	1074
Others Direct	144	507
Others Indirect	168	793
Others Household Deaths	56	56
Others Life Table	56	56

TABLE 1

Summary of the U5MR data series and observations in the UN IGME 2013 database by source type. *Note: "Other DHS" refers to non-standard demographic and health surveys, i.e. Special, Interim and National DHS, Malaria Indicator Surveys, AIDS Indicator Surveys and World Fertility Surveys.*

3.2.1. Data model

VR data The error distribution for observations from complete VR or SVR is given by

$$\delta_i \sim N(0, \tau_i^2/u_i^2),$$

where τ_i/u_i is the stochastic standard error. These errors are calculated using a Poisson approximation (using live birth numbers from the World Population Prospects (United Nations, Department of Economic and Social Affairs, Population Division 2011)) and set to a minimum of 0.025 (i.e. 2.5%).

If the Poisson sampling standard error cannot be calculated (e.g. for SVR data where the number of sampled live births is not given), it is set to 0.1 (i.e. 10%).

VR observations are typically calculated for single-year periods but longer periods were used for smaller countries in instances where the coefficient of variation of the observation was larger than 10% (due to small numbers of births and deaths).

Non-VR data For non-VR data, the data model needs to account for (i) sampling and non-sampling errors, (ii) potential biases in trends and levels of U5MR data series, and (iii) possibility of outliers. This is incorporated into the following model set-up:

$$\delta_i = E_i + S_i \cdot X_i,$$

where E_i is the mean bias, S_i the scale parameter and X_i determines the distribution for observation i.

For all non-VR data series with repeated observations, mean biases were modeled as a linear function of the retrospective period of the observation in the survey (the difference between the observation reference date and the date of the survey/census), motivated by known problems with retrospective data, such as the occurrence of recall biases and violations of modeling assumptions when calculating indirect U5MR observations:

$$E_i = \beta_{0,s[i]} + \beta_{1,s[i]} \cdot \pi_i, \tag{6}$$

where $\beta_{0,s[i]} + \beta_{1,s[i]} \cdot \pi_i$ represents the bias in level and trend as a function of the retrospective period π_i for observation i (centered at 10 years) in data series $s[i]$. The bias in the level of the series, $\beta_{0,s}$ is estimated with a multilevel model:

$$\beta_{0,s} \sim N(\mu_{0,d[s]}, \gamma^2_{0,d[s]}), \tag{7}$$

where $d[s]$ refers to the source type of series s, based on data source and U5MR calculation method (the source types with multiple observations per series are given by (Standard) DHS Direct, Other DHS Direct (including Special, Interim and National DHS, Malaria Indicator Surveys, AIDS Indicator Surveys and World Fertility Surveys), MICS Indirect, Census Indirect, Others Direct and Others Indirect), and $\mu_{0,d}$ and $\gamma^2_{0,d}$ represent source type-specific mean bias and between-series variance respectively. A similar approach is used to estimate the slope $\beta_{1,s}$:

$$\beta_{1,s} \sim N(\mu_{1,d[s]}, \gamma^2_{1,d[s]}), \tag{8}$$

where $\mu_{1,d}$ and $\gamma^2_{1,d}$ represent the mean slope and the between-series variance for source type d. For single observations constructed from reported household deaths, and single observations obtained from reported life tables, we assume that $E_i = \mu_{0,d[s[i]]}$.

Scale parameter S_i is modeled as a combination of sampling variance τ_i^2/u_i^2 (based on sampling variance τ_i^2 for U5MR) and non-sampling variance $\omega^2_{d'[s[i]]}$:

$$S_i^2 = \omega^2_{d'[s[i]]} + \tau_i^2/u_i^2, \tag{9}$$

where source type $d'[s]$ for series s refers to a further breakdown of source types to distinguish between DHS, Other DHS and MICS surveys with and without reported sampling errors for their observations (as indicated in Table 1). Where the sampling variance for non-VR data is not reported, we assume a sampling standard error of 2.5% for Census Indirect observations and 10% otherwise.

Finally, the distribution for δ_i is given by:

$$X_i \sim \begin{cases} N(0,1) & \text{for Standard and Other DHS direct,} \\ t_\nu & \text{otherwise, with } \nu \sim U(2,30). \end{cases} \tag{10}$$

A t-distribution with ν degrees of freedom is used for observations that are not obtained from Standard or Other DHS sources because posterior predictive checks suggested the presence of outliers.

All model parameters in Eq.(6)–(10) were assigned spread out prior distributions, with the exception of the mean bias $\mu_{0,d}$ for the DHS Direct series: an informative prior distribution was used, based on an analysis of these biases in the previous 2012 round of UN IGME estimates.

3.3. Countries with high HIV prevalence, conflicts or natural disasters

The estimation approach for countries with high HIV prevalence, crises or natural disasters follows the IGME 2012 approach. In short, the spline regression model is fitted to the "extreme event-free" observations, which are observed (and bias-adjusted for high HIV prevalence countries) U5MR values minus the extreme events U5MR. The B3 model is not fit to the (bias-adjusted) U5MR values directly to avoid situations where estimates or extrapolations may contradict trends related to the crisis mortality (e.g. AIDS U5MR changes as the AIDS epidemic progresses and with changing availability of anti-retroviral therapy). The uncertainty assessment for U5MR for the crises-years is based on the uncertainty in crisis-free U5MR: the relative uncertainty bounds for total U5MR are assumed to be equal to the relative uncertainty bounds for non-AIDS U5MR.

3.4. Country-specific model used by UN IGME

The Bayesian B-spline model was accepted by UN IGME to evaluate countries' progress and performance in reducing U5MR. For this purpose, a computationally cheaper and more user-friendly country-specific model was implemented, with non-country-specific parameters fixed at the posterior medians from the global model run, which resulted in very similar estimates. After reviewing the estimates, two model adjustments were incorporated in the country-specific models to consistently adjust the level of under- or over-smoothing in a subset of countries.

The first adjustment was applied to 1) countries with both VR and non-VR data, 2) countries that have VR data with gaps more than 5 years in the data, as well as 3) small countries with less than 10,000 live births in 2012 (based on World Population Prospects 2012 (United Nations, Department of Economic and Social Affairs, Population Division 2013). In such countries, the country-specific variance parameter σ_c tended to be large (there was little penalization of changes in spline coefficients, or in other words, there was little smoothing of the spline regression model fit) to be able to capture VR data precisely. This resulted in unrealistic short-term changes for a subset of country-years without VR data. For these countries, instead of using country-specific smoothing determined by variance parameter σ_c^2 from Eq.(5), we set $\sigma_c = \exp(\hat{\chi})$, where $\hat{\chi}$ refers to the posterior median of χ, which is referred to as the global smoothing level.

The second adjustment relates to countries with data series based on summary birth histories that were collected before 1975. Generally, these countries exhibit a number of characteristics: 1) only a very limited number of series were available in the early period, 2) there were data gaps between these early series and more recent ones and 3) these older series were much higher than the more recent ones. For several countries, the B3 estimate from the global model often ended up below the earlier series, even if these series were not considered to be more biased upwards than more recent series of this type. The estimates were pulled down because of the penalization in the spline model; often the drop in U5MR between the oldest and more recent data series was unusually steep as compared to changes in later periods. We tested out various approaches to avoid unrealistically low estimates in earlier periods and found that setting the level bias, $\beta_{0,s}$ from Eq. (7), of the older data series at the median estimated bias for a data series of that source type ($\beta_{0,s} = \hat{\mu}_{0,d[s]}$, where $\hat{\mu}_{0,d}$ refers to the posterior median of $\mu_{0,d}$) resulted in more plausible estimates for years in the past without affecting the more recent estimates. This approach was implemented consistently for all countries with data series collected before 1975.

A different set of adjustments was applied to 10 countries in the regional grouping of the Central and Eastern Europe/Commonwealth of Independent States (CEE/CIS), namely Armenia, Azerbaijan, Georgia, Kazakhstan, Kyrgyzstan, Moldova, Tajikistan, Turkmenistan, Ukraine and Uzbekistan. In these countries, VR data were deemed incomplete and generally excluded from the estimation procedure in previous rounds of UN IGME estimation, and thus excluded from the global B3 model. However, although not informative about the level of U5MR, these observations were deemed to provide information on U5MR trends in the early 1990s as well as for recent years. To make use of this information, we incorporated the option to use VR data to inform trend estimates in the country-specific B3 model.

During the early 1990s, in several CEE/CIS countries, data from the VR suggested a plateauing of or even an increase in U5MR. This is illustrated in Figure 1 for Moldova. This observed trend is deemed to reflect a true stagnation in progress in reducing U5MR. To use the observed trend in VR data to inform the U5MR estimates, we first selected the VR observation in 1990 and the maximum observed

VR observation from 1990 to 1995 in each country. We then included the selected observations into the model, with a country-specific bias parameter p_c assigned to both of them such that the observations would inform the trend in U5MR estimates but not the level as follows:

$$\begin{aligned} \delta_i &\sim N(\log(p_{c[i]}), \tau_i^2/u_i^2), \\ p_c &\sim U(0,1), \end{aligned}$$

where index $i = n_{c[i],1}, n_{c[i],2}$ in error δ_i refers to the two selected observations in country c.

For the most recent period starting from 2005, for a subset of CEE/CIS countries, U5MR extrapolations based on the global model either decreased below incomplete VR observations, or resulted in estimates far above VR observations for which a minimum level of completeness (ranging from 50% to 90%) was assumed. We fixed the U5MR discrepancies between the B3 extrapolations and VR data by including the VR observations as a minimum U5MR value into the model (accounting for stochastic errors). For selected countries where minimum levels of completeness could be determined, we also included an upper bound based on the estimated minimum level of completeness. An overview of completeness assumptions is included in the Appendix. VR-based upper and lower bounds were incorporated into the model by excluding any U5MR estimates for country c that fell outside the interval $(L^*_{c,t}, U^*_{c,t})$ in year t (the likelihood for these estimates is set to zero), where VR-based lower bound $L^*_{c,t}$ and upper bound $U^*_{c,t}$ are defined as functions of the VR observation y_i for the corresponding country-year:

$$\begin{aligned} y_i^* &\sim N(y_i, \tau_i^2/u_i^2), \\ L^*_{c[i],t[i]} &= y_i^*, \\ U^*_{c[i],t[i]} &= y_i^* - log(M_{c,t}), \end{aligned}$$

and $M_{c,t}$ refers to the minimum completeness.

Lastly, adjustments were applied to the Democratic Republic of Congo and Somalia, where the U5MR data are not deemed to be representative of the country's past. Specifically, B-splines corresponding to conflict periods where the U5MR is unlikely to have declined were combined such that only one spline coefficient is estimated for each conflict period. The resulting fit is constant for the conflict periods.

3.5. Computation

A Markov Chain Monte Carlo (MCMC) algorithm was employed to sample from the posterior distribution of the parameters in the global and country-specific models with the use of the software JAGS (Plummer 2003). For the global run which supplied global parameters for the country-specific runs, 6 parallel chains with different starting points were run with a total of 50,000 iterations in each chain. Of these, the first 10,000 iterations in each chain were discarded as burn-in and every 20th iteration after was retained. The resulting chains contained 2,000 samples each. For the country-specific runs, we ran 6 chains with a total of 35,000 iterations in each chain. Of these, the first 10,000 iterations in each chain were discarded as burn-in and every 20th iteration after was retained. The resulting chains contained 1,250 samples each.

Standard diagnostics checks (using trace plots, the Raftery and Lewis diagnostic (Raftery and Lewis 1992; Raftery and Lewis 1996) and the Gelman and Rubin diagnostic (Gelman and Rubin 1992)) were used to check convergence.

Estimates of relevant quantities are given by the posterior medians while 90% credible intervals (CIs) were constructed from the 5% and 95% percentiles of the posterior sample. Given the inherent uncertainty in U5MR estimates, 90% CIs are used by UN IGME instead of the more conventional 95% ones.

3.6. Extrapolation using a logarithmic pooling approach

While the penalized B-spline regression model as discussed in Section 3.1 performed well to estimate past U5MRs, the extrapolations based on the country-specific predictive distributions for the changes in spline coefficients were found to result in very wide projection intervals for countries lacking recent data and in potentially unrealistic U5MR projections when the most recently observed change in the $\alpha_{c,k}$'s corresponded to an extremely high or low rate of change (which is generally unlikely to be maintained over a longer projection period).

We implemented a logarithmic pooling procedure to combine country-specific posterior predictive distributions (PPDs) for changes in spline coefficients with a global PPD and verified whether this approach improved out-of-sample projections. This procedure was applied to modify the PPDs for $\alpha_{c,k}$ for $k = K_c, K_c+1, \ldots, P_c$, where P_c refers to the last spline in the projection period of interest. While α_{c,K_c} was among the spline coefficients that were included in the observation period up to year t_{n_c}, it was included in the set of "projected" coefficients to be pooled because its estimate is based mainly on an extrapolation of past changes ($B_{c,K_c}(t)$ is negligible as compared to $B_{c,K_c-2}(t)$ and $B_{c,K_c-1}(t)$ in $t = t_{n_c}$ based on a predefined knot placement).

The approach is summarized as follows: Let $\alpha^{(j)}_{c,k}$ denote j-th posterior sample of spline coefficient k for country c, $j = 1, \ldots, J$ and let $\Gamma^{(j)}_{c,k+1} = \Delta\alpha^{(j)}_{c,k+1} = \alpha^{(j)}_{c,k+1} - \alpha^{(j)}_{c,k}$, the j-th posterior sample of the differences between two adjacent spline coefficients. In line with the terminology used in the Bayesian melding approach (Poole and Raftery 2000), we refer to the country-specific PPD for $\Gamma^{(j)}_{c,K_c}$ defined by the penalized splines model as the model-induced PPD. The model-induced PPD is pooled with a global PPD for future changes in the spline coefficients. The global PPD is referred to as the direct (as opposed to induced) PPD on $\Gamma^{(j)}_{c,K_c}$. This PPD was based on the set of posterior median estimates of the $\Gamma^{(j)}_{c,k}$'s, $\hat{\Gamma}_{c,k}$ for $c = 1, \ldots C$ and $k = 2, \ldots, K_c - 1$ (during the observation period for each country). We used country-projection-step specific logarithmic pooling weights to obtain the same extent of pooling for all countries. For $P \geq 0$, the resulting draw $\tilde{\Gamma}^{(j)}_{c,K_c+P}$ from the pooled PPD for $\Gamma^{(j)}_{c,K_c+P}$ is given by:

$$\tilde{\Gamma}^{(j)}_{c,K_c+P} \quad \sim \quad N\left(\tilde{\mu}^{(j)}_{c,K_c+P}, (\tilde{\sigma}^{(j)}_{c,K_c+P})^2\right),$$

where

$$\begin{aligned}
\tilde{\mu}^{(j)}_{c,K_c+P} &= \kappa \cdot \mu_W + (1-\kappa) \cdot \tilde{\Gamma}^{(j)}_{c,K_c+P-1}, \\
(\tilde{\sigma}^{(j)}_{c,K_c+P})^2 &= \kappa \cdot \varphi^2_W + (1-\kappa) \cdot (\tilde{\sigma}^{(j)}_{c,K_c+P-1})^2,
\end{aligned}$$

with μ_W and φ^2_W equal to the median and variance of the $\hat{\Gamma}_{c,k}$'s respectively, and for $P = 0$, $\tilde{\Gamma}^{(j)}_{c,K_c-1} = \Gamma^{(j)}_{c,K_c-1}$ and $\tilde{\sigma}^{(j)}_{c,K_c+P-1} = \sigma^{(j)}_c$. The overall pooling weight $0 \leq \kappa \leq 1$ was chosen through an out-of-sample validation exercise (described in Section 3.7). Further details of the logarithmic pooling procedure are given in the Appendix.

3.7. Model validation

Model performance was assessed through an out-of-sample validation exercise. Given the retrospective nature of U5MR data and the occurrence of data in series, the training set was not constructed by leaving out observations at random, but based on all available data in some year in the past (Alkema et al. 2012); here 2006 was chosen. To construct the training set, all data that were collected in or after 2006 were removed. For example, if a DHS was carried out in 2006, all (retrospective) observations from that DHS were left out of the training set. Fitting the B3 model to the training set resulted in point estimates and CIs that would have been constructed in 2006 based on the proposed method. To validate model performance, we calculated various validation measures (mean/median errors, coverage and interval scores) based on the left-out observations and based on the estimates obtained from the full dataset and the estimates obtained from the training dataset.

For the left-out observations, errors are defined as $e_i = u_i - \tilde{u}_i$, where $\tilde{u}_i$ denotes the posterior median of the predictive distribution for a left-out observation u_i based on the training set. Coverage is given by $1/n \sum 1[u_i \geq l_{c[i]}(t[i])] \cdot 1[u_i \leq r_{c[i]}(t[i])]$, where n denotes the total number of left-out observations considered and $l_{c[i]}(t[i])$ and $r_{c[i]}(t[i])$ the lower and upper bounds of the 90% predictions intervals for the i-th observation. The (negatively oriented) interval score s_i for observation i is given by (Gneiting and Raftery 2007):

$$\begin{aligned}
s_i \quad = \quad & \left(\log(r_{c[i]}) - \log(l_{c[i]})\right) + 2/x \left(\log(l_{c[i]}) - y_i\right) \cdot 1\left[u_i < l_{c[i]}\right] \\
& + 2/x \left(y_i - \log(r_{c[i]})\right) \cdot 1\left[u_i > r_{c[i]}\right],
\end{aligned}$$

with significance level $x = 0.1$. This score combines the width of the prediction interval with a penalty for any intervals that do not contain the left-out observation. The validation measures were calculated for 100 sets of left-out observations, where each set consisted of a random sample of one left-out observation per country. Reported results include the median and standard deviation of the validation measures based on the outcomes in the 100 sets.

"Updated" estimates, denoted by $\widehat{U}_c(t)$ for country c in year t, refer to the U5MR estimates obtained from the full data set. The error in the estimate based on the training sample is defined as $e_{c,t} = \widehat{U}_c(t) - \widetilde{U}_c(t)$, where $\widetilde{U}_c(t)$ refers to the posterior median estimate based on the training sample, while relative error is defined as $e_{c,t}/\widehat{U}_c(t) \cdot 100$. Coverage and interval scores were calculated based on the lower and upper bound of the 90% CIs for U5MR obtained from the training set, denoted by $L_c(t)$ and $R_c(t)$ respectively. Coverage is defined as $1/m \sum 1[\widehat{U}_c(t) \geq L_c(t)] \cdot 1[\widehat{U}_c(t) \leq R_c(t)]$, where m denotes the total number of estimates considered. The (negatively oriented) interval score $S_{c,t}$ for country c in year t is given by:

$$\begin{aligned} S_{c,t} &= (\log(R_c(t)) - \log(L_c(t))) + 2/x\left(\log(L_c(t)) - \log(\widehat{U}_c(t))\right) \cdot 1\left[\widehat{U}_c(t) < L_c(t)\right] \\ &\quad +2/x\left(\log(\widehat{U}_c(t)) - \log(R_c(t))\right) \cdot 1\left[\widehat{U}_c(t) > R_c(t)\right], \end{aligned}$$

again with significance level $x = 0.1$. Coverage, mean/median errors and interval scores were also evaluated for the annual rate of reduction (ARR) from 1990 to 2005, whereby the interval score was calculated based on the (untransformed) ARR.

Particular attention was paid to the performance of the B3 model for the group of high mortality countries, where high here refers to a U5MR greater than 40 deaths per 1,000 births in 1990. This set was selected because of the importance of the UN IGME U5MR estimates for tracking progress in reducing child mortality. Crises years and HIV adjustments were not considered in the out-of-sample model validation because the calculation of crisis U5MR is not included in the B3 method (so it is not possible to reconstruct these estimates).

4. Results

4.1. Model validation and choice of pooling weight

To set the pooling weight κ (to combine the PPD of country-specific changes in spline coefficients with the global PPD), validation measures were obtained for $\kappa = 0, 0.1, \ldots, 0.6$, where $\kappa = 0$ corresponds to the "no-pooling" (country-specific) variant. In general, the effects of the pooling on posterior median estimates were small with respect to the changes in point estimates. An illustration of the default (unpooled) and pooled projections (using $\kappa = 0.5$) are shown in Figure 4 for Cambodia, Ghana and Papua New Guinea (PNG). The introduction of the pooling procedure increases the projections upwards in Cambodia and decreases them in Ghana and PNG but differences in point estimates are minor. Projection intervals do vary more across countries; the bounds are similar for weights 0 and 0.5 for Ghana but narrow down in Cambodia and change the level in PNG.

Model validation results based on the left-out observations and the comparison between estimates based on the training and full data set are shown in Tables 3 and 4 in the Appendix for the range of pooling weights. Differences in mean/median (absolute) errors are small. While for median errors, the comparison across the different weights varies by indicator, mean errors tend to decrease with increasing pooling weights. Coverage and interval width scores for left-out observations tend to improve slightly with increasing pooling weight (the pooled prediction intervals skew towards the direction of the observations). For the estimated U5MR, the percentage of updated estimates that fell below the 90% credible intervals tends to increase slightly with the pooling weight but mean interval scores for U5MR decrease significantly with increasing pooling weight. In particular, the mean interval score for the U5MR in 2005 and 2000 decreased from 2.2 to 1.6.

Based on these findings, we chose to apply the pooling. Because differences in validation outcomes were small when comparing the results for $\kappa = 0.5$ to those with $\kappa = 0.6$, and because of the convenient interpretation of $\kappa = 0.5$ (the projected mean and variance of the differences in the spline coefficients are the simple average of the country-specific and global estimates), we set $\kappa = 0.5$.

With this choice of κ, the model validation results for the B3 model show an improvement over those for the UN IGME 2012 estimation approach. In a similar validation exercise carried out for the UN

IGME 2012 estimation approach (Alkema and New 2012), the updated estimate of ARR for 1990-2005 (based on the full data set) was above the training 90% CI for 16% of the high mortality countries (11 out of 70 countries), and below for only 6% of those countries. This indicates that declines in U5MR were underestimated for a substantial proportion of high mortality countries. The same effect is observed in the validation results for the B3 model but to a much lesser extent, with only 9% of the updated upper bounds for the ARR being too low and 3% of the updated lower bounds being too high. Overall, the calibration measures are better with the B3 model. Specifically, the percentage of updated estimates falling below and above the 90% uncertainty intervals was 4% and 5% respectively for the U5MR in 2000 and 8% and 1% for the U5MR in 2005 in the B3 model. These percentages were 10% and 6% for the U5MR in 2000 and 17% and 7% for the U5MR in 2005 in the IGME 2012 estimation approach.

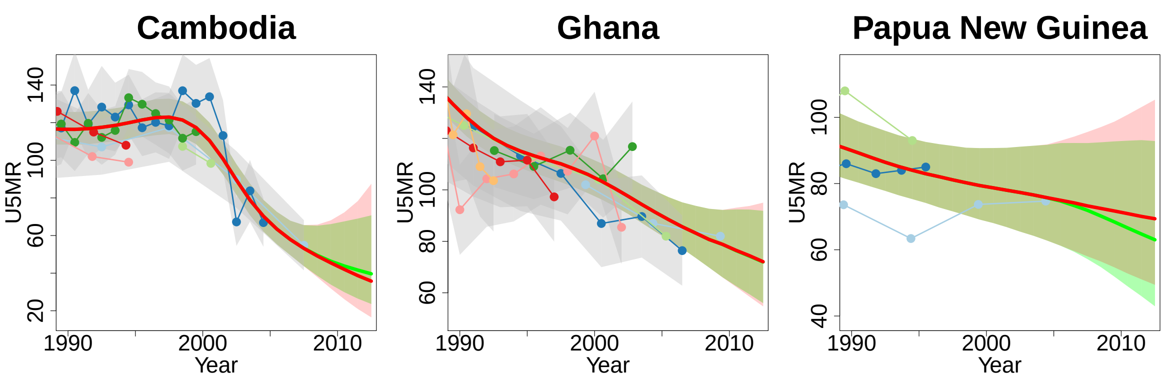

Fig 4: **Illustration of differences in estimates and projections for Cambodia, Ghana and Papua New Guinea between the unpooled (country-specific) and pooled B-spline model projection approach.** Estimates based on the unpooled (country-specific) approach are displayed in red, and results based on the pooling approach, using a weight of 0.5, are displayed in green. Solid lines denote posterior medians and shaded areas denote 90% CIs.

4.2. Data model biases

Mean biases in U5MR levels and trends, as well as 90% prediction intervals for the expected range of U5MR values for a "new" U5MR data series, given a "true" U5MR of 100 deaths per 1,000 live births, are shown in Figure 5 for the different source types, for retrospective periods of 5 and 15 years. For indirect series, the 90% prediction intervals based on uncertainty in biases alone (the dark blue horizontal lines) are wide, indicating substantial variability in biases across data series. For example, the prediction interval ranges from about 87 to about 143 deaths per 1,000 live births for an observation from a MICS Indirect series, with a retrospective period of 5 years. The error variance tends to contribute less to the width of the 90% prediction intervals, implying that there is significant variability in data series that is not attributed to random error. For retrospective periods of 5 years, mean biases are slightly positive for indirect series, but almost zero or negative for direct series: observations from direct series tend to be below indirect series for these retrospective periods.

4.3. U5MR estimates

B3 estimates for the selected illustrative countries in Figures 1 and 2 are displayed in the country-specific figures, together with the estimates that would have been obtained using the default Loess estimation approach used for constructing the IGME 2012 estimates.

Point estimates from the B3 model and default Loess are almost identical for the Netherlands during the entire observation period, but differ for all or a subset of observation years in the other countries. For Mexico, the trend in the Loess estimates for the late 2000s contradicts the observed trend in VR data. B3 estimates take into account the small stochastic error in the VR and follow the

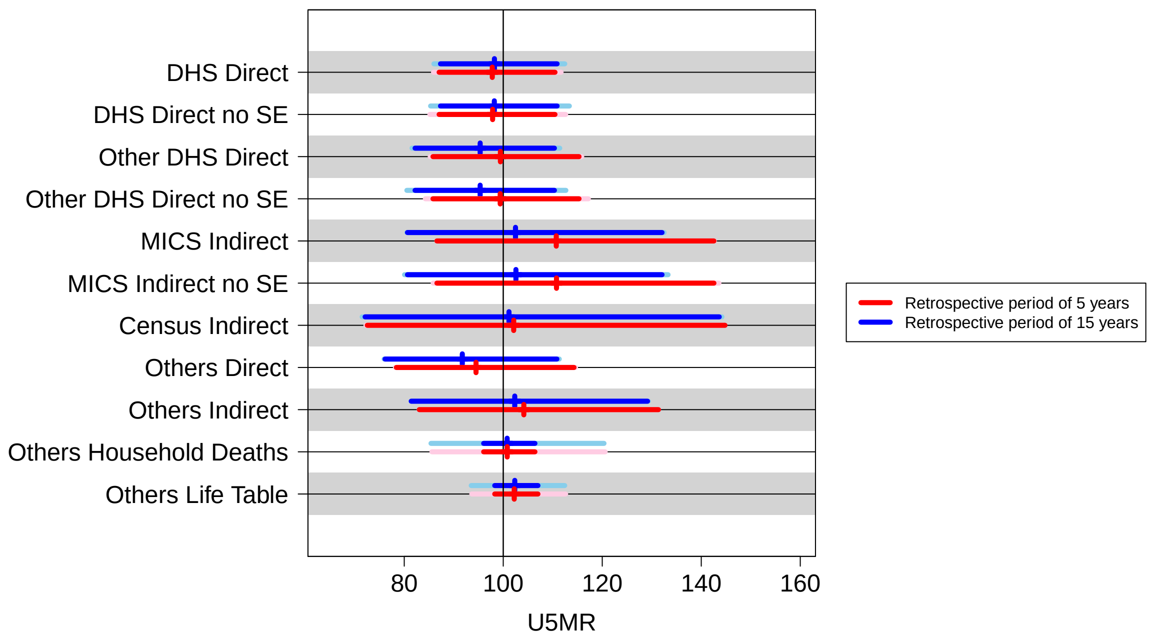

Fig 5: **Visualization of 90% prediction intervals for new data points by source type and retrospective period.** For a "true" U5MR of 100 deaths per 1,000 live births (represented by the black line), the 90% prediction interval for a U5MR observation with a retrospective period of 5/15 years is shown in light blue/pink (excluding the sampling variability) and the predicted mean observed U5MR is represented by the dark blue/red vertical line (the difference between the mean U5MR and 100 represents the mean bias). The dark blue/red horizontal line represents the 90% prediction interval for an observation based on uncertainty in the bias parameters only (excluding sampling and non-sampling variability).

data points closely. For Moldova, the inclusion of the VR observations in the early 1990s with a VR bias parameter for those years results in U5MR estimates that capture the VR-indicated trend. The inclusion of VR data for recent years guarantees that the point estimates and credible intervals do not cross through the VR. In future revisions for Moldova, a further extension could be to include all incomplete VR observations as a minimum to avoid the situation in the early 1980s, when the lower bound of the CI is below the incomplete VR.

For Ghana, B3 estimates and Loess estimates are similar. Small differences are observed in the years with VR data, where the B3 estimates capture these points while the Loess does not. In more recent years, the extrapolated decline is slightly steeper for the B3 model, as indicated by the decline in the most recent observations. Differences between B3 and Loess estimates are much larger in the other countries in the figure. In Cambodia, the B3 estimates follow the trend as observed in the data series, including the stagnation of child mortality decline in the 1980s and 1990s and the more recent acceleration in the decline of child mortality. The default Loess fit does not capture these fluctuations. In the IGME 2012 method, this country would be a candidate for an expert-based adjustment of the Loess smoothing parameter to better capture the trend. In the B3 penalized spline model approach, such expert adjustments are not necessary.

In Pakistan, the B3 estimates follow the registration data. The DHS from 2006–2007 does not bias downwards the estimates (as observed in the Loess estimates) because of the inclusion of bias parameters for survey data; we estimate that the DHS direct series is biased downwards. Lastly, in PNG, B3 estimates suggest a slightly flatter trend in U5MR than the Loess during the 1980s and 1990s based on the lack of downwards trends in all individual series during that period.

5. Discussion

The estimation of child mortality is challenging for the great majority of developing countries without well-functioning vital registration systems due to issues with data quantity and quality. In this paper, we described a Bayesian penalized B-spline regression model to evaluate levels and trends in the U5MR for all countries in the world. This model estimates biases in data series for all non-VR source types using a multilevel model to improve upon the limitations of current methods. Improved spline extrapolations are obtained via logarithmic pooling of the posterior predictive distribution of country-specific changes in spline coefficients with observed changes on the global level. The proposed model can flexibly capture changes in U5MR over time, provides point estimates and credible intervals that take into consideration potential biases in data series and indicates better model validation than the UN IGME 2012 estimation approach.

The differences between the B3 estimates and the default Loess fits as discussed in Section 4.3 highlight the need for more attention for appropriate data models in U5MR estimation. When treating all observations equally, U5MR estimates can end up below (incomplete) VR observations, or follow a trend in U5MR that is dictated by the (lack of) overlap of different data series with potentially different level biases.

While our data model overcomes the main limitations of the previous UN IGME estimation methods, there remains room for improvement. The primary issue with child mortality estimation is data quality. In the B3 data model, we incorporated source-specific bias parameters, that are drawn from a source type-specific distribution based on the assumptions that biases are comparable across data series of the same source type. However, large variation exists across series; ideally external information on data quality should be included to distinguish between the more or less reliable series in the database. A residual analysis conducted with respect to a number of data quality predictors (region that country belongs to, series source type, series year, observation year, retrospective period, level of U5MR in observation year, total fertility rate in the series year and change in the total fertility rate in the last 15 years before the series) suggested that the linear model for biases in direct and indirect series seemed to work reasonably well. However, for some DHS Direct series, an additional negative bias for observations with retrospective periods shorter than 5 years may be present. This may be due to birth transference, whereby dates of birth are incorrectly reported to avoid answering more questions pertaining to those births in the DHS questionnaires (Sullivan 2008). Given the importance of the observations with short retrospective periods in driving recent estimates and short-term projections, this issue needs to be investigated more in future work.

Ultimately, the issue of data quality is one that can only be resolved by implementing fully-functioning VR systems that can provide accurate data on births and deaths in every country. However, currently only about 50 countries have such VR systems in place; implementation of VR systems for all countries remains an ambitious and long-term goal (United Nations Children's Fund and USAID 2012). In the short term, the B3 model allows for inclusion of information from incomplete vital registration systems, as illustrated for Moldova. The inclusion of data from alternative data sources and the implementation of novel data collection methods, that can provide accurate and timely child mortality data (e.g. see Clark et al. (2012) and Amouzou (2011)), could further aid child mortality estimation. The advantage of the use of the Bayesian framework in the B3 model is that the model can be readily extended to incorporate such information into the estimation process.

To assess progress towards MDG 4, much focus is placed on the point estimates of the U5MR and annual rate of reduction despite the large uncertainty in estimates because communication of uncertainty in U5MR estimates is challenging (Oestergaard et al. 2013). To provide a straightforward inclusion of the uncertainty assessment into the MDG 4 progress assessment, countries could be categorized by whether the attainment of the MDG target of an ARR of 4.4% is considered to be unlikely, not clear, or likely based on the uncertainty intervals of ARR estimate (Alkema and New 2012).

Moving beyond the MDGs, the issue of inequality is likely to feature prominently in the post-2015 development agenda. While the MDGs have focused much attention on national, regional and global averages of key indicators, they have also potentially masked growing disparities at the intra-national level (UN System Task Team on the Post-2015 UN Development Agenda 2012). In light of this, disaggregated estimates of child mortality (e.g. by state, wealth quintile, residence) will be increasingly important to evaluate progress for all population groups to better address inequalities. Further work can be carried out to extend the B3 model so that this growing body of disaggregated data can be fully utilized to produce disaggregated estimates in the future.

Acknowledgements

The authors are very grateful to all members of the (Technical Advisory Group of the) United Nations Inter-agency Group for Child Mortality Estimation for passionate discussions about U5MR data and preliminary B3 estimates which have greatly improved this work. Additional thanks to Danzhen You, Patrick Gerland, Simon Cousens, Kenneth Hill, Kirill Andreev, François Pelletier, Bruno Masquelier, David Nott, Andrew Lover and Jakub Bijak for specific comments and suggestions related to the B3 model, and Jon Pederson, Jing Liu, Philip Bastian and Jingxian Wu for database construction and assistance on data issues. We also thank the numerous survey participants and the staff involved in the collection and publication of the data that we analysed. The research was funded by a research grant at the National University of Singapore and the United Nations Children's Fund. The views expressed in this paper are those of the authors and do not necessarily reflect the views of the United Nations Children's Fund.

References

Alkema, L. and J. R. New (2012, December). Progress toward global reduction in under-five mortality: A bootstrap analysis of uncertainty in millennium development goal 4 estimates. *PLoS Med 9*(12), e1001355.

Alkema, L., M. B. Wong, and P. R. Seah (2012, June). Monitoring progress towards Millennium Development Goal 4: a call for improved validation of under-5 mortality rate estimates. *Statistics, Politics and Policy 3*(2).

Amouzou, A. (2011, November). Real-time results tracking: Technical note. Technical note. Available at http://www.jhsph.edu/departments/international-health/centers-and-institutes/institute-for-international-programs/_documents/RRT_Technical_Note.pdf.

Brass, W. (1964). Uses of census or survey data for the estimation of vital rates. Paper presented at the African Seminar on Vital Statistics.

Clark, S. J., J. Wakefield, T. McCormick, and M. Ross (2012, September). Hyak mortality monitoring system: Innovative sampling and estimation methods. Working Paper 118. Available at http://www.csss.washington.edu/Papers/wp118.pdf.

Currie, I. D. and M. Durban (2002, December). Flexible smoothing with P-splines: a unified approach. *Statistical Modelling 2*(4), 333–349.

Eilers, P. H. C. (1999). Discussion of: Verbyla, A. P., B. R. Cullis, M. G. Kenward, and S. J. Welham. "The Analysis of Designed Experiments and Longitudinal Data Using Smoothing Splines.". *Journal of the Royal Statistical Society Series C*(48), 300–311.

Eilers, P. H. C. and B. D. Marx (1996). Flexible Smoothing with B-splines and Penalties. *Statistical Science 11*(2), 89–121.

Eilers, P. H. C. and B. D. Marx (2010, November). Splines, knots, and penalties. *Wiley Interdisciplinary Reviews: Computational Statistics 2*(6), 637–653.

Gelman, A. and D. Rubin (1992). Inference from iterative simulation using multiple sequences. *Statistical Science 7*, 457–511.

Gneiting, T. and A. E. Raftery (2007, March). Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association 102*(477), 359–378.

Hill, K., D. You, M. Inoue, and M. Z. Oestergaard (2012, August). Child mortality estimation: accelerated progress in reducing global child mortality, 1990-2010. *PLoS Medicine 9*(8), e1001303.

Oestergaard, M. Z., L. Alkema, and J. E. Lawn (2013). Millennium development goals national targets are moving targets and the results will not be known until well after the deadline of 2015. *International journal of epidemiology 42*(3), 645–647.

Pedersen, J. and J. Liu (2012, August). Child Mortality Estimation: Appropriate Time Periods for Child Mortality Estimates from Full Birth Histories. *PLoS Medicine 9*(8), e1001289.

Plummer, M. (2003). JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling. In *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003), March 20-22, Vienna, Austria. ISSN 1609-395X. http://mcmc-jags.sourceforge.net/.*

Poole, D. and A. E. Raftery (2000, December). Inference for Deterministic Simulation Models: The Bayesian Melding Approach. *Journal of the American Statistical Association 95*(452), 1244–1255.

Raftery, A. E. and S. M. Lewis (1992). How many iterations in the Gibbs sampler? In J. M. Bernardo et al. (Ed.), *Bayesian Statistics 4*, pp. 763–773. Oxford University Press.

Raftery, A. E. and S. M. Lewis (1996). Implementing MCMC. In W. R. Gilks, D. J. Spiegelhalter, and S. Richardson (Eds.), *Markov Chain Monte Carlo in Practice*, pp. 115–130. London: Chapman and Hall.

Rajaratnam, J. K., J. Marcus, A. Flaxman, H. Wang, A. Levin-Rector, L. Dwyer, M. Costa, A. Lopez, and C. Murray (2010). Neonatal, postneonatal, childhood, and under-5 mortality for 187 countries, 1970-2010: a systematic analysis of progress towards millennium development goal 4. *The Lancet 375 Issue 9730*, 1988–2008.

Sullivan, J. M. (2008, October). An assessment of the credibility of child mortality declines estimated from dhs mortality rates. Working Paper 1. Available at http://www.childinfo.org/files/Overall_Results_of_Analysis.pdf.

UN System Task Team on the Post-2015 UN Development Agenda (2012, May). Addressing inequalities: The heart of the post-2015 agenda and the future we want for all. Available at http://www.un.org/millenniumgoals/pdf/10_inequalities_20July.pdf.

United Nations (1983). *Manual X, indirect techniques for demographic estimation*, Chapter 2. United Nations, New York. http://www.un.org/esa/population/publications/Manual_X/Manual_X.htm.

United Nations Children's Fund and USAID (2012, December). *Real-Time Child Mortality Monitoring Meeting*, New York, New York. United Nations Children's Fund and USAID. Available at http://www.childrenandaids.org/newsletter/real-time-child-mortality-monitoring-meeting-december-19-2012/.

United Nations Children's Fund, Division of Policy and Strategy (2012). *Committing to Child Survival: A Promise Renewed Progress Report 2012*. United Nations Children's Fund.

United Nations, Department of Economic and Social Affairs, Population Division (2011). *World Population Prospects. The 2010 Revision.*

United Nations, Department of Economic and Social Affairs, Population Division (2013). *World Population Prospects. The 2012 Revision.*

United Nations Inter-agency Group for Child Mortality Estimation (2012). *Levels & Trends in Child Mortality: Report 2012*. New York: United Nations Children's Fund.

Wang, H., L. Dwyer-Lindgren, K. T. Lofgren, J. K. Rajaratnam, J. R. Marcus, A. Levin-Rector, C. E. Levitz, A. D. Lopez, and C. J. Murray (2012, December). Age-specific and sex-specific mortality in 187 countries, 1970–2010: a systematic analysis for the Global Burden of Disease Study 2010. *The Lancet 380*(9859), 2071–2094.

Appendix A

The full model specification is given as follows.

Splines model Let $\boldsymbol{f}_c = (f_{c,t_1}, \ldots, f_{c,T_c})$; the splines model is given by:

$$\boldsymbol{f}_c = \boldsymbol{B}_c \boldsymbol{G}_c \boldsymbol{b}_c + \boldsymbol{Z}_c \boldsymbol{e}_c,$$

where the first part describes the linear trend over the observation period, and the second part describes the fluctuations around the main trend. We assumed for the second part that

$$\begin{aligned} e_{c,q} &\sim N(0, \sigma_c^2), \\ \log(\sigma_c) &\sim N(\chi, \varphi_\sigma^2), \end{aligned}$$

such that the variance is modelled hierarchically.

Priors are given by:

$$\begin{aligned} \exp(b_{c,1}) &\sim U(1, 1000), \\ b_{c,2}/I &\sim U(-0.25, 0.2), \\ \chi &\sim N(-3, 10), \\ \varphi &\sim U(0, 5), \end{aligned}$$

where $\exp(b_{c,1})$ represents the level of U5MR in the approximate midyear of the observation period, and $b_{c,2}/I$ is the average ARR over the observation period (I is the interval length between knots).

Data model For VR data, we assumed

$$\delta_i \sim N(0, \tau_i^2/u_i^2),$$

where τ_i/u_i is the stochastic standard error.
For non-VR data, the data model is given by:

$$\begin{aligned} \delta_i &= E_i + S_i \cdot X_i, \\ E_i &= \begin{cases} \mu_{d[s[i]]} & \text{for single observations from reported household deaths and life tables,} \\ \beta_{0,s[i]} + \beta_{1,s[i]} \cdot \pi_i & \text{otherwise,} \end{cases} \\ S_i^2 &= \tau_i^2/u_i^2 + \omega^2_{d'[s[i]]}, \\ X_i &\sim \begin{cases} N(0,1) & \text{for DHS direct and Other DHS Direct,} \\ t_\nu & \text{otherwise, with } \nu \sim U(2, 30). \end{cases} \end{aligned}$$

Spread out prior distributions were assigned to all data model parameters, with the exception of the mean bias $\mu_{0,d}$ for the DHS Direct series, which has an informative prior distribution:

$$\begin{aligned} \boldsymbol{\mu}_d &\sim N(\boldsymbol{\mu_0}, \boldsymbol{\Sigma}_0), \\ \gamma_{0,d[s]} &\sim U(0, 5), \\ \gamma_{1,d[s]} &\sim U(0, 5), \\ \omega_{d'} &\sim U(0, 0.5), \end{aligned}$$

where $\boldsymbol{\mu_0} = (-0.0123, 0)$ and $\boldsymbol{\Sigma}_0 = \text{Diag}(0.00556^2, 0.02^2)$ for DHS Direct and $\boldsymbol{\mu_0} = \boldsymbol{0}$ and $\boldsymbol{\Sigma}_0 = \text{Diag}(0.15^2, 0.02^2)$ otherwise.

Assumed minimum levels of VR completeness For a subset of CEE/CIS countries, a minimum level of completeness was assumed for VR observations from the year 2005 onwards. The assumed level was 50% for Armenia, Georgia and Moldova and 90% for Kazakhstan, Kyrgyzstan and Ukraine.

Logarithmic pooling approach The posterior predictive distribution (PPD) for $\Gamma^{(j)}_{c,K_c} = \Delta\alpha^{(j)}_{c,k+1} = \alpha^{(j)}_{c,k+1} - \alpha^{(j)}_{c,k}$, based on (4), is given by:

$$\Gamma^{(j)}_{c,K_c} \sim N\left(\Gamma^{(j)}_{c,K_c-1}, (\sigma^2_{c,K_c-1})^{(j)}\right), \tag{11}$$

where $(\sigma^2_{c,K_c-1})^{(j)} = (\sigma^2_c)^{(j)}$. In line with the terminology used in the Bayesian melding approach (Poole and Raftery 2000), we refer to the PPD for $\Gamma^{(j)}_{c,K_c}$, defined by Eq.(11) as the model-induced posterior predictive distribution, with density function denoted by (leaving out superscripts to denote the posterior sample for notational convenience)

$$p^*(\Gamma_{c,K_c}) = h(\Gamma_{c,K_c}|\Gamma_{c,K_c-1}, \sigma^2_{c,K_c-1}),$$

where $h(\Gamma|\mu, \sigma^2)$ denotes the probability density function for a normal random variable with mean μ and variance σ^2. The model-induced PPD is pooled with a global PPD for future changes in the spline coefficients. The global PPD is referred to as the direct (as opposed to induced) PPD on Γ_{c,K_c}. The global PPD was based on the set of posterior median estimates of the $\Gamma^{(j)}_{c,k}$'s, $\hat{\Gamma}_{c,k}$ for $c = 1, \dots C$ and $k = 2, \dots, K_c - 1$ (during the observation period for each country):

$$p(\Gamma) = h(\Gamma|\mu_W, \varphi^2_W), \tag{12}$$

where μ_W and φ^2_W were given by the median and variance of the $\hat{\Gamma}_{c,k}$'s respectively.

Logarithmic pooling is used to combine both density functions:

$$\begin{aligned}\tilde{p}(\Gamma_{c,K_c}) &\propto p^*(\Gamma_{c,K_c})^{1-w_{c,K_c}} \cdot p(\Gamma_{c,K_c})^{w_{c,K_c}},\\ &= h(\Gamma_{c,K_c}|\tilde{\mu}_{c,K_c}, (\tilde{\sigma}_{c,K_c})^2),\end{aligned}$$

where w_{c,K_c} is the logarithmic pooling weight that determines the extent of pooling, and

$$\begin{aligned}\tilde{\mu}_{c,K_c} &= \tilde{\sigma}^2_{c,K_c} \cdot \left(w_{c,K_c} \cdot \mu_W/\varphi^2_W + (1 - w_{c,K_c}) \cdot \Gamma_{c,K_c}/\sigma^2_{c,K_c-1}\right),\\ \tilde{\sigma}^2_{c,K_c} &= \left(w_{c,K_c}/\varphi^2_W + (1 - w_{c,K_c})/\sigma^2_{c,K_c-1}\right)^{-1}.\end{aligned}$$

For $0 \le \kappa \le 1$, logarithmic pooling weights are given by:

$$w_{c,K_c} = \frac{\kappa \cdot \varphi^2_W}{\kappa \cdot \varphi^2_W + (1-\kappa)\sigma^2_{c,K_c-1}},$$

such that

$$\begin{aligned}\tilde{\mu}_{c,K_c} &= \kappa \cdot \mu_W + (1-\kappa) \cdot \Gamma_{c,K_c-1},\\ \tilde{\sigma}^2_{c,K_c} &= \kappa \cdot \varphi^2_W + (1-\kappa) \cdot \sigma^2_{c,K_c-1}.\end{aligned}$$

The logarithmic pooling weights (that depend on σ^2_{c,K_c-1}) were proposed to obtain the same extent of pooling for all countries; the mean of the pooled PPD's are weighted averages of the global mean and country mean, with weights given by κ and $1-\kappa$ respectively for all countries (regardless of the uncertainty in the most recent Γ_{c,K_c}). The same result holds true for the variance.

For $P \ge 1$, the induced PPD is defined as

$$p^*(\Gamma_{c,K_c+P}) = h(\Gamma_{c,K_c+P}|\tilde{\Gamma}_{c,K_c+P-1}, \tilde{\sigma}^2_{c,K_c+P-1}),$$

where $\tilde{\Gamma}_{c,K_c+P-1}$ refers to Γ_{c,K_c+P-1} drawn from its pooled PPD. With the global distribution from Eq.(12) and logarithmic pooling weights

$$w_{c,K_c+P} = \frac{\kappa \cdot \varphi^2_W}{\kappa \cdot \varphi^2_W + (1-\kappa)\tilde{\sigma}^2_{c,K_c+P-1}},$$

the pooled distribution for Γ_{c,K_c+P} is given by:

$$\begin{aligned}\tilde{p}(\Gamma_{c,K_c+P}) &\propto p^*(\Gamma_{c,K_c+P})^{1-w_{c,K_c+P}} \cdot p(\Gamma_{c,K_c+P})^{w_{c,K_c+P}},\\ &= h(\Gamma_{c,K_c+P}|\tilde{\mu}_{c,K_c+P}, \tilde{\sigma}^2_{c,K_c+P}),\\ \tilde{\mu}_{c,K_c+P} &= \kappa \cdot \mu_W + (1-\kappa) \cdot \tilde{\Gamma}_{c,K_c+P-1},\\ \tilde{\sigma}^2_{c,K_c+P} &= \kappa \cdot \varphi^2_W + (1-\kappa) \cdot \tilde{\sigma}^2_{c,K_c+P-1}.\end{aligned}$$

Validation results Validation results are described in Tables 2, 3 and 4.

	Year $\leq$ 2005		Year $>$ 2005	
κ	% Below	% Above	% Below	% Above
0	8.5 (2.6)	7.0 (1.9)	9.2 (1.2)	4.6 (1.9)
0.1	7.0 (2.6)	7.0 (2.0)	6.2 (1.2)	3.1 (1.3)
0.2	7.0 (2.6)	7.0 (2.0)	6.2 (1.2)	3.1 (1.2)
0.3	7.0 (2.5)	7.0 (2.0)	6.2 (1.2)	1.5 (1.0)
0.4	7.0 (2.5)	7.0 (1.9)	6.2 (1.3)	1.5 (1.0)
0.5	7.0 (2.4)	7.0 (1.8)	6.2 (1.5)	1.5 (1.0)
0.6	7.0 (2.5)	7.0 (1.7)	6.2 (1.5)	1.5 (1.0)

TABLE 2

Validation results based on left-out observations I. *Results refer to the median (and standard deviation) of outcomes based on 100 sets of left-out observations, where each set contains one randomly selected observation per included country (before/including 2005, and after 2005). Included countries are given by high mortality countries (high means U5MR greater than 40 deaths per 1,000 births in 1990) without crises or HIV adjustments, with data in both training and test set and left-out observations in the period of interest, 71 and 65 countries in total for the indicators left-out observations before and including 2005, and left-out observations after 2005 respectively. The outcome measures are: % of observations below and above the 90% prediction interval based on the training set.*

Year ≤ 2005					
	Median				
κ	ME	MAE	MRE	MARE	S
0	-2.2 (1.4)	11.3 (1.2)	-2.2 (1.8)	13.2 (1.3)	0.6 (0.0)
0.1	-2.2 (1.3)	11.2 (1.2)	-2.2 (1.7)	13.2 (1.3)	0.6 (0.0)
0.2	-1.9 (1.3)	10.9 (1.3)	-2.1 (1.7)	12.9 (1.4)	0.6 (0.0)
0.3	-1.9 (1.4)	10.8 (1.3)	-2.1 (1.7)	12.9 (1.4)	0.6 (0.0)
0.4	-1.9 (1.3)	10.8 (1.3)	-2.0 (1.7)	12.9 (1.4)	0.6 (0.0)
0.5	-1.7 (1.3)	10.7 (1.3)	-1.9 (1.7)	12.9 (1.4)	0.6 (0.0)
0.6	-1.5 (1.3)	10.6 (1.3)	-1.9 (1.7)	12.9 (1.5)	0.6 (0.0)
	Mean				
κ	ME	MAE	MRE	MARE	S
0	-2.9 (1.6)	16.6 (1.2)	-4.4 (1.8)	18.6 (1.5)	1.0 (0.1)
0.1	-2.5 (1.5)	16.4 (1.2)	-4.2 (1.7)	18.3 (1.4)	1.0 (0.1)
0.2	-2.3 (1.5)	16.2 (1.2)	-4.0 (1.7)	18.1 (1.4)	1.0 (0.1)
0.3	-2.2 (1.4)	16.1 (1.2)	-3.9 (1.7)	17.9 (1.4)	1.0 (0.1)
0.4	-2.0 (1.4)	16.0 (1.2)	-3.9 (1.6)	17.7 (1.4)	1.0 (0.1)
0.5	-1.9 (1.4)	15.8 (1.1)	-3.8 (1.6)	17.6 (1.3)	1.0 (0.1)
0.6	-1.9 (1.4)	15.6 (1.1)	-3.7 (1.6)	17.4 (1.3)	1.0 (0.1)
Year > 2005					
	Median				
κ	ME	MAE	MRE	MARE	S
0	-3.6 (0.4)	10.4 (0.6)	-10.2 (1.1)	17.6 (0.7)	0.9 (0.0)
0.1	-3.6 (0.3)	9.1 (1.0)	-9.6 (0.7)	17.9 (0.8)	0.9 (0.0)
0.2	-3.7 (0.2)	8.4 (1.5)	-8.8 (1.2)	17.3 (0.7)	0.9 (0.0)
0.3	-3.5 (0.2)	7.6 (1.6)	-9.2 (1.3)	17.1 (0.9)	0.9 (0.0)
0.4	-3.6 (0.1)	7.3 (1.4)	-10.0 (0.9)	17.2 (1.3)	0.9 (0.0)
0.5	-3.7 (0.1)	7.6 (1.2)	-10.7 (1.1)	17.5 (1.5)	0.9 (0.0)
0.6	-3.8 (0.2)	7.9 (1.0)	-10.3 (1.0)	18.2 (1.3)	0.8 (0.0)
	Mean				
κ	ME	MAE	MRE	MARE	S
0	-8.1 (0.5)	18.3 (0.4)	-15.7 (1.7)	30.0 (1.6)	1.4 (0.1)
0.1	-7.1 (0.5)	17.0 (0.5)	-14.8 (1.5)	28.3 (1.4)	1.3 (0.1)
0.2	-6.6 (0.5)	16.0 (0.5)	-14.6 (1.4)	27.2 (1.3)	1.3 (0.1)
0.3	-6.2 (0.5)	15.2 (0.5)	-14.6 (1.3)	26.6 (1.2)	1.3 (0.1)
0.4	-6.1 (0.5)	14.7 (0.5)	-14.7 (1.3)	26.1 (1.2)	1.3 (0.1)
0.5	-6.0 (0.5)	14.2 (0.5)	-14.8 (1.2)	25.8 (1.1)	1.3 (0.1)
0.6	-5.9 (0.5)	13.9 (0.5)	-14.8 (1.2)	25.6 (1.1)	1.3 (0.1)

TABLE 3

Validation results based on left-out observations II. *Results refer to the median (and standard deviation) of outcomes based on 100 sets of left-out observations, where each set contains one randomly selected observation per included country (before/including 2005, and after 2005). Included countries are given by high mortality countries (high means U5MR greater than 40 deaths per 1,000 births in 1990) without crises or HIV adjustments, with data in both training and test set and left-out observations in the period of interest, 71 and 65 countries in total for the indicators left-out observations before and including 2005, and left-out observations after 2005 respectively. The outcome measures are: median or mean relative error (MRE), median or mean absolute relative error (MARE), median or mean and interval score (S) based on the training set.*

U5MR 2000								
	Median			Mean				
κ	MRE	MARE	S	MRE	MARE	S	% Below	% Above
0	-2.4	4.5	1.5	-4.8	9.9	2.2	3.8	5.1
0.1	-2.4	4.5	1.6	-4.6	9.7	2.1	3.8	5.1
0.2	-2.4	4.5	1.5	-4.5	9.6	1.9	3.8	5.1
0.3	-2.4	4.5	1.4	-4.3	9.3	1.8	3.8	5.1
0.4	-2.5	4.5	1.3	-4.2	9.2	1.7	3.8	5.1
0.5	-2.4	4.4	1.2	-4.1	9.0	1.6	3.8	5.1
0.6	-2.5	4.5	1.1	-4.0	8.8	1.5	3.8	5.1
U5MR 2005								
	Median			Mean				
κ	MRE	MARE	S	MRE	MARE	S	% Below	% Above
0	-5.0	10.4	1.5	-11.0	18.9	2.2	6.4	5.1
0.1	-4.7	9.4	1.6	-10.2	17.5	2.1	6.4	3.8
0.2	-4.8	8.9	1.5	-9.6	16.4	1.9	6.4	2.6
0.3	-5.3	8.1	1.4	-9.3	15.7	1.8	7.7	2.6
0.4	-5.3	8.1	1.3	-9.0	15.2	1.7	7.7	2.6
0.5	-6.1	8.1	1.2	-8.9	14.7	1.6	7.7	1.3
0.6	-6.0	8.5	1.1	-8.7	14.4	1.5	7.7	2.6
ARR 1990-2005								
	Median			Mean				
κ	ME	MAE	S	ME	MAE	S	% Below	% Above
0	0.2	0.7	3.5	0.3	1.0	5.7	5.1	7.7
0.1	0.2	0.6	3.7	0.3	1.0	5.5	5.1	6.4
0.2	0.2	0.5	3.7	0.3	0.9	5.3	3.8	7.7
0.3	0.1	0.6	3.7	0.3	0.9	5.2	3.8	7.7
0.4	0.1	0.5	3.8	0.3	0.8	5.1	2.6	9.0
0.5	0.2	0.5	3.7	0.3	0.8	5.0	2.6	9.0
0.6	0.2	0.5	3.7	0.3	0.8	5.0	2.6	9.0

TABLE 4

Validation results for U5MR and ARR estimates. *Results refer to high mortality countries (high means U5MR greater than 40 deaths per 1,000 births in 1990) without crises or HIV adjustments, with data in both training and test set, 78 countries in total. Median and mean outcome measures are reported for the U5MR in 2000 and 2005, and the annual rate of reduction (ARR) from 1990 to 2005. Outcome measures are given by: median/mean relative error (MRE) and median/mean absolute relative error (MARE) for the U5MR, median or mean error (ME) and median/mean absolute error (MAE) for the ARR, and median/mean interval score S as well as % of countries below and above the 90% uncertainty intervals based on the training set.*