

DISCUSSION PAPER SERIES

Discussion paper No. 266

Bayesian analysis of verbal autopsy data using factor models with age- and sex-dependent associations between symptoms

Tsuyoshi Kuniham

(Department of Economics, Kwansei Gakuin University)

Zehang Richard Li

(Department of Statistics, University of California, Santa Cruz)

Samuel J. Clark

(Department of Sociology, Ohio State University)

Tyler H. McCormick

(Department of Statistics and Department of Sociology, University of Washington)

March 2024



SCHOOL OF ECONOMICS

KWANSEI GAKUIN UNIVERSITY

1-155 Uegahara Ichiban-cho
Nishinomiya 662-8501, Japan

Bayesian analysis of verbal autopsy data using factor models with age- and sex-dependent associations between symptoms

Tsuyoshi Kuniyama¹, Zehang Richard Li², Samuel J. Clark³, and Tyler H. McCormick⁴

¹Department of Economics, Kwansei Gakuin University

²Department of Statistics, University of California, Santa Cruz

³Department of Sociology, Ohio State University

⁴Department of Statistics and Department of Sociology, University of Washington

January, 2024

Abstract

Verbal autopsies (VAs) are extensively used to investigate the population-level distributions of deaths by cause in low-resource settings without well-organized vital statistics systems. Computer-based methods are often adopted to assign causes of death to deceased individuals based on the interview responses of their family members or caregivers. In this article, we develop a new Bayesian approach that extracts information about cause-of-death distributions from VA data considering the age- and sex-related variation in the associations between symptoms. Its performance is compared with that of existing approaches using gold-standard data from the Population Health Metrics Research Consortium. In addition, we compute the relevance of predictors to causes of death based on information-theoretic measures.

Key words: Bayesian factor models; Causes of death distribution; Multivariate data; Verbal autopsies; Survey data.

1 Introduction

Cause-of-death distributions provide fundamental information about the current dynamics of a population, which is crucial for designing, monitoring, and evaluating public health actions. However, this information is inaccessible to decision-makers in low-resource regions, where many deaths occur outside hospitals and are not officially registered. A survey-based method called verbal autopsy (VA) has been developed as a practical solution; it

involves inferring the cause of death without medical certification and estimating cause-specific mortality rates in populations without complete-coverage civil registration and vital statistics systems (Maher et al., 2010; Sankoh and Byass, 2012; Nkengasong et al., 2020). In VA, information about the cause of death of an individual is collected by interviewing a person close to them with questions about their demographics, medical history, and signs, and symptoms in the period leading to their death. The VA interview needs to be analyzed to assign a probable cause of death, and a common practice is to have a panel of physicians interpret the interview data. However, this consumes physicians' time for patients and is difficult to scale up for massive amounts of data. In recent years, as an alternative, algorithmic and statistical methods have been developed for cause-of-death assignment using VA data (Byass et al., 2003; Serina et al., 2015; McCormick et al., 2016). Computer-based approaches have the advantages of being scalable for large-scale data and becoming free once their programming codes are publicly distributed. For example, popular methods are easily implemented using the openly available R package `openVA` (Li et al., 2023). Chandramohan et al. (2021) conducted a comprehensive review of the VA developments over the last several decades.

Regarding VA interview questions as predictors, statistical methods assign a cause of death based on the probability of causes given these predictors $P(\text{cause} | \text{predictors})$. Standard parametric models, such as multinomial probit/logit regression, can be applied to this conditional probability. However, VA surveys usually contain large amounts of missing values and thus require the challenging task of imputing high-dimensional data with complex interactions. The common approach is to use an indirect expression based on Bayes theorem as $P(\text{cause} | \text{predictors}) \propto P(\text{predictors} | \text{cause})P(\text{cause})$, which makes it easy to accommodate abundant missing values under the missing-at-random assumption. Popular methods assume the conditional independence of predictors given a cause (Byass et al., 2012, 2019; Miasnikof et al., 2015; McCormick et al., 2016), but VA surveys can contain more than a hundred of questionnaire items and may violate this assumption. Recent studies investigated approaches to expressing complex dependent structures among predictors (Li et al., 2020; Kunihama et al., 2020; Moran et al., 2021).

VA surveys collect various information relevant to causes of death, ranging from the

demographic backgrounds to the symptoms, signs, or medical history (henceforth collectively called symptoms) of deceased individuals. All survey items are treated equally as predictors in existing methods, although some of them may play different roles in cause-of-death assignment. Demographic characteristics, especially age and sex, are key factors in many public health settings. For example, worldwide differences by sex and age have been identified in disability-adjusted life years (World Health Organization, 2020) and the cancer burden (Ferlay et al., 2020). In addition, World Health Organization (2022) reported inequalities in the distribution of COVID-19 cases and deaths across different demographic profiles such that males are more likely than females to die and older people are more likely to develop serious illness. The report also shows age and sex differences in preventive behaviors during the pandemic, such as social distancing, and mask wearing. Considering the key roles of age and sex in public health, we explicitly differentiate them from other predictors and assume they affect the associations between symptoms for each cause of death. Then, assuming predictors consist of symptoms, age and sex, we consider the expression

$$\begin{aligned}
P(\text{cause} \mid \text{predictors}) &\propto P(\text{predictors} \mid \text{cause})P(\text{cause}) \\
&\propto P(\text{symptoms, age, sex} \mid \text{cause})P(\text{cause}) \\
&\propto P(\text{symptoms} \mid \text{age, sex, cause})P(\text{age, sex} \mid \text{cause})P(\text{cause}). \tag{1}
\end{aligned}$$

Moran et al. (2021) developed a framework where the distribution of symptoms differs by covariates, namely, $P(\text{cause} \mid \text{predictors}) \propto P(\text{symptoms} \mid \text{covariates, cause})P(\text{cause})$. In comparison to (1), with age and sex set as covariates, the assumption of independence between (age, sex) and causes of death is implicitly imposed; that is, $P(\text{age, sex} \mid \text{cause}) = P(\text{age, sex})$. Given that age and sex are often associated with causes of death (World Health Organization, 2020), avoiding the independence assumption is appropriate.

1.1 PHMRC VA survey data

Using gold-standard VA data, we examine the following points: 1. The associations between symptoms vary by age and sex. 2. The distributions of age and sex differ by cause of death. We analyze neonatal VA data collected by the Population Health Metrics Research

Consortium (PHMRC) project (Murray et al., 2011; PHMRC, 2013). These data consist of medically certified causes of death and VA survey responses collected by interviewers who were blinded to the causes of death assigned in hospitals. We use five neonatal causes of death and 97 predictors for 1,619 babies, and all predictors are binarized based on the transformation by Murray et al. (2011). Figure 1 shows a histogram of the causes of death in our study, and the predictors are listed in the supplementary materials.

As for the associations between symptoms, we compute Cramér’s V (Cramér, 1946) that measures strength of associations between two nominal variables, producing a value between 0 (no association) and 1 (complete association). Figure 2 shows Cramér’s V values of all pairs of symptoms for deaths by birth asphyxia between age groups and between sexes. If age and sex do not affect the associations between symptoms, then the corresponding points will be close to the diagonal lines. Figure 3 displays the absolute difference from each of the points to the diagonal line in Figure 2, indicating that some symptom pairs show relatively large differences between age groups and between sexes. Therefore, it is worth pursuing a direction of modeling the age- and sex-dependent associations between symptoms to estimate the causes-of-death distribution.

Figure 4 shows the proportions of age groups and sexes for each cause of death. The bar patterns differ between causes. For example, most of the babies who died due to birth asphyxia are in the early age group, and the ratio of females is higher than that of males for pneumonia and preterm delivery. Hence, age and sex are considered not independent of causes of death.

The remainder of this paper is organized as follows. Section 2 shows the proposed Bayesian approach to estimating cause-of-death distributions using factor models with age- and sex-dependent associations between symptoms. Section 3 describes the Markov chain Monte Carlo (MCMC) algorithm of the proposed method. In Section 4, we assess the proposed approach in comparison to existing methods and evaluate the relevance of the predictors to the causes of death using the PHMRC neonate dataset. Section 5 concludes this article.

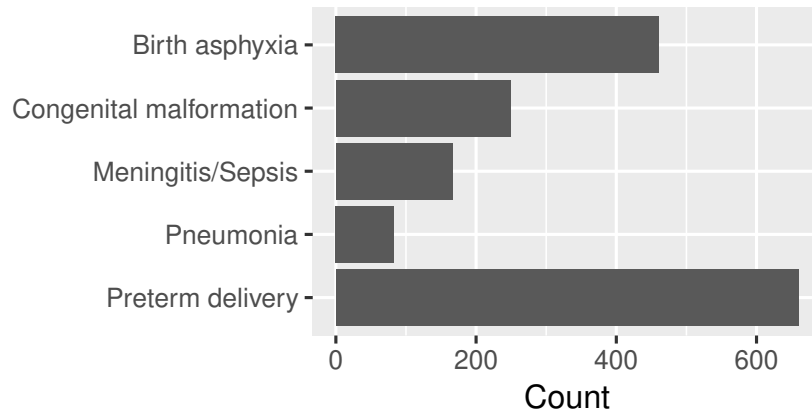


Figure 1: Histogram of causes of death in our PHMRC neonate dataset.

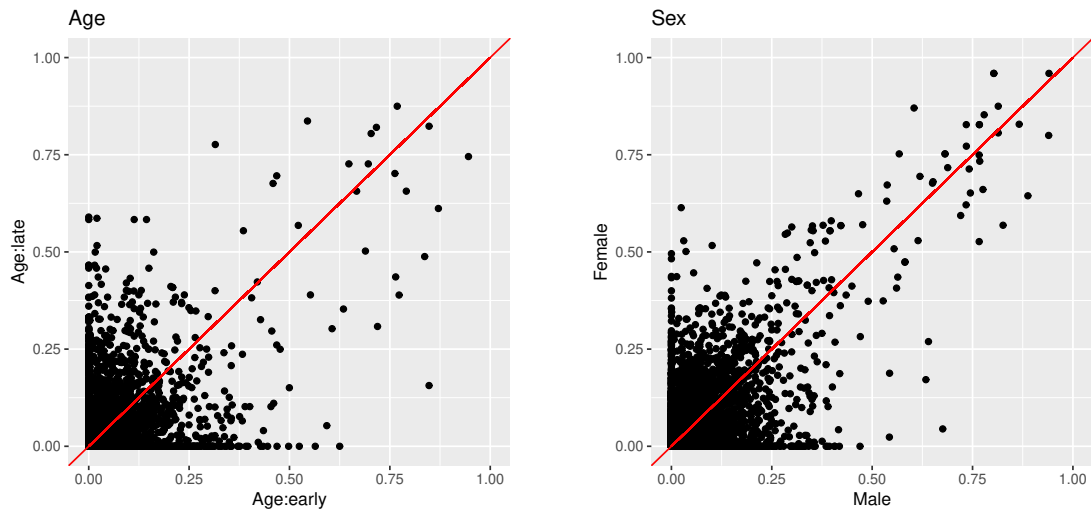


Figure 2: Plots of Cramér's V values for each pair of symptoms for birth asphyxia between age groups (left) and between sexes (right).

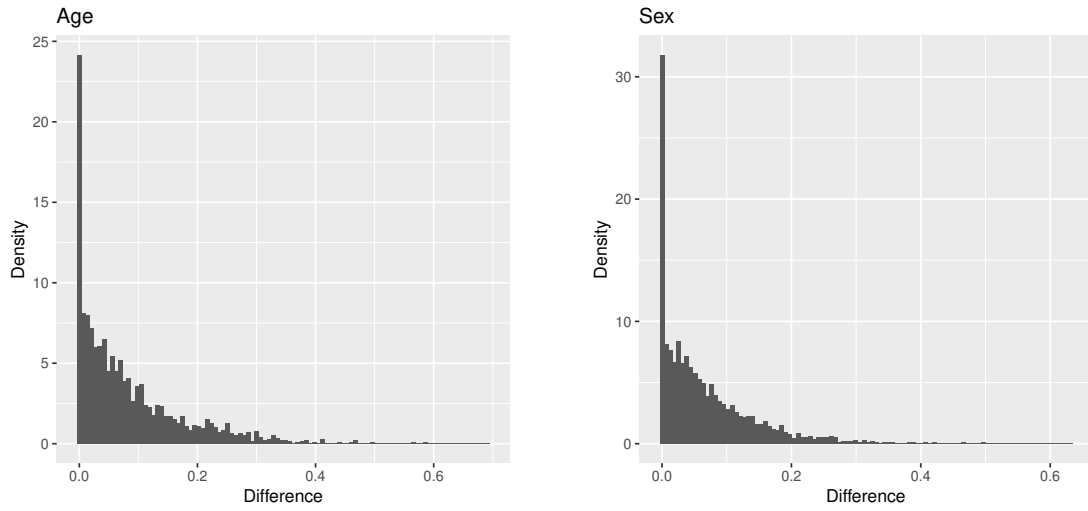


Figure 3: Histograms of absolute difference in Cramér's V values for each pair of symptoms for birth asphyxia between age groups (left) and between sexes (right).

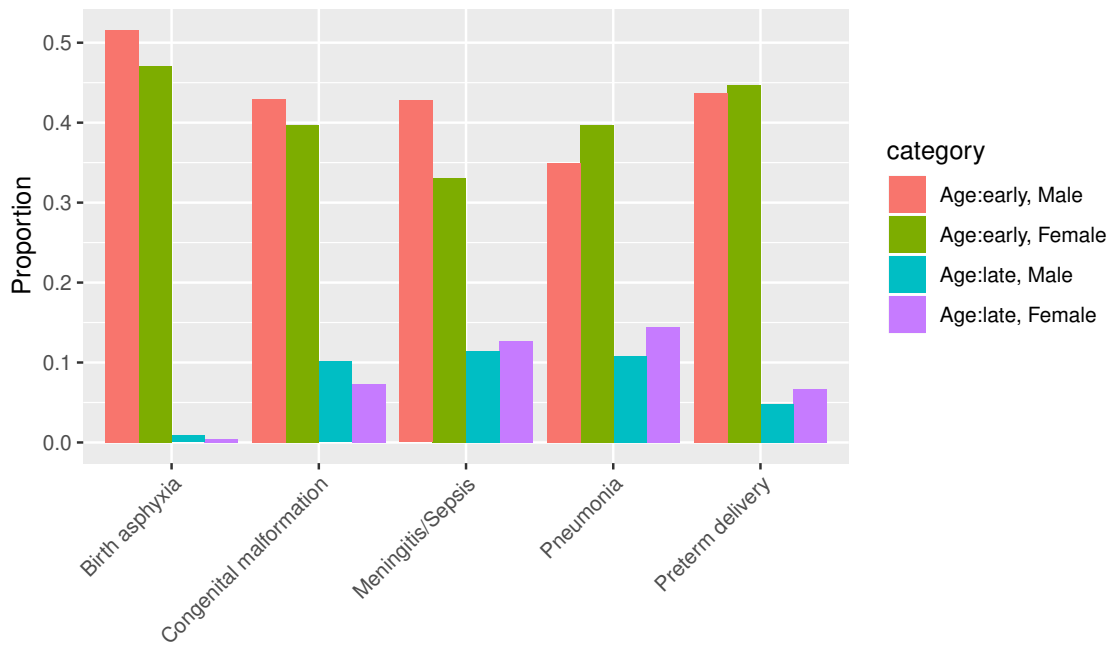


Figure 4: Proportions of age (early and late stages) and sex (male and female) for each cause of death in our PHMRC neonate dataset.

2 Proposed Bayesian approach

We develop the Bayesian models in (1) for estimating the population-level distribution of the causes of death using the VA survey data. The proposed approach divides the VA predictors into basic demographic information (age and sex) and the other questionnaire items (symptoms), and the associations between symptoms differs by age and sex. In addition, we measure the relevance of each predictor to the causes of death using the proposed models. For the notations below, let $y_i \in \{1, \dots, L\}$ be the cause of the i th person's death, where $i = 1, \dots, n$, and $x_i = (x_{i1}, \dots, x_{ip})'$ be the binary indicators of the symptoms, where $x_{ij} \in \{0, 1\}$ for $j = 1, \dots, p$. Moreover, age_i and sex_i indicate the age and sex, respectively, of the i th baby.

2.1 Modeling of $P(\text{symptoms} \mid \text{age, sex, cause})$ in (1)

First, we develop models for symptoms in which the dependence structure varies between the age and sex groups for each cause of death. Let $z_i = (z_{i1}, \dots, z_{ip})'$ be a latent continuous variable for x_i . The Bayesian factor (BF) model (Kunihama et al., 2020) is expressed as

$$\begin{aligned} x_{ij} &= 1(z_{ij} > 0), \quad j = 1, \dots, p, \\ z_i &= \mu_{y_i} + \Lambda_{y_i} \eta_i + \varepsilon_i, \quad \varepsilon_i \sim N(0, I_p), \end{aligned}$$

where $1(\cdot)$ is an indicator function; μ_y and Λ_y denote the mean and the loading matrix, respectively, for cause y ; and η_i is a latent factor, where $\eta_i \sim N(0, I_K)$. By integrating out the factor η_i , the dependence is induced with $\text{cov}(z_i) = \Lambda_y \Lambda_y' + I_p$. By borrowing the idea of covariance regression (Hoff and Niu, 2012; Niu and Hoff, 2019), we place additional factors in z_i , leading to age- and sex-dependent associations between the symptoms.

$$x_{ij} = 1(z_{ij} > 0), \quad j = 1, \dots, p, \tag{2}$$

$$z_i = B_{y_i} w_i + \Lambda_{y_i} \eta_i + C_{y_i} (w_i \otimes \gamma_i) + \varepsilon_i, \quad \varepsilon_i \sim N(0, I_p), \tag{3}$$

where $w_i = (1, \text{age}_i, \text{sex}_i)'$ and $C_y = [C_y^{(1)} \ C_y^{(2)} \ C_y^{(3)}]$, where $C_y^{(q)}$ is a $p \times G$ matrix for $q = 1, 2, 3$; γ_i and η_i are latent factors, where $\eta_i \sim N(0, I_K)$ and $\gamma_i \sim N(0, I_G)$, respectively.

The Kronecker product is expanded, and (3) is expressed as

$$z_i = B_{y_i} w_i + \tilde{\Lambda}_{y_i} \tilde{\eta}_i + \varepsilon_i, \quad \varepsilon_i \sim N(0, I_p), \quad (4)$$

where $\tilde{\eta}_i = (\eta'_i, \gamma'_i)'$ is a $K + G$ -dimensional latent factor, where $\tilde{\eta}_i \sim N(0, I_{K+G})$, and $\tilde{\Lambda}_y = [\Lambda_y \ D_y]$ is a factor loading, where Λ_y is constant against age and sex and D_y is an age- and sex-dependent matrix given by

$$D_y = C_y^{(1)} + C_y^{(2)} \text{age}_i + C_y^{(3)} \text{sex}_i.$$

For example, if $\text{age}_i \in \{0, 1\}$ and $\text{sex}_i \in \{0, 1\}$ correspond to $\{\text{young, old}\}$ and $\{\text{male, female}\}$, respectively, then

$$D_y = \begin{cases} C_y^{(1)} & (\text{young male}) \\ C_y^{(1)} + C_y^{(2)} & (\text{old male}) \\ C_y^{(1)} + C_y^{(3)} & (\text{young female}) \\ C_y^{(1)} + C_y^{(2)} + C_y^{(3)} & (\text{old female}) \end{cases}$$

By integrating $\tilde{\eta}_i$ out in (4), we obtain the following covariance matrix,

$$\text{cov}(z_i) = \tilde{\Lambda}_y \tilde{\Lambda}'_y + I_p = \Lambda_y \Lambda'_y + D_y D'_y + I_p,$$

where $\Lambda_y \Lambda'_y$ denotes rank- K associations shared between all age and sex groups and $D_y D'_y$ denotes rank- G associations changing by age and sex. To avoid complexity, we assume $K = G$ for the dimensions of the latent factors and select an appropriate value via cross-validation in our application. In general, the loading matrix need to be constrained to identify latent factors, but we use factor modeling to reduce the dimensions of the covariance matrix, in which case identification is unnecessary.

2.2 Modeling of $P(\text{age, sex} \mid \text{cause})$ and $P(\text{cause})$ in (1)

In terms of modeling of a cause of death, age and sex, it may be standard to develop a probability of a cause given age and sex, $P(\text{cause} \mid \text{age, sex})$, because a cause is the variable of main interest in the VA analysis. However, our objective is not in proposing an exact data generating process of a cause given only age and sex but in developing a predictive function of a cause given all predictors. Therefore, we consider modeling $P(\text{age, sex} \mid \text{cause})$, which is a part of $P(\text{cause} \mid \text{predictors})$ derived from the expression (1).

We analyze the dichotomized version of VA data in our application and apply the Dirichlet distribution to $P(\text{age, sex} \mid \text{cause})$ independently for each cause.

$$\{P(\text{age}_i = a, \text{sex}_i = b \mid y_i), a, b \in \{0, 1\}\} \sim \text{Dirichlet}(a_1, \dots, a_4),$$

where a_1, \dots, a_4 are concentration parameters. Likewise, regarding the prior distribution of causes $P(\text{cause})$, we assume the Dirichlet distribution

$$\{P(y_i = l), l \in \{1, \dots, L\}\} \sim \text{Dirichlet}(a_1, \dots, a_L).$$

In cases with little prior information, we assume a uniform prior with the concentration parameters that are equal to one. For example, we assume $P(y_i = l) \propto 1$ for $l = 1, \dots, L$ if little prior information is available about the distribution of causes in a new VA study site.

2.3 Measurement of predictor relevance

VA surveys consist of many questionnaire items, which are conceptually useful for predicting a cause of death. However, the degree of association with causes of death can vary much from question to question, so it is of interest to investigate the relevance of each predictor to causes of death in VA survey data.

Using the proposed models in (1), we compute the strength of the associations between a cause of death and each predictor based on information-theoretic measures (MacKay, 2003; Cover and Thomas, 2006). Let $H(y)$ denote the entropy of y , which quantifies the expected amount of information about y ; $H(y) = -\sum_y P(y) \log P(y)$. Mutual information (MI) is a measure of the mutual dependence between two random variables. Letting $\tilde{x} = (\text{age, sex}, x)$,

MI is defined by

$$I(y; \tilde{x}_j) = \sum_y \sum_{\tilde{x}_j} P(y, \tilde{x}_j) \log \frac{P(y, \tilde{x}_j)}{P(y)P(\tilde{x}_j)}. \quad (5)$$

This denotes the amount of information obtained about a cause y by observing the j th predictor \tilde{x}_j . Theoretically, (5) takes a nonnegative value and equals zero if and only if the two random variables are independent. To improve our interpretation, we normalize the MI to take a value in the $[0, 1]$ interval, we normalize MI by dividing it by the entropy of y ; that is $I(y; \tilde{x}_j)/H(y)$, which indicates the amount of information about causes added by the j th predictor relative to all the information about causes.

VA questionnaires often have multiple questions on each topic under symptoms. For example, the PHMRC survey on neonatal death asks six questions each about suckling and breathing. If these questions provide highly overlapped information about causes, then some of them may show high MI values despite adding little new information about causes given the other questions. Considering this point, we compute the conditional MI (CMI) along with the MI. The CMI indicates the amount of information each predictor can add given the other items.

$$I(y; \tilde{x}_j | \tilde{x}_{-j}) = \sum_y \sum_{\tilde{x}} P(y, \tilde{x}) \log \frac{P(y, \tilde{x}_j | \tilde{x}_{-j})}{P(y | \tilde{x}_{-j})P(\tilde{x}_j | \tilde{x}_{-j})}, \quad (6)$$

where \tilde{x}_{-j} indicates all the predictors except the j th one. We normalize the CMI to $I(y; \tilde{x}_j | \tilde{x}_{-j})/H(y)$, which assumes a value in the $[0, 1]$ interval and indicates the amount of additional information about causes we can obtain by knowing x_j given all the other predictors relative to the total amount of information related to causes. The CMI can capture the strength of the conditional association between the predictors and causes of death, but relying only on CMI may result in a misleading conclusion. For example, if two predictors are strongly associated with causes but share most of their information, then both of them will show small CMI values despite being useful for predicting the causes of death. Therefore, the relevance of predictors should be evaluated by investigating both the MI and CMI values. The computation of the MI and CMI using the proposed models is detailed in Section 3.

3 Posterior computation

We develop an Markov-chain Monte Carlo (MCMC) algorithm for the proposed method in Section 2. Let $m_i = (m_{i1}, \dots, m_{ip})'$ be a vector of indicators denoting missing values for the i th person such that $m_{ij} = 1$ if x_{ij} is missing and $m_{ij} = 0$ if x_{ij} is observed, where $j = 1, \dots, p$. We define the notation $[m_i]$ such that $a_{[m_i]}$ and $A_{[m_i]}$ (where a and A are a vector and a matrix with p rows, respectively) denote a subvector and a submatrix consisting of components with $m_{ij} = 0$ for $j = 1, \dots, p$. In addition, A_j shows the j th row of the matrix A .

For the prior distributions on B_y , Λ_y and C_y , we use Cauchy distributions. It has high density around zero and heavy tails, which reduce the effects of redundant elements while capturing meaningful signals. Based on the expression of the Cauchy distribution via normal-gamma distributions, we assume $B_{yjq} \sim N(0, \phi_{Bjq}^{-1})$, $\phi_{Bjq} \sim Ga(0.5, 0.5)$, $\Lambda_{yjk} \sim N(0, \phi_{\Lambda_j}^{-1})$, $\phi_{\Lambda_j} \sim Ga(0.5, 0.5)$, $C_{yjk}^{(q)} \sim N(0, \phi_{Cjq}^{-1})$, $\phi_{Cjq} \sim Ga(0.5, 0.5)$ where $q = 1, 2, 3$, $j = 1, \dots, p$, $k = 1, \dots, K$, and $Ga(a, b)$ denotes a gamma distribution with a mean a/b . The latent variables ϕ_{Bjq} , ϕ_{Λ_j} and ϕ_{Cjq} are shared between causes and factors to reduce the number of model parameters.

The proposed MCMC algorithm is below.

1. Update $\beta_{yj} \equiv (B_{yj}, \Lambda_{yj}, C_{yj}^{(0)}, C_{yj}^{(1)}, C_{yj}^{(2)})'$ from $N(\mu_*, \Sigma_*)$ for $y = 1, \dots, L$ and $j = 1, \dots, p$ with

$$\mu_* = \Sigma_* \left(\sum_{i:y_i=y, m_{ij}=0} a_i z_{ij} \right), \quad \Sigma_* = \left(\sum_{i:y_i=y, m_{ij}=0} a_i a_i' + \Sigma_0^{-1} \right)^{-1}$$

where $a_i = (w_i', \eta_i', \gamma_i', \text{age}_i \gamma_i', \text{sex}_i \gamma_i')'$ and

$\Sigma_0^{-1} = \text{diag}(\phi_{Bj1}, \phi_{Bj2}, \phi_{Bj3}, \phi_{\Lambda_j} \mathbf{1}_K, \phi_{Cj1} \mathbf{1}_K, \phi_{Cj2} \mathbf{1}_K, \phi_{Cj3} \mathbf{1}_K)$, where $\mathbf{1}_K$ denote a K -dimensional vector with all elements one.

2. Update $\tilde{\eta}_i = (\eta_i', \gamma_i')'$ from $N(\tilde{\mu}, \tilde{\Sigma})$ for $i = 1, \dots, n$ with

$$\tilde{\mu} = \tilde{\Sigma} \tilde{\Lambda}'_{y_i[m_i]} (z_i - B_{y_i} w_i)_{[m_i]}, \quad \tilde{\Sigma} = \left(\tilde{\Lambda}'_{y_i[m_i]} \tilde{\Lambda}_{y_i[m_i]} + I_{2K} \right)^{-1}.$$

3. Update ϕ_{Bjq} from $Ga(0.5(L+1), 0.5(\sum_{y=1}^L B_{yjq}^2 + 1))$ for $q = 1, 2, 3$ and $j = 1, \dots, p$.
4. Update ϕ_{Λ_j} from $Ga(0.5(LK+1), 0.5 \sum_{y=1}^L \sum_{k=1}^K \Lambda_{yjk}^2 + 1))$ for $j = 1, \dots, p$.
5. Update ϕ_{Cjq} from $Ga(0.5(LK+1), 0.5 \sum_{y=1}^L \sum_{k=1}^K C_{yjk}^{(q)2} + 1))$ for $q = 1, 2, 3$ and $j = 1, \dots, p$.
6. Update z_{ij} with $m_{ij} = 0$ for $i = 1, \dots, n$ and $j = 1, \dots, p$ from

$$\begin{cases} N_+(B_{yij} \cdot w_i + \tilde{\Lambda}_{yij} \cdot \tilde{\eta}_i, 1) & \text{if } x_{ij} = 1, \\ N_-(B_{yij} \cdot w_i + \tilde{\Lambda}_{yij} \cdot \tilde{\eta}_i, 1) & \text{if } x_{ij} = 0, \end{cases}$$

where N_+ and N_- denote truncated normals with support $[0, \infty)$ and $(-\infty, 0]$.

7. Update $\{P(\text{age}_i = a, \text{sex}_i = b | y_i = y), a, b \in \{0, 1\}\}$ from Dirichlet($b_{00}, b_{01}, b_{10}, b_{11}$) where $b_{ab} = 1 + \sum_{i=1}^n 1(y_i = y, \text{age}_i = a, \text{sex}_i = b)$ for $y = 1, \dots, L$.
8. Impute age_i and sex_i if missing from

$$\begin{aligned} P(\text{age}_i = a | y_i, x_i, \text{sex}_i, \tilde{\eta}_i) &= \frac{P(x_i | y_i, \text{age}_i = a, \text{sex}_i, \tilde{\eta}_i) P(\text{age}_i = a, \text{sex}_i | y_i)}{\sum_{l=0}^1 P(x_i | y_i, \text{age}_i = l, \text{sex}_i, \tilde{\eta}_i) P(\text{age}_i = l, \text{sex}_i | y_i)}, \quad a = 0, 1, \\ P(\text{sex}_i = b | y_i, x_i, \text{age}_i, \tilde{\eta}_i) &= \frac{P(x_i | y_i, \text{age}_i, \text{sex}_i = b, \tilde{\eta}_i) P(\text{age}_i, \text{sex}_i = b | y_i)}{\sum_{l=0}^1 P(x_i | y_i, \text{age}_i, \text{sex}_i = l, \tilde{\eta}_i) P(\text{age}_i, \text{sex}_i = l | y_i)}, \quad b = 0, 1. \end{aligned}$$

where $P(x_i | y_i, \text{age}_i, \text{sex}_i, \tilde{\eta}_i)$ corresponds to the probability from (2) and (4).

9. For $i \in S$ where S is target data, predict y_i with

$$P(y_i = y | x_i, \text{age}_i, \text{sex}_i) = \frac{P(x_i | y_i = y, \text{age}_i, \text{sex}_i) P(\text{age}_i, \text{sex}_i | y_i = y) P(y_i = y)}{\sum_{l=1}^L P(x_i | y_i = l, \text{age}_i, \text{sex}_i) P(\text{age}_i, \text{sex}_i | y_i = l) P(y_i = l)},$$

with $y = 1, \dots, L$ where $P(x_i | y_i, \text{age}_i, \text{sex}_i) = \int P(x_i | y_i, \text{age}_i, \text{sex}_i, \tilde{\eta}) f(\tilde{\eta}) d\tilde{\eta}$ is evaluated via a Monte Carlo approximation with $\tilde{\eta}_{(r)} \sim N(0, I_{2K})$ for $r = 1, \dots, R$,

$$P(x_i | y_i, \text{age}_i, \text{sex}_i) \approx \frac{1}{R} \sum_{r=1}^R \left\{ \prod_{j: m_{ij}=0} P(x_{ij} | \tilde{\eta}_{(r)}, y_i, \text{age}_i, \text{sex}_i) \right\}. \quad (7)$$

Then, compute the population distribution of the causes of death as

$$\left(\frac{1}{n_S} \sum_{i \in S} 1(y_i = 1), \dots, \frac{1}{n_S} \sum_{i \in S} 1(y_i = L) \right)$$

where n_S is the number of observations in the test data. For the computation of the measures of importance of the predictors, Step 9 is replaced as follows:

9. Update $\{P(y_i = y), y \in \{1, \dots, L\}\}$ from

$$\text{Dirichlet} \left(1 + \frac{1}{n} \sum_{i=1}^n 1(y_i = 1), \dots, 1 + \frac{1}{n} \sum_{i=1}^n 1(y_i = L) \right).$$

10. Compute the MI and CMI for each predictor in (5) and (6). Because the summation in (6) is intractable for a nonsmall p , it is approximated using the Monte Carlo sample $(y_{(r)}, \tilde{x}_{(r)}) \sim P(y, \tilde{x}) = P(x | y, \text{age}, \text{sex})P(\text{age}, \text{sex} | y)P(y)$, where $r = 1, \dots, \tilde{R}$.

$$I(y; \tilde{x}_j | \tilde{x}_{-j}) \approx \frac{1}{\tilde{R}} \sum_{r=1}^{\tilde{R}} \log \frac{P(y_{(r)}, \tilde{x}_{(r)j} | \tilde{x}_{(r)-j})}{P(y_{(r)} | \tilde{x}_{(r)-j})P(\tilde{x}_{(r)j} | \tilde{x}_{(r)-j})}. \quad (8)$$

The programming codes for the proposed method are distributed on GitHub (<https://github.com/kunihama/BF-AS>).

4 Results

In this section, we evaluate the performance of the proposed approach using the PHMRC neonatal VA data. Our primary interest is in estimating the distribution of deaths by cause in a target population. In VA studies, it is often extremely challenging to obtain gold-standard training data from a target site where the VA survey questions are tied to the medically certified cause of death of each individual. Therefore, the cause-of-death distribution in a target site is predicted using training data collected from different geographic locations, so the obtained fractions of death by cause vary widely between testing and training data. Considering this point, we consider realistic scenarios where each PHMRC study site is treated as a target site and all other sites are jointly regarded as training data. Thus,

we have six cases where the targets are Andhra Pradesh, India (case 1), Bohol, Philippines (case 2), Dar es Salaam, Tanzania (case 3), Mexico City, Mexico (case 4), Pemba Island, Tanzania (case 5) and Uttar Pradesh, India (case 6).

We assess VA classification methods based on the accuracy of their cause-specific mortality fractions (CSMFs) (Murray et al., 2011).

$$\text{CSMF accuracy} = 1 - \frac{\sum_{y=1}^L |P_0(y_i = y) - P(y_i = y)|}{2\{1 - \min_{1 \leq y \leq L} P_0(y_i = y)\}}$$

where P_0 and P are the true and estimated distributions, respectively, of the causes of death in a target site. The CSMF measures the closeness between the two distributions, taking a value in the $[0,1]$ interval, and a larger value means better performance. In our study, the true distribution P_0 , is approximated using the empirical distribution of causes in the test data, whereas P corresponds to the distribution of causes estimated using a classification tool.

The proposed method (BF models with age and sex dependence [BF-AS]) is compared with recent VA classification tools. Popular methods in the VA literature are available in the R package, such as openVA (Li et al., 2023) with InterVA (Byass et al., 2012), the tariff method (James et al., 2011; Serina et al., 2015), the naive Bayes classifier (NBC) (Miasnikof et al., 2015) and InSilicoVA (McCormick et al., 2016). These approaches are based on the conditional independence of symptoms given a cause, whereas the BF models of Kuniyama et al. (2020) and the Bayesian hierarchical factor regression (FARVA) model of Moran et al. (2021) describe the dependence between symptoms for each cause. FARVA allows the associations between symptoms to vary by covariate, which we set as age and sex. The settings of the competitor methods are detailed in the supplementary materials. For the proposed method, we standardized age_i and sex_i in w_i in (3) with the mean zero and the standard deviation one. And we selected the number of latent factors K from the set $\{1, 2, 3, 4, 5\}$ via fivefold cross-validation and set $R = 1,000$ in (7) and $\tilde{R} = 10,000$ in (8) as the number of random samples for the Monte Carlo approximation. We generated 10,000 MCMC samples after the initial 1,000 burn-in iterations, and every 20th sample was saved. We observed that the sample paths were stable, and the sample autocorrelations dropped smoothly. The trace plot and the autocorrelation plot of the CSMF in the target

Method	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
BF-AS	0.757	0.932	0.769	0.893	0.845	0.822
BF	0.741	0.942	0.742	0.788	0.792	0.801
FARVA	0.734	0.891	0.768	0.806	0.740	0.746
InsilicoVA	0.671	0.810	0.706	0.707	0.628	0.884
Tariff	0.777	0.880	0.811	0.566	0.800	0.740
InterVA	0.471	0.683	0.463	0.612	0.659	0.539
NBC	0.759	0.840	0.716	0.837	0.685	0.877

Table 1: CSMF accuracy. The bold numbers indicate the top three results in each case. BF-AS: BF models with age and sex dependence (proposed method), BF: Bayesian factor models, FARVA: Bayesian hierarchical factor regression model, NBC: naive Bayes classifier. The posterior means are used as the estimate of the CSMF for BF-AS, BF, FARVA and InsilicoVA.

site are in the supplementary materials.

First, as an illustrative example, Figure 5 shows the CSMF estimation results for case 2. All methods except InterVA work reasonably well, capturing the true mortality fractions closely; InterVA underestimates birth asphyxia while overestimating pneumonia. The figures for the other cases are in the supplementary materials. Table 1 reports the CSMF accuracy of all methods, with the bold numbers indicating the top three results in each case. No single method dominates its competitors all the time; the approach with the highest accuracy differs from case to case. However, as the bold numbers show, the proposed method consistently performs well in all cases. In addition, Figure 6 shows the average CSMF accuracy of each method over the six cases. InterVA lags by a large margin, and although the BF models produce the highest accuracy among the competing methods, BF-AS outperforms them.

Regarding uncertainty, the coverage rate of the true fractions by the proposed method in a 95% interval is 0.6. Although it is larger than those of the BF models (0.43), FARVA (0.4), and InSilicoVA (0.3) (Tariff, InterVA, and NBC produce only point estimates), certain gaps need to be addressed in future work, as discussed in Section 5.

About the relevance of the predictors, Figure 7 shows the estimated standardized MI for the top 15 predictors. The first two predictors, weight and size at the time of delivery, have more than 10% of the information about the causes of death. Early/late delivery, physical abnormality at birth and several items related to crying and suckling also have high MI values. As for the demographic variables, age is on the list, but sex is out.

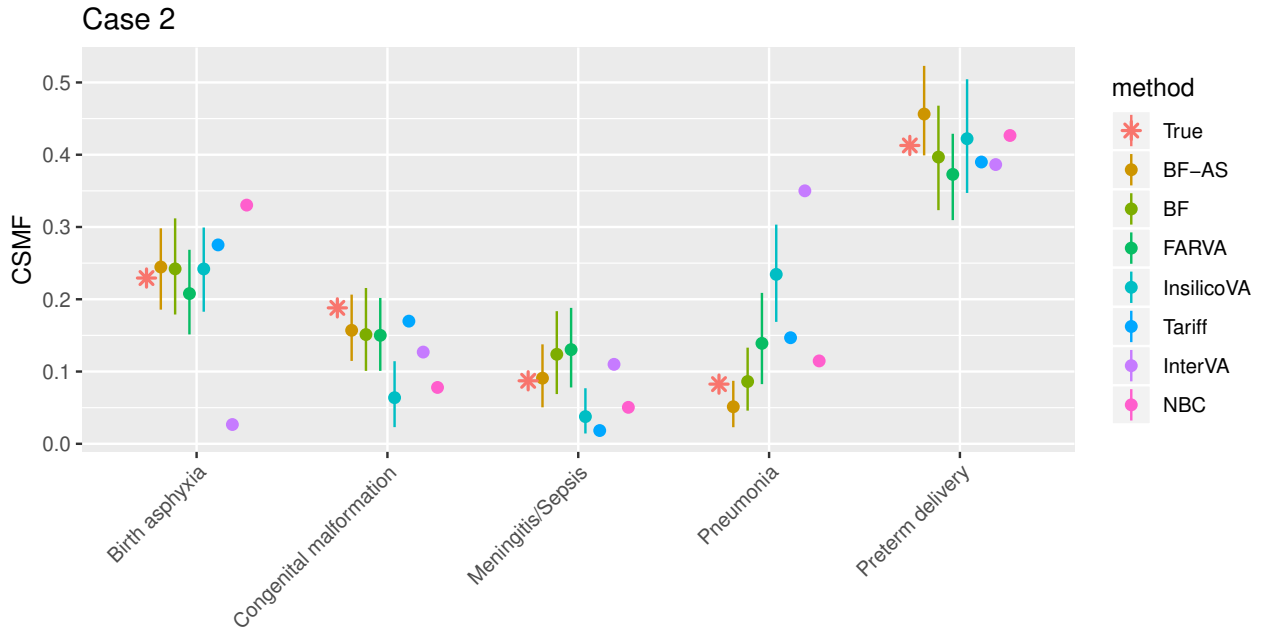


Figure 5: CSMF estimation results for case 2. The red asterisks show the true values, and each circle with an interval corresponds to the mean and the 95% interval, respectively, of each statistical method. BF-AS: BF models with age and sex dependence (proposed method), BF: Bayesian factor models, FARVA: Bayesian hierarchical factor regression model, NBC: naive Bayes classifier. Tariff, InterVA, and NBC produce point estimates only.

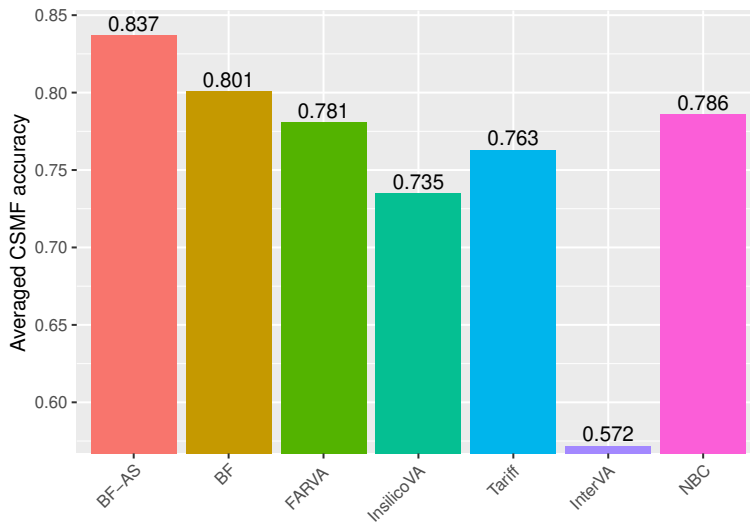


Figure 6: Average CSMF accuracy of each method in six cases.

As discussed in Section 2.3, the MI of a predictor does not reflect the effects of other predictors. Figure 8 reports the estimated CMI for the top 15 predictors. Many of the predictors with large MI values appear again on the CMI list of CMI, such as weight, size, early/late delivery, physical abnormality, birth order and symptoms related to crying and suckling. The estimated values of CMI values for these predictors are smaller than the estimated MI values, indicating that a certain amount of their information about the causes of death is shared with the other predictors. In addition, we observe a remarkable difference –the position of sex– between Figures 7 and 8. Sex shows a small posterior mean of the MI (0.15%), ranking 96th out of the 97 predictors (not shown in this figure), but leads the CMI list by a wide margin. It is also the only predictor whose CMI value is larger than its MI value. Thus, although one does not obtain so much information about causes only by knowing the sex of a baby, it can indirectly deliver unique information about the cause of death through the other predictors.

Finally, we investigate the variation in the association between a cause and each symptom across different age and sex groups. The mutual information, conditional on age and sex, is expressed as the expected value of the Kullback–Leibler (KL) divergence between the joint distribution and the product of marginal distributions,

$$\begin{aligned}
& I(\text{cause}; \text{symptom} \mid \text{age}, \text{sex}) \\
&= E_{\text{age}, \text{sex}} [D_{\text{KL}}\{P(\text{cause}, \text{symptom} \mid \text{age}, \text{sex}) \parallel P(\text{cause} \mid \text{age}, \text{sex})P(\text{symptom} \mid \text{age}, \text{sex})\}], \\
&= \sum_{\text{age}} \sum_{\text{sex}} P(\text{cause}, \text{symptom}, \text{age}, \text{sex}) \log \frac{P(\text{cause}, \text{symptom} \mid \text{age}, \text{sex})}{P(\text{cause} \mid \text{age}, \text{sex})P(\text{symptom} \mid \text{age}, \text{sex})}.
\end{aligned}$$

To examine the differences between age and sex groups, Figure 9 presents a comparison of the KL divergence for symptoms that are highly relevant to causes of death, revealing diverse patterns. For instance, the KL divergence indicates larger values for older age groups (pregnant end early/late, stop suckling and first cry within 5 minutes) and for females (small or very small). Conversely, there is less variation in the KL divergence by age and sex concerning the physically abnormal symptom. Additionally, females in the early age stage are associated with relatively smaller KL divergence values (birth order and first cry more than 30 minutes). These findings suggest that the degree of association between a cause of

death and each symptom varies significantly by age and sex.

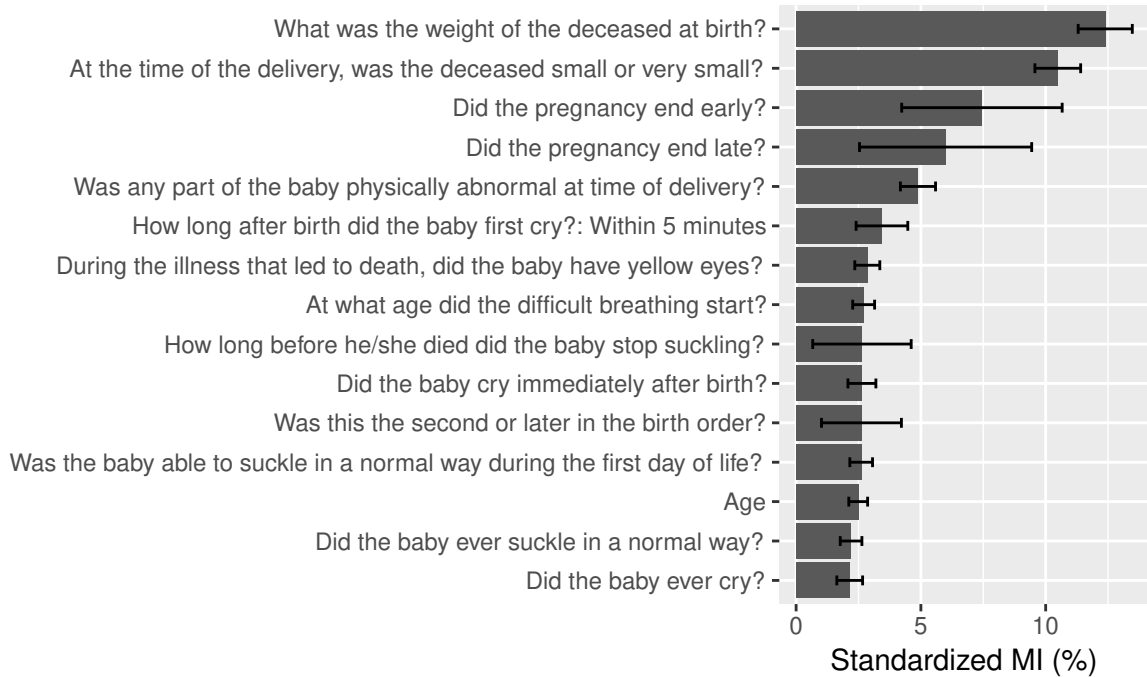


Figure 7: Posterior mean of standardized MI for top 15 predictors. The error bars correspond to standard deviations.

5 Conclusion

In this article, we develop an approach to estimating cause-of-death distributions using VA survey data. Considering the differences between basic demographic information (age and sex) and the other types of predictors, we propose BF models with age- and sex-dependent associations between symptoms for each cause of death. In addition, the relevance of predictors to causes of death is computed based on information-theoretic measures. In our analysis of PHMRC neonate data, BF-AS performs well in estimating the cause-of-death distributions in the target sites compared with existing methods in the VA literature. However, as the proposed approach has low coverage rates of the true fractions of causes in the 95% interval, it needs further development for a better understanding of the relations between causes of death and VA survey items.

One research direction is to allow cause-predictor relations to differ flexibly by location. In the proposed method, the distribution of predictors given a cause is independent of the

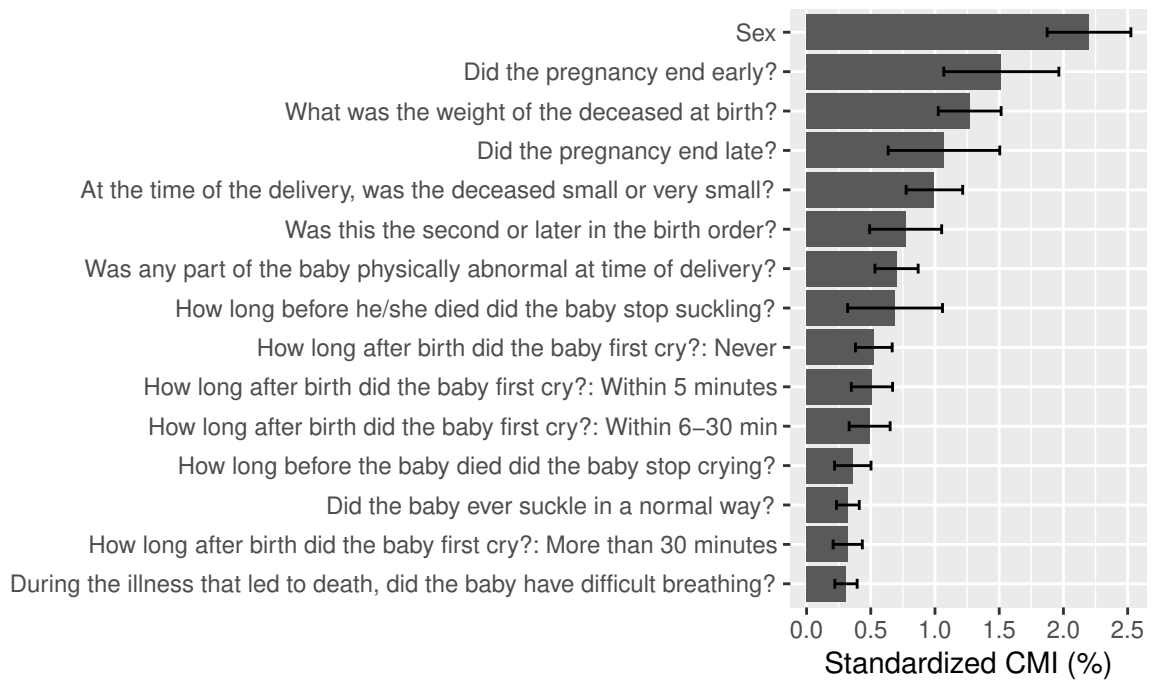


Figure 8: Posterior mean of standardized CMI for top 15 predictors. The error bars correspond to standard deviations.

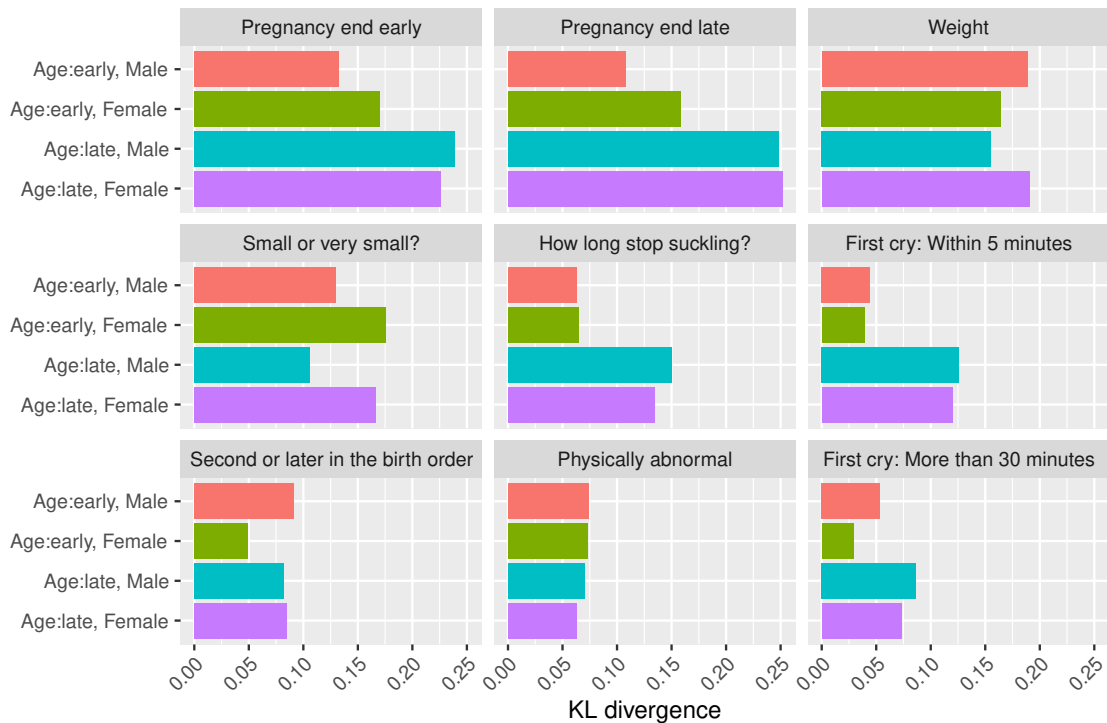


Figure 9: Posterior means of the KL divergence for cause of death and symptom associations across age and sex groups.

VA survey site, but a recent work stated that this conditional probability can be affected by the domain (Wu et al., 2021; Li et al., 2021). One approach may be to treat the domain information similarly to age and sex in the proposed model by adding site-related dummy variables as covariates. However, this is challenging because target sites often do not have training data, so domain effects on symptom distributions can not be estimated directly. Domains may have similar patterns of the relations between causes and VA questions, so borrowing information from other sites for incorporation into the factor models is key.

Another study direction is to investigate whether the estimation of cause-of-death distributions will be improved by allowing symptom associations to vary by other demographic variables along with age and sex. For example, educational backgrounds, income levels and marital status may affect the symptom distribution for each cause. Adding information will increase the flexibility of factor models, but sparsity may have to be incorporated into the models to avoid overfitting, such as by using variable selection methods for covariance regression (Niu and Hoff, 2019).

Acknowledgment

This work was supported by JSPS KAKENHI Grant Numbers JP21K13274, JP20KK0292, JP19H00588. The results are generated mainly using Ox (Doornik, 2007). The programming codes for the proposed method and the Bayesian factor model by Kuniyama et al. (2020) are distributed on the GitHub website (<https://github.com/kuniyama/BF-AS>).

References

- Byass, P., D. Chandramohan, S. J. Clark, L. D’Ambruoso, E. Fottrell, W. J. Graham, A. J. Herbst, A. Hodgson, S. Hounton, K. Kahn, et al. (2012). Strengthening standardised interpretation of verbal autopsy data: the new InterVA-4 tool. *Global Health Action* 5, 19281.
- Byass, P., D. L. Huong, and H. V. Minh (2003). A probabilistic approach to interpreting verbal autopsies: methodology and preliminary validation in Vietnam. *Scandinavian Journal of Public Health. Supplement* 62, 32–37.

- Byass, P., L. Hussain-Alkhateeb, L. D’Ambruoso, S. Clark, J. Davies, E. Fottrell, J. Bird, C. Kabudula, S. Tollman, K. Kahn, et al. (2019). An integrated approach to processing WHO-2016 verbal autopsy data: the interVA-5 model. *BMC Medicine* 17, 102.
- Chandramohan, D., E. Fottrell, J. Leitao, E. Nichols, S. J. Clark, C. Alsokhn, D. C. Munoz, C. AbouZahr, A. D. Pasquale, R. Mswia, et al. (2021). Estimating causes of death where there is no medical certification: evolution and state of the art of verbal autopsy. *Global Health Action* 14, 1982486.
- Cover, T. M. and J. A. Thomas (2006). *Elements of information theory*. Wiley-Interscience.
- Cramér, H. (1946). *Mathematical Methods of Statistics*. Princeton landmarks in mathematics and physics. Princeton University Press.
- Doornik, J. A. (2007). *Object-Oriented Matrix Programming Using Ox, 3rd ed.* London: Timberlake Consultants Press and Oxford.
- Ferlay, J., M. Ervik, F. Lam, M. Colombet, L. Mery, M. Pineros, A. Znaor, I. Soerjomataram, and F. Bray (2020). Global cancer observatory: Cancer today. lyon, france: International agency for research on cancer. <https://gco.iarc.fr/today>.
- Hoff, P. D. and X. Niu (2012). A covariance regression model. *Statistica Sinica* 22, 729–753.
- James, S. L., A. D. Flaxman, and C. J. Murray (2011). Performance of the tariff method: validation of a simple additive algorithm for analysis of verbal autopsies. *Population Health Metrics* 9, 31.
- Kunihama, T., Z. R. Li, S. J. Clark, and T. H. McCormick (2020). Bayesian factor models for probabilistic cause of death assessment with verbal autopsies. *The Annals of Applied Statistics* 14, 241–256.
- Li, Z. R., T. H. McComick, and S. J. Clark (2020). Using Bayesian latent Gaussian graphical models to infer symptom associations in verbal autopsies. *Bayesian Analysis* 15, 781–807.
- Li, Z. R., J. Thomas, E. Choi, T. H. McCormick, and S. J. Clark (2023). The r journal: The openva toolkit for verbal autopsies. *The R Journal*, 1. <https://journal.r-project.org/articles/RJ-2023-020/>.

- Li, Z. R., Z. Wu, I. Chen, and S. Clark (2021). Bayesian nested latent class models for cause-of-death assignment using verbal autopsies across multiple domains. Working paper. <https://arxiv.org/abs/2112.12186>.
- MacKay, D. J. C. (2003). *Information theory, inference, and learning algorithms*. Copyright Cambridge University Press.
- Maher, D., S. Biraro, V. Hosegood, R. Isingo, T. Lutalo, P. Mushati, B. Ngwira, M. Nyirenda, J. Todd, and B. Zaba (2010). Translating global health research aims into action: the example of the alpha network. *Tropical Medicine & International Health* 15, 321–328.
- McCormick, T. H., Z. R. Li, C. Calvert, A. C. Crampin, K. Kahn, and S. J. Clark (2016). Probabilistic cause-of-death assignment using verbal autopsies. *Journal of the American Statistical Association* 111, 1036–1049.
- Miasnikof, P., V. Giannakeas, M. Gomes, L. Aleksandrowicz, A. Y. Shestopaloff, D. Alam, S. Tollman, A. Samarikhalaj, and P. Jha (2015). Naive Bayes classifiers for verbal autopsies: comparison to physician-based classification for 21,000 child and adult deaths. *BMC Medicine* 13, 286.
- Moran, K. R., E. L. Turner, D. Dunson, and A. H. Herring (2021). Bayesian hierarchical factor regression models to infer cause of death from verbal autopsy data. *Journal of the Royal Statistical Society Series C: Applied Statistics* 70, 532–557.
- Murray, C. J., A. D. Lopez, R. Black, R. Ahuja, S. M. Ali, A. Baqui, L. Dandona, E. Dantzer, V. Das, U. Dhingra, et al. (2011). Population Health Metrics Research Consortium gold standard verbal autopsy validation study: design, implementation, and development of analysis datasets. *Population Health Metrics* 9, 27.
- Murray, C. J., R. Lozano, A. D. Flaxman, A. Vahdatpour, and A. D. Lopez (2011). Robust metrics for assessing the performance of different verbal autopsy cause assignment methods in validation studies. *Population Health Metrics* 9, 28.
- Niu, X. and P. Hoff (2019). Joint mean and covariance modeling of multiple health outcome measures. *Annals of Applied Statistics* 13, 321–339.

- Nkengasong, J., E. Gudo, I. Macicame, X. Maunze, A. Amouzou, K. Banke, S. Dowell, and I. Jani (2020). Improving birth and death data for African decision making. *The Lancet Global Health* 8, e35–e36.
- PHMRC (2013). Population health metrics research consortium gold standard verbal autopsy data 2005–2011. <https://ghdx.healthdata.org/record/ihme-data/population-health-metrics-research-consortium-gold-standard-verbal-autopsy-data-2005-2011>.
- Sankoh, O. and P. Byass (2012). The INDEPTH network: filling vital gaps in global epidemiology. *International Journal of Epidemiology* 41, 579–588.
- Serina, P., I. Riley, A. Stewart, S. L. James, A. D. Flaxman, R. Lozano, B. Hernandez, M. D. Mooney, R. Luning, R. Black, et al. (2015). Improving performance of the tariff method for assigning causes of death to verbal autopsies. *BMC Medicine* 13, 291.
- World Health Organization (2020). Global health estimates 2020: Deaths by cause, age, sex, by country and by region, 2000-2019. geneva. <https://www.who.int/data/global-health-estimates>.
- World Health Organization (2022). World health statistics 2022: monitoring health for the sdgs, sustainable development goals. <https://www.who.int/publications/i/item/9789240051157>.
- Wu, Z., Z. R. Li, I. Chen, and M. Li (2021). Tree-informed Bayesian multi-source domain adaptation: cross-population probabilistic cause-of-death assignment using verbal autopsy. Working paper. <https://arxiv.org/abs/2112.10978>.

Supplementary materials for “Bayesian analysis of verbal autopsy data using factor models with age- and sex-dependent associations between symptoms”

1 Lists of predictors in our analysis

Tables 1-3 show the lists of 97 predictors in our analysis ($n = 1,619$). We selected these questions from sections 1, 2 and 3 in the VA questionnaire for neonate and child in PHMRC (2013), removing overlapped items and very rare symptoms only less than 10 babies exhibited. Following the approach by Murray et al. (2011), the non-binary predictors are transformed into dichotomous variables.

2 Setting of the competitors

We estimate InterVA, Tariff, the naive Bayes classifier and InSilicoVA using the R package openVA (Li et al., 2023) with the default setting discussed in McCormick et al. (2016). In addition, we estimate the Bayesian hierarchical factor regression (FARVA) with age- and sex-dependent associations between symptoms using the farva package (<https://github.com/kelrenmor/farva>). Following the settings in the example file in the package, the number of factors was set as 10 and we ran the MCMC for 10,000 iterations and saved every 10th iteration after discarding the first half of the chain.

3 MCMC convergence

To investigate posterior convergence in our applications in Section 4, the sample paths and the autocorrelations of the MCMC sample of the CSMF in the target site by the proposed model are shown in Figures 1 (case 1 and case 2), 2 (case 3 and case 4) and 3 (case 5 and case 6).

4 CSMF estimation results

Figures 4, 5 and 6 show the CSMF estimation results for cases 1 and 2 (Figure 4), 3 and 4 (Figure 5) and 5 and 6 (Figure 6), respectively.

Table 1: List of predictors: 1-40

No.	Predictors
1	Age (How old was the deceased at the time of death?)
2	Sex
3	Was the deceased a singleton or multiple birth?
4	Was this the first, second, or later in the birth order?
5	Is the mother still alive?
6	Was the deceased not born in a health facility?
7	At the time of the delivery, was the deceased small or very small?
8	What was the weight of the deceased at birth?
9	Was the child born alive or dead?
10	Did the baby ever cry?
11	Did the baby ever move?
12	Did the baby ever breathe?
13	For questions 10, 11 and 12, all three responses are No?
14	How old was the baby/child when the fatal illness started?
15	How long did the illness last?
16	Did the deceased die at home or on route to a health facility?
17	Was the late part of the pregnancy, labor or delivery complicated by convulsions?
18	... complicated by high blood pressure?
19	... complicated by severe anemia?
20	... complicated by diabetes?
21	... complicated by child delivered not head first?
22	... complicated by cord delivered first?
23	... complicated by cord around child's neck?
24	... complicated by excessive bleeding?
25	... complicated by fever during labor?
26	Did the pregnancy end early?
27	Did the pregnancy end late?
28	Was the baby moving in the last few days before the birth?
29	When did the mother last feel the baby move?
30	Did the water break before labor or during labor?
31	How much time before labor did the water break?
32	Was the color of the liquid not clear when it broke?
33	Was the liquor foul smelling?
34	How much time did the labor and delivery take?
35	Did the mother receive any vaccinations since reaching adulthood including during this pregnancy?
36	Did the mother receive fewer than 3 doses of vaccine since reaching adulthood?
37	Did the delivery not occur in a health facility?
38	Was the delivery carried out by someone other than a health professional?
39	Was the delivery...? Vaginal with Forceps
40	Was the delivery...? Vaginal without Forceps

Table 2: List of predictors: 41-80

No.	Predictors
41	Was the delivery...? Vaginal Don't Know
42	Was the delivery...? C-Section
43	During labor but before delivery, did the mother receive any kind of injection?
44	Were there any bruises or signs of injury on the baby's body at birth?
45	Was any part of the baby physically abnormal at time of delivery?
46	What were the abnormalities? Head size very small at time of birth
47	What were the abnormalities? Head size very large at time of birth?
48	What were the abnormalities? Mass defect on the back of head or spine
49	What were the abnormalities? Other
50	Did the baby breathe immediately after birth?
51	Did the baby have difficulty breathing?
52	Was anything done to try to help the baby breathe at birth?
53	Did the baby cry immediately after birth?
54	How long after birth did the baby first cry? Within 5 minutes
55	How long after birth did the baby first cry? Within 6-30 minutes
56	How long after birth did the baby first cry? More than 30 minutes
57	How long after birth did the baby first cry? Never
58	Did the baby stop being able to cry?
59	How long before the baby died did the baby stop crying? One day or more?
60	Was the baby able to suckle in a normal way during the first day of life?
61	Did the baby ever suckle in a normal way?
62	Did the baby stop being able to suckle in a normal way?
63	How long after birth did the baby stop suckling?
64	How long before he/she died did the baby stop suckling? One day or more?
65	Was the baby able to open his/her mouth at the time he/she stopped sucking?
66	During the illness that led to death, did the baby have difficult breathing?
67	At what age did the difficult breathing start?
68	For how many days did the difficult breathing last?
69	During the illness that led to death, did the baby have fast breathing?
70	At what age did the fast breathing start?
71	For how many days did the fast breathing last?
72	During the illness that led to death, did the baby have indrawing of the chest?
73	During the illness that led to death, did the baby have grunting?
74	During the illness that led to death did the baby have spasms or convulsions?
75	During the illness that led to death, did the baby have fever?
76	At what age did the fever start?
77	How many days did the fever last?
78	During the illness that led to death, did the baby become cold to touch?
79	At what age did the baby start feeling cold to touch?
80	How many days did the baby feel cold to touch?

Table 3: List of predictors: 81-97

No.	Predictors
81	During the illness that led to death, did the baby become lethargic after a period of normal activity?
82	During the illness that led to death, did the baby become unresponsive or unconscious?
83	During the illness that led to death, did the baby have a bulging fontanelle?
84	During the illness that led to death, did the baby have pus drainage from the umbilical cord stump?
85	During the illness that led to death, did the baby have redness of the umbilical cord stump?
86	Did the redness of the umbilical cord stump extend onto the abdominal skin?
87	During the illness that led to death, did the baby have skin bumps containing pus or a single large area with pus?
88	During the illness that led to death, did the baby have ulcer(s) (pits)?
89	During the illness that led to death, did the baby have an area(s) of skin with redness and swelling?
90	During the illness that led to death, did he/she have areas of the skin that turned black?
91	During the illness that led to death, did the baby bleed from anywhere?
92	During the illness that led to death, did he/she have more frequent loose or liquid stools than usual?
93	How many stools did the baby have on the day that diarrhea/loose liquid stools were most frequent?
94	During the illness that led to death, did he/she vomit everything?
95	During the illness that led to death, did he/she have yellow skin?
96	During the illness that led to death, did the baby have yellow eyes?
97	Did the infant appear to be healthy and then just die suddenly?

References

- Byass, P., D. Chandramohan, S. J. Clark, L. D’Ambruoso, E. Fottrell, W. J. Graham, A. J. Herbst, A. Hodgson, S. Hounton, K. Kahn, et al. (2012). Strengthening standardised interpretation of verbal autopsy data: the new InterVA-4 tool. *Global Health Action* 5, 19281.
- James, S. L., A. D. Flaxman, and C. J. Murray (2011). Performance of the tariff method: validation of a simple additive algorithm for analysis of verbal autopsies. *Population Health Metrics* 9, 31.
- Li, Z. R., J. Thomas, E. Choi, T. H. McCormick, and S. J. Clark (2023). The r journal: The openva toolkit for verbal autopsies. *The R Journal*, 1. <https://journal.r-project.org/articles/RJ-2023-020/>.
- McCormick, T. H., Z. R. Li, C. Calvert, A. C. Crampin, K. Kahn, and S. J. Clark (2016). Probabilistic cause-of-death assignment using verbal autopsies. *Journal of the American Statistical Association* 111, 1036–1049.
- Murray, C. J., A. D. Lopez, R. Black, R. Ahuja, S. M. Ali, A. Baqui, L. Dandona, E. Dantzer, V. Das, U. Dhingra, et al. (2011). Population Health Metrics Research Consortium gold standard verbal autopsy validation study: design, implementation, and development of analysis datasets. *Population Health Metrics* 9, 27.
- PHMRC (2013). Population health metrics research consortium gold standard verbal autopsy data 2005–2011. <https://ghdx.healthdata.org/record/ihme-data/population-health-metrics-research-consortium-gold-standard-verbal-autopsy-data-2005-2011>.
- Serina, P., I. Riley, A. Stewart, S. L. James, A. D. Flaxman, R. Lozano, B. Hernandez, M. D. Mooney, R. Luning, R. Black, et al. (2015). Improving performance of the tariff method for assigning causes of death to verbal autopsies. *BMC Medicine* 13, 291.

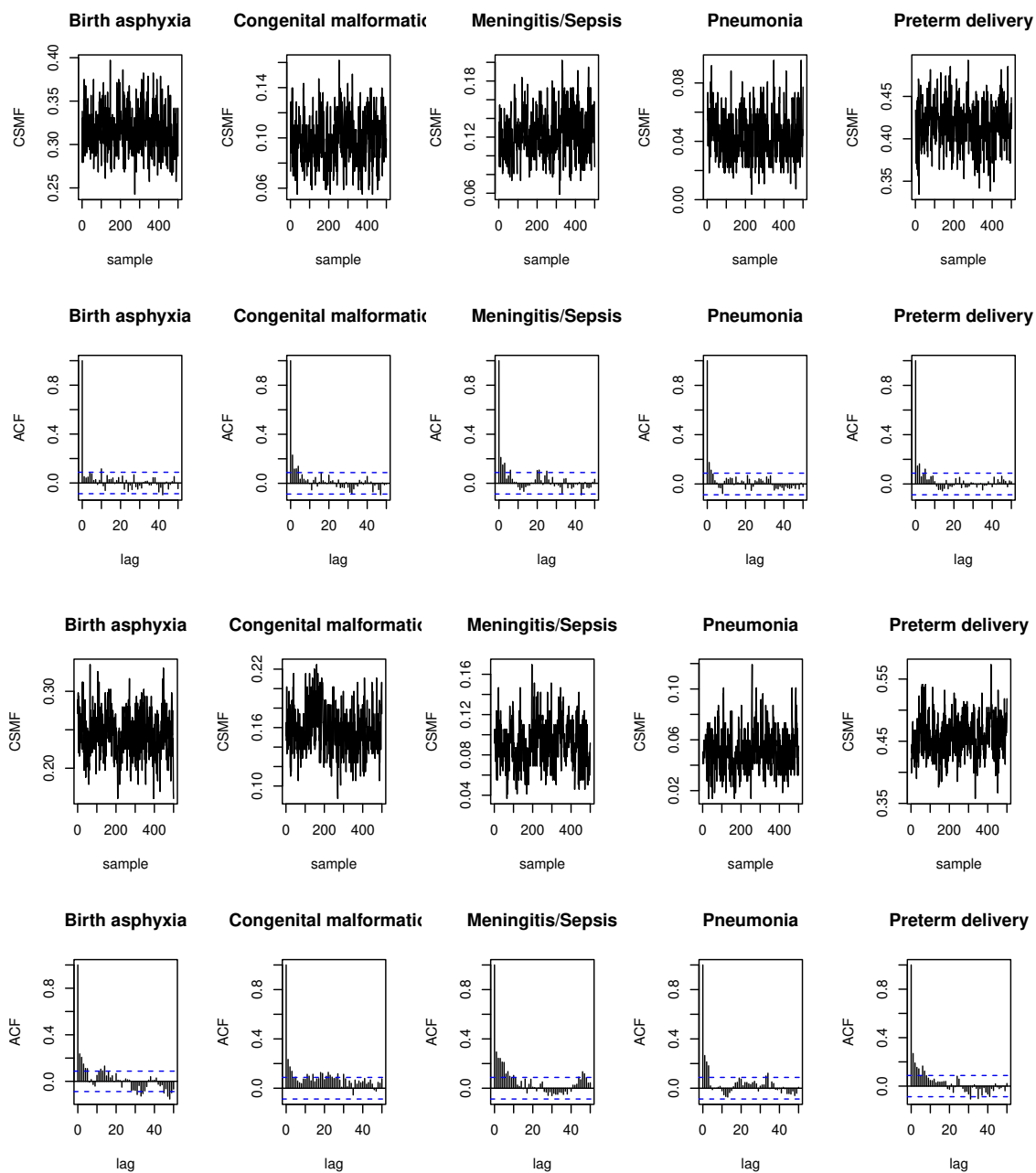


Figure 1: Sample paths and autocorrelations of the MCMC sample of the CSMF in the target site for the proposed model in case 1 (upper half) and case 2 (lower half).

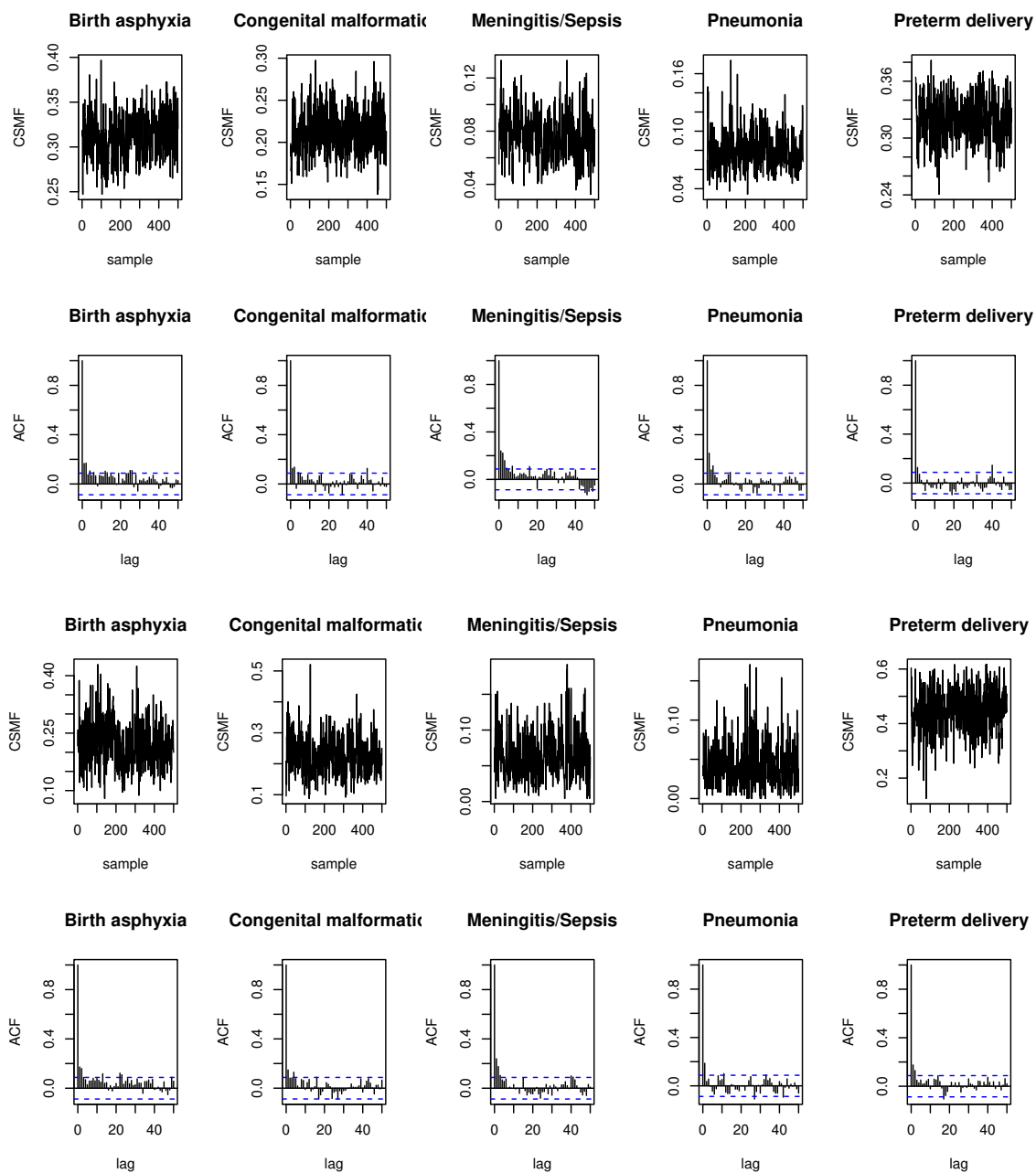


Figure 2: Sample paths and autocorrelations of the MCMC sample of the CSMF in the target site for the proposed model in case 3 (upper half) and case 4 (lower half).

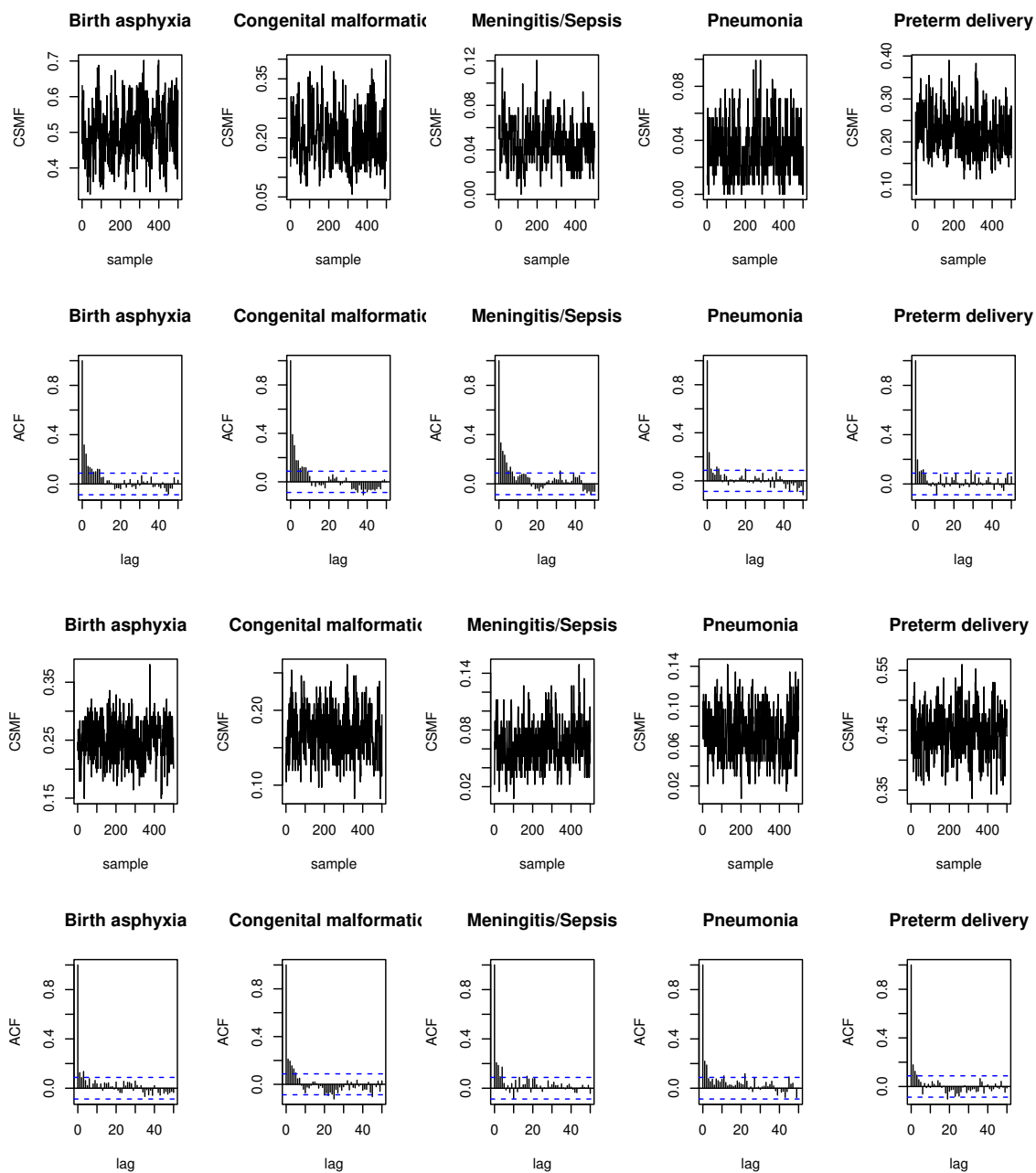


Figure 3: Sample paths and autocorrelations of the MCMC sample of the CSMF in the target site for the proposed model in case 5 (upper half) and case 6 (lower half).

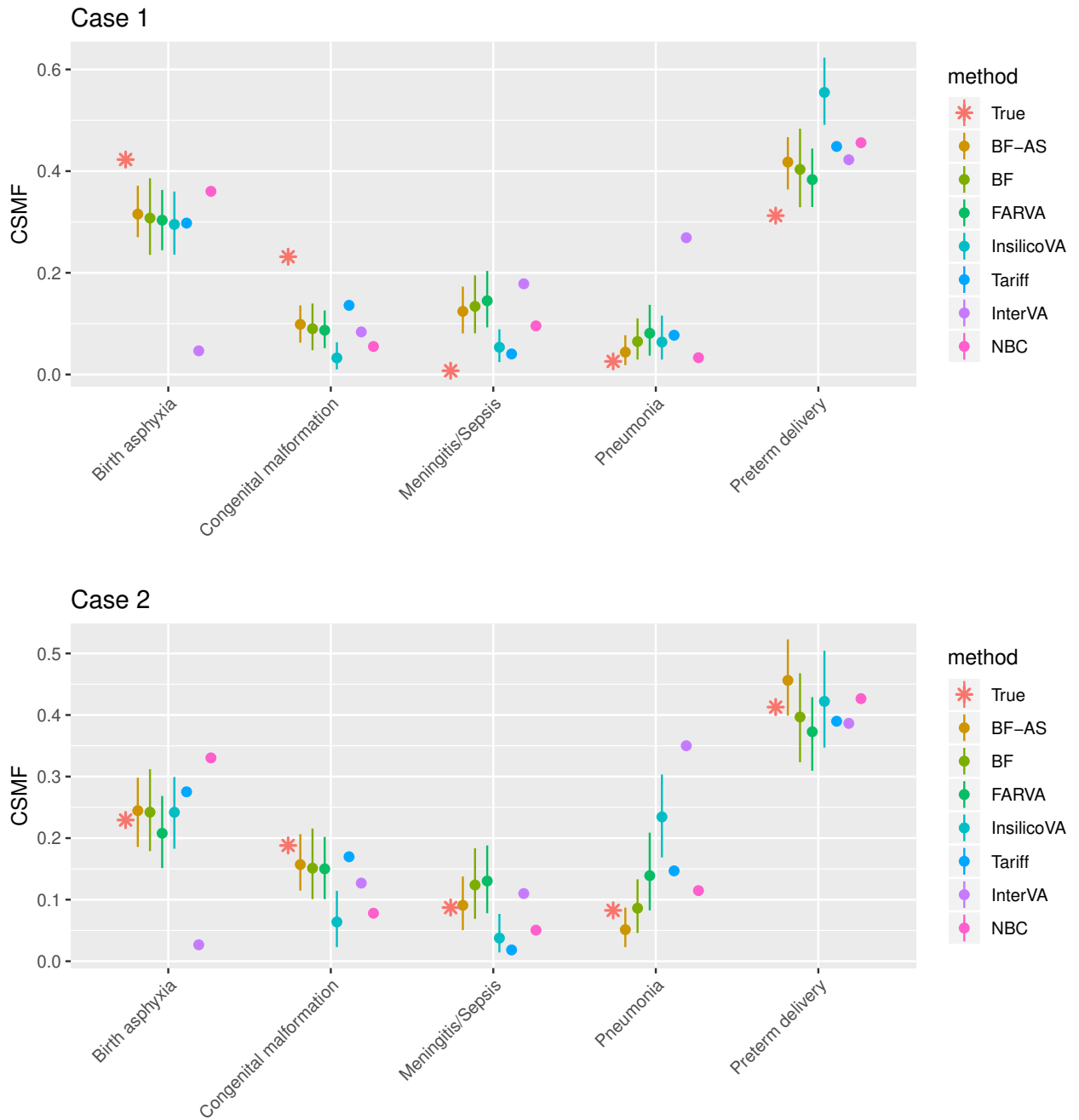


Figure 4: Estimation result of CSMF in Case 1 (above) and Case 2 (below). Red asterisk shows true values, and circle and interval correspond to mean and 95% interval of each statistical method. BF-AS, BF, FARVA, NBC mean Bayesian factor model with age- and sex-dependence (proposed method), Bayesian factor model, Bayesian hierarchical factor regression model and Naive Bayes Classifier. Tariff, InterVA and NBC produce only the point estimates.

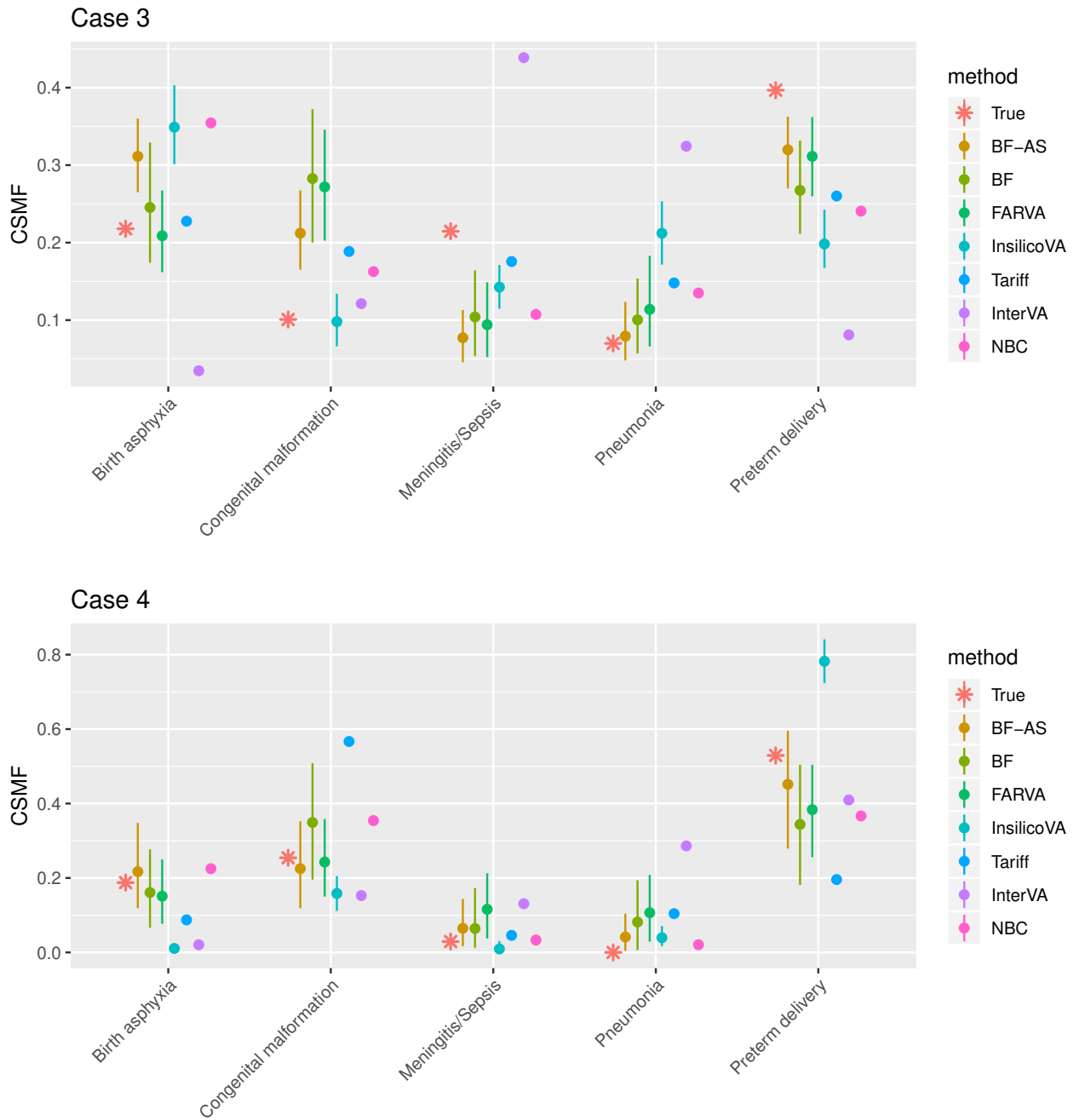


Figure 5: Estimation result of CSMF in Case 3 (above) and Case 4 (below). Red asterisk shows true values, and circle and interval correspond to mean and 95% interval of each statistical method. BF-AS, BF, FARVA, NBC mean Bayesian factor model with age- and sex-dependence (proposed method), Bayesian factor model, Bayesian hierarchical factor regression model and Naive Bayes Classifier. Tariff, InterVA and NBC produce only the point estimates.

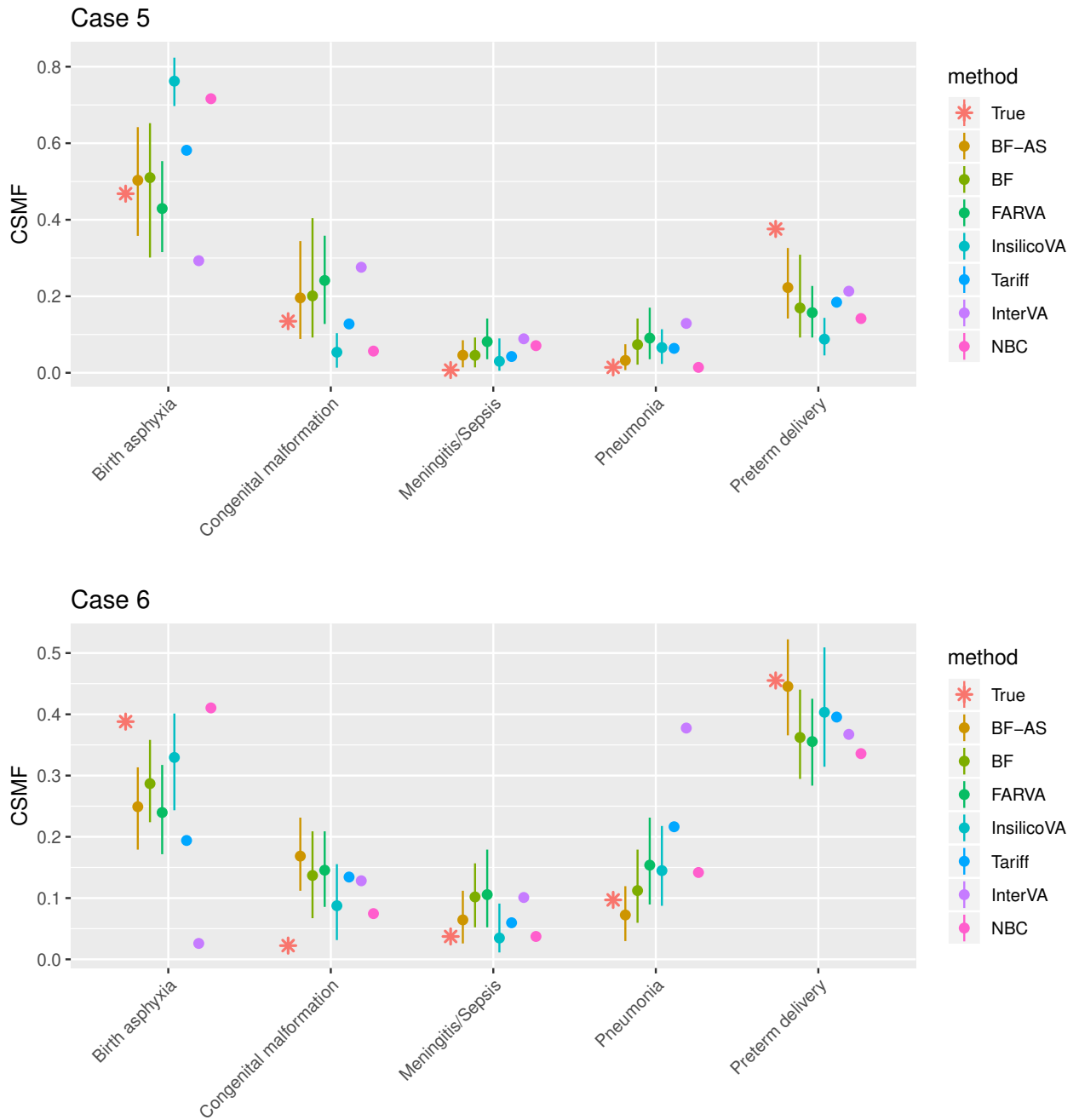


Figure 6: Estimation result of CSMF in Case 5 (above) and Case 6 (below). Red asterisk shows true values, and circle and interval correspond to mean and 95% interval of each statistical method. BF-AS, BF, FARVA, NBC mean Bayesian factor model with age- and sex-dependence (proposed method), Bayesian factor model, Bayesian hierarchical factor regression model and Naive Bayes Classifier. Tariff, InterVA and NBC produce only the point estimates.