

RESEARCH ARTICLE

Estimating misclassification errors in the reporting of maternal mortality in national civil registration vital statistics systems: A Bayesian hierarchical bivariate random walk model to estimate sensitivity and specificity for multiple countries and years with missing data

Emily Peterson¹  | Doris Chou² | Ann-Beth Moller² | Alison Gemmill³ |
Lale Say² | Leontine Alkema⁴

¹Department of Biostatistics and Bioinformatics, Emory University, Atlanta, Georgia

²Department of Sexual and Reproductive Health and Research, World Health Organization, Geneva, Switzerland

³Department of Population, Family, and Reproductive Health, Johns Hopkins University, Baltimore, Maryland, USA

⁴Department of Biostatistics and Epidemiology, University of Massachusetts Amherst, Amherst, Massachusetts, USA

Correspondence

Emily Peterson, Department of Biostatistics and Bioinformatics, Emory University, Atlanta, Georgia.
Email: emily.nancy.peterson@emory.edu

Funding information

WHO; USAID; University of Massachusetts Amherst

Civil registration vital statistics (CRVS) systems provide data on maternal mortality that can be used for monitoring trends and to inform policies and programs. However, CRVS maternal mortality data may be subject to substantial reporting errors due to misclassification of maternal deaths. Information on misclassification is available for selected countries and periods only. We developed a Bayesian hierarchical bivariate random walk model to estimate sensitivity and specificity for multiple populations and years and used the model to estimate misclassification errors in the reporting of maternal mortality in CRVS systems. The proposed Bayesian misclassification (BMis) model captures differences in sensitivity and specificity across populations and over time, allows for extrapolations to periods with missing data, and includes an exact likelihood function for data provided in aggregated form. Validation exercises using maternal mortality data suggest that BMis is reasonably well calibrated and improves upon the CRVS-adjustment approach used until 2018 by the UN Maternal Mortality Inter-Agency Group (UN-MMEIG) to account for bias in CRVS data resulting from misclassification error. Since 2019, BMis is used by the UN-MMEIG to account for misclassification errors when estimating maternal mortality using CRVS data.

KEYWORDS

Bayesian hierarchical models, maternal mortality, misclassification

1 | INTRODUCTION

National civil vital registration (CRVS) systems record the number of deaths in a population, as well as the cause associated with each death using ICD coding.¹ Under ideal circumstances, when all deaths are captured and all causes

License: CC BY IGO. The views expressed in this article are those of the authors and do not necessarily reflect the views of the WHO, UNICEF, UNFPA, the World Bank Group, or the United National Population Division.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 World Health Organization; licensed by John Wiley & Sons Ltd. *Statistics in Medicine* published by John Wiley & Sons Ltd.

are accurately classified, CRVS systems provide perfect information on the number of deaths associated with a given course. However, cause-specific deaths may be reported incorrectly if deaths are unregistered or misclassified, where misclassification of deaths refers to incorrect coding in vital registration systems, due either to error in the medical certification of cause of death or error in applying the correct ICD code. In the presence of under-registration or misclassification, the estimation of cause-specific mortality based on CRVS data needs to account for bias associated with reporting errors.

This article focuses on the estimation of misclassification errors, as summarized in sensitivity and specificity of reporting, for multiple populations and periods, in the presence of missing data. This work is motivated by the assessment of maternal mortality based on CRVS data for countries with CRVS systems. A maternal death is “the death of a woman whilst pregnant or within 42 days of termination of pregnancy, irrespective of the duration and site of the pregnancy, from any cause related to or aggravated by the pregnancy or its management but not from accidental or incidental causes.”¹ The reporting of maternal mortality in CRVS systems is often subject to substantial misclassification errors.²⁻⁴ Prior work comparing the ratio of the proportion of deaths to women of reproductive age that are maternal (PM) as obtained from specialized studies—rigorous assessments of maternal mortality for a given country-period—to CRVS-based PMs found that these ratios are around 150%.² Given that substantial misclassification errors may be present in CRVS recording of maternal deaths, accounting for such errors is necessary to obtain unbiased estimates of maternal mortality. Unfortunately, information on misclassification errors in CRVS systems is typically available for a limited number of years only. In addition, information may be available in aggregated form only, providing only the corrected and CRVS-reported total number of maternal deaths, as opposed to the number of deaths that were misclassified.

We developed a Bayesian hierarchical misclassification model (referred to as the BMis model), which produces estimates of sensitivity and specificity for multiple population-periods in presence of missing data. Our contributions to the statistical literature related to estimating sensitivity and specificity are twofold. First, we introduce a bivariate hierarchical temporal model to capture differences in sensitivity and specificity across populations and over time. This extends the work on bivariate modeling of sensitivity and specificity by Reitsma et al⁵ and the hierarchical modeling approach used by Chu and Cole⁶ to allow for extrapolations to periods with missing data and account for temporal autocorrelation in the estimation of sensitivity and specificity.^{5,6} Second, we introduce an exact likelihood function for data provided in aggregated form.

The BMis model is used by the UN Maternal Mortality Inter-Agency Group (UN-MMEIG) to estimate and account for sensitivity and specificity in the reporting of maternal mortality from 1985 until 2017 for all WHO member states. The UN-MMEIG publishes internationally comparable estimates of maternal mortality for UN reporting. The group uses a Bayesian hierarchical time series regression model (BMat) to estimate the proportion of maternal deaths out of all cause deaths (PM).^{3,7,8} In BMat, estimates of the PM are produced based on the available input data for the respective country-period, taking account of data quality issues in reporting. In prior rounds of constructing estimates, the UN-MMEIG adjusted CRVS data based on simplifying assumptions, that is, keeping constant the ratio of bias-corrected to CRVS-reported PM in country-periods with missing data on misclassification. We show that BMis improves upon the prior UN-MMEIG approach in validation exercises.

This article is organized as follows: Section 2 introduces the data available to inform estimation of the extent of incorrect reporting. Section 3 introduces a Bayesian model to estimate the extend of misclassification in the reporting of maternal deaths in CRVS systems. The estimation is based on summarizing misclassification in terms of sensitivity and specificity, and modeling these two indicators for all country-years with CRVS data using a bivariate hierarchical random walk model. Finally, we present findings and the validation exercises.

2 | DATA

2.1 | CRVS data on maternal mortality

The WHO Mortality Database maintains data from CRVS systems.⁹ CRVS systems collect information based on death certification data, which provides ICD coded certified causes of death. Further information regarding CRVS systems and how they vary across countries can be found in Mikkelsen et al.¹⁰ For each country-year c, t , this database provides the number of maternal deaths reported in the CRVS, denoted by $y_{c,t}^{(\text{mat})}$ and the number of deaths to women aged 15 to 49 reported in the CRVS, denoted as $y_{c,t}^{(\text{crvs})}$, referred to as the CRVS envelope.

In this study, we focus on assessing misclassification errors among deaths reported within CRVS systems. Deaths reported within CRVS systems are all deaths to women of reproductive ages for systems that capture all deaths, and a

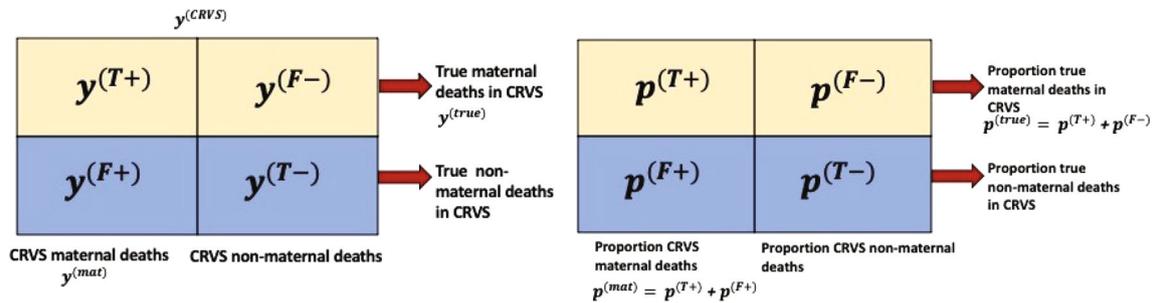


FIGURE 1 Diagram of breakdown of CRVS reporting of deaths to women of reproductive age for a country-year, by CRVS-assigned maternal cause (columns) and true maternal cause (rows). (Left) $y^{(T+)}$ and $y^{(F-)}$ deaths refer to maternal deaths that are correctly registered as maternal deaths, and incorrectly registered as non-maternal deaths, respectively. Similarly, $y^{(F+)}$ and $y^{(T-)}$ maternal deaths refers to non-maternal deaths that are incorrectly registered as maternal deaths, and correctly registered as non-maternal deaths, respectively. (Right) Corresponding observed proportions denoted by p referring to counts $y^{(T+)}$, $y^{(F-)}$, $y^{(F+)}$, and $y^{(T-)}$ out of total CRVS-reported deaths to women of reproductive age $y^{(CRVS)}$

subset of such deaths for incomplete systems that do not capture all deaths. Completeness of the reporting of deaths into the CRVS system was assessed by comparing CRVS-reported deaths to WHO estimates of deaths to women of reproductive age, obtained from life tables for WHO Member States.^{7,11} In this study, we focus on data from country-years for which CRVS completeness is either deemed complete, or for which we have specialized studies data referring to the deaths as reported in the CRVS.

2.2 | Specialized studies data

In the context of this study, we define specialized studies as those studies that provide accurate information on the number of maternal deaths among CRVS-reported deaths of women of reproductive age. Specialized studies include a review of cause of death for all deaths to women of reproductive age registered in the CRVS system. These studies may be carried out independently of CRVS-reported data or based on the checking of CRVS-reported deaths.

We constructed a data base with information on maternal mortality from specialized studies. Specialized studies were obtained through (1) a literature review, (2) the UN MMEIG 2015 maternal mortality data base,³ and (3) information provided by countries based on a follow-up survey, sent to countries in response to discussions with the Pan-American Health Organization (PAHO), and during country consultation. Details on the literature review and inclusion criteria are provided in Supplementary Appendix Section S6.2.

2.3 | Data on misclassification errors in CRVS reporting of maternal mortality

Misclassification errors in CRVS reporting of maternal mortality refers to the inaccurate attribution of a maternal cause of death, due to error in the medical certification of cause of death, and/or error in applying the correct code. Misclassification errors occur in the form of either false negative ($F-$) or false positive ($F+$) maternal deaths. False negative maternal deaths refer to deaths that are of a maternal cause but not reported as such in the CRVS while false positives refer to non-maternal deaths that are reported as maternal in the CRVS.

The diagram in Figure 1 illustrates misclassification errors in CRVS systems by breaking down total deaths to women of reproductive age by CRVS-reporting status (columns) and true maternal cause (rows). For country-year c, t , the total number of false negative and false positive maternal deaths is denoted as $y_{c,t}^{(F-)}$ and $y_{c,t}^{(F+)}$, respectively. From the individual categories in Figure 1, cumulative totals are calculated summing across rows and columns, that is, CRVS-reported maternal deaths, denoted by $y^{(mat)}$, are the sum of $y_{c,t}^{(T+)}$ and $y_{c,t}^{(F+)}$, whereas, the true number of maternal deaths within the CRVS, $y_{c,t}^{(true)}$, is the sum of $y_{c,t}^{(T+)}$ and $y_{c,t}^{(F-)}$. For a complete list of definitions, refer to Table S4 in Supplementary Appendix S6.1.

Information on CRVS misclassification errors was obtained from comparing information from CRVS-reported deaths to data from specialized studies that refer to the same envelope of deaths. Data availability is summarized in Table 1. We have information from studies on 204 unique country-periods (observations) in 32 countries. Reported information varied

TABLE 1 Overview of data available from combining specialized studies and CRVS data

Reported counts	# of observations	# of countries
True maternal in CRVS only	147	26
$F-$ and $F+$	18	24
$F-$ only	39	5
Total	204	32

across observations. While some studies reported a detailed breakdown of false positive and/or false negative maternal deaths, the majority of studies reported only the confirmed total number of maternal deaths for a given country-period (147 observations, 26 countries). Information on both false negative and false positive breakdowns was available for 18 observations (4 countries). Most studies with breakdown information solely reported on false negative breakdowns, 39 observations from 5 countries.

3 | METHODS

3.1 | Notation and definitions of parameters

We refer to observed quantities using lower-case Roman letters, with $y_{c,t}^{(b)}$ and $p_{c,t}^{(b)}$ referring to observed counts and proportions of deaths in the CRVS for country c in year t in category b , respectively, with $b \in B = \{T+, T-, F+, F-\}$, see Figure 1. CRVS-reported maternal deaths are denoted by $y_{c,t}^{(\text{mat})} = y_{c,t}^{(T+)} + y_{c,t}^{(F+)}$, CRVS-reported total deaths are denoted by $y_{c,t}^{(\text{CRVS})}$, and the true number of maternal deaths in the CRVS is denoted by $y_{c,t}^{(\text{true})} = y_{c,t}^{(T+)} + y_{c,t}^{(F-)}$.

We refer to unknown parameters using lower-case Greek letters. With the same notation for sub- and superscripts as introduced above for y , we let $\gamma_{c,t}^{(b)}$ refer to the probability that a death that is reported in the CRVS is classified into category b , such that $\sum_{b \in B} \gamma_{c,t}^{(b)} = 1$, see Figure 2. For example, $\gamma^{(F-)}$ refers to the probability of a false negative maternal death within the CRVS system, and $\gamma^{(F+)}$ refers to the probability of a false positive maternal death within the CRVS system, respectively. With this definition of counts y and probabilities γ , it follows that,

$$y_{c,t} | y_{c,t}^{(\text{CRVS})}, \gamma_{c,t} \sim \text{Multinom} \left(y_{c,t}^{(\text{CRVS})}, \gamma_{c,t} \right), \quad (1)$$

where (leaving out subscripts (c, t) for improve readability):

$$\begin{aligned} \mathbf{y} &= (y^{(T-)}, y^{(T+)}, y^{(F-)}, y^{(F+)}) , \\ \boldsymbol{\gamma} &= (\gamma^{(T-)}, \gamma^{(T+)}, \gamma^{(F-)}, \gamma^{(F+)}) . \end{aligned}$$

We define sensitivity of CRVS reporting of maternal mortality as the probability that a maternal death in the CRVS is classified as such: $\lambda^{(+)} = \gamma_{c,t}^{(T+)} / \gamma_{c,t}^{(\text{true})}$, and specificity as a non-maternal death classified as such: $\lambda^{(-)} = \gamma_{c,t}^{(T-)} / (1 - \gamma_{c,t}^{(\text{true})})$. Based on the definitions of sensitivity $\lambda^{(+)}$ and specificity $\lambda^{(-)}$, we can write the cell-specific probabilities $(\gamma_{c,t}^{(T+)}, \gamma_{c,t}^{(F-)}, \gamma_{c,t}^{(F+)}, \gamma_{c,t}^{(T-)})$ as a function of both measures and the true probability of a maternal death $\gamma_{c,t}^{(\text{true})}$, see Figure 2. The relation between the CRVS-reported probability of a maternal death $\gamma_{c,t}^{(\text{mat})}$ and the true probability $\gamma_{c,t}^{(\text{true})}$ is as follows:

$$\gamma_{c,t}^{(\text{mat})} = \lambda^{(+)} \gamma_{c,t}^{(\text{true})} + (1 - \lambda^{(-)}) (1 - \gamma_{c,t}^{(\text{true})}) . \quad (2)$$

3.2 | Objective and summary of modeling approach

We aim to estimate sensitivity $\lambda_{c,t}^{(+)}$ and specificity $\lambda_{c,t}^{(-)}$ in the reporting of maternal mortality for *all* country-years with CRVS data, including country-years for which no specialized study data are available, such that CRVS data can be used to inform the estimation of maternal mortality while accounting for CRVS misclassification errors.

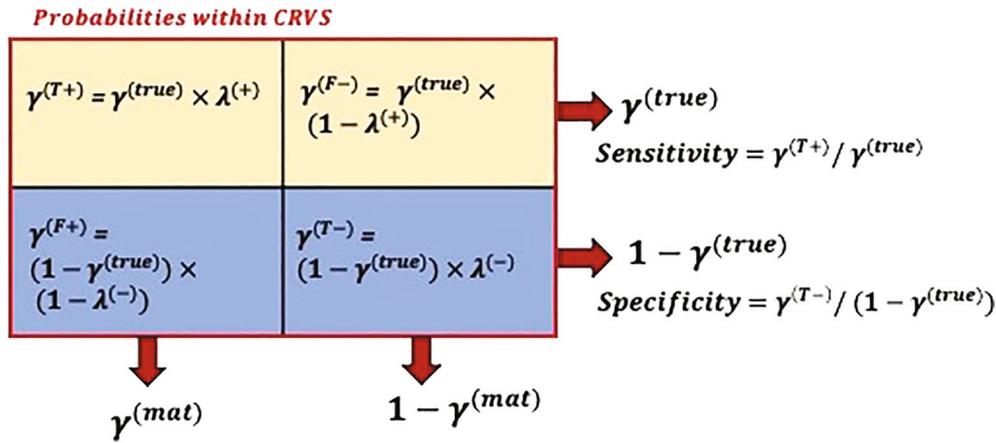


FIGURE 2 Diagram of breakdown of probabilities within CRVS systems for a given country-year, by CRVS-reporting status (columns) and true maternal cause (rows), where $\gamma^{(b)}$ refers to the probability of a CRVS-reported death being in category $b \in B = \{T+, T-, F+, F-\}$. The sum of $\gamma^{(T+)}$ and $\gamma^{(F-)}$ probabilities equal the probability of a true maternal deaths within CRVS ($\gamma^{(true)}$). Similarly, $\gamma^{(F+)}$ and $\gamma^{(T-)}$ equal the probability of a reported maternal death within CRVS ($\gamma^{(mat)}$). The four probabilities are parameterized into metrics of sensitivity = $\frac{\gamma^{(T+)}}{\gamma^{(true)}}$ and specificity = $\frac{\gamma^{(T-)}}{1 - \gamma^{(true)}}$

We developed a bivariate hierarchical random walk model for estimating sensitivity and specificity for all country-years. The process model for sensitivity and specificity is introduced in Section 3.3. We obtained estimates of sensitivity and specificity for maternal mortality reporting in CRVS systems using all available data for model fitting, including data on the total number of maternal deaths only. Model fitting is explained in Section 3.4, including the specification of the likelihood function.

3.3 | Bivariate hierarchical random walk model for sensitivity and specificity

We developed a bivariate hierarchical random walk model to estimate sensitivity $\lambda_{c,t}^{(+)}$ and specificity $\lambda_{c,t}^{(-)}$ for all countries c with CRVS data for some year(s) t . Estimates of sensitivity and specificity were constrained to be within 0 and 1 using probit-transformations.

The model set-up used is a bivariate hierarchical random walk process on probit-transformed sensitivity (se) and specificity (sp). In brief, we model the country-year specific probit-transformed se and sp as the sum of country-specific average levels over the period of interest (1985-2017) plus a linear combination of first order differences. The model captures correlation in sensitivity and specificity, which may arise based on efforts to improve the classification of maternal deaths. Specifically, if a CRVS system attempts to capture more true maternal deaths then there will be an increase in sensitivity, that is, more true maternal deaths are captured due to increase in classified maternal deaths. However, there may also be a decrease in specificity, that is, more non-maternal deaths will be classified as maternal, which increases the rate of false positive maternal deaths. Therefore, accounting for this correlation is important for estimating misclassification parameters, that is, see Chu et al.^{5,6} Details on the model specification are provided in the remainder of this section.

Let probit-transformed se and sp be denoted by $\eta_{c,t}^{(+)} = \Phi(\lambda_{c,t}^{(+)})$ and $\eta_{c,t}^{(-)} = \Phi(\lambda_{c,t}^{(-)})$, respectively, where $\Phi()$ denotes the probit-transform. With the definition of country-specific averages $\eta_c^0 = \frac{1}{T} \sum \eta_{c,t}^0$ and first-order differences $\xi_{c,t}^0 = \eta_{c,t}^0 - \eta_{c,t-1}^0$, we can write the η 's as linear combinations of the average levels and first order differences,

$$\begin{aligned} \eta_{c,1:T}^{(+)} &= \eta_c^{(+)} + \mathbf{D} \cdot \xi_{c,1:T-1}^{(+)} \\ \eta_{c,1:T}^{(-)} &= \eta_c^{(-)} + \mathbf{D} \cdot \xi_{c,1:T-1}^{(-)} \end{aligned} \tag{3}$$

where matrix $\mathbf{D} = \mathbf{W}' \times (\mathbf{W} \cdot \mathbf{W}')^{-1}$, with

$$\mathbf{W} = \begin{bmatrix} -1 & 1 & 0 & 0 & \dots & 0 \\ 0 & -1 & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & -1 & 1 \end{bmatrix}$$

We assume a hierarchical bivariate distribution for country-specific levels of average probit-transformed sensitivity and specificity given by:

$$\begin{pmatrix} \eta_c^{(+)} \\ \eta_c^{(-)} \end{pmatrix} \sim N_2 \left(\begin{bmatrix} \eta_{\text{global}}^{(+)} \\ \eta_{\text{global}}^{(-)} \end{bmatrix}, \begin{bmatrix} \sigma^{(+)^2} & \rho \cdot \sigma^{(+)} \cdot \sigma^{(-)} \\ \rho \cdot \sigma^{(+)} \cdot \sigma^{(-)} & \sigma^{(-)^2} \end{bmatrix} \right), \quad (4)$$

in which the country-specific levels of transformed sensitivity and specificity $(\eta_c^{(+)}, \eta_c^{(-)})$ are distributed bivariate normal centered on their respective global levels $(\eta_{\text{global}}^{(+)}, \eta_{\text{global}}^{(-)})$. The variance and correlation terms, associated with global levels of probit-transformed sensitivity and specificity, are denoted by $\sigma^{(+)}$, $\sigma^{(-)}$, and ρ .

The first-order differences, $\xi_{c,t}$ are modeled with a zero-mean bivariate normal given by:

$$\begin{pmatrix} \xi_{c,t}^{(+)} \\ \xi_{c,t}^{(-)} \end{pmatrix} \sim N_2 \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \delta^{(+)^2} & \phi \cdot \delta^{(+)} \cdot \delta^{(-)} \\ \phi \cdot \delta^{(+)} \cdot \delta^{(-)} & \delta^{(-)^2} \end{bmatrix} \right). \quad (5)$$

We assigned uniform prior distributions to the inverse-probit transformed global mean values of probit-transformed sensitivity and specificity. Specifically, we define $\lambda_{\text{global}}^{(0)} = \Phi^{-1}(\eta_{\text{global}}^{(0)})$, or equivalently, $\eta_{\text{global}}^{(0)} = \Phi(\lambda_{\text{global}}^{(0)})$, and assign uniform priors to the $\lambda_{\text{global}}^{(0)}$'s:

$$\lambda_{\text{global}}^{(+)} \sim \text{Unif}(0, 1),$$

$$\lambda_{\text{global}}^{(-)} \sim \text{Unif}(0, 1).$$

Prior distributions for the correlation and standard deviations of the random walk are given below, using vague half-normal distributions for variance parameters:¹²

$$\phi \sim \text{Unif}(-1, 1), \quad (6)$$

$$\rho \sim \text{Unif}(-1, 1), \quad (7)$$

$$\sigma^0 \sim N_{T(0,\infty)}(0, 1), \quad (8)$$

$$\delta^0 \sim N_{T(0,\infty)}(0, 1), \quad (9)$$

where $N_{T(0,\infty)}(0, 1)$ denotes a half-normal distribution (a truncated normal distribution with lower bound at 0).

We explored the use of indicators gross domestic product (GDP), the general fertility rate (GFR), the proportion of ill-defined causes, CRVS completeness, and ICD coding (ICD10 or earlier) as possible covariates to inform estimates of sensitivity and specificity. However, exploratory analyses suggested no substantially meaningful relations. Covariates were not included in the final model (see Supplementary Appendix Section S6.3 for exploratory plots shown in Figures S7 and S8).

3.4 | Model fitting

3.4.1 | Process model

For each country-year with data, we define the four unknown CRVS-based probabilities $\gamma_{c,t}^{(b)}$ in terms of the two misclassification parameters $\lambda_{c,t}^{(+)}$ and $\lambda_{c,t}^{(-)}$, and the true CRVS-based probability of a maternal death $\gamma_{c,t}^{(\text{true})}$ as follows:

$$\begin{aligned} \gamma_{c,t}^{(T+)} &= \lambda_{c,t}^{(+)} \cdot \gamma_{c,t}^{(\text{true})}, \\ \gamma_{c,t}^{(F-)} &= \gamma_{c,t}^{(\text{true})} - \gamma_{c,t}^{(T+)}, \\ \gamma_{c,t}^{(T-)} &= \lambda_{c,t}^{(-)} \cdot (1 - \gamma_{c,t}^{(\text{true})}), \\ \gamma_{c,t}^{(F+)} &= (1 - \gamma_{c,t}^{(\text{true})}) - \gamma_{c,t}^{(T-)}. \end{aligned} \quad (10)$$

Misclassification parameters are defined through the bivariate hierarchical random model on $\lambda_{c,t}^{(+)}$ and $\lambda_{c,t}^{(-)}$, as explained in the previous section. We use vague independent priors on $\gamma_{c,t}^{(\text{true})}$:

$$\gamma_{c,t}^{(\text{true})} \sim U(0.0001, 1).$$

3.4.2 | Likelihood function

We define $z_i^{(b)}$ to refer to the observed death count for category b in study $i = 1, \dots, n$. We let $c[i]$ denote the country to which study i refers. The start calendar year for study i is indexed by $t1[i]$, its end year by $t2[i]$, and its observation period midyear is given by $t[i]$. As per earlier notation, let $y_{c,t}^{(b)}$ denote the number of deaths in category b , for country c in year t , such that we can write $z_i^{(b)} = \sum_{t=t1[i]}^{t2[i]} y_{c[i],t}^{(b)}$.

We relate multinomial study counts $\mathbf{z}_i = (z_i^{(T-)}, z_i^{(T+)}, z_i^{(F-)}, z_i^{(F+)})$ to the corresponding within CRVS probabilities $\boldsymbol{\gamma}_{c,t} = (\gamma_{c,t}^{(T-)}, \gamma_{c,t}^{(T+)}, \gamma_{c,t}^{(F-)}, \gamma_{c,t}^{(F+)})$, assuming a multinomial data generation process as follows:

$$\mathbf{z}_i | z_i^{(\text{CRVS})}, \boldsymbol{\gamma}_{c[i],t[i]} \sim \text{Multinom} \left(z_i^{(\text{CRVS})}, \boldsymbol{\gamma}_{c[i],t[i]} \right), \tag{11}$$

with $z_i^{(\text{CRVS})} = \sum_b z_i^{(b)}$, and unknown probability vector $\boldsymbol{\gamma}_{c,t} = (\gamma_{c,t}^{(T-)}, \gamma_{c,t}^{(T+)}, \gamma_{c,t}^{(F-)}, \gamma_{c,t}^{(F+)})$.

The corresponding density for $\mathbf{z}_i | z_i^{(\text{CRVS})}, \boldsymbol{\gamma}_{c[i],t[i]}$ is given by,

$$p_z(\mathbf{z}_i | z_i^{(\text{CRVS})}, \boldsymbol{\gamma}_{c[i],t[i]}) = \frac{z_i^{(\text{CRVS})}!}{\prod_b z_i^{(b)}!} \prod_b \gamma_{c[i],t[i]}^{(b) z_i^{(b)}}. \tag{12}$$

Study reported data fall into two categories of data reporting. The first category consists of studies which report a specific set of non-overlapping death counts $(z_i^{(T-)}, z_i^{(T+)}, z_i^{(F-)}, z_i^{(F+)})$ or some non-overlapping subset or aggregation of these counts. The second category consists of studies which report overlapping counts, that is, studies reporting true maternal counts $z_i^{(\text{true})} = z_i^{(T+)} + z_i^{(F-)}$, which overlaps with the CRVS-reported maternal deaths for the corresponding country-period, $z_i^{(\text{mat})} = z_i^{(T+)} + z_i^{(F+)}$. We give a simple illustration of data in both categories and details on the corresponding likelihood function in the Supplementary Appendix, see Section S6.4.1. The likelihood functions are summarized in the remainder of this section.

For studies that report on a specific set of non-overlapping categories, that is, the number of false positive maternal deaths and/or the number of true positive maternal deaths, the corresponding likelihood function was obtained directly using the multinomial data generating process in Equation (11). Specifically, the likelihood function for studies that report $F+$ and $F-$ deaths, all four categories, is given directly by $p_z(\mathbf{z} | z_i^{(\text{CRVS})}, \boldsymbol{\gamma}_{c[i],t[i]})$, where p_z refers to the multinomial density function for the four CRVS-based categories from Equation (11). For studies that report $(z_i^{(F-)}, z_i^{(T-)}, z_i^{(\text{mat})})$, the likelihood function follows from aggregating the relevant non-overlapping categories in the multinomial:

$$\left(z_i^{(F-)}, z_i^{(T-)}, z_i^{(\text{mat})} \right) | z_i^{(\text{CRVS})}, \boldsymbol{\gamma}_{c[i],t[i]} \sim \text{Multinom} \left(z_i^{(\text{CRVS})}, \left(\gamma_{c[i],t[i]}^{(F-)}, \gamma_{c[i],t[i]}^{(T-)}, \gamma_{c[i],t[i]}^{(\text{mat})} \right) \right).$$

The second category of studies consists of the majority of studies, which only reported information on the number of true maternal deaths within the CRVS (see Table 1). For these studies, data are given by $\mathbf{d}_i = (z_i^{(\text{true})}, z_i^{(\text{mat})})$ and the study-reported count of maternal deaths, $z_i^{(\text{true})} = z_i^{(T+)} + z_i^{(F-)}$, overlaps with the CRVS-reported maternal deaths for the corresponding country-period, $z_i^{(\text{mat})} = z_i^{(T+)} + z_i^{(F+)}$. Details on the corresponding likelihood function are given in Section S6.4.2. In brief, for each study period with information on overlapping categories, we obtained the exact likelihood function for the available death counts by summing over multinomial densities evaluated at each unique combination that satisfied the observed set of counts. We obtain the complete set of unique combinations denoted by $\tilde{\mathbf{z}}_i^{(s)} = (z_i^{(T+)^{(s)}}, z_i^{(F+)^{(s)}}, z_i^{(T-)^{(s)}}, z_i^{(F-)^{(s)})}$ for $s = 1, \dots, S[i]$, which satisfies the observed marginals, given by,

$$\tilde{z}_i^{(mat)(s)} = \tilde{z}_i^{(T+)(s)} + \tilde{z}_i^{(F+)(s)} = z_i^{(mat)}, \tag{13}$$

$$\tilde{z}_i^{(true)(s)} = \tilde{z}_i^{(T+)(s)} + \tilde{z}_i^{(F-)(s)} = z_i^{(true)}. \tag{14}$$

The likelihood function for data \mathbf{d}_i is given by:

$$p(\mathbf{d}_i | \boldsymbol{\gamma}_{c[i],t[i]}) = \sum_{s=1}^{S[i]} p(\tilde{\mathbf{z}}_i^{(s)} | \boldsymbol{\gamma}_{c[i],t[i]}) \cdot k_i^{(s)}, \tag{15}$$

with density $p(\tilde{\mathbf{z}}_i^{(s)} | \boldsymbol{\gamma}_{c[i],t[i]})$ given in Equation (11). Indicator function $k_i^{(s)}$ is added to exclude combinations with negligible probability of being the true combination: $k_i^{(s)} = 0$ for combination s with negligible probability of being the true combination, and $k_i^{(s)} = 1$ otherwise. This indicator function was added to improve computational efficiency when fitting the model to studies with large numbers of unique combinations. Exclusions are carried out for combinations that are based on sensitivity less than 0.97, accounting for stochastic uncertainty. The indicator function is defined as follows:

$$k_i^{(s)} = 1 \left(\tilde{z}_i^{(T-)(s)} > q_i^{(0.97)} \right), \tag{16}$$

where $q_i^{(0.97)}$ refers to the 2.5th percentile of a Binomial distribution with sample size $z_i^{(CRVS)} - z_i^{(true)}$ and probability 0.97. For further information on exclusion of combinations with negligible probability, see Appendix Section S6.4.2.1.

3.4.3 | Computation

A Markov chain Monte Carlo (MCMC) algorithm was employed to sample from the posterior distribution of the parameters with the use of the software JAGS.^{13,14} Ten parallel chains were run with a total of 40 000 iterations in each chain. Of these, the first of 10 000 iterations in each chain were discarded as burn-in and every 20th iteration after was retained. The resulting chains contained 1500 samples each, with a total of 15 000 posterior samples. Standard diagnostic checks (using trace plots and Gelman and Rubin diagnostics) were used to check convergence.¹⁵

3.5 | Comparison of estimates of misclassification bias in CRVS-based PM between the BMis model and the UN MMEIG 2015 approach

3.5.1 | UN-MMEIG2015 approach

The BMat model, which estimates the proportion of maternal deaths out of all-cause deaths, uses CRVS adjustment factors, denoted $\Omega_{c,t}$ for country c and year t , to correct for bias in CRVS-based reported PM associated with misclassification of maternal deaths.⁸ In the UN MMEIG 2015 approach, the CRVS adjustment factor was summarized using the ratio of the observed PM as reported in specialized studies to the PM reported in the CRVS, resulting in point estimate $\hat{\Omega}_{c,t}^{(UN-MMEIG2015)}$. The adjustment ratios were obtained for all country-years with CRVS data and used in a CRVS bias-adjustment preprocessing step.⁸ For countries with specialized studies for a subset of country-years only, linear interpolation was used to obtain adjustments in years in between observed adjustments. For forward extrapolation, the CRVS adjustment was kept constant at the level of the most recent observed CRVS adjustment. For backward extrapolations in countries with studies, the CRVS adjustment was assumed to increase or decrease linearly to the same global adjustment factor of 1.5 in 5 years. The uncertainty of the adjustment was expert-based: the standard deviation of the log-transformed adjustment was set equal to the variability associated with $\Omega_{c,t}$, defined as follows:

$$\log(\Omega_{c,t}^{(UN-MMEIG2015)}) \sim N \left(\log(\hat{\Omega}_{c,t}^{(UN-MMEIG2015)}), 0.25^2 \right).$$

3.5.2 | BMis approach

Our approach to assess bias and associated CRVS adjustments differs from the UN MMEIG 2015 approach; the CRVS adjustment factor in the BMis model is obtained from MCMC samples of sensitivity and specificity and varies with the true PM. Specifically, the model-based sample estimate of the CRVS adjustment factor $\Omega_{c,t}^{(BMis)^{(s)}}$ for country c in year t , s th sample, with complete CRVS is given by:

$$\Omega_{c,t}^{(BMis)^{(s)}} = \frac{\gamma_{c,t}^{(true)}}{\lambda_{c,t}^{(+)(s)} \cdot \gamma_{c,t}^{(true)} + (1 - \lambda_{c,t}^{(-)(s)}) \cdot (1 - \gamma_{c,t}^{(true)})}, \tag{17}$$

in which $\lambda_{c,t}^{(s)}$ refers to the s th sample estimates of sensitivity and specificity from BMis, and the true PM $\gamma_{c,t}^{(true)}$ is estimated within the BMat model. In using MCMC samples, we propagate the uncertainty associated with sensitivity and specificity into the derived samples for the CRVS adjustment factor.

3.6 | Model validation

We carried out model validation to check the performance of the proposed approach in terms of predicting the relative bias associated with CRVS-reported PMs and compared it to the UN-MMEIG (2015) approach. Model performance was assessed through the two out-of-sample validation exercises. In the first exercise (repeated 20 times), 20% of the observations were left out at random to form a training data set. In the second exercise, we left out the observation corresponding to the most recent study period in each country. In each exercise, we fitted the BMis model to the training set and obtained posterior samples for sensitivity and specificity in the country-years with left-out specialized studies. We combined samples of sensitivity and specificity with the reported study-based PMs to obtain samples of predicted CRVS-based PMs. We summarized the difference between the model-based predictions and the reported CRVS-based PM in terms of error, that is, the difference between the observed CRVS-based PM and its point estimate, and coverage of 80% prediction intervals. The procedure is described in detail in Figure 3.

The UN MMEIG 2015 approach was validated using the same training and test sets and results were compared between the two approaches. For the UN-MMEIG approach, observed CRVS adjustment factors were calculated based on the training set, and interpolation and extrapolation was carried out as described in Section 3.5.

Calculation of outcome measures in the validation exercise

1. Model fitting: Fit the BMis model to the training data and obtain posterior samples $\lambda_{c,t}^{(+)(s)}$ and $\lambda_{c,t}^{(-)(s)}$ for posterior samples $s = 1, 2, \dots, S$ for country-years with left-out data in the test set.
2. Predict CRVS-based PM $p_i = z_i^{(mat)} / z_i^{(CRVS)}$: For each left-out observations with index i , derive S samples of the CRVS-based reported number of maternal deaths using the S samples for sensitivity and specificity, and the observed true PM $p_i^{(true)}$:

$$p_i^{(mat)(s)} = p_i^{(true)} \cdot \lambda_{c[i],t[i]}^{(+)(s)} + (1 - p_i^{(true)}) \cdot (1 - \lambda_{c[i],t[i]}^{(-)(s)})$$

3. Error calculation: Calculate the difference between observed CRVS-based PM and the samples from its posterior predictive distribution:

$$error_i^{matCRVS(s)} = (p_i^{(mat)} - p_i^{(mat)(s)})$$

Various summaries of the errors are reported.

4. Calibration: Calculate the proportion of CRVS-based PMs $p_i^{(mat)}$ above and below their respective 80% prediction interval.

FIGURE 3 Overview of calculation of errors and coverage of prediction intervals in out-of-sample validation exercises

4 | RESULTS

4.1 | Validation results

The BMis model performs well in out-of-sample validation exercises, see Table 2. Median and relative errors are small in both exercises, and absolute errors are around 8% when predicting the CRVS-based PM. The model is well calibrated, the coverage of the 80% prediction intervals is around 80%, with around 10% falling below (above) the lower (upper) bounds.

The BMis model improves upon the UN-MMEIG 2015 approach in terms of prediction error: mean errors and mean absolute errors are smaller in magnitude in the BMis model as compared to the UN-MMEIG 2015 approach. For example, the median absolute relative error (MARE) decreases from 15.9% to 4.6% between the UN-MMEIG 2015 and the BMis approach when predicting the observations left out at random. Similarly, in the leave-out last observation validation, there is a reduction in median absolute relative error from 14.4% to 6.1% between the old and new model.

4.2 | Global findings

Table 3 lists the posterior estimates of the hyperparameters and their respective 80% credible intervals (CI). Global estimates of sensitivity and specificity were estimated to be 0.593 with 80%CI given by (0.523, 0.659) and 0.9995 (0.9992, 0.9997), respectively. The correlation of changes in sensitivity and specificity in countries over time was estimated to be negative at -0.737 (-0.370 , 0.096). Correlation between average levels of country-specific sensitivity and specificity was estimated to be uncertain with CI ranging from -0.409 to 0.786 .

TABLE 2 Validation results

Error in CRVS-PM								
Validation	Model	# left-out obs	Median errors		Relative error (%)		Outside 80% PI	
			ME	MAE	MRE	MARE	Prop below	Prop above
Leave-out 20% at random	BMis model	43	-0.00006	0.0004	-0.6	4.6	0.13	0.10
	UN MMEIG 2015	43	-0.00010	0.0009	-1.8	15.9	0.08	0.05
Leave-out last observation	BMis model	20	0.0001	0.0007	3.2	6.1	0.10	0.10
	UN MMEIG 2015	20	-0.0003	0.0010	-4.0	14.4	0.10	0.10

Note: The outcome measures are: median error (ME), median absolute error (MAE), relative error (MRE), absolute relative error (MARE), as well as the % of left-out observations below and above their respective 80% prediction intervals (PI) based on the training set. For the leave-out 20% exercise, reported errors and coverage summarize within-country mean errors and coverage.

TABLE 3 Posterior estimates of global parameters; median estimate (50%) and lower (10%) and upper (90%) bounds of 80% credible intervals

	10%	50%	90%
Global sensitivity $\lambda_{\text{global}}^{(+)}$	0.523	0.593	0.659
Global specificity $\lambda_{\text{global}}^{(-)}$	0.9992	0.9995	0.9997
Correlation (global) ρ	-0.409	0.307	0.786
SD (global) sensitivity $\sigma^{(+)}$	0.450	0.599	0.806
SD (global) specificity $\sigma^{(-)}$	0.0344	0.157	0.326
Correlation (first-order differences) ϕ	-0.737	-0.370	0.096
SD (first-order differences) sensitivity $\delta^{(+)}$	0.095	0.117	0.146
SD (first-order differences) specificity $\delta^{(-)}$	0.099	0.131	0.162

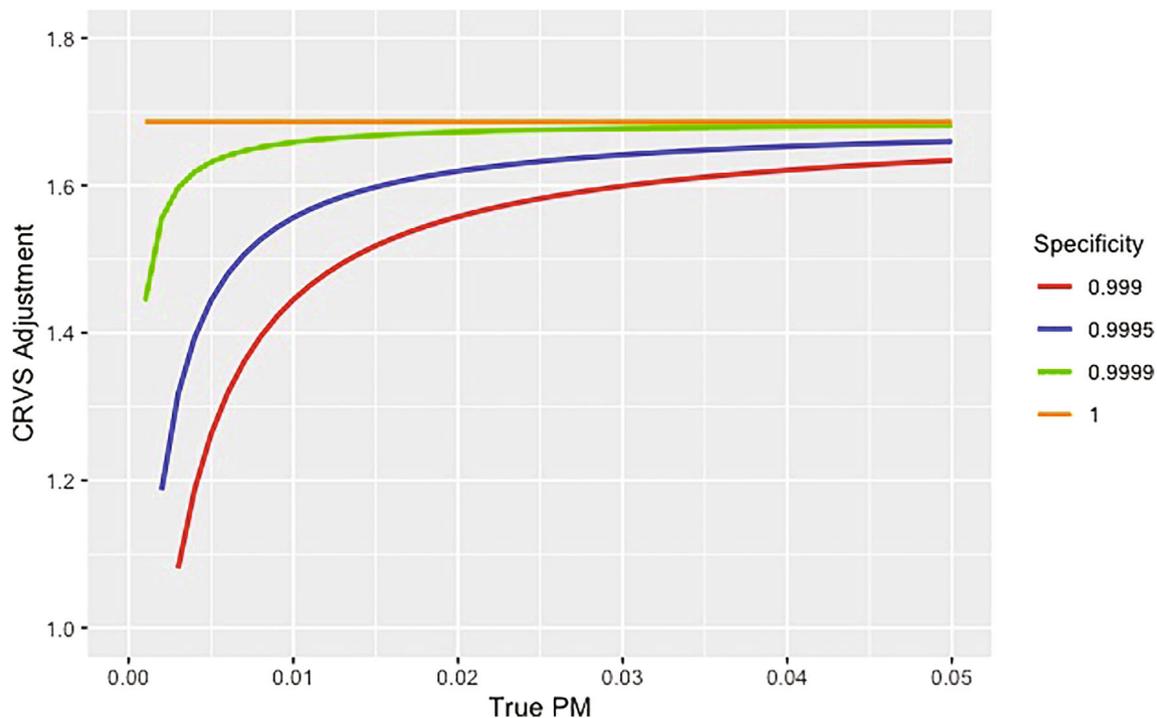


FIGURE 4 CRVS adjustment for different values of specificity, calculated at different levels of true PM when sensitivity is fixed at the global estimate of 0.593

Figure 4 shows the relationship between true PM and the estimated CRVS adjustment factors, for specific values of specificity to illustrate their effect on the CRVS adjustment factor. When specificity equals one, the CRVS adjustment factor equals one over sensitivity, hence lower sensitivity results in a higher adjustment; conversely higher sensitivity results in a lower adjustment. When specificity is less than one, while keeping sensitivity fixed, the adjustment factor decreases with decreasing true PM. This effect is due to an increasing share of false positive maternal deaths among all deaths, and a decreasing share of false negative deaths. In other words, as the true PM decreases, the number of non-maternal deaths reported as maternal increases while the number of maternal deaths reported as non-maternal decreases. This relationship implies that keeping specificity and sensitivity constant in extrapolations will result in changing adjustment factors as the true PM changes. Specifically, the adjustment factor will decrease if the true PM decreases in forward projections. Similarly, when using a fixed value of sensitivity and specificity, the adjustment factor associated with these values will depend on the value of the true PM.

Figure 4 also highlights that small differences in specificity, for a given value of sensitivity, result in notable differences in CRVS adjustment factor, especially at low PM. These differences are due to the large share of non-maternal deaths, and thus the relatively larger share of false positives, among all deaths.

4.3 | Country estimates

Sensitivity, specificity and CRVS adjustment estimates are shown for selected countries in Figure 5. Results for all countries are included in Appendix Section S6.5. In the plots, posterior estimates (blue) are shown with observed data (red) during the estimation period.

For the selected countries in Figure 5, given by Australia, Brazil, the United Kingdom, Republic of Korea, and New Zealand, estimates of sensitivity range from 0.468 in Brazil to 0.87 in New Zealand, and estimates of specificity range from 0.997 in New Zealand to 0.9998 in Australia. The results illustrates how uncertainty in estimates of sensitivity and specificity depends on (i) what information is available, (ii) the number of deaths in the country, and (iii) the observation years. In some countries, breakdown information on false positive and false negative deaths is

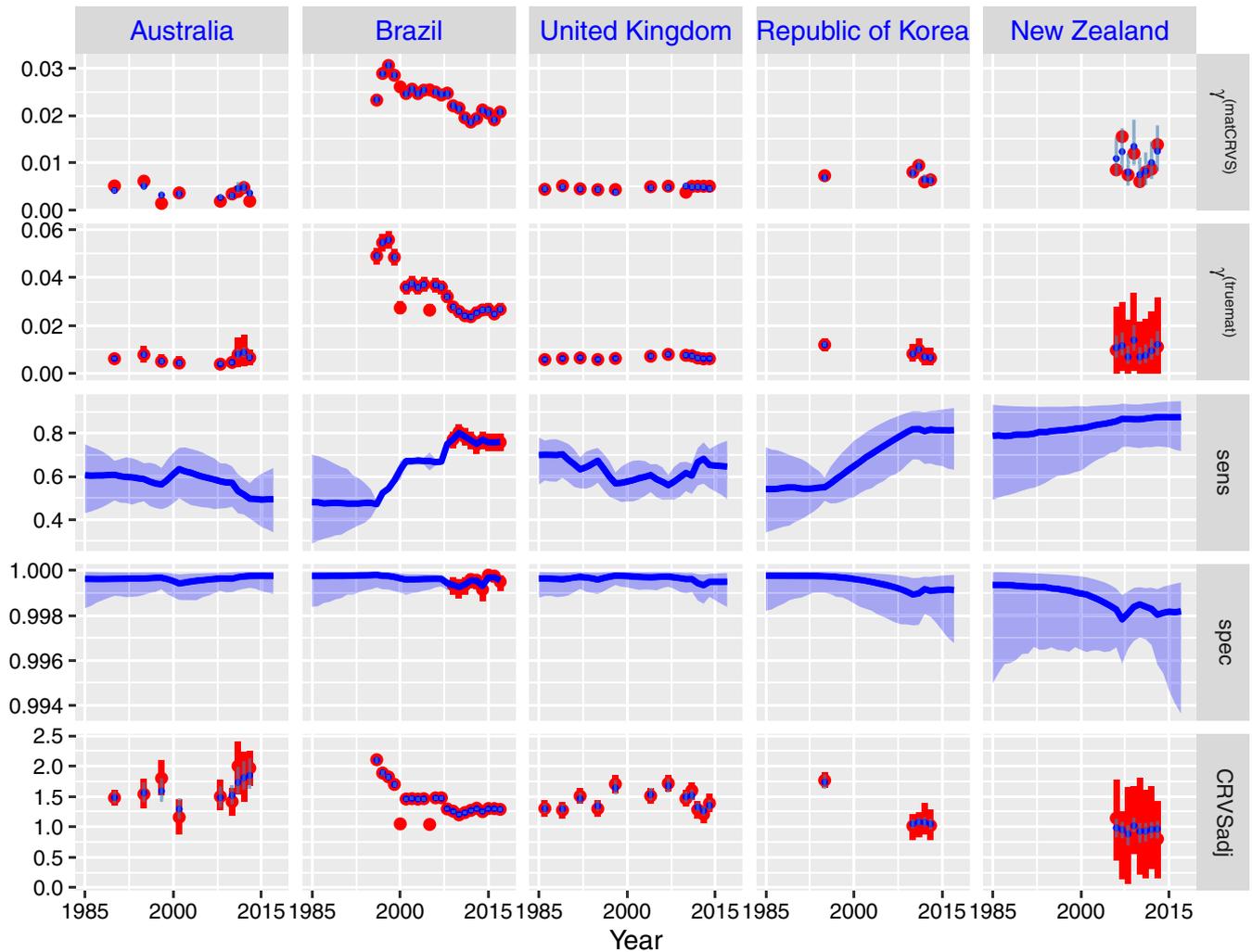


FIGURE 5 Illustration of BMis model data and estimates for Australia, Brazil, the United Kingdom, Republic of Korea, and New Zealand. Parameters plotted are CRVS-based PM, true PM, sensitivity, specificity, and CRVS adjustment factors. The plots include: (1) Observed data with associated observation-based 80% confidence intervals (red), (2) posterior estimates with 80% credible intervals (blue)

available. An example country is Brazil, where sensitivity and specificity are recorded for recent years. Resulting estimates in those years are more certain than estimates in Brazil for earlier years (1996-2008). In addition to availability of data, the number of deaths in the country also determines the uncertainty in estimated sensitivity and specificity. For example, data in New Zealand is very uncertain due to the small number of maternal deaths and total number of deaths to women of reproductive age. Lastly, uncertainty in sensitivity and specificity increases in years further away from years with data. For example in Brazil and New Zealand, uncertainty increases in backprojections prior to the earliest observation year. In Australia and the Republic of Korea, we see more uncertainty in long periods without data.

5 | CONCLUSION

In this article, we presented a Bayesian hierarchical random walk model to assess maternal mortality misclassification errors in the CRVS with uncertainty. The model is based on the assessment of sensitivity and specificity of maternal mortality reporting, and captures differences therein between countries and within countries over time. Validation exercises suggest that the model performs well in terms of predicting CRVS-based PM for country-periods without specialized studies.

The new model improves upon limitations of the 2015 UN MMEIG approach. In the UN MMEIG 2015 round of estimation, for countries with specialized studies that overlapped with CRVS data, adjustments were calculated directly from available data (ie, the study's reported PM to CRVS-based PM) and kept constant in extrapolations. The rationale for keeping adjustments constant in the 2015 approach for countries with studies was to implement “no change in quality of reporting.” However, when measuring quality of reporting in terms of sensitivity and specificity, the adjustment is not constant but varies with the true PM when keeping quality metrics constant, as illustrated in Figure 4. The BMis model-based approach to obtaining adjustment factors improves upon this limitation of the UN MMEIG 2015 approach because its projections, which are based on constant sensitivity and specificity, are aligned with the assumption of constant quality of reporting. Model validation suggests that the new approach reduces errors in predicted CRVS-based PM, in both validation exercises, that is, 20% random exclusions and leaving out the last observation. Finally, uncertainty assessments differ between the old and new approach. In the old approach, uncertainty in adjustments was assumed to be around 50% for all country-periods. In the new approach, uncertainty in the adjustment factor follows from the uncertainty in the estimates for sensitivity and specificity and resulting adjustments are more certain in settings with recent information about quality of reporting. Validation exercises suggests that the new model is well calibrated.

Substantively, we estimate that sensitivity of maternal death reporting globally is 0.593 (0.523, 0.659)—in other words, that more than half the number of maternal deaths are reported as such—and that there are substantial differences in sensitivity and specificity across countries and within countries over time. More information is needed to understand why some CRVS systems have high or low estimates of sensitivity and specificity, to guide improvements in such systems for improved monitoring and to allow for more precise predictions of misclassification for country-years without specialized studies.

Our work contributes to the statistics literature by proposing a bivariate hierarchical temporal model that captures differences in sensitivity and specificity across populations and over time and a likelihood function for data provided in aggregated form. The modeling approach allows for extrapolations to periods with missing data and accounts for temporal autocorrelation in the estimation of sensitivity and specificity. In this article, we applied the approach to the estimation of misclassification errors in maternal mortality in CRVS systems. Its application would extend naturally to the estimation of misclassification in other cause fractions where golden-standard data are available for a subset of years. The approach can also be considered for other sets of populations or data systems, such as subnational estimation, where populations are defined by geographical regions within countries, and health information systems.

ACKNOWLEDGEMENTS

The authors are very grateful to all members of the United Nations Maternal Mortality Estimation Inter-Agency Group for the support of this work and technical advisors Saiffuddin Ahmed, Peter Byass, Thomas Pullum, Tim Colbourn, Jeff Eaton, Marie Klingberg Alvin, Helena Nordenstedt, Laina Mercer, and Jon Wakefield. We are also grateful to the PAHO regional office collaborators Antonio Sanhueza, Patricia Lorena Ruiz Luna and Adrienne Lavita Cox. Additional thanks to Michael Lavine, Laura Balzer, and Ken Kleinman for specific comments and suggestions related to the BMis model. We thank the numerous participants and staff involved in the collection and publication of the data that we analyzed. We thank Maria Barriex for help with the systematic review. Financial support was provided by WHO, through the Department of Reproductive Health and Research and HRP (the UNDP-UNFPA-UNICEF-WHO-World Bank Special Programme of Research, Development and Research Training in Human Reproduction), USAID, and University of Massachusetts Amherst.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available upon request from the corresponding author and the World Health Organization.

ORCID

Emily Peterson  <https://orcid.org/0000-0002-7606-8445>

REFERENCES

1. World Health Organization. International classification of diseases for mortality and morbidity statistics (10th revision); May 2010. <https://icd.who.int/en>

2. Wilmoth J, Mizohguchi N, Oestergaard M, et al. A new method for deriving global estimates of maternal mortality. *Stat Polit Policy*. 2012;3(2):2151-7509. 1038.
3. United Nations Maternal Mortality Estimation Inter-Agency Group. Trends in maternal mortality: 1990 to 2015. Estimates by WHO, UNICEF, UNFPA, World Bank and the United Nations Population Division; 2015. <https://www.who.int/reproductivehealth/publications/monitoring/maternal-mortality-2015/en/>
4. Alkema L, Chou D, Hogan D, et al. Global, regional, and national levels and trends in maternal mortality between 1990 and 2015, with scenario-based projections to 2030: a systematic analysis by the UN Interagency group for maternal mortality estimation. *Lancet*. 2016;387(10017):462-474.
5. Reitsma J, Glass A, Rutjes A, Scholten R, Bossuyt P, Zwinderman A. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol*. 2005;58(10):982-990.
6. Chu H, Cole S. Bivariate meta-analysis of sensitivity and specificity with sparse data: a generalized linear mixed model approach. *J Clin Epidemiol*. 2006;59(12):1331-1333.
7. United Nations Maternal Mortality Estimation Inter-Agency Group. Trends in maternal mortality: 2000 to 2017. Estimates by WHO, UNICEF, UNFPA, World Bank and the United Nations Population Division; September 19, 2019. <https://unfpa.org/featured-publication/trends-maternal-mortality-2000-2017>
8. Alkema L, Zhang S, Chou D, et al. A Bayesian approach to the global estimation of maternal mortality. *Ann Appl Stat*. 2017;11(3):1245074.
9. World Health Organization. WHO maternal mortality database; November 2015; Geneva. <https://www.who.int/reproductivehealth/publications/monitoring/maternal-mortality-2015/en/>
10. Mikkelsen L, Phillips DE, AbouZahr C, et al. A global assessment of civil registration and vital statistics systems: monitoring data quality and progress. *Lancet*. 2015;386(10001):1395-1406.
11. Peterson E, Chou D, Moller AB, Gemmill A, Say L, Alkema L. Estimation of maternal mortality using data from national civil registration vital statistics systems: a Bayesian hierarchical bivariate random walk model to estimate sensitivity and specificity of reporting; 2019. arXiv:1909.08578.
12. Gelman A. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Anal*. 2006;1(3):515-534.
13. Plummer M. JAGS: a program for analysis of Bayesian graphical models using Gibbs sampling; October 13, 2021. <https://sourceforge.net/projects/mcmc-jags>
14. Su YS, Yajima M. R2Jags: using R to run 'JAGS' 2020; 2020. <https://CRAN.R-project.org/package=R2jags>
15. Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. *Stat Sci*. 1992;7(4):457-472.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Peterson E, Chou D, Moller A-B, Gemmill A, Say L, Alkema L. Estimating misclassification errors in the reporting of maternal mortality in national civil registration vital statistics systems: A Bayesian hierarchical bivariate random walk model to estimate sensitivity and specificity for multiple countries and years with missing data. *Statistics in Medicine*. 2022;41(14):2483-2496. doi: 10.1002/sim.9335