

# ESTIMATING THE STILLBIRTH RATE FOR 195 COUNTRIES USING A BAYESIAN SPARSE REGRESSION MODEL WITH TEMPORAL SMOOTHING

BY ZHENGFAN WANG<sup>1,a</sup>, MIRANDA J. FIX<sup>2,c</sup>, LUCIA HUG<sup>3,e</sup>, ANU MISHRA<sup>3,f</sup>,  
DANZHEN YOU<sup>3,g</sup>, HANNAH BLENCOWE<sup>4,h</sup>, JON WAKEFIELD<sup>2,d</sup> AND  
LEONTINE ALKEMA<sup>1,b</sup>

<sup>1</sup>*Department of Biostatistics and Epidemiology, University of Massachusetts Amherst, <sup>a</sup>zhengfanwang@umass.edu, <sup>b</sup>lalkema@umass.edu*

<sup>2</sup>*Department of Biostatistics, University of Washington, <sup>c</sup>miranda.fix@gmail.com, <sup>d</sup>jonno@uw.edu*

<sup>3</sup>*Division of Data, Analytics, Planning and Monitoring, UNICEF, <sup>e</sup>lhug@unicef.org, <sup>f</sup>a.mishra@imperial.ac.uk, <sup>g</sup>dyou@unicef.org*

<sup>4</sup>*London School of Hygiene and Tropical Medicine, University of London, <sup>h</sup>annah-jayne.blencowe@lshtm.ac.uk*

Estimation of stillbirth rates globally is complicated because of the paucity of reliable data from countries where most stillbirths occur. We compiled data and developed a Bayesian hierarchical temporal sparse regression model for estimating stillbirth rates for 195 countries from 2000 to 2019. The model combines covariates with a temporal smoothing process so that estimates are data-driven in country-periods with high-quality data and determined by covariates for country-periods with limited or no data. Horseshoe priors are used to encourage sparseness. The model adjusts observations with alternative stillbirth definitions and accounts for various sources of uncertainty. In-sample goodness of fit and out-of-sample validation results suggest that the model is reasonably well calibrated. The model is used by the UN Inter-agency Group for Child Mortality Estimation to monitor the stillbirth rate for 195 countries.

**1. Introduction.** The United Nations Inter-agency Group for Child Mortality Estimation (UN IGME) defines a stillbirth as a baby born with no signs of life at 28 weeks or more of gestation (UN Inter-agency Group for Child Mortality Estimation (2020)), consistent with the International Classification of Diseases (ICD-11, World Health Organization (2019)) definition of a “late gestation fetal death.” Prior estimates highlighted the large global burden of stillbirths with an estimated 2.6 million stillbirths for the year 2015 (Blencowe et al. (2016)). Ending preventable stillbirths is one of the core goals of the UN’s Global Strategy for Women’s, Children’s and Adolescents’ Health from 2016 until 2030 (Kuruville et al. (2016)) and the Every Newborn Action Plan (ENAP, World Health Organization (2014)). These global initiatives aim to reduce the stillbirth rate (SBR, the number of stillbirths per 1000 total births) to 12 or fewer stillbirths per 1000 births in every country by 2030.

Monitoring of SBRs is challenging because of data paucity in countries where most stillbirths occur. Estimates of SBRs for a country can be derived from administrative data from registration systems (e.g., civil registration and vital statistics (CRVS) and medical birth and death registries). The reliability of SBR estimates from such data sources depends on the accuracy and completeness of reporting and recording of stillbirths and live births. Not all countries maintain an accurate, timely, and complete registration system for stillbirths. Moreover, in many low- and middle-income countries (LMICs), stillbirths are not reported in registration systems at all. For such countries, stillbirth data can be obtained from health management information systems (HMIS) with limitations similar to the registration systems: stillbirth

data from registries and HMIS may be reported in different stillbirth definitions and may be biased due to underreporting, misclassification, and other data quality issues. Lastly, SBR data can be obtained from household surveys and population-based studies but—in addition to limitations similar to the other data sources regarding definitions—these data are typically not available for all years of interest and may be subject to potentially large biases and/or nonsampling errors.

Blencowe et al. (2016) produced estimates of the SBR for all countries from 2000 to 2015. Yearly estimates for developed countries with high-quality data were obtained from the data directly, using a Loess smoother. Estimates for all other countries were obtained from a regression model with country-specific intercepts and global regression coefficients. The main limitation of this work is the use of the regression model for countries with limited data: resulting trend estimates are covariate-driven, even if available data suggest deviations away from covariate-predicted trends. In addition, a stepwise approach was taken to carry out variable selection which underestimates uncertainty since the model selection process is not accounted for.

In this paper we propose a new approach to estimating the SBR for all countries, using a Bayesian hierarchical temporal sparse regression model (BHTSRM). The model is used by the UN IGME to monitor the SBR globally (UN Inter-agency Group for Child Mortality Estimation (2020), Hug et al. (2021)). Our approach updates and extends the work of Blencowe et al. (2016). As its name implies, BHTSRM combines a hierarchical regression model with a temporal smoothing process. This type of model produces estimates that track high-quality data while producing covariate-driven trend estimates for countries with limited or no SBR data. While this kind of model has been used for estimating global health indicators in other settings, for example, in Alkema et al. (2017), prior work does not address sparsity. Here, we extend upon previous work by introducing sparsity-inducing priors for estimating regression coefficients. In particular, we use horseshoe priors (Piiironen and Vehtari (2017a)) to shrink the less important coefficients toward zero which makes BHTSRM an approach that can deal with a large number of covariates.

As compared to Blencowe et al. (2016), our proposed model also introduces new statistical approaches to address various data quality issues. First, we propose a statistical procedure for data exclusion based on comparing observed ratios of SBR to the neonatal mortality rate (NMR). Second, we add to the model an estimation approach to incorporate observations with alternative definitions of a stillbirth (e.g., based on 22 weeks gestational age or 1000 grams birthweight) while accounting for the additional uncertainty associated with such observations.

This paper is organized as follows: in Section 2 we provide an overview of data sources and definitions that are available for measuring SBR. We introduce the exclusion of data based on the ratio of SBR to NMR in Section 3. We describe the SBR estimation model in Section 4, including the BHTSRM. In Section 5 we present estimates of SBR, data quality parameters, and validation results. Last, we conclude with a discussion of limitations and future research directions in Section 6.

## 2. Data.

2.1. *Database construction.* SBR data were compiled by the UN IGME from various sources for the year 2000 and onward. The majority of data collected on stillbirths were obtained from administrative data systems and health management information systems (HMIS). UN IGME conducts an annual country consultation to solicit up-to-date administrative data on stillbirths from ministries of health or national statistics offices. Population-based study data were obtained from a review of the academic literature and a WHO data call to maternal-newborn health experts. Nationally representative household surveys (e.g., demo-

graphic and health surveys, multiple indicator cluster surveys, reproductive health surveys) are another source of stillbirth data.

After data were compiled, general exclusion rules were applied. The evaluation and assessment for data quality were applied to all data sources based on predefined exclusion criteria. Data were excluded if they lacked information on definition or data collection systems, if the proportion of reported stillbirths with unknown gestational age or birthweight was above 50 per cent, if data were internally inconsistent, or if coverage of live births in administrative data systems was estimated as below 80 per cent. Vital registration data with incomplete coverage of child deaths were also excluded, where incompleteness was taken from the WHO CRVS completeness assessment (WHO Department of Information, Evidence, and Research (2018)).

*2.2. Notation.* We use lowercase Greek letters for unknown parameters and uppercase Greek letters for variables which are functions of unknown parameters (modeled estimates). Roman letters indicate variables that are known or fixed, including data (in lowercase) and estimates provided by other sources or the literature (in uppercase).

Data compilation and general exclusion resulted in a global database of observed SBR values. Observations are available across countries over time and are indexed by  $i$ ; For each  $i$ ,  $c[i]$  refers to the country for which the  $i$ th observation was recorded and  $t[i]$  to the calendar year of observation  $i$ . Index  $j[i]$  is used to refer to the source category of observation  $i$ . We define an observed value  $y_i$  as the SBR calculated from the number of reported stillbirths  $z_i$  and number of live births  $q_i$  from a given source for a country-period with  $y_i = z_i / (z_i + q_i)$ . Periods refer to calendar years when available, or longer if the source does not provide information on annual SBR. In the database, data source types are categorized as: (1) administrative data, (2) HMIS data, (3) household survey data, and (4) population-based study data. Among population-based studies, we distinguish between population-based prospectively-collected data, with recruitment prior to 28 weeks of gestation, and follow-up to, at least, 28 days for live births, referred to here as PopPros data (Bose et al. (2015), Ahmed et al. (2018)) and additional data (PopLR).

We denote the set of all available observations resulting after the general exclusion step as  $\mathcal{B}$ . The data set  $\mathcal{B}$  forms the basis of all analyses, as outlined in Figure 1. First, an exclusion procedure is introduced for observations in the global data set  $\mathcal{B}$  based on the ratio of SBR  $y_i$  to NMR  $o_i$ . The NMR  $o_i$  is calculated from the number of neonatal deaths  $m_i$  and number of live births  $q_i$  with  $o_i = m_i / q_i$ . The ratio of SBR to NMR is analyzed using the PopPros data set  $\mathcal{P}$ . The details of the exclusion are described in Section 3. Other subsets of data set  $\mathcal{B}$  are used for fitting the definition adjustment model and the SBR estimation model.

To allow for international comparison, we focus on estimating SBRs reported using the standard definition (gestational age  $\geq 28$  weeks). In fitting the SBR model, we used data based on the standard definition when available. However, for a subset of country-periods in  $\mathcal{B}$ , stillbirths were reported using an alternative definition only, based on birthweight or a different gestational age cut-off. Four kinds of alternative definitions are incorporated in the analysis: definitions referring to a baby born with no signs of life at: (1) 24 weeks or more of gestation, (2) 22 weeks or more, (3) birthweight  $\geq 1000$  grams, and (4) birthweight  $\geq 500$  grams. To use these observations for estimating the SBR, we estimated adjustments and uncertainties associated with alternative definition  $d$  using the definition adjustment data set  $\mathcal{D}^d$ . The data set and definition adjustment model are given in Section 4.3.

We denote the subset of observations used for SBR estimation by  $\mathcal{B}^-$ . This database is obtained after: (i) excluding observations that are identified as outlying based on the SBR to NMR ratio exclusion approach and (ii) selecting a subset of country-period-specific data in cases where multiple observations are available for the same country-period; see Figure 1.

**Data Sets and Exclusion Steps:**

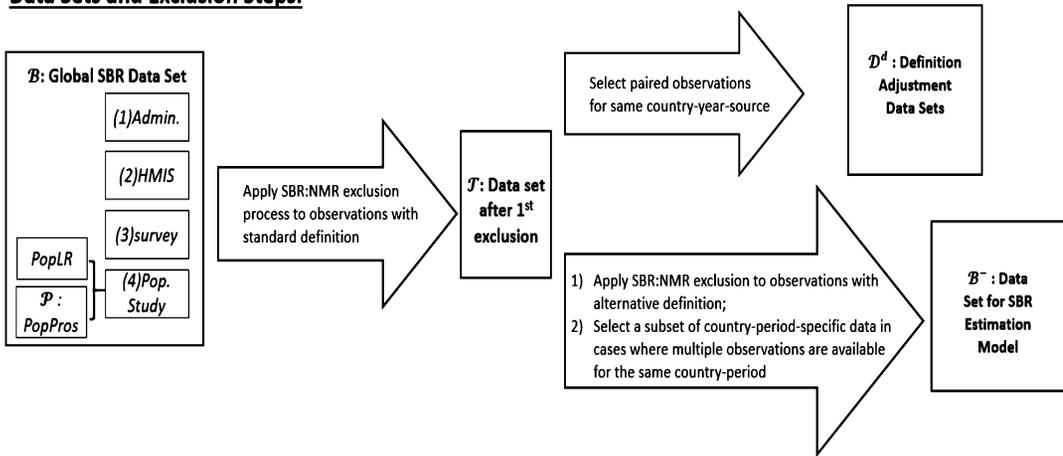


FIG. 1. *Data sets and exclusion steps.* This chart summarizes the data sets used for estimating the SBR. Data sets are indicated in rectangle boxes, and the processing steps are summarized by the thick arrows. The global SBR data set  $\mathcal{B}$  consists of administrative data (“Admin.”), HMIS, survey and population studies (“Pop. study”), including population-based prospectively data (“PopPros”), and additional data (“PopLR”).

The approach in (ii) is as follows: if observations are recorded in multiple definitions, we select only one definition based on the following order of preference: (1) standard definition, (2) birthweight  $\geq 1000$  grams, (3) 22 weeks or more of gestation, (4) 24 weeks or more of gestation, and (5) birthweight  $\geq 500$  grams. There are 1531 observations from 133 countries in this SBR model data set  $\mathcal{B}^-$ . Table 1 summarizes the breakdown of observations based on definition and source.

Data availability is illustrated for selected countries in Figure 2. Data availability ranges in the selected countries from no included data in Afghanistan to an annual time series of national administrative data based on the standard 28 weeks definition for Ireland. Botswana,

TABLE 1  
*Data set  $\mathcal{B}^-$  used for fitting the SBR estimation model by source and definition for countries in 2000–2019. For example, there are 75 countries with administrative data. “28 weeks” represents the standard definition. “22 weeks” and “24 weeks” represent 22/24 weeks or more of gestation; “500 grams” and “1000” grams represent birthweight  $\geq 500/1000$  grams*

Data Source	Number of Countries	Number of Obs
Administrative	75	1157
HMIS	26	162
Household Survey	44	95
Population Study	23	117

Definition	Number of Countries	Number of Obs
28 weeks	124	1220
24 weeks	3	44
22 weeks	15	85
1000 grams	20	146
500 grams	5	36

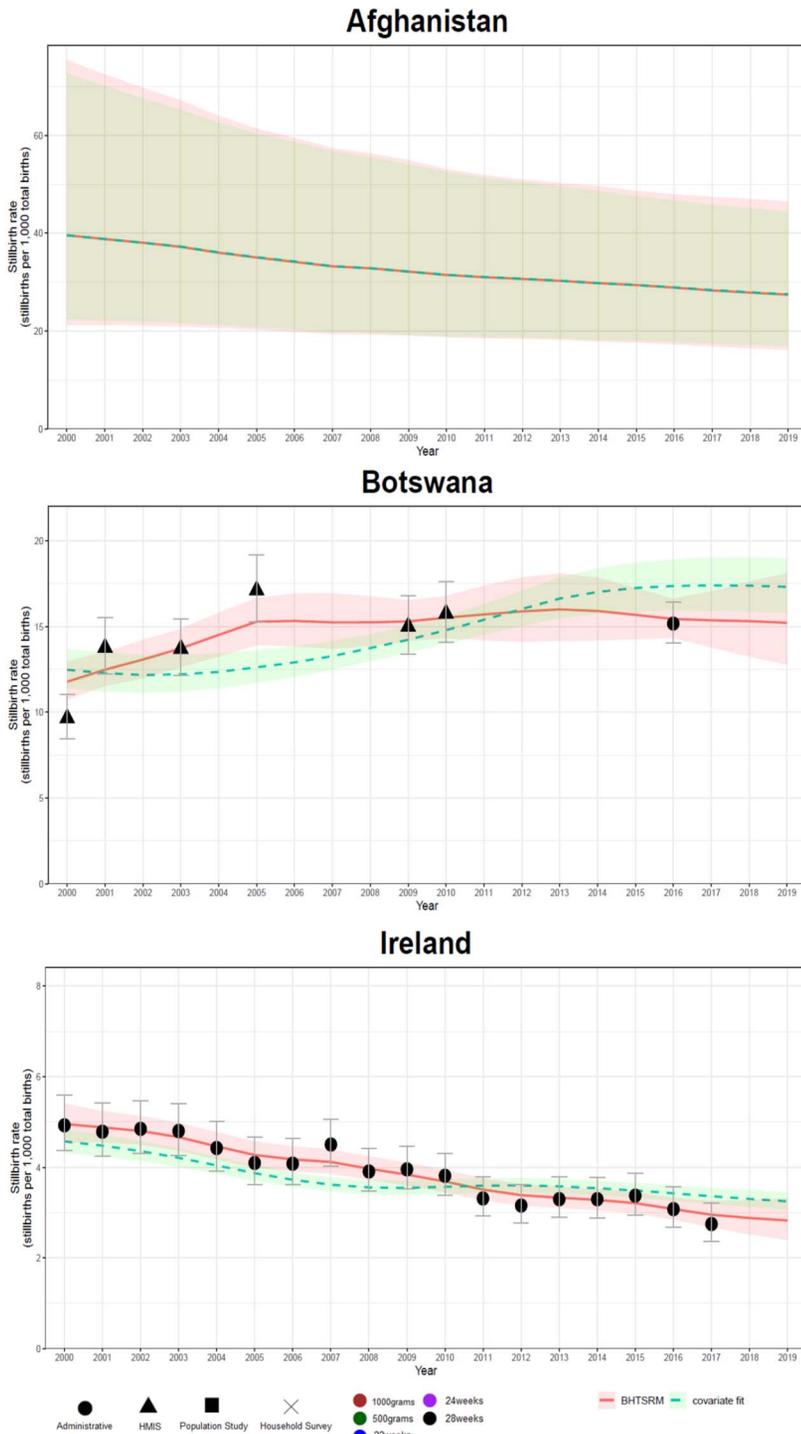


FIG. 2. SBR data and estimates for 2000–2019 for selected countries. Posterior median point estimates from BHTSRM (red line) with 90% credible intervals (red area) and covariate-based estimates (dashed green line) with 90% credible intervals (green area) are shown. Observed but unadjusted observations are displayed as hollow symbols. Adjusted data (based on definition adjustments and accounting for survey biases where applicable) and data that do not require adjustments (nonsurvey data with standard definition) are shown for all source types. Colors indicate the definition of the observation. Error bars displayed with adjusted observations indicate 95% confidence interval of the SBR, based on the observation, accounting for its estimated bias and error variance. Note that the y-axis varies across countries, and that data excluded based on the data quality assessment are not shown.

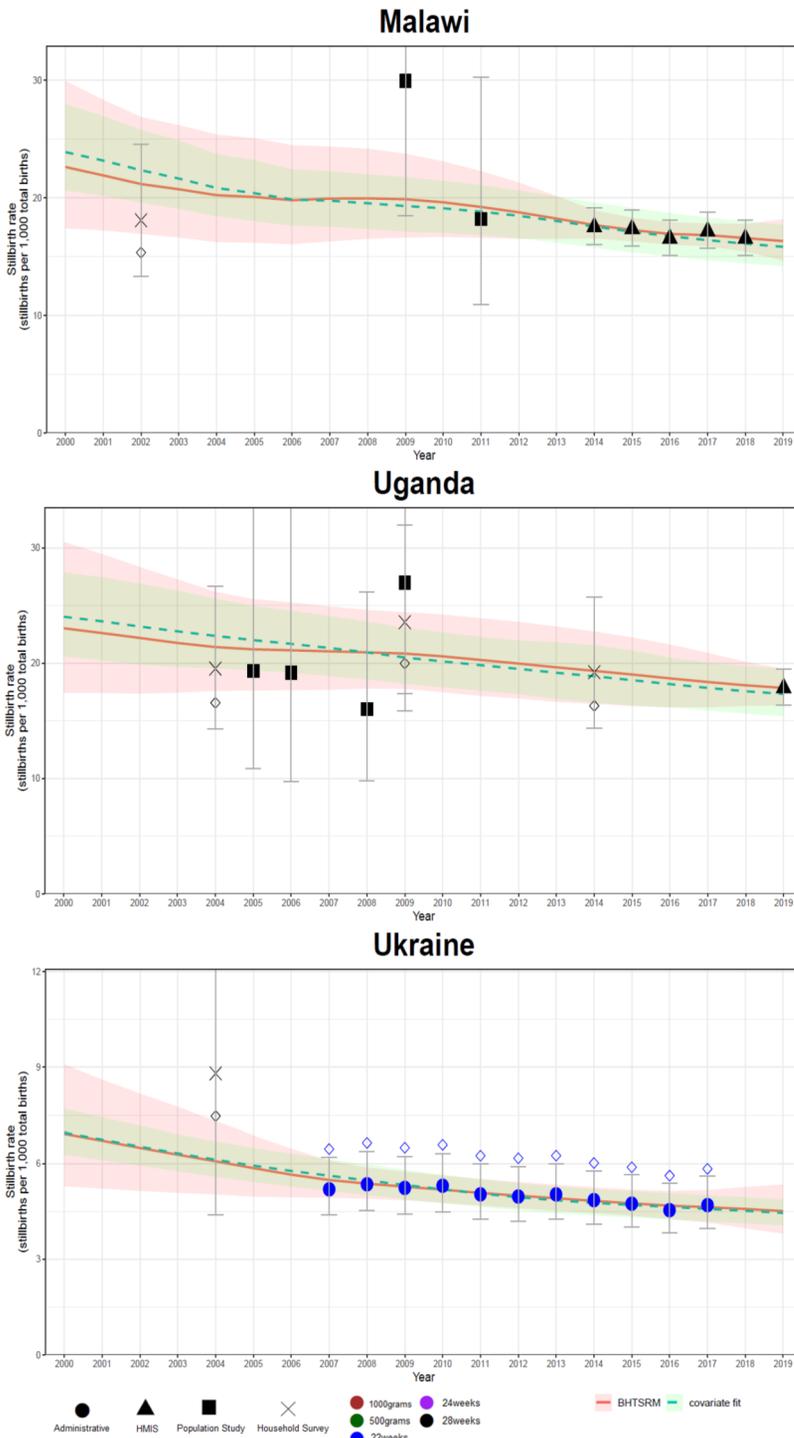


FIG. 2. (Continued).

Malawi, Uganda, and Ukraine are examples of countries with SBR data from multiple sources, available for selected periods only. In Ukraine, SBR data are available from 2007 to 2017 from administrative systems but recorded using 22 weeks definition. In Uganda, the only available data comes from surveys and population-based studies. In Malawi, available data sources are HMIS, population-based studies, and household surveys.

2.3. *Covariates.* Blencowe et al. (2016) identified a large number of candidate covariates for estimating SBR based on a conceptual framework. The framework includes distal determinants, such as socioeconomic factors, demographic and biomedical factors, associated perinatal outcome markers, and access to health care. Covariate database  $\mathcal{C}$  contains information on the 16 covariates for all country-years; further details are given in Table 3 in the Supplementary Material (Wang et al. (2022)).

**3. Exclusion based on the ratio of SBR to NMR.** Stillbirths are typically more poorly recorded than deaths of liveborn neonates which are themselves underrecorded in many settings (Stanton et al. (2006) and Woods (2008)). We exclude data points whose stillbirths are likely to be underreported based on the ratio of observed SBR to NMR, making use of the fact that, in settings where stillbirth case ascertainment is poor, the ratio of SBR to NMR is expected to be low.

We describe the approach in detail in the remainder of this section. In summary, we assume that each observed log-ratio is the sum of a setting-specific expected log-ratio and random error. We use the PopPros database  $\mathcal{P}$  to build a model for the expected log-ratio. We then calculate observed log-ratios for all observations in the global data set  $\mathcal{B}$  and exclude observations that, based on a comparison between the observation and its predictive distribution using the model for the expected log-ratio, are deemed subject to underreporting. The exclusion process is summarized in Figure 3.

The proposed approach improves upon the approach used previously for SBR estimation in Blencowe et al. (2016). In the previously used approach, observations were excluded based on a percentile of the observed distribution of SBR to NMR ratios. This approach did not account for varying uncertainty associated with the observed ratios, and, contrary to our approach, the previous approach did not make explicit the probability of a false exclusion.

3.1. *Predictive model for the SBR to NMR ratio.* In the predictive model for the SBR to NMR ratio, we assume that each observed log-ratio is the sum of a setting-specific expected log-ratio and random error. This model is specified as follows. Let  $r_i = y_i/o_i$  denote the observed ratio of SBR  $y_i$  to NMR  $o_i$ . We assume that

$$(1) \quad \log(r_i)|\theta_i \sim N(\theta_i, v_i^2),$$

**SBR:NMR Exclusion Process:**

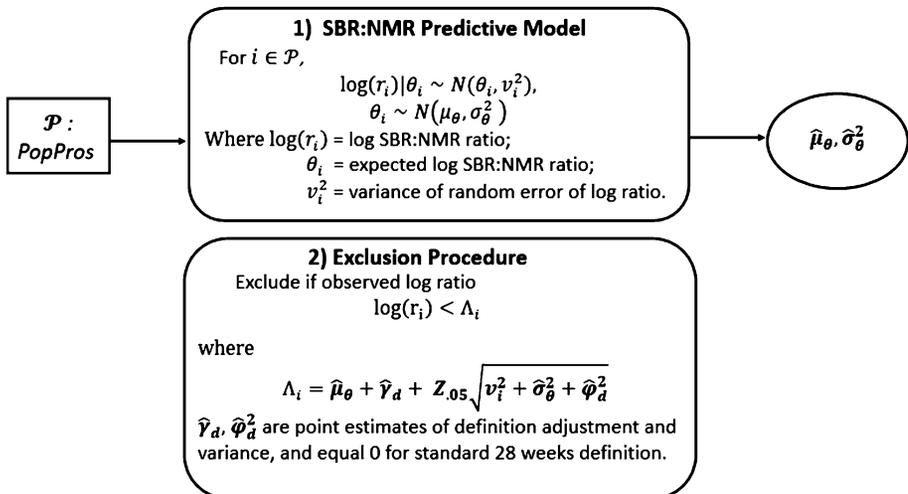


FIG. 3. SBR to NMR ratio exclusion process. This chart summarizes the two-step exclusion process based on SBR:NMR ratios. The thin arrows indicate the flow of data and parameters.

where  $\theta_i = E(\log(r_i))$  refers to the expected log-ratio of SBR to NMR and  $v_i^2$  refers to the error variance.

The error variance  $v_i^2$  is calculated using a Monte Carlo approximation. Specifically, denote  $z_i$  as the number of observed stillbirths and  $m_i$  as neonatal deaths. Then, we have

$$\begin{aligned} z_i | y_i &\sim \text{Bin}(g_i, y_i), \\ m_i | o_i &\sim \text{Bin}(q_i, o_i), \end{aligned}$$

where  $g_i$  refers to total births and  $q_i$  refers to the number of live births. Assuming independence between stillbirths and neonatal deaths, we obtain samples  $(z_i^{(s)}, m_i^{(s)})$  and calculate the associated ratio  $r_i^{(s)}$ ,

$$r_i^{(s)} = \frac{z_i^{(s)} / g_i}{m_i^{(s)} / q_i}.$$

The variance  $v_i^2$  is given by the empirical variance of the samples  $\log(r_i^{(s)})$ .

We specify the distribution of the expected log-ratios  $\theta_i$  as follows: assuming conditionally independence and a normal distribution, we set

$$(2) \quad \theta_i | \mu_\theta, \sigma_\theta^2 \sim N(\mu_\theta, \sigma_\theta^2)$$

with  $\mu_\theta$  referring to the mean log-ratio across different SBR and NMR settings and  $\sigma_\theta^2$  referring to variability across settings. We assign vague priors to  $\mu_\theta$  and  $\sigma_\theta^2$ .

The model is fitted to data from PopPros data set  $\mathcal{P}$ . Based on the data collection procedure used by the studies in this data set, data are assumed to be based on complete reporting of stillbirths. The data set contains 73 data points from 10 LMICs in different years. Based on the data set, the estimated mean ratio on the log scale is  $\hat{\mu}_\theta = -0.180$  ( $-0.250, -0.111$ ) and variance across settings is estimated as  $\hat{\sigma}_\theta^2 = 0.083$ . The estimates of  $\theta_i$  are shown in Figure 1 in the Supplementary Material (Wang et al. (2022)).

**3.2. Exclusion procedure.** If stillbirths are underreported relative to neonatal deaths for a specific observation, its associated observed log-ratio of SBR to NMR  $\log(r_i)$  is biased downward, as compared to the true log-ratio  $\theta_i$ . We calculate observed SBR to NMR ratios for all observations in data set  $\mathcal{B}$  and use the fitted model described above to construct a predictive distribution for each log-ratio. We exclude an observation if its observed ratio is less than the 5% lower bound of its corresponding predictive distribution of the SBR to NMR ratio. Specifically, the predictive distribution of the SBR to NMR ratio for the  $i$ th observation follows from equation (1) and is given by

$$\log(r_i) \sim N(\hat{\mu}_\theta, \hat{\sigma}_\theta^2 + v_i^2),$$

where  $\hat{\mu}_\theta$  and  $\hat{\sigma}_\theta^2$  refer to point estimates for the mean and across-setting variance of  $\theta$  and  $v_i^2$  to the error variance of the log-ratio specific to that observation. Let  $\Lambda_i$  denote the lower 5% quantile of the predictive distribution for observation  $i$ ,  $\Lambda_i = \hat{\mu}_\theta + z_{.05} \sqrt{\hat{\sigma}_\theta^2 + v_i^2}$ . We exclude observation  $i$  if its observed log ratio  $\log(r_i) < \Lambda_i$ . Based on the point estimates of  $\mu_\theta$  and  $\sigma_\theta^2$ , the 5% lower bound of the predictive distribution of the SBR to NMR ratio is  $\exp(\Lambda_i) = 0.52$  for observations with variance  $v_i = 0$ . For the data with alternative stillbirth definitions, we apply the exclusion procedure after definition adjustment (see Section 4.3).

**SBR Estimation:**

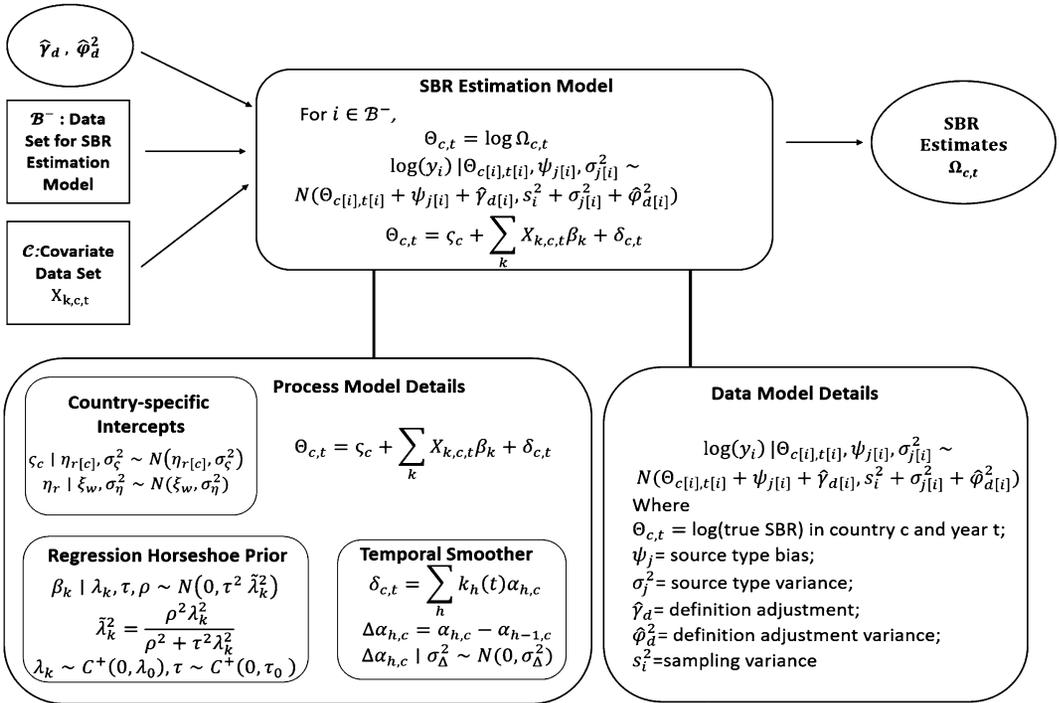


FIG. 4. SBR estimation model overview. This chart summarizes the inputs and set up of the SBR estimation model. The input data includes SBR data set  $\mathcal{B}^-$ , covariate data set  $\mathcal{C}$ , and point estimates from the definition adjustment model (see Section 4.3).

**4. Methods for SBR estimation.**

4.1. *SBR estimation model summary.* The SBR estimation model is summarized in Figure 4. We let  $\Omega_{c,t}$  denote the main outcome of interest which is the SBR for country  $c$  in year  $t$  using the standard definition. The process model specification, referring to the specification of  $\Theta_{c,t} = \log(\Omega_{c,t})$ , is explained in Section 4.4.

$\Omega_{c,t}$  is estimated using data set  $\mathcal{B}^-$ . Following earlier notation, observations are available across countries over time and are indexed by  $i$ ;  $c[i]$  refers to the country for which the  $i$ th observation was recorded,  $t[i]$  the calendar year of the observation,  $j[i]$  the data source type of the observation, and  $d[i]$  to its stillbirth definition. The index  $r[c]$  refers to the region of country  $c$ . The data model is

$$(3) \quad \log(y_i) | \Theta_{c[i],t[i]}, \psi_{j[i]}, \sigma_{j[i]}^2 \sim N(\Theta_{c[i],t[i]} + \psi_{j[i]} + \hat{\gamma}_{d[i]}, s_i^2 + \hat{\varphi}_{d[i]}^2 + \sigma_{j[i]}^2),$$

where  $\Theta_{c,t} = \log(\Omega_{c,t})$  refers to the log-transformed true SBR  $\Omega_{c,t}$  for that country-year,  $s_i^2$  to variance of  $\log(y_i)$  (see Section 4.2.1),  $\psi_{j[i]}$  and  $\sigma_{j[i]}^2$  refer to its source type-specific bias and variance, respectively (see Section 4.2.2), and  $\hat{\gamma}_d$  and  $\hat{\varphi}_d^2$  to definition-specific adjustment and variance for observations that are reported using alternative definitions.

Definition adjustment parameters are estimated prior to model fitting. As compared to the approach used previously in Blencowe et al. (2016), we have made two improvements. First, we developed predictive models for the differences in SBRs that capture how stillbirths based on the alternative definition relate to stillbirths reported according to the standard definition. Second, we assess the variability in the relationship between standard and alternative SBRs and account for this uncertainty in the SBR estimation model. The approach is described in Section 4.3.

4.2. Estimation of data quality parameters.

4.2.1. Variance of  $\log(y_i)$ . The term  $s_i^2$  in the data model equation (3) refers to the variance of  $\log(y_i)$ . For observations administrative data, HMIS, and population studies, we assume a Poisson data-generating process to obtain  $s_i^2$ . Specifically, for SBR rate  $y_i = z_i/g_i$ , with stillbirth  $z_i$  and total births  $g_i$  for the  $i$ th observation, we assume  $z_i | \Omega_i \sim \text{Poisson}(g_i \cdot \Omega_i)$ . Then,  $\text{var}(y_i) = z_i/g_i^2$ , and by using the delta method, we obtain

$$\hat{\text{var}}(\log(y_i)) = \frac{1}{z_i \cdot y_i}.$$

Therefore, the variance  $s_i^2$  for the  $i$ -th observation is set to  $\frac{1}{z_i \cdot y_i}$ . For observations from surveys, sampling error  $s_i$  is pre-calculated using a jackknife method (Pedersen and Liu (2012)), to reflect the survey sampling design.

4.2.2. Source type bias  $\psi_j$  and measurement error variance term  $\sigma_j^2$ . Source type bias terms  $\psi_j$  are included in model fitting to capture systematic biases associated with specific source types. We assume there is no source type biases for administrative, HMIS, and population-based studies, that is,  $\psi_j = 0$  for  $j$  referring to these three source types

$$\psi_{1,2,4} = 0.$$

Liu et al. (2016) and Bradley, Winfrey and Croft (2015) suggest that stillbirths tend to be underreported in surveys, so we assume that data from surveys have a negative bias term and estimate this bias term, assigning a a half-normal vague prior to  $\psi_3$ ,

$$(4) \quad \psi_3 \sim N^-(0, 5^2).$$

Due to errors introduced in reporting, the measurement error variance term  $\sigma_j^2$  captures nonsystematic errors. These variance parameters are estimated and assigned vague priors

$$\sigma_j \sim N^+(0, 1), \quad j = 1, \dots, 4.$$

4.3. Definition adjustment. To estimate the definition-specific adjustment  $\gamma_d$  and variance  $\varphi_d^2$  in equation (3), we use data sources that reported stillbirths using multiple definitions. Specifically, we construct definition adjustment data set  $\mathcal{D}^d$  for each definition  $d$ , which contains all available paired observations of stillbirth counts  $(z_i^{(d)}, z_i)$ , where  $z_i^{(d)}$  is the number of stillbirths under the alternative definition  $d$ ,  $z_i$  is the number of stillbirths under standard definition, and the pair refers to the same source, country, and year. We use the paired counts to estimate  $\gamma_d$  and  $\varphi_d^2$  for definition  $d$ , without controlling for year and source, but separately for high-income countries (HICs) and LMICs. Due to lack of data, in LMICs we assume that 500 grams birthweight is equivalent to 22 weeks of gestational age, and 1000 grams birthweight is equivalent to 28 weeks of gestational age. Table 2 summarizes the data used for the analysis of the definition and income group combinations.

We define  $\kappa_i^{(d)}$  as the log-ratio of the SBR as per alternative definition  $d$  to standard definition for observation  $i$ :  $\kappa_i^{(d)} = \log\left(\frac{\Omega_{c[i],t[i]}^{(d)}}{\Omega_{c[i],t[i]}}\right)$ . With this definition of  $\kappa$ , the true log-transformed SBR for observation  $i$ , under definition  $d[i]$ ,  $\Theta_{c[i],t[i]}^{(d)}$ , can be written as

$$\Theta_{c[i],t[i]}^{(d)} = \Theta_{c[i],t[i]} + \kappa_i^{(d)},$$

where  $\Theta_{c,t}$  refers to the log-transformed SBR under the standard definition. We use this relation to define the adjustment term  $\gamma_d$  and variance  $\varphi_d^2$  in equation (3): we set the adjustment

TABLE 2

Data availability in definition adjustment data sets  $\mathcal{D}^d$ . “22 weeks” and “24 weeks” represent 22/24 weeks or more of gestation; “500 grams” and “1000” grams represent birthweight  $\geq 500/1000$  grams

Definition	Income Group	Number of Countries	Number of Obs
22 weeks	Low&Middle	14	59
22 weeks	High	34	369
24 weeks	High	8	28
1000 grams	High	34	477
500 grams	High	30	355

$\gamma_d$  and variance  $\varphi_d^2$  in the SBR data model equal to the posterior median and variance of the predictive distribution for  $\kappa_i^{(d)}$  for each alternative definition  $d$ .

In the derivation of the predictive density of  $\kappa$ , we approximate the log-ratio of SBRs  $\kappa_i^{(d)}$  by the ratio of stillbirths, justified by the fact that the number of stillbirths are small relative to live births. Specifically, the true SBR for alternative definition  $d$  can be written as follows:  $\Omega_{c,t}^{(d)} = \frac{\Upsilon_{c,t}^{(d)}}{q_{c,t} + \Upsilon_{c,t}^{(d)}}$ , where  $\Upsilon_{c,t}^{(d)}$  refers to the “true” stillbirth count associated with the true SBR, under alternative definition  $d$ , and  $q_{c,t}$  the number of live births. Given that  $\Upsilon_{c,t}^{(d)} \ll q_{c,t}$ , we approximate  $\kappa$  as follows:

$$(5) \quad \kappa_i^{(d)} = \log\left(\frac{\Omega_{c[i],t[i]}^{(d)}}{\Omega_{c[i],t[i]}}\right) \approx \log\left(\frac{\Upsilon_{c[i],t[i]}^{(d)}}{\Upsilon_{c[i],t[i]}}\right).$$

The assumptions made to obtain the predictive distribution for  $\kappa$  varies by the definition. Alternative definitions fall into two categories: definitions containing the standard definition and definitions overlapping with the standard definition. We consider each of these below. In each setup we work toward providing a predictive distribution for  $\kappa_i^{(d)}$  by introducing probabilities that relate the survival based on the alternative definition to that based on the standard definition.

*Definitions containing the standard 28 weeks definition.* Stillbirths  $z_i$ , recorded using the 28 weeks definition, are a subset of stillbirths recorded using the 22 or 24 weeks definitions,  $z_i \leq z_i^{(d)}$  for  $d$  referring to 22 and 24 weeks. Given that stillbirths based on 22 or 24 weeks definitions contain those with 28 weeks definitions, we may assume

$$(6) \quad z_i | \omega_i^{(d)} \sim \text{Binomial}(z_i^{(d)}, \omega_i^{(d)}),$$

where  $\omega^{(d)}$  is the definition-specific probability of a stillbirth with gestational age beyond 28 weeks conditional on being dead after 22 or 24 weeks. The probability  $\omega_i^{(d)}$  relates to  $\kappa_i^{(d)}$  as follows (as per equation (5) and the definition of  $\omega_i^{(d)}$ ),

$$\kappa_i^{(d)} \approx \log(\Upsilon_{c[i],t[i]}^{(d)} / \Upsilon_{c[i],t[i]}) = -\log(\omega_i^{(d)}).$$

Based on this equation, we estimate the adjustment  $\hat{\gamma}_d$  and variance  $\hat{\varphi}_d^2$  in equation (3) by the median and variance of the predictive distribution for  $-\log(\omega_i^{(d)})$ . This predictive distribution is based on the following assumption:

$$(7) \quad \text{logit}(\omega_i^{(d)}) | \mu_{\omega,d}, \sigma_{\omega,d}^2 \sim N(\mu_{\omega,d}, \sigma_{\omega,d}^2),$$

where  $\mu_{\omega,d}$  is the mean of the logit-transformed probabilities and  $\sigma_{\omega,d}$  the standard deviation. We use vague prior for the mean and variance parameters,

$$\begin{aligned} \sigma_{\omega,d} &\sim N^+(0, 1), \\ \text{expit}(\mu_{\omega,d}) &\sim U(0, 1). \end{aligned}$$

*Definitions overlapping with the standard 28 weeks definition.* Stillbirths  $z_i^{(d)}$  recorded using the 1000 or 500 grams definitions are overlapping with the stillbirths  $z_i$  using the standard definition.

In this setting, let  $N_i = n_i^{(r\&d)} + n_i^{(r)} + n_i^{(d)}$  refer to the total number of stillbirth based on the standard definition *or* an alternative definition, with  $n_i^{(r\&d)}$  the count of stillbirths that satisfy the 28-week and alternative definition,  $n_i^{(r)}$  the count of stillbirth with standard definition rather than alternative definition, and, finally,  $n_i^{(d)}$  the counts of stillbirth with alternative definition rather than standard definition. We can assume

$$(n^{(r\&d)}, n^{(r)}, n^{(d)}) | (\omega_i^{(r\&d)}, \omega_i^{(r)}, \omega_i^{(d)}) \sim \text{Multinom}(N_i, (\omega_i^{(r\&d)}, \omega_i^{(r)}, \omega_i^{(d)})),$$

where  $\omega_i^{(r\&d)}$ ,  $\omega_i^{(r)}$  and  $\omega_i^{(d)}$  refer to the probabilities of a stillbirth satisfying both definitions, the standard definition only, and the alternative definition only, respectively.

Based on the expression for  $\kappa_i$  in equation (5) and the definitions of the  $\omega_i^{(\cdot)}$ s, we obtain the following relation:

$$\kappa_i^{(d)} \approx \log\left(\frac{\Upsilon_{c[i],t[i]}^{(d)}}{\Upsilon_{c[i],t[i]}}\right) = \log\left(\frac{\omega_i^{(r\&d)} + \omega_i^{(d)}}{\omega_i^{(r\&d)} + \omega_i^{(r)}}\right).$$

Based on this equation, we estimate the adjustment  $\hat{\gamma}_d$  and variance  $\hat{\phi}_d^2$  in equation (3) by the median and variance of the predictive distribution for log-ratio of the definition-specific probabilities  $\Gamma_i^{(d)} = \log\left(\frac{\omega_i^{(r\&d)} + \omega_i^{(d)}}{\omega_i^{(r\&d)} + \omega_i^{(r)}}\right)$ .

We assume that the  $\Gamma_i^{(d)}$ s are normally distributed,

$$(8) \quad \Gamma_i^{(d)} | \mu_{\Gamma,d}, \sigma_{\Gamma,d}^2 \sim N(\mu_{\Gamma,d}, \sigma_{\Gamma,d}^2),$$

with  $\mu_{\Gamma,d}$  and  $\sigma_{\Gamma,d}^2$  referring to the across-setting mean and variance of the log-ratios. To guarantee that the estimation results in sets of  $\omega_i^{(r\&d)}$ ,  $\omega_i^{(r)}$ , and  $\omega_i^{(d)}$  that add up to one, we introduce the constraint  $\frac{1}{1+\exp(\Gamma_i^{(d)})} < \omega_i^{(r)} + \omega_i^{(d)} < \frac{1}{\max\{1, \exp(\Gamma_i^{(d)})\}}$  and incorporate this constraint through a prior on the sum,

$$(\omega_i^{(r)} + \omega_i^{(d)}) | \Gamma_i^{(d)} \sim U\left(\frac{1}{1 + \exp(\Gamma_i^{(d)})}, \frac{1}{\max\{1, \exp(\Gamma_i^{(d)})\}}\right).$$

Vague priors are used for the mean and variance parameters of  $\Gamma_i^{(d)}$ ,

$$\begin{aligned} \sigma_{\Gamma,d} &\sim N^+(0, 1), \\ \mu_{\Gamma,d} &\sim N(0, 20). \end{aligned}$$

When fitting the model to the database  $\mathcal{D}^d$  for the overlapping definition, we typically have available data pairs  $(z_i, z_i^{(d)})$ , as opposed to  $n_i^{(r\&d)}$ ,  $n_i^{(r)}$  and  $n_i^{(d)}$ . It follows that  $z_i = n_i^{(r\&d)} + n_i^{(r)}$  and  $z_i^{(d)} = n_i^{(r\&d)} + n_i^{(d)}$ . We implemented an exact likelihood function to estimate the  $\omega$ s from the overlapping data sets.

4.4. *Bayesian hierarchical temporal sparse regression model.* We developed a Bayesian hierarchical temporal regression model (BHTRM) to estimate the SBR for all country-years. It combines country-specific intercept  $\zeta_c$ , linear regression function  $\sum_k X_{k,c,t}\beta_k$ , and a temporal smoothing process  $\delta_{c,t}$ ,

$$(9) \quad \Theta_{c,t} = \zeta_c + \sum_k X_{k,c,t}\beta_k + \delta_{c,t}.$$

Country-specific intercepts  $\zeta_c$  are estimated hierarchically, with

$$\begin{aligned} \zeta_c | \eta_{r[c]}, \sigma_\zeta^2 &\sim N(\eta_{r[c]}, \sigma_\zeta^2), \\ \eta_r | \xi_w, \sigma_\eta^2 &\sim N(\xi_w, \sigma_\eta^2), \end{aligned}$$

where  $\eta_{r[c]}$  refers to the regional mean,  $\sigma_\zeta^2$  to the across-country variance within regions,  $\xi_w$  to the global mean, and  $\sigma_\eta^2$  to the across-region variance. Vague priors were used for the global mean and variances,

$$\begin{aligned} \xi_w &\sim N(2.5, 2^2), \\ \sigma_\zeta, \sigma_\eta &\sim N^+(0, 1). \end{aligned}$$

A penalized spline regression model is used for  $\delta_{c,t}$ ,

$$(10) \quad \delta_{c,t} = \sum_{h=1}^H k_h(t)\alpha_{h,c},$$

where  $k_h(t)$  refers to the  $h$ th spline function, evaluated at time  $t$ , and  $\alpha_{h,c}$  to its regression coefficient for country  $c$ .

We use equally spaced quadratic B-splines, with knots spaced one year apart and placed at each integer year (Eilers and Marx (1996), Currie and Durban (2002)). The spline regression coefficients are modeled with a first-order random walk process with a sum-to-zero constraint  $\frac{1}{H} \sum_h \alpha_{h,c} = 0$  to ensure identifiability. For each country, we define first-order difference  $\Delta\alpha_{h,c}$ ,

$$\Delta\alpha_{h,c} = \alpha_{h,c} - \alpha_{h-1,c}.$$

First-order differences are penalized as follows:

$$\Delta\alpha_{h,c} | \sigma_\Delta^2 \sim N(0, \sigma_\Delta^2),$$

where the variance term  $\sigma_\Delta^2$  determines the smoothness of the fit. We address the sensitivity to these settings in Section 5.4.

*Estimating regression coefficients using sparsity-inducing priors.* Blencowe et al. (2016) identified 16 candidate covariates for estimating SBR, based on a conceptual framework, and used a stepwise approach variable selection. In this study we refrain from stepwise selection methods and instead use regularized horseshoe priors on regression coefficients (Piironen and Vehtari (2017b)) to impose sparsity by allowing shrinkage of coefficients to zero. We expand upon BHTRM by introducing sparsity-inducing priors for estimating regression coefficients  $\beta_k$  and refer to the resulting model set up as a Bayesian hierarchical temporal sparse regression model (BHTSRM) which can be applied when the number of candidate covariates is large.

Regularized horseshoe priors for the regression coefficients are defined as follows:

$$\begin{aligned} \beta_k | \lambda_k, \tau, \rho &\sim N(0, \tau^2 \tilde{\lambda}_k^2), \\ \tilde{\lambda}_k^2 &= \frac{\rho^2 \lambda_k^2}{\rho^2 + \tau^2 \lambda_k^2}, \end{aligned}$$

where  $\tau$  and  $\rho$  are global shrinkage parameters and the  $\lambda_k$ s are local (coefficient-specific) parameters. Priors are set as follows:

$$\begin{aligned}\lambda_k &\sim C^+(0, \lambda_0), \\ \tau &\sim C^+(0, \tau_0), \\ \rho^2 &\sim \text{Inv-Gamma}(\rho_1, \rho_2),\end{aligned}$$

where  $C^+(0, s)$  refers to a half-Cauchy distribution with location parameter 0 and scale parameter  $s$ ;  $\lambda_0$ ,  $\tau_0$ ,  $\rho_1$ , and  $\rho_2$  are fixed. The Cauchy distribution, which—compared to a normal distribution—has greater density around 0 and a heavier tail, allows the global hyperparameter  $\tau$  to shrink all the parameters toward zero, while the heavy tail allows the coefficient-specific parameters  $\lambda_k$ 's to make some coefficients escape from the global shrinkage. This setup allows for the inclusion of a larger set of candidate covariates and encourages sparseness by shrinking irrelevant covariates toward zero. It is not a variable selection method because it does not shrink all posterior samples to zero.

We set  $\lambda_0 = \tau_0 = 1$ ,  $\rho_1 = 2$ , and  $\rho_2 = 8$ , as per the recommended defaults in Piironen and Vehtari (2017a), Carvalho, Polson and Scott (2009), and Gelman (2006). We address the sensitivity to these settings in Section 5.4.

**4.5. Computation.** A Hamiltonian Monte Carlo (HMC) algorithm is employed to sample from the posterior distribution of the parameters of the SBR estimation model with the use of Stan (Carpenter et al. (2017)) and R package Rstan (Stan Development Team (2018)). Six parallel chains are run with a total of 6000 iterations in each chain. The first 2000 iterations in each chain are discarded as burn-in so that the resulting chains contain 4000 samples each. Point estimates are given by medians of the posterior samples. Standard diagnostic checks are used to check convergence and sampling efficiency. These checks are based on trace plots, the improved Rhat diagnostic using rank-normalized draws (Gelman and Rubin (1992), Vehtari et al. (2021)), and various calculations of effective sample size (ESS), including the bulk ESS and the tail ESS, giving the minimum of the effective sample sizes of the 5% and 95% quantiles.

**4.6. Model validation and comparison.** Performance of the SBR estimation model is assessed through two out-of-sample validation exercises. In the first exercise, we randomly leave out 20% of the observations and repeat this exercise 20 times (leaving out 306 observations each time). In the second exercise, we leave out the last observation for each country to check the predictive performance. To evaluate model performance, we calculate various measures based on a comparison between left-out observations and their predictive distributions. We define prediction errors  $e_i$  as the difference between the left-out observation and the median of its predictive posterior distribution based on the training set,

$$e_i = (\log(y_i) - \log(\tilde{y}_i))/S_i,$$

where  $y_i$  is the left-out observations and  $\tilde{y}_i$  and  $S_i$  refer to the estimated median and standard deviation of the predictive distribution for  $y_i$  based on the training set. Coverage of prediction intervals is given by  $N^{-1} \sum_{i=1}^N 1[l_i \leq y_i \leq u_i]$ , where  $N$  denotes the total number of left-out observations considered and  $l_i$  and  $u_i$  are the lower and upper bounds of the prediction interval for the  $i$ th observation. We also carry out approximate leave-one-out cross-validation (LOO) which is implemented in the loo package in R (Vehtari et al. (2019)).

For comparing models, we consider the expected log pointwise predictive density (ELPD) and Pareto K diagnostic (Vehtari, Gelman and Gabry (2017)). The ELPD is the log pointwise predictive density for a new data set which can be used to evaluate the performance of the model to predict the future data. The Pareto K diagnostic refers to the estimates of the shape parameter  $k$  of the generalized Pareto distribution. Values larger than one may indicate that the observation is outlying and influential.

TABLE 3  
Adjustments and standard deviation of alternative definition vs. the 28-week definition

Definition	Income group	$\hat{\gamma}_d$ (95% CI)	$e^{\hat{\gamma}_d}$ (95% CI)	$\hat{\varphi}_d$
22 weeks	Low	0.214 (0.101, 0.426)	1.239 (1.106, 2.031)	0.084
22 weeks	High	0.389 (0.175, 0.777)	1.476 (1.192, 2.175)	0.156
24 weeks	High	0.222 (0.058, 0.709)	1.248 (1.060, 2.031)	0.172
1000 grams	High	-0.065 (-0.074, -0.056)	0.937 (0.929, 0.946)	0.073
500 grams	High	0.244 (0.232, 0.257)	1.277 (1.261, 1.293)	0.087

## 5. Results.

5.1. *Data quality and data adjustments.* Adjustments  $\hat{\gamma}_d$  and standard deviations  $\hat{\varphi}_d$  associated with alternative definitions are given in Table 3. For example, adjustments on the log-scale for 1000 grams definition is  $-0.065$  ( $-0.074$ ,  $-0.056$ ), suggesting that the 1000 grams definition data are on average 0.937 (0.929, 0.946) times lower than the standard definition.

Table 4 summarizes the differences in error standard deviation  $\sigma_j$  associated with the different source types, ranging from a standard deviation of 0.017 for national administrative data to 0.239 for population study data. The bias  $\psi_j$  for survey data is estimated at  $-0.165$  ( $-0.229$ ,  $-0.100$ ) on the log-transformed scale, suggesting that survey data are on average 0.848 (0.795, 0.905) times lower than the truth.

5.2. *Illustrative findings.* Estimates for selected countries<sup>1</sup> are given in Figure 2, with final estimates displayed in red and underlying covariate-based estimates (obtained by removing the smoother term  $\delta_{c,t}$  from the model) in green. As highlighted earlier in the paper, data availability ranges in the selected countries from no data (Afghanistan) to an annual time series of national administrative data based on the standard definition for Ireland. The BHT-SRM produces estimates for both countries. Point estimates for Ireland track the observed SBR from administrative system closely, and credible intervals are close to the uncertainty associated with each observed SBR. Estimates for Afghanistan are driven by covariates and the estimates are uncertain due to the absence of data.

Botswana, Malawi, Ukraine, and Uganda are examples of countries with SBR data that are either subject to bias, substantial error variance, or missing for periods of interest. In Ukraine, SBR data are available from 2007 to 2017 from administrative systems but recorded using a 22 week definition. SBR estimates are informed by the adjusted observations and uncertainty increases in extrapolations past the observation period. The survey data point has

TABLE 4  
Source type bias and source type standard deviation

Source type	$\hat{\psi}_j$	$\hat{\sigma}_j$
Administrative	-	0.017
HMIS	-	0.045
Population study	-	0.239
Survey	$-0.165$ ( $-0.229$ , $-0.100$ )	0.135

<sup>1</sup>Estimates for all countries see [childmortality.org](http://childmortality.org)

a large associated uncertainty and has little influence on the resulting model fit. In Uganda, the only available data come from HMIS, surveys, and population-based studies. There is substantial uncertainty associated with survey and population-based study data and resulting SBR estimates reflect this. There are four different data sources in Botswana and Malawi. Resulting estimates are more certain in years with administrative or HMIS data, as compared to population-based, survey, or no data.

The effect of adding the smoother to the regression model on point estimates is most visible in Ireland where final point estimates differ from the covariate-driven ones. In general, credible intervals are wider for the model that includes the smoother, as shown in Figure 2. Exceptions include countries where data are limited except for a short period with low-variance data, such as Malawi: in such countries the addition of the smoother results in reduced uncertainty in the period with low-variance data (when the estimates are data-driven).

**5.3. Covariates.** Table 5 summarizes the estimates for regression coefficients, ordered by absolute point estimates of the coefficients. Given that covariates were standardized, the coefficients are measured in units of standard deviation of the covariate which are added to the table for reference. In the analysis by [Blencowe et al. \(2016\)](#), NMR, low birthweight, gross national income, mean years of female education, and coverage of four antenatal care visits ( $\log(\text{nmr})$ ,  $\log(\text{lbw})$ ,  $\log(\text{gni})$ , edu, and anc4 in Table 5) were selected for inclusion in the regression model. Here, we find that these covariates are ranked among the top in terms of their absolute regression coefficient along with C-section (csec). Comparisons between the model with horseshoe priors and additional models for sensitive checks are given in Section 5.4.

**5.4. Model validation, comparison and sensitivity analyses.** Validation results for the BHTSRM are given in Table 6. For all scenarios, mean residuals are close to zero, and the mean of the absolute residuals are around 0.1. The approximate leave-one-out validation exercise suggests that predictive distributions are overdispersed, as compared to the left-out observations, with the percentages outside of 80% and 90% prediction intervals being lower

TABLE 5

*Estimates for regression coefficients under BHTSRM fit. Point estimates of regression coefficients, 95% credible interval given by the 2.5th and 97.5th percentiles of the posterior, and the standard deviation of the covariate prior to standardization. Details on covariates are given in Table 3 in the Supplementary Material*

Covariates	Estimate $\hat{\beta}$	2.5%	97.5%	SD (covariate)
$\log(\text{nmr})$	0.414	0.336	0.492	0.999
$\log(\text{gni})$	-0.102	-0.212	0.001	1.20
$\log(\text{lbw})$	0.078	0.009	0.141	0.439
edu	-0.037	-0.104	0.007	3.41
csec	-0.027	-0.082	0.008	11.9
anc4	-0.025	-0.094	0.014	21.8
pab	-0.018	-0.050	0.006	11.6
abr	-0.017	-0.109	0.023	46.5
urban	-0.012	-0.087	0.024	23.1
gini	0.010	-0.017	0.061	8.17
sab	-0.010	-0.083	0.026	0.215
anc1	-0.009	-0.067	0.021	14.7
mmr	0.003	-0.057	0.109	288.5
pfpr	-0.002	-0.045	0.030	0.118
gdp	0.001	-0.047	0.063	$207 \cdot 10^2$
gfr	0.000	-0.057	0.054	0.049

TABLE 6

Validation results for SBR estimates. “N.test” represent the number of observations in the test set. Validation exercises “Recent,” “Random,” and “In-sample” represent leaving out recent observations, randomly leaving out 20% of all observations, and approximate leave-one-out validation, respectively. The outcome measures are as follows: mean of error, mean absolute error, and % of left-out observations below and above their respective 90% and 80% prediction intervals. Desirable\* refers to outcomes for models that are unbiased and well calibrated

Validation	N.test	Mean err.	Mean abs. err.	below 5%	below 10%	above 90%	above 95%
Desirable*	–	0	N/A	5%	10%	10%	5%
BHTSRM Recent	112	–0.001	0.091	3.5%	6.2%	2.7%	1.8%
BHTSRM In-sample	1531	–0.002	0.090	1.8%	3.5%	4.2%	1.6%
BHTSRM Random	306	–0.001	0.090	1.8%	3.3%	4.3%	1.8%

than expected. The out-of-sample exercises suggest that the model is reasonably well calibrated with slightly less left-out observations falling below their respective predictive intervals than expected.

We compare the performance of the BHTSRM, using sparse priors, to that of a model with vague priors on the regression coefficient, labeled BHTRM. Regression coefficients estimates for both the BHTSRM and BHTRM are given in Figure 2 and Table 1 in the Supplementary Material. Some of the coefficients are closer to zero in the BHTSRM, as compared to in the BHTRM, due to the shrinkage by the regularized horseshoe prior. We compare predictive performance between the BHTSRM and the BHTRM in Table 2 in the Supplementary Material and find that the mean error and mean absolute error are close to each other. Validation results by income group do not suggest difference in model performance either. The ELPD is higher for our reference BHTSRM, as compared to the BHTRM, the 95% CI for the difference is (–12.6, –0.06) (see Table 7), suggesting improved predictive performance due to the horseshoe prior.

We also compare the reference model to another BHTSRM that is fitted using an alternative choice of hyperparameters for the horseshoe prior based on Piironen and Vehtari (2017a). For standard regression models with  $y_i \sim N((X_i\beta, \sigma^2)$ , Piironen and Vehtari propose to set the scale parameter  $\tau_0$  in the prior for  $\tau$  as follows:

$$\tau \sim C^+(0, \tau_0),$$

$$\tau_0 = \frac{p_0}{D - p_0} \frac{\sigma}{\sqrt{n}},$$

TABLE 7

Model comparison based on expected log pointwise predictive density and Pareto K diagnostic values. BHTSRM is our proposed model, and BHTRM is model with vague prior on covariates. The HS  $\tau_0 = 0.001$  model stands for BHTSRM with  $\tau_0 = 0.001$ . Smooth1 and Smooth2 are two models with different settings of smoothers described in the text. When comparing models, larger ELPD value suggests better model performance. The percentage of high influential points (Pareto K values > 1) for all models are presented in the “Pareto K diag.” column, lower outcomes are preferred

Model	ELPD			Pareto K diag.
	estimate	SE	95% CI for difference	
BHTSRM	1194.5	40.6	reference model	0.5%
HS $\tau_0 = 0.001$	1192.7	40.6	(–6.62, 3.07)	0.7%
BHTRM	1188.2	40.7	(–12.59, –0.06)	0.8%
smooth1	1185.9	40.5	(–13.08, –4.15)	0.5%
smooth2	1176.7	40.9	(–11.1, 3.0)	0.3%

where  $p_0$  is the guess of number of relevant predictors,  $D$  is the total number of predictors,  $\sigma$  is the standard deviation of observation  $\log(y)$ , and  $n$  is the number of observations. We cannot directly follow this recommendation because our modeling context differs from the one where this setting was explored, that is, our setting includes heteroskedasticity of observations and the regression model is combined with a temporal smoothing term. We obtain a model fit based on the recommendation as a sensitivity test. Specifically, we obtain the fit for  $p_0 = 5$ ,  $D = 16$ ,  $\sigma = 0.094$  (the median standard deviation across observations), and  $n = 1531$ , corresponding to  $\tau_0 = 0.001$ . Its ELPD is lower than the reference BHTSRM, but the difference is not significant according to the 95% CI  $(-6.62, 3.07)$ .

We checked the sensitivity of the choice of the splines in the smoothing term  $\delta_{c,t}$  in equation (10) by comparing the reference model fit to the fits obtained from two alternative models. Model specifications were the same for the three models, except for the specification of the smoothing term. In the reference model a quadratic B-spline model was used, with knots spaced one year apart and placed at each integer year. In model “smooth1,” a cubic B-splines model was used while in “smooth2” the knots were spaced two years apart. Table 7 summarizes the differences in ELPD and Pareto K values for different models. There are no improvements when comparing the alternative smoothers with our reference SBR model.

**6. Discussion.** We developed a Bayesian hierarchical temporal sparse regression model (BHTSRM) for estimating SBRs for all countries from 2000 until 2019. Estimating SBRs is challenging because of data paucity, especially for many LMICs, where most stillbirths occur, and the substantial uncertainty associated with observations due to reporting issues and errors associated with the observations. Our BHTSRM extends the approach previously proposed by [Blencowe et al. \(2016\)](#) to produce estimates that are informed by a covariate model and available data, accounting for different definitions and uncertainty associated with the available data. Model validation exercises suggest that the model is reasonably well calibrated.

The BHTSRM extends upon previous applications of Bayesian hierarchical temporal regression models through the introduction of sparsity-inducing priors and new statistical approaches to addresses data quality issues. Sparsity-inducing priors allow for the inclusion of larger sets of (potentially correlated) candidate covariates into the model. While validation exercises do not indicate improved performance of the model with the horseshoe prior over a model with vague priors, improved predictive performance was suggested by higher ELPD in our application.

To address data quality issues, we developed a statistical procedure for data exclusion based on comparing observed ratios of SBR to NMR for the population of interest to a reference distribution of such ratios. This approach improves upon the previously used approach for data exclusion by defining a predictive distribution for the ratio and a decision rule that makes explicit the probability of a false exclusion. Second, we developed a new approach to adjust and estimate additional uncertainty associated with observations using a different definition of stillbirths. In the model fitting we used a data model that accounts for bias and varying sources of random error associated with the observations.

While our approach to estimating the SBR improves upon existing approaches, there are limitations related to the model and data availability. Limited data availability restricted the analyses we are able to carry out and result in stricter modeling assumptions. For example, we excluded data based on observed SBR to NMR ratios. In this analysis, we combined data across settings when constructing a predictive distribution for the expected ratio and chose 5% as the threshold for data exclusion. We acknowledge that the choice of a higher (or lower) threshold would have resulted in the exclusion of more (or less) data. Additional data related to the quality of reporting would allow for more detailed analyses and may allow for avoiding

having to set a threshold. Relative differences in SBRs associated with the use of different definitions, that is, gestational age, may vary across settings. Data limitations resulted in the use of a simple dichotomy of high income and low income countries to capture this difference. With additional data this relationship can be studied in more detail. Lastly, although the horseshoe prior allows for the inclusion of a larger set of candidate covariates and shrinkage toward zero of irrelevant covariates, it is not a variable selection method because it does not shrink all posterior samples to zero.

The BHTRM as described in this paper is used by the UN IGME to generate estimates for the SBR globally (UN Inter-agency Group for Child Mortality Estimation (2020), Hug et al. (2021)). While the modeling approach allows for the construction of estimates for all countries, we find that uncertainty associated with the estimates is substantial in many settings, including countries with high SBRs. This highlights the need for additional data collection to produce more precise information for monitoring and program planning, especially in high-burden settings.

**Acknowledgments.** We thank all members of the Core Stillbirth Estimation Group (Leontine Alkema, Dianna M. Blau, Hannah Blencowe, Simon Cousens, Andreea Creanga, Trevor Croft, Kenneth Hill, KS Joseph, Salome Maswime, Elizabeth M. McClure, Robert Pattinson, Jon Pedersen, Lucy Smith, Jennifer Zeitlin) and the members of the Technical Advisory Group of the UN Inter-agency Group for Child Mortality Estimation (Leontine Alkema, Robert Black, Simon Cousens, Trevor Croft, Michel Guillot, Kenneth Hill, Bruno Masquelier, Colin Mathers, Jon Pedersen, Jon Wakefield, Neff Walker) for providing input and feedback on estimating stillbirths. We thank the members of the UN Interagency Group for Child Mortality Estimation (Allisyn Moran, Emi Suzuki, Victor Gaigbe-Togbe, Danzhen You) for additional guidance. We also thank Remy Wang, Serena Wang, and Zitong Wang for their contributions to the modeling in an earlier stage of the project.

**Funding.** The stillbirth estimation work was supported by the Bill & Melinda Gates Foundation (OPP1180460 OP190601 and INV-001395), and the National Institute of Environmental Health Sciences of the National Institutes of Health under award number T32ES015459.

The content is solely the responsibility of the authors and does not necessarily represent the official views of the funders.

## SUPPLEMENTARY MATERIAL

**Supplement to “Estimating the stillbirth rate for 195 countries using a Bayesian sparse regression model with temporal smoothing”** (DOI: [10.1214/21-AOAS1571SUPP](https://doi.org/10.1214/21-AOAS1571SUPP); .pdf). Supplementary tables and figure.

## REFERENCES

- AHMED, I., ALI, S. M., AMENGA-ETEGO, S., ARIFF, S., BAHL, R., BAQUI, A. H., BEGUM, N., BHANDARI, N., BHATIA, K. et al. (2018). Population-based rates, timing, and causes of maternal deaths, stillbirths, and neonatal deaths in south Asia and sub-Saharan Africa: A multi-country prospective cohort study. *The Lancet Global Health* **6** e1297–e1308. [https://doi.org/10.1016/S2214-109X\(18\)30385-1](https://doi.org/10.1016/S2214-109X(18)30385-1)
- ALKEMA, L., ZHANG, S., CHOU, D., GEMMILL, A., MOLLER, A.-B., FAT, D. M., SAY, L., MATHERS, C. and HOGAN, D. (2017). A Bayesian approach to the global estimation of maternal mortality. *Ann. Appl. Stat.* **11** 1245–1274. <https://doi.org/10.1214/16-AOAS1014>
- BLENCOWE, H., COUSENS, S., JASSIR, F. B., SAY, L., CHOU, D., MATHERS, C., HOGAN, D., SHIEKH, S., QURESHI, Z. U. et al. (2016). National, regional, and worldwide estimates of stillbirth rates in 2015, with trends from 2000: A systematic analysis. *The Lancet Global Health* **4** e98–e108. [https://doi.org/10.1016/S2214-109X\(15\)00275-2](https://doi.org/10.1016/S2214-109X(15)00275-2)

- BOSE, C. L., BAUSERMAN, M., GOLDENBERG, R. L., GOUDAR, S. S., MCCLURE, E. M., PASHA, O., CARLO, W. A., GARCES, A., MOORE, J. L. et al. (2015). The global network maternal newborn health registry: A multi-national, community-based registry of pregnancy outcomes. *Reproductive Health* **12** S1. <https://doi.org/10.1186/1742-4755-12-S2-S1>
- BRADLEY, S. E. K., WINFREY, W. and CROFT, T. N. (2015). Contraceptive use and perinatal mortality in the DHS: An assessment of the quality and consistency of calendars and histories. DHS Methodological Reports No. 17. ICF International, Rockville, Maryland, USA.
- CARPENTER, B., GELMAN, A., HOFFMAN, M. D., LEE, D., GOODRICH, B., BETANCOURT, M., BRUBAKER, M., GUO, J., LI, P. et al. (2017). Stan: A probabilistic programming language. *J. Stat. Softw.* **76**. <https://doi.org/10.18637/jss.v076.i01>
- CARVALHO, C. M., POLSON, N. G. and SCOTT, J. G. (2009). Handling sparsity via the horseshoe. *Artificial Intelligence and Statistics* **8**.
- CURRIE, I. D. and DURBAN, M. (2002). Flexible smoothing with  $P$ -splines: A unified approach. *Stat. Model.* **2** 333–349. MR1951589 <https://doi.org/10.1191/1471082x02st039ob>
- EILERS, P. H. C. and MARX, B. D. (1996). Flexible smoothing with  $B$ -splines and penalties. *Statist. Sci.* **11** 89–121. With comments and a rejoinder by the authors. MR1435485 <https://doi.org/10.1214/ss/1038425655>
- GELMAN, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Anal.* **1** 515–533. MR2221284 <https://doi.org/10.1214/06-BA117A>
- GELMAN, A. and RUBIN, B. (1992). Inference from iterative simulation using multiple sequences. *Statist. Sci.* **7** 457–72.
- HUG, L., YOU, D., BLENCOWE, H., MISHRA, A., WANG, Z., FIX, M. J., WAKEFIELD, J., MORAN, A. C., GAIGBE-TOGBE, V. et al. (2021). Global, regional, and national estimates and trends in stillbirths from 2000 to 2019: A systematic assessment. *Lancet* **398** 772–785.
- KURUVILLA, S., BUSTREO, F., KUO, T., MISHRA, C., TAYLOR, K., FOGSTAD, H., GUPTA, G. R., GILMORE, K., TEMMERMAN, M. et al. (2016). The *Global strategy for women's, children's and adolescents' health (2016–2030)*: A roadmap based on evidence and country experience. *Bull. World Health Organ.* **94** 398–400. <https://doi.org/10.2471/BLT.16.170431>
- LIU, L., KALTER, H. D., CHU, Y., KAZMI, N., KOFFI, A. K., AMOUZOU, A., JOOS, O., MUNOS, M. and BLACK, R. E. (2016). Understanding misclassification between neonatal deaths and stillbirths: Empirical evidence from Malawi. *PLoS ONE* **11** e0168743. <https://doi.org/10.1371/journal.pone.0168743>
- PEDERSEN, J. and LIU, J. (2012). Child mortality estimation: Appropriate time periods for child mortality estimates from full birth histories. *PLoS Med.* **9** e1001289. <https://doi.org/10.1371/journal.pmed.1001289>
- PIIRONEN, J. and VEHTARI, A. (2017a). Sparsity information and regularization in the horseshoe and other shrinkage priors. *Electron. J. Stat.* **11** 5018–5051. MR3738204 <https://doi.org/10.1214/17-EJS1337SI>
- PIIRONEN, J. and VEHTARI, A. (2017b). Comparison of Bayesian predictive methods for model selection. *Stat. Comput.* **27** 711–735. MR3613594 <https://doi.org/10.1007/s11222-016-9649-y>
- STANTON, C., LAWN, J. E., RAHMAN, H., WILCZYNSKA-KETENDE, K. and HILL, K. (2006). Stillbirth rates: Delivering estimates in 190 countries. *Lancet* **367** 1487–1494. [https://doi.org/10.1016/S0140-6736\(06\)68586-3](https://doi.org/10.1016/S0140-6736(06)68586-3)
- STAN DEVELOPMENT TEAM (2018). RStan: The R interface to Stan. R package version 2.18.2.
- UN INTER-AGENCY GROUP FOR CHILD MORTALITY ESTIMATION (2020). A neglected tragedy: The global burden of stillbirths. Report of the UN Inter-agency Group for Child Mortality Estimation, 2020. United Nations Children's Fund, New York.
- VEHTARI, A., GELMAN, A. and GABRY, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat. Comput.* **27** 1413–1432. MR3647105 <https://doi.org/10.1007/s11222-016-9696-4>
- VEHTARI, A., GABRY, J., MAGNUSSON, M., YAO, Y. and GELMAN, A. (2019). loo: Efficient leave-one-out cross-validation and WAIC for Bayesian models. R package version 2.2.0.
- VEHTARI, A., GELMAN, A., SIMPSON, D., CARPENTER, B. and BÜRKNER, P.-C. (2021). Rank-normalization, folding, and localization: An improved  $\hat{R}$  for assessing convergence of MCMC (with discussion). *Bayesian Anal.* **16** 667–718. Includes comments and discussions by seven discussants and a rejoinder by the authors. MR4298989 <https://doi.org/10.1214/20-ba1221>
- WANG, Z., FIX, M. J., HUG, L., MISHRA, A., YOU, D., BLENCOWE, H., WAKEFIELD, J. and ALKEMA, L. (2022). Supplement to “Estimating the stillbirth rate for 195 countries using a Bayesian sparse regression model with temporal smoothing.” <https://doi.org/10.1214/21-AOAS1571SUPP>
- WHO DEPARTMENT OF INFORMATION, EVIDENCE, AND RESEARCH (2018). WHO methods and data sources for life tables 1990–2016.
- WOODS, R. (2008). Long-term trends in fetal mortality: Implications for developing countries. *Bull. World Health Organ.* **2008** 460–466. <https://doi.org/10.2471/BLT.07.043471>

- WORLD HEALTH ORGANIZATION (2014). *Every Newborn: An Action Plan to End Preventable Deaths*. World Health Organization, Geneva. OCLC: 918973731.
- WORLD HEALTH ORGANIZATION (2019). *International Classification of Disease (ICD), Revised Definition for ICD-11*.