Bayesian Inference for High Dimensional Cox Models with Gaussian and Diffused-Gamma Priors: A Case Study of Mortality in COVID-19 Patients Admitted to the ICU

Jiyeon Song[1*], Subharup Guha[2] and Yi Li[1]

[1*]Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA.

[2]Department of Biostatistics, University of Florida, Gainesville, FL, USA.

*Corresponding author(s). E-mail(s): jiyeons@umich.edu;
Contributing authors: s.guha@ufl.edu; yili@umich.edu;

Abstract

Bayesian approaches have been utilized to address the challenge of variable selection and statistical inference in high-dimensional survival analysis. However, the discontinuity of the ℓ_0-norm prior, including the useful spike-and-slab prior, may lead to computational and implementation challenges, potentially limiting the widespread use of Bayesian methods. The Gaussian and diffused-gamma (GD) prior has emerged as a promising alternative due to its continuous-and-differentiable ℓ_0-norm approximation and computational efficiency in generalized linear models. In this paper, we extend the GD prior to semi-parametric Cox models by proposing a rank-based Bayesian inference procedure with the Cox partial likelihood. We develop a computationally efficient algorithm based on the iterative conditional mode (ICM) and Markov chain Monte Carlo methods for posterior inference. Our simulations demonstrate the effectiveness of the proposed method, and we apply it to an electronic health record dataset to identify risk factors associated with COVID-19 mortality in ICU patients at a regional medical center.

Keywords: Bayesian variable selection, Risk assessment, Iterative conditional mode algorithm, Markov chain Monte Carlo sampling, Highest posterior density intervals, Urgent care

1 Introduction

Identifying risk factors relevant to mortality is of paramount importance for building reliable predictive models to profile patient risks and to effectively triage patients in an urgent care setting. Recently, electronic health record (EHR) data have emerged as a major resource to identify such risk factors. For example, the University of Michigan Health System (or Michigan Medicine), as a primary regional center managing the care of COVID-19 patients during the pandemic, has collected a wealth of demographic and clinical data via EHR. The rich EHR data have been routinely used for monitoring patients in need of urgent care at Michigan Medicine. A total of 1,265 COVID-19 patients were admitted to the Intensive Care Unit (ICU) in Michigan Medicine between March 10, 2020 (the date of the first case in that state) and December 31, 2021 (the cutoff date of the released EHR data). These patients were extremely vulnerable with a high mortality rate of 20%. It is crucial to assess the associations of demographic and clinical factors with in-hospital mortality among this vulnerable population, a primary endpoint for patients most severely attacked by COVID-19. The high dimensionality of EHR data, ranging from demographics, to lab measurements, to comorbidity conditions, poses challenges in extracting and assessing these candidate features using Cox proportional hazards models [1]. These models have been widely used for exploring associations between patient attributes and time-to-event outcomes, such as mortality. As demonstrated through the management of COVID-19 patients at Michigan Medicine, the utilization of EHR data has become essential for comprehending the multitude of factors that play a role in influencing in-hospital mortality. Insights from such analyses can aid healthcare professionals in devising targeted interventions and personalized treatment strategies to mitigate mortality risks effectively, making efficient usage of EHR data indispensable for evidence-based decision-making in urgent care settings and transformative medical discoveries.

For extracting relevant features from a massive candidate set of features, numerous penalization-based variable selection approaches have been proposed. [2] proposed an ℓ_0-norm penalty to induce sparsity for regression coefficients. To address the discontinuity and non-convexity of the ℓ_0 penalty, an alternative ℓ_1-norm penalty, the least absolute shrinkage and selection operator (Lasso) was presented by [3]. Lasso was extended to the Cox model by [4] and [5], who implemented the LARS algorithm [6] to approximate the Lasso in Cox regression models. Other commonly used regularized regression methods for the Cox model include the smoothly clipped absolute deviation penalty [SCAD; 7], the minimax concave penalty [MCP; 8] and the Elastic-net penalty [E-net; 9]. Further, [10] developed a debiased lasso (DLASSO) method for drawing inference on Cox models with a diverging number of predictors, and [11] proposed a method for hypothesis testing under Cox models with high dimensional covariates.

From the Bayesian perspective, spike-and-slab [12] priors and shrinkage priors are the most common variable selection techniques. The spike-and-slab prior, an analog of ℓ_0-norm penalties, is a mixing distribution with a point mass of zero and a continuous distribution for all non-zero values, and was adopted by [13] for variable selection and further extended by [14] to form a spike-and-slab mixture double-exponential prior. Other useful Bayesian shrinkage priors, with a performance comparable with the spike-and-slab prior, include normal-gamma mixture priors [15], relevance vector machine

[16], horseshoe prior [17], and Dirichlet-Laplace priors [18]. These priors have empowered various Bayesian variable selection methods for Cox models. For example, [19] developed a semi-automatic Bayesian variable selection method for up to 20 covariates, and [20] built a penalized semiparametric method for high dimensional survival data. [21] proposed a product inverse moment non-local prior density (piMOM) on non-zero coefficients, and [22] incorporated the horseshoe prior in Cox models. However, even when p is relatively large, Markov chain Monte Carlo (MCMC) sampling algorithms for the spike-and-slab prior methods with discontinuous ℓ_0-norm penalties often fail to converge to the target posterior distribution [23, 24], and, in our experience (as reported in our simulations), may produce a large number of false positives.

To address these limitations, we explore the idea of approximating ℓ_0-norm penalty from an empirical Bayes perspective. Specifically, we consider the use of the Gaussian and diffused-gamma (GD) prior [24], a novel Bayesian shrinkage prior which produces a differentiable ℓ_0-norm approximation under maximum *a posteriori* (MAP) estimation. Compared to other Bayesian methods, the GD prior is computationally efficient and effective in selecting important predictors and eliminating noise variables in generalized linear models [24]. However, since the GD prior was primarily proposed for fully parametric models, it is unclear about its utility in a semiparametric setting, particularly for Cox models. In this work, we extend the GD prior-based Bayesian strategy to Cox models. To avoid specifying priors on the (infinite dimensional) baseline hazard functions, we focus on the ranks of the observed survival outcomes, whose distribution turns out to be equal to the partial likelihood of a Cox model under independent censoring [25]; in a similar context, [21] proposed to use the partial likelihood as the sampling distribution for Bayesian model selection but did not provide justifications or draw inference. For quantifying the uncertainty of the estimated coefficients of all the predictors (regardless of being selected or not), we further propose a computationally efficient inference procedure via the iterative conditional mode (ICM) algorithm and Markov chain Monte Carlo methods.

Our work makes the following contributions. First, with independent censoring [26], we build the likelihood based on the ranks of the observed survival outcomes, which is equal to the commonly used partial likelihood of a Cox model and is free of the baseline hazard function; thus, our sampling distribution in the proposed Bayesian procedure does not need to involve complicated priors for baseline hazards. Second, our proposed GD prior can be described by a continuous and convex Gaussian and gamma mixture model to approximate the ℓ_0 penalty, which further reduces computational costs. Finally, our Bayesian framework facilitates simultaneous variable selection and inference, which offers a distinct advantage in comparison to other methodologies dedicated solely to variable selection or high-dimensional inference, such as the piMOM method [21].

The paper is organized as follows. Section 2 introduces the rank-based Bayesian approach for inferring on Cox Models. Section 3 proposes the GD prior for the regression coefficients, and Section 4 designs a posterior computational algorithm. Section 5 conducts simulations to evaluate the performance of the proposed method in finite sample settings. Section 6 applies the method to analyze the aforementioned Michigan Medicine EHR data to assess risk factors associated with the mortality of COVID

patients admitted to the ICU. Section 7 concludes the paper with a discussion. Additional results are presented in online Supporting Information.

2 Formulation of the Problem

2.1 Cox Models with High Dimensional Covariates

Let T_i and C_i respectively denote the survival time (e.g., the time from an ICU admission to death) and censoring time for the ith individual. The observed responses consist of $\tilde{T}_i = \min\{T_i, C_i\}$ representing the observed survival time, and $\delta_i = \mathbb{I}(T_i \leq C_i)$ representing the censoring indicator. Here, $\mathbb{I}(a)$ equals 1 if condition a holds, and equals 0 otherwise. As commonly assumed in the literature, we assume independent censoring, that is, T_i and C_i are conditionally independent given $\mathbf{x}_i$, where $\mathbf{x}_i \in \mathbb{R}^p$ is the covariate vector of individual i. Also, we assume that the T_i's are continuous, implying that ties occur with a probability of 0. The observed data, $(\mathbf{x}_i, \tilde{T}_i, \delta_i), i = 1, \ldots, n$, are independently and identically distributed copies of $(\mathbf{x}, \tilde{T}, \delta)$. Finally, denote by $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)^\top \in \mathbb{R}^{n \times p}$ the covariate matrix, by $\boldsymbol{X}_j$ $(j = 1, \ldots, p)$ the jth column of $\mathbf{X}$, and by $n_d = \sum_{i=1}^n \delta_i$ the number of observed events.

The Cox proportional hazards model [26] stipulates that the hazard rate for a subject with covariate $\mathbf{x}$ is $h(t \mid \mathbf{x}) = h_0(t) \exp\left(\mathbf{x}^\top \boldsymbol{\beta}\right)$, where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^\top$ is a vector of regression coefficients and $h_0(t)$ is the baseline hazard function. Associated with the specified Cox model is the partial likelihood [26]:

$$L(\boldsymbol{\beta}) = \prod_{k=1}^{n} \left\{ \frac{\exp\left(\mathbf{x}_k^\top \boldsymbol{\beta}\right)}{\sum_{l \in \mathscr{R}_k} \exp\left(\mathbf{x}_l^\top \boldsymbol{\beta}\right)} \right\}^{\delta_k}$$

where $\mathscr{R}_k$ is the set of indices of subjects at risk at $\tilde{T}_k$. Denote by $l(\boldsymbol{\beta}) = \log L(\boldsymbol{\beta})$ the log partial likelihood. When p is fixed and small, the frequentist estimates of $\boldsymbol{\beta}$ are routinely obtained by maximizing $L(\boldsymbol{\beta})$ or $l(\boldsymbol{\beta})$. However, when $p > n_d$ or even p/n_d is not negligible, the high dimensionality of predictors may defy traditional frequentist methods, causing inaccurate inferences [10].

Although our initial theory assumes the absence of ties, it is imperative to consider real-world scenarios in which ties can occur. Should ties indeed happen, the partial likelihood function we have presented represent the Breslow approximation [27] of likelihood for such tied observations, which is an effective and popular strategy for addressing ties [28]. Alternatively, we can employ either the exact likelihood [29] or Efron's approximation [30] if "heavy" ties are present.

2.2 Rank-based Bayesian Inference on Cox Models

As $L(\boldsymbol{\beta})$ depends on only the relative rankings of the survival data and not on their specific values [26], we propose a rank-based Bayesian approach to estimate and draw inference on $\boldsymbol{\beta}$, even when p is large. The key idea is that, regardless of the dimension of $\boldsymbol{\beta}$, the partial likelihood, $L(\boldsymbol{\beta})$, can be used as the sampling distribution in a Bayesian

procedure because $L(\boldsymbol{\beta})$ is the marginal likelihood for all possible underlying ranks consistent with the observed data, subject to independent censoring [25, 31].

Specifically, introduce $y_{ij} = \delta_i \mathbb{I}(\tilde{T}_i \leq \tilde{T}_j) - \delta_j \mathbb{I}(\tilde{T}_j < \tilde{T}_i)$, which specifies the possible ordering of subjects i and j; it is 1 if subject j is observed to survive longer than i, is -1 if subject i is observed to survive longer than j, or is 0 if the ordering between them cannot be determined with the observed data. As such, $\mathbf{y} = (y_{ij}, 1 \leq i < j \leq n)$ contain all possible rankings consistent with the observed survival outcomes, $(\tilde{T}_i, \delta_i), i = 1, \ldots, n$, and thus the density $f(\mathbf{y}|\boldsymbol{\beta}) = L(\boldsymbol{\beta})$ under independent censoring [25]. The approach is an effective alternative to the common Bayesian survival approaches [32], by avoiding specifying priors for the baseline hazard functions and relieving much computational burden.

Our procedure enables variable selection, for which we introduce the marginal probability of observing the rank data, $\mathbf{y}$, under $L(\boldsymbol{\beta})$, i.e., $m(\mathbf{y}) = \int L(\boldsymbol{\beta})\pi(\boldsymbol{\beta})d\boldsymbol{\beta}$, and use it for selecting tuning parameters. As $\pi(\boldsymbol{\beta})$, the prior of $\boldsymbol{\beta}$, impacts the overall performance of the selection procedure and the amount of sparsity, we adapt a GD version of $\pi(\boldsymbol{\beta})$ to Cox models.

3 GD Prior Adapted to Cox Models

The GD prior [24] is a Bayesian shrinkage prior, which is differentiable and approximates the ℓ_0-norm, i.e., $\sum_{j=1}^{p} \mathbb{I}(\beta_j \neq 0)$, with the maximum *a posteriori* (MAP) estimation. We adapt this prior, originally designed for generalized linear models, to Cox models. We start by proposing an ℓ_0-norm approximation to obtain

$$\tilde{\ell}_0(\boldsymbol{\beta} \mid \kappa, \tau_0) = \kappa \sum_{j=1}^{p} \frac{\boldsymbol{X}_j^\top \boldsymbol{X}_j \beta_j^2}{\tau_0 + \boldsymbol{X}_j^\top \boldsymbol{X}_j \beta_j^2}, \tag{1}$$

where $\kappa \geq 0$ is a hyperparameter controlling the level of sparsity, τ_0 is predetermined to be sufficiently small, and $\boldsymbol{X}_j$ is the jth column of $\mathbf{X}$. Moreover, writing $g(\tau \mid x) = x^2/\left(\tau^2 + x^2\right)$, as

$$\lim_{\tau \to 0} g(\tau \mid x) = \mathbb{I}(x \neq 0),$$

we have that $\tilde{\ell}_0(\boldsymbol{\beta} \mid \kappa, \tau_0) \to \kappa \sum_{j=1}^{p} \mathbb{I}(\beta_j \neq 0)$ as $\tau_0 \to 0$, for any given κ and $\boldsymbol{\beta}$ [24]. Therefore, if τ_0 is sufficiently small, penalty function (1) provides a good approximation to the ℓ_0-norm penalty. The behavior of the GD prior to an appropriate tuning parameter τ_0 is illustrated in Figure 1. As τ_0 decreases, the shrinkage rate towards zero increases, suggesting that non-important variables can be quickly excluded for small τ_0.

Consequently, we derive a prior that attains this form of ℓ_0-norm approximation. We set independent Gaussian priors for the β_j's with precision parameter d_j, for $j = 1, \ldots, p$. We further propose that the d_j's have independent gamma priors with shape parameter $\kappa + \frac{1}{2}$ and rate parameter $\tau_0/(2\boldsymbol{X}_j^\top \boldsymbol{X}_j)$. Let $\boldsymbol{d} = (d_1, \ldots, d_p)^\top$. Then the joint GD prior for $\boldsymbol{\beta}$ and auxiliary statistic $\boldsymbol{d}$ can be specified as

$$\pi_{GD}(\boldsymbol{\beta}, \boldsymbol{d}) \propto \pi_G(\boldsymbol{\beta} \mid \boldsymbol{d})\, \pi_D(\boldsymbol{d}), \tag{2}$$

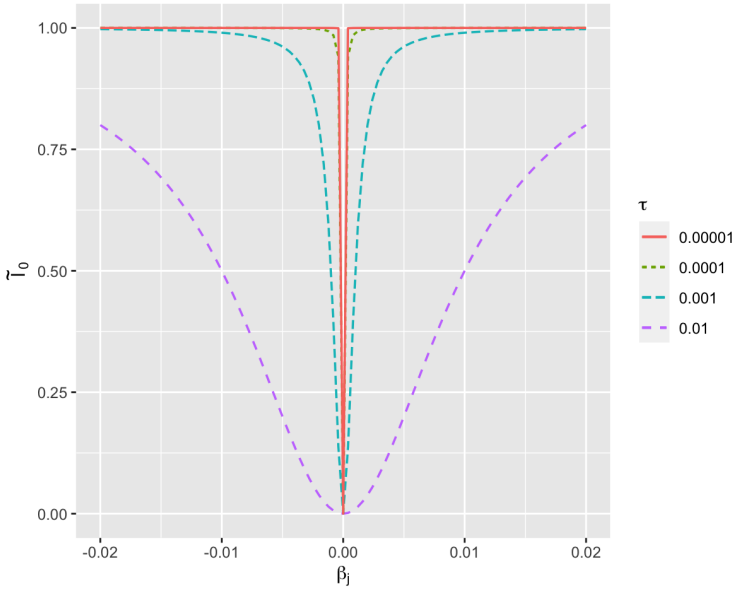

Fig. 1 Selection performance of the Gaussian and Diffused-gamma (GD) prior based on the values of τ_0.

where

$$\pi_G(\boldsymbol{\beta} \mid \boldsymbol{d}) \propto \prod_{j=1}^{p} \left\{ d_j^{1/2} \exp\left(-\frac{d_j}{2} \beta_j^2 \right) \right\}, \tag{3}$$

$$\pi_D(\boldsymbol{d}) \propto \prod_{j=1}^{p} \left\{ d_j^{\kappa - 1/2} \exp\left(-\frac{\tau_0^2}{2 \boldsymbol{X}_j^\top \boldsymbol{X}_j} d_j \right) \right\}. \tag{4}$$

One could potentially integrate $\boldsymbol{d}$ out of (2) to obtain the "marginal prior" for $\boldsymbol{\beta}$. However, the integration would incur high computational costs. An alternative strategy is to replace $\boldsymbol{d}$ in (2) by its MAP estimator, $\hat{\boldsymbol{d}}_{MAP} = \arg\max_{\boldsymbol{d}} [\pi_{GD}(\boldsymbol{\beta}, \boldsymbol{d})] = \left(\frac{2\kappa}{\tau_1^2 + \beta_1^2}, ..., \frac{2\kappa}{\tau_p^2 + \beta_p^2} \right)$, where $\tau_j^2 = \frac{\tau_0^2}{\boldsymbol{X}_j^\top \boldsymbol{X}_j}$ and $j = 1, \cdots, p$. We refer to $\pi_{\text{GD}}(\boldsymbol{\beta}, \hat{\boldsymbol{d}}_{MAP})$ as the *MAP-based prior* for $\boldsymbol{\beta}$. Applying (3) and (4), and since $\pi_{GD}(\boldsymbol{\beta}, \hat{\boldsymbol{d}}_{MAP}) \propto \pi_G(\boldsymbol{\beta} \mid \hat{\boldsymbol{d}}_{MAP})$, we have

$$\arg\max_{\boldsymbol{\beta}} \left\{ L(\boldsymbol{\beta})\, \pi_{\text{GD}}(\boldsymbol{\beta}, \hat{\boldsymbol{d}}_{MAP}) \right\} = \arg\min_{\boldsymbol{\beta}} \left\{ -l(\boldsymbol{\beta}) + \tilde{\ell}_0 \left(\boldsymbol{\beta} \mid \kappa, \tau_0 \right) \right]. \tag{5}$$

The resulting estimate is termed the MAP estimate of $\boldsymbol{\beta}$, aligning with that a penalized estimator can be considered as MAP estimator of the posterior distribution in a Bayesian analysis [3]; for a given τ_0 (i.e., $\tau_0 = 10^{-4}$, based on Figure 1 and our numerical experience), the MAP-based prior for $\boldsymbol{\beta}$ helps attain the desired continuous and differentiable ℓ_0-norm penalty (1), and the results are not sensitive to the choice of τ_0 when it is sufficiently small. Furthermore, a Bayesian framework complements the MAP estimates with the estimation of a posterior distribution of $\boldsymbol{\beta}$, laying the groundwork for posterior inference.

We have utilized a prior that integrates data information and shares similarities with an ℓ_0 norm penalty, which might lead to concerns about the double usage of the data. To address this, we have introduced a hyperprior (τ_0), which facilitates the specification of a higher-level distribution, permitting the incorporation of additional external knowledge or expert insights. By integrating this hyperprior, we may improve the model's resilience and adaptability, avoiding overfitting the data and mitigating biases linked to data-driven priors.

4 Algorithms

We adopt a component-wise updating ICM algorithm to compute the MAP estimate of $\boldsymbol{\beta}$ and further develop an MCMC sampler to approximate the posterior distribution of $\boldsymbol{\beta}$.

4.1 MAP estimation via component-wise updating ICM

It follows that the MAP estimates of $\boldsymbol{\beta}$ and $\boldsymbol{d}$, separately obtained from (2) and (5), can be jointly obtained by maximizing the following posterior density of $\boldsymbol{\beta}$ and $\boldsymbol{d}$ given $\boldsymbol{y}$. That is,

$$(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{d}}) = \arg\max_{\boldsymbol{\beta},\boldsymbol{d}} \pi(\boldsymbol{\beta}, \boldsymbol{d}|\boldsymbol{y}) = \arg\max_{\boldsymbol{\beta},\boldsymbol{d}} L(\boldsymbol{\beta})\pi_G(\boldsymbol{\beta}|\boldsymbol{d})\pi_D(\boldsymbol{d}),$$

where $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{d}}$ denote the MAP estimates of $\boldsymbol{\beta}$ and $\boldsymbol{d}$ (without the subscript of "MAP" for simplicity). This observation facilitates the use of a component-wise updating ICM algorithm:

$$\begin{aligned}
\widehat{\boldsymbol{d}} &\leftarrow \arg\max_{\boldsymbol{d}} \left\{ \pi_G(\widehat{\boldsymbol{\beta}}|\mathbf{d})\pi_{\mathrm{D}}(\boldsymbol{d}) \right\}, \\
\widehat{\boldsymbol{\beta}} &\leftarrow \arg\max_{\boldsymbol{\beta}} \{ L(\boldsymbol{\beta})\pi_{\mathrm{G}}(\boldsymbol{\beta}|\widehat{\boldsymbol{d}}) \},
\end{aligned}$$

which can be implemented by using *Algorithm 1*. Based on our numerical studies, ϵ_0 in the algorithm is set to be 0.0001, and for choosing the initial estimates of $\hat{\boldsymbol{\beta}}^{(0)}$, we recommend using the estimates from the ridge regression. Following [24], the hard-thresholding parameter $\zeta_j^{(t)}$ is chosen to be $1.96/\sqrt{\frac{2\kappa}{\tau_0^2/\boldsymbol{X}_j^\top \boldsymbol{X}_j + (\tilde{\beta}_j^{(t)})^2}}$, corresponding to the bound of a 95% prior credible interval when $\beta_j = 0$, which turns out to work well in our experience.

Algorithm 1 The component-wise ICM algorithm

1. Set an initial $\hat{\boldsymbol{\beta}}^{(0)} = (\hat{\beta}_1^{(0)}, \ldots, \hat{\beta}_p^{(0)})^\top$.
2. For $t = 1, 2, \ldots,$

$$\hat{d}_j^{(t)} = \frac{2\kappa}{\tau_0^2 / \boldsymbol{X}_j^\top \boldsymbol{X}_j + (\hat{\beta}_j^{(t-1)})^2},$$

$$\tilde{\beta}_j^{(t)} = \arg\min_{\beta_j} \left[- l(\hat{\beta}_1^{(t-1)}, \ldots, \hat{\beta}_{j-1}^{(t-1)}, \beta_j, \hat{\beta}_{j+1}^{(t-1)}, \ldots, \hat{\beta}_p^{(t-1)}) + \frac{\hat{d}_j^{(t)}}{2} \beta_j^2 \right],$$

$$\hat{\beta}_j^{(t)} = \tilde{\beta}_j^{(t)} I\{|\tilde{\beta}_j^{(t)}| > \zeta_j^{(t)}\},$$

   where $j = 1, 2, \cdots, p$.
3. Stop when $\max_{j=1,\cdots,p} |\widehat{\beta}_j^{(t)} - \widehat{\beta}_j^{(t-1)}| < \epsilon_0$, where ϵ_0 is sufficiently small (e.g., 0.0001).
4. Following convergence, set $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1^{(t)}, \ldots, \hat{\beta}_p^{(t)})^\top$ and $\hat{\boldsymbol{d}} = (\hat{d}_1^{(t)}, \ldots, \hat{d}_p^{(t)})^\top$.

4.2 MCMC sampling method

For Bayesian inference, it is important to obtain the posterior distribution of the regression coefficient. With no closed form for the full conditional of $\boldsymbol{\beta}$ or its components β_j, we apply the Metropolis-Hasting algorithm. At the $(s+1)$th MCMC iteration, using $\beta_j^{(s)}$ sampled at the sth iteration, we sample β_j with $\beta_j^{(s+1)}$ from a Gaussian proposal density $\phi\left(\cdot|\beta_j^{(s)}\right)$ with mean $\beta_j^{(s)}$ and variance $\omega\big(-\frac{\partial^2}{\partial \beta_j^2} l(\boldsymbol{\beta})|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} + \hat{d}_j\big)^{-1}$ for some $\omega > 0$, where $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{d}}$ are obtained from *Algorithm 1*. We then compute the acceptance probability for the proposed move as

$$\alpha = \min \left\{ 1, \frac{L * \pi_{\mathrm{GD}}(\beta_1^{(s)}, \ldots, \beta_{j-1}^{(s)}, \beta_j^{(s+1)}, \beta_{j+1}^{(s)}, \ldots, \beta_p^{(s)}) \phi\left(\beta_j^{(s)} | \beta_j^{(s+1)}\right)}{L * \pi_{\mathrm{GD}}(\beta_1^{(s)}, \ldots, \beta_{j-1}^{(s)}, \beta_j^{(s)}, \beta_{j+1}^{(s)}, \ldots, \beta_p^{(s)}) \phi\left(\beta_j^{(s+1)} | \beta_j^{(s)}\right)} \right\},$$

for $j = 1, \cdots, p$, where for notational ease $L * \pi_{\mathrm{GD}}(\boldsymbol{\beta}) = L(\boldsymbol{\beta}) \pi_{\mathrm{GD}}(\boldsymbol{\beta}, \hat{\boldsymbol{d}})$. This way, the MAP estimates, $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{d}}$, from the ICM algorithm are incorporated into the proposal density. Further, ω controls the acceptance rate of the MCMC procedure and we aimed to have 40% - 60% acceptance rates. Setting ω equal to 1 or 1.5 in the simulation studies yielded reasonable acceptance rates. For inference, we use the MCMC samples to construct the 95% highest posterior density (HPD) intervals for the coefficients of all the potential predictors, regardless of them being selected.

4.3 Tuning parameter selection

As parameter κ controls the degree of sparsity in the MAP estimate, we detail its selection process. For a given κ, denote by $\boldsymbol{\beta}_\kappa$ the vector of non-zero coefficients, and by p_κ the dimension of $\boldsymbol{\beta}_\kappa$. Motivated by BIC, we consider an independent "working"

normal prior so that $\boldsymbol{\beta}_\kappa \sim N\left(\mathbf{0}, \frac{1}{p_\kappa}\mathbf{I}\right)$. When there are a few non-zero coefficients, the normal prior variance is the reciprocal of p_κ, making the prior less informative. As the marginal distribution of the data is $m(\mathbf{y}|\kappa) = \int L(\boldsymbol{\beta}_\kappa)\pi(\boldsymbol{\beta}_\kappa)d\boldsymbol{\beta}_\kappa$, we propose a Monte Carlo (MC) estimator of $m(\mathbf{y}|\kappa)$, i.e., $\widehat{m}(\mathbf{y}|\kappa) = S^{-1}\sum_{i=1}^{S}\frac{L(\boldsymbol{\beta}_\kappa)\pi(\boldsymbol{\beta}_\kappa)}{N(\boldsymbol{\beta}_\kappa|\boldsymbol{\mu}_\kappa,\Sigma_\kappa)}$, where S denotes the number of simulations, $\boldsymbol{\mu}_\kappa = \arg\max_{\boldsymbol{\beta}_\kappa}\log\{L(\boldsymbol{\beta}_\kappa)\pi(\boldsymbol{\beta}_\kappa)\}$, and $\Sigma_\kappa = -\left[\frac{\partial^2}{\partial\boldsymbol{\beta}_\kappa\partial\boldsymbol{\beta}_\kappa^\top}\log L\{(\boldsymbol{\beta}_\kappa)\pi(\boldsymbol{\beta}_\kappa)\}|_{\boldsymbol{\beta}_\kappa=\boldsymbol{\mu}_\kappa}\right]^{-1}$. Finally, we choose $\hat{\kappa} = \arg\max_\kappa \widehat{m}(\mathbf{y}|\kappa)$, which can be implemented using a grid search. Justifications for choosing κ in this fashion were given by [24] in the context of generalized linear models, and our simulations confirmed its utility in the setting of Cox models by detecting sparse models.

5 Simulation Studies

We conducted simulation studies to compare our proposed GD prior method with the competing methods, including ridge, E-net, LASSO, MCP, SCAD, the spike-and-slab mixture double-exponential [BLasso; 14, 33] known as Bayesian spike-and-slab mixture lasso, product moment [pMOM; 21], product inverse moment nonlocal priors[piMOM; 21], and the debiased Lasso [DLASSO; 10] for Cox models. The competing frequentist methods and the Bayesian Lasso were respectively implemented by using `glmnet`, `ncvreg`, `BhGLM`, and `BVSNLP` packages in R. The selection of the tuning parameter, κ, for our proposed GD prior method was performed as detailed in the last section. For the ridge, E-net, LASSO, MCP, and SCAD penalized regression models that involved regularization parameters, we conducted 10-fold cross-validation over a grid of values of these parameters and chose those that minimized the cross-validation error, which in our case was the negative log-partial likelihood [34–36]. We have made the code publicly available on GitHub.

We evaluated selection and prediction performance using false positive rate (FPR), false negative rate (FNR), false discovery rate (FDR), and the average number of selected covariates:

$$\text{FPR} = \frac{\text{FP}}{\text{TN}+\text{FP}}, \quad \text{FNR} = \frac{\text{FN}}{\text{TP}+\text{FN}}, \quad \text{FDR} = \frac{\text{FP}}{\text{FP}+\text{TP}},$$

where a false positive incorrectly identifies a variable as being relevant for the prediction when it is actually irrelevant or has no true association with the target outcome, and a false negative selects a variable as being irrelevant for the prediction when it is actually irrelevant or has a true association with the target outcome. To generate the survival outcomes and covariates, we considered the following three scenarios, similar to those adopted by [37].

The covariate $\mathbf{x} = (x_1, \ldots, x_p)^\top$ was generated from multivariate Gaussian, marginally standard normal, but with the following correlation structures and model sizes (i.e., numbers of non-zero coefficients):

Scenario 1: $\text{Corr}(x_i, x_j) = 0$ for all $i \neq j \in \{1, \cdots, p\}$. The true regression coefficients, $\boldsymbol{\beta}_0 = (1, 1, 0, 0, 1, 1, 0, \cdots, 0)^\top$, consisted of four non-zero regression coefficients.

Scenario 2: $\text{Corr}(x_i, x_j) = 0.5^{|i-j|}$ for all $i, j \in \{1, \cdots, p\}$. The true regression coefficients,

$$\boldsymbol{\beta}_0 = (1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 0 \cdots, 0)^\top,$$

had 10 non-zero regression coefficients.

Scenario 3: $\text{Corr}\,(x_i, x_4) = 1/\sqrt{2.2}$ for all $i \neq 4$, and $\text{Corr}\,(x_i, x_j) = 1/2$ for all $i \neq j \in \{1, \cdots, p\} \backslash \{4\}$. Only the first 5 predictors of $\boldsymbol{\beta}_0 = (1, 1, 1, -1.5\sqrt{2}, 1/3, 0, \cdots, 0)^\top$ were non-zero.

We varied n to be 100 and 1,250, and set p equal to 20, 100, and 200. When $n = 100$, we generated the survival and censoring times independently and respectively from an exponential distribution with the rate of $\exp(\mathbf{x}^\top \boldsymbol{\beta}_0)$ and 0.1, corresponding to an expected censoring rate of 30%. When $n = 1,250$, we generated the survival times the same way, but generated the censoring times independently from an exponential distribution with a rate of $\exp(\mathbf{x}^\top \boldsymbol{\beta}_0)$ and $4\exp(\mathbf{x}^\top \boldsymbol{\beta}_0)$, corresponding to a roughly 50% and 80% censoring rate respectively; the same size and a censoring rate of 80%, in this case, were meant to mimic our real data. A total of 100 independent datasets were generated for simulation configuration. We ran MCMC simulations for 3,000 iterations with a burn-in of 500 for each dataset. In all three scenarios, the tuning parameter κ was obtained by performing a grid search in the interval $[0.1, 3]$, as described in Section 4.3.

Table 1 presents the selection performance of each method when $n = 100$ with an approximately 30% censoring rate, i.e., n_d was about 70. We note that DLASSO is only applicable with $p < n_d$. The GD prior provided the lowest FPR, FNR, and FDR in most cases considered, outperforming the other methods as the number of predictors increased. It produced a sparse model overall, though pMOM and piMOM provided sparser models in all cases, and MCP yielded a smaller FPR for $n = 200$. However, pMOM failed to select significant variables with more complicated covariance structures as in Scenario 3. When evaluating the estimation of each coefficient, we found that the performance of each method varied. The prediction performance was assessed by using the mean squared error (MSE) and the concordance index [C-index; 38]. Table 1 also presents the prediction performance of the different methods, showing that the proposed method provided a smaller MSE than most of the other methods, except for pMOM and piMOM in some cases. Using FPR, FNR and FDR as the criteria, our proposed method was comparable with pMOM and piMOM and was better than the other methods. The MCP method was competitive in prediction for a larger number of predictors and under more complicated correlation structures; in contrast, the proposed method did provide a stable performance.

We repeated the same experiment with a larger sample size ($n = 1,250$) but moderate and severe censoring rates ($\sim 50\%$ and $\sim 80\%$). When the censoring rate was $\sim 80\%$, n_d was about 250, and both the sample size and the number of events were roughly equal to those of our real example. Tables 2 and 3 summarize the performance of each method. With $n = 1,250$, the pMOM algorithm failed to converge in most cases, whereas piMOM, the improved version of pMOM, performed quite well. Therefore, we did not include pMOM, but still kept piMOM in the table. The results were consistent with what was observed when $n = 100$ and the censoring rate was

$\sim 30\%$ and revealed that our proposed method remained effective even with a moderate or high censoring rate, and in many cases was comparable, if not better than, with piMOM.

We next compared the time required to obtain the MAP estimates using Bayesian methods presented in Table 4 under the three aforementioned scenarios for sample sizes of n=100 and n=1250. Our findings indicate that our method is faster than BLASSO and comparable to pMOM and piMOM. This is reasonable because BLASSO requires the MCMC sampling to compute the MAP estimates, which is more time-consuming. On the other hand, our GD approach, pMOM, and piMOM methods have closed-form solutions, making them more computationally efficient.

Table 5 displays a summary of MAP estimates of the regression coefficients obtained by MCP, BLASSO, and DLASSO over 100 repetitions of Scenarios 1–3 with $n = 100$ and $p = 20$. The proposed method yielded smaller standard errors and MSEs than the other methods, indicating better accuracy and efficiency. Summaries of the MAP estimate for the $p = 100$ and $p = 200$ cases supported that the important variables detected by MCP were substantially underestimated compared to our method. The underestimation worsened as predictors increased (Web Tables S1 and S2). We further find that even when $p > n$, the GD prior showed stable estimation results without suffering from underestimation. Moreover, Web Figures S3 and S4 illustrate the grid search results for κ as specified in Section 4.3.

Finally, to verify the accuracy of the Bayesian interval estimates produced by the GD prior technique, we post-processed the Scenario 1 MCMC samples to compute the MC coverage of the 95% HPD intervals for the variables x_j, $j = 1, \ldots, p$. These coverage probabilities are displayed in Figure 2. Our GD approach achieved high coverage probabilities for the non-zero coefficients, even for relatively modest-sized data ($p = 100$) as well as high dimensional data ($p = 200$). The MC coverage plot of Scenario 1 with $p = 20$ is presented in Web Figure S1, and the plots for Scenarios 2 and 3 are shown in Web Figure S2.

6 Analysis of EHR with COVID-19 Patients Admitted to ICU

We applied our proposed GD prior method and the competing methods listed in the simulation studies to analyze the Michigan Medicine EHR data to assess risk factors associated with in-hospital mortality of COVID-19 patients admitted to ICU. The data consist of $n = 1,265$ COVID-19 patients admitted between March 10, 2020, and December 31, 2021. In-hospital survival time was defined as the time from ICU admission until in-hospital death, possibly censored by discharge or termination of the study. During the follow-up, a total of $n_d = 250$ deaths were observed.

The analytical goal was to estimate the associations of common risk factors with in-hospital mortality. We extracted and created a set of demographic, socioeconomic, and clinical risk factors identified in the literature relevant to COVID-19 health outcomes, including age, sex, race (Black or non-Black), ethnicity (Hispanic or non-Hispanic), smoking status, alcohol use, drug use, and prevalent comorbidity conditions. We obtained physiologic measurements within 24 hours of admission, such as body

Table 1 Comparisons of mean squared error (MSE) and concordance index (C-index), false positive rate (FPR), false negative rate (FNR), false discovery rate (FDR), and an average number of selected variables across methods applied to Scenarios 1–3 when $n = 100$ and the censoring rate is 30%.

		Scenario 1						Scenario 2						Scenario 3					
p	Method	MSE	CINDEX	FPR	FNR	FDR	NUM.SEL	MSE	CINDEX	FPR	FNR	FDR	NUM.SEL	MSE	CINDEX	FPR	FNR	FDR	NUM.SEL
20	ORACLE	0.005	0.793	0.000	0.000	0.000	4	0.036	0.861	0.000	0.000	0.000	10	0.010	0.752	0.000	0.000	0.000	5
	RIDGE	0.025	0.778	100.000	0.000	80.000	20	0.053	0.858	100.000	0.000	50.000	20	0.089	0.720	100.000	0.000	75.000	20
	ENET	0.019	0.788	56.062	0.000	67.609	13	0.048	0.859	64.700	0.000	38.537	16	0.069	0.733	64.067	0.400	65.032	15
	LASSO	0.015	0.790	39.062	0.000	58.023	10	0.046	0.859	50.000	0.000	32.140	15	0.044	0.740	48.133	0.200	56.994	12
	MCP	0.007	0.792	2.625	0.000	6.570	4	0.062	0.857	10.900	1.800	8.719	11	0.019	0.747	4.867	2.600	9.119	6
	SCAD	0.006	0.793	6.188	0.000	15.494	5	0.062	0.858	18.100	1.000	14.006	12	0.017	0.746	8.467	1.200	16.031	6
	DLASSO	0.017	0.784	100.000	0.000	80.000	20	0.050	0.857	100.000	0.000	50.000	20	0.034	0.735	100.000	0.000	75.000	20
	BLASSO	0.009	0.790	44.938	0.000	62.199	11	0.033	0.858	58.800	0.000	36.300	16	0.028	0.742	49.067	0.200	58.020	12
	pMOM	0.001	0.796	2.750	0.750	7.819	4	0.005	0.863	5.815	1.064	5.195	10	0.069	0.593	0.067	74.178	1.853	1
	piMOM	0.006	0.791	0.125	0.000	0.400	4	0.028	0.859	1.502	0.500	1.336	10	0.017	0.742	0.600	11.600	1.619	5
	GD	0.006	0.792	0.625	0.000	2.000	4	0.028	0.860	0.400	0.400	0.373	10	0.015	0.744	2.133	9.200	5.408	5
100	ORACLE	0.001	0.796	0.000	0.000	0.000	4	0.008	0.865	0.000	0.000	0.000	10	0.003	0.750	0.000	0.000	0.000	5
	RIDGE	0.035	0.677	100.000	0.000	96.000	100	0.087	0.799	100.000	0.000	90.000	100	0.071	0.603	100.000	0.000	95.000	100
	ENET	0.009	0.781	18.885	0.000	80.686	22	0.023	0.857	24.878	0.000	67.940	32	0.046	0.684	18.916	6.200	76.573	23
	LASSO	0.006	0.785	15.354	0.000	76.207	19	0.017	0.858	16.978	0.000	58.773	25	0.029	0.705	18.316	4.000	76.573	22
	MCP	0.002	0.794	1.177	0.000	17.432	5	0.037	0.834	1.856	20.700	15.507	10	0.040	0.672	1.379	37.600	25.550	4
	SCAD	0.002	0.794	3.740	0.000	41.865	8	0.040	0.835	4.622	12.300	29.135	13	0.024	0.705	3.937	14.800	42.200	8
	BLASSO	0.003	0.788	15.188	0.000	77.262	19	0.018	0.857	15.467	0.200	57.054	24	0.020	0.713	14.663	5.200	72.069	19
	pMOM	0.001	0.796	0.000	0.000	0.000	4	0.005	0.863	0.085	1.282	0.699	10	0.069	0.593	0.029	79.444	2.778	1
	piMOM	0.002	0.794	0.083	0.000	1.533	4	0.010	0.859	0.144	4.558	1.296	10	0.005	0.734	0.095	17.193	1.967	4
	GD	0.002	0.791	1.115	0.000	15.487	5	0.007	0.863	0.189	0.500	1.514	10	0.010	0.725	2.547	11.400	27.939	7
200	ORACLE	0.000	0.797	0.000	0.000	0.000	4	0.003	0.862	0.000	0.000	0.000	10	0.001	0.751	0.000	0.000	0.000	5
	RIDGE	0.018	0.628	100.000	0.000	98.000	200	0.044	0.755	100.000	0.000	95.000	200	0.037	0.560	100.000	0.000	97.500	200
	ENET	0.007	0.778	10.153	0.000	81.687	24	0.015	0.849	15.011	0.000	73.211	39	0.029	0.673	9.333	16.200	78.531	22
	LASSO	0.004	0.783	8.939	0.000	79.159	22	0.011	0.850	11.353	0.000	67.065	32	0.021	0.691	10.041	12.600	79.163	24
	MCP	0.001	0.796	0.949	0.000	26.829	6	0.025	0.808	0.747	34.400	15.212	8	0.030	0.633	0.672	58.600	29.375	3
	SCAD	0.001	0.794	3.071	0.000	54.542	10	0.025	0.816	2.163	17.200	27.801	12	0.023	0.666	2.344	31.000	50.992	8
	BLASSO	0.001	0.791	9.077	0.000	80.523	22	0.012	0.847	10.163	0.700	64.842	29	0.014	0.705	7.067	13.200	72.303	18
	pMOM	0.001	0.798	0.025	0.000	0.976	4	0.003	0.857	0.038	3.273	0.674	10	0.035	0.589	0.029	81.143	5.714	1
	piMOM	0.001	0.796	0.026	0.250	1.427	4	0.008	0.848	0.137	11.200	2.605	9	0.003	0.734	0.031	18.000	1.315	4
	GD	0.002	0.784	1.510	0.000	32.644	7	0.003	0.858	0.126	0.800	2.188	10	0.008	0.709	2.682	12.400	45.616	10

Table 2 Comparisons of mean squared error (MSE) and concordance index (C-index), false positive rate (FPR), false negative rate (FNR), false discovery rate (FDR), and an average number of selected variables across methods applied to Scenarios 1–3 when $n = 1,250$ and the censoring rate is 50%.

		Scenario 1						Scenario 2						Scenario 3					
p	Method	MSE	CINDEX	FPR	FNR	FDR	NUM.SEL	MSE	CINDEX	FPR	FNR	FDR	NUM.SEL	MSE	CINDEX	FPR	FNR	FDR	NUM.SEL
20	ORACLE	0.001	0.827	0.000	0.000	0.000	4	0.072	0.918	0.000	0.000	0.000	10	0.001	0.778	0.000	0.000	0.000	5
	RIDGE	0.004	0.825	100.000	0.000	80.000	20	0.027	0.918	100.000	0.000	50.000	20	0.027	0.773	100.000	0.000	75.000	20
	ENET	0.002	0.826	77.188	0.000	75.148	16	0.004	0.918	90.300	0.000	47.231	19	0.006	0.776	84.667	0.000	71.496	18
	LASSO	0.002	0.826	54.125	0.000	65.930	13	0.004	0.919	76.700	0.000	42.899	18	0.005	0.776	61.600	0.000	63.433	14
	MCP	0.001	0.827	3.062	0.000	7.695	4	0.002	0.919	5.700	0.000	3.849	11	0.002	0.777	3.000	0.000	4.963	5
	SCAD	0.001	0.827	4.750	0.000	9.376	5	0.002	0.919	7.100	0.000	4.807	11	0.002	0.777	5.400	0.000	8.292	6
	DLASSO	0.001	0.829	100.000	0.000	80.000	20	0.098	0.875	100.000	0.000	50.000	20	0.352	0.764	100.000	0.000	75.000	20
	BLASSO	0.001	0.827	58.312	0.000	68.176	13	0.003	0.919	85.600	0.000	45.837	19	0.003	0.776	64.133	0.000	64.727	15
	piMOM	0.001	0.827	0.000	0.000	0.000	4	0.002	0.919	0.000	0.000	0.000	10	0.002	0.776	0.116	0.000	0.451	5
	GD	0.001	0.827	0.438	0.000	1.400	4	0.003	0.919	0.000	0.000	0.000	10	0.002	0.777	0.133	0.000	0.333	5
100	ORACLE	0.000	0.829	0.000	0.000	0.000	4	0.016	0.918	0.000	0.000	0.000	10	0.000	0.779	0.000	0.000	0.000	5
	RIDGE	0.003	0.815	100.000	0.000	96.000	100	0.007	0.913	100.000	0.000	90.000	100	0.008	0.758	100.000	0.000	95.000	100
	ENET	0.002	0.825	40.021	0.000	90.312	42	0.004	0.917	57.994	0.000	83.738	62	0.007	0.771	50.505	0.000	90.365	53
	LASSO	0.001	0.827	23.969	0.000	84.007	27	0.003	0.918	40.779	0.000	78.111	47	0.004	0.776	27.705	0.000	82.956	31
	MCP	0.000	0.829	1.135	0.000	13.195	5	0.000	0.919	0.736	0.000	4.801	11	0.000	0.779	0.505	0.000	6.430	5
	SCAD	0.000	0.829	1.604	0.000	16.648	6	0.000	0.919	0.606	0.000	3.994	11	0.000	0.779	1.368	0.000	14.738	6
	DLASSO	0.002	0.817	100.000	0.000	96.000	100	0.023	0.865	100.000	0.000	90.000	100	0.075	0.764	100.000	0.000	95.000	100
	BLASSO	0.000	0.827	25.385	0.000	85.011	28	0.001	0.918	43.636	0.000	79.351	49	0.001	0.777	29.337	0.000	83.876	33
	piMOM	0.000	0.828	0.000	0.000	0.000	4	0.001	0.918	0.013	0.000	0.094	10	0.000	0.779	0.134	0.000	1.015	5
	GD	0.000	0.828	0.479	0.000	8.191	4	0.001	0.919	0.014	0.000	0.118	10	0.000	0.779	0.168	0.000	2.667	5
200	ORACLE	0.000	0.829	0.000	0.000	0.000	4	0.008	0.918	0.000	0.000	0.000	10	0.000	0.778	0.000	0.000	0.000	5
	RIDGE	0.003	0.795	100.000	0.000	98.000	200	0.005	0.906	100.000	0.000	95.000	200	0.015	0.720	100.000	0.000	97.500	200
	ENET	0.001	0.825	26.027	0.000	92.406	55	0.003	0.915	42.325	0.000	88.765	90	0.006	0.766	31.933	0.000	92.309	67
	LASSO	0.001	0.827	15.164	0.000	86.462	34	0.002	0.917	27.455	0.000	83.459	62	0.003	0.773	19.096	0.000	87.216	42
	MCP	0.000	0.829	0.426	0.000	8.728	5	0.000	0.919	0.573	0.000	7.072	11	0.000	0.778	0.130	0.000	3.505	5
	SCAD	0.000	0.829	1.737	0.000	22.931	7	0.000	0.919	1.438	0.000	13.504	13	0.000	0.778	0.879	0.000	17.590	7
	DLASSO	0.001	0.821	100.000	0.000	98.000	200	0.013	0.860	100.000	0.000	95.000	200	0.029	0.750	100.000	0.000	97.500	200
	BLASSO	0.000	0.827	16.559	0.000	87.850	36	0.001	0.917	30.864	0.000	85.215	69	0.000	0.775	20.000	0.000	87.913	44
	piMOM	0.000	0.828	0.000	0.000	0.000	4	0.000	0.919	0.062	0.000	0.838	10	0.000	0.777	0.054	0.000	0.721	5
	GD	0.000	0.829	0.420	0.000	9.261	5	0.000	0.919	0.063	0.000	1.085	10	0.000	0.778	0.138	0.000	4.346	5

Table 3 Comparisons of mean squared error (MSE) and concordance index (C-index), false positive rate (FPR), false negative rate (FNR), false discovery rate (FDR), and an average number of selected variables across methods applied to Scenarios 1–3 when $n = 1,250$ and the censoring rate is 80%.

		Scenario 1						Scenario 2						Scenario 3					
p	Method	MSE	CINDEX	FPR	FNR	FDR	NUM.SEL	MSE	CINDEX	FPR	FNR	FDR	NUM.SEL	MSE	CINDEX	FPR	FNR	FDR	NUM.SEL
20	ORACLE	0.001	0.827	0.000	0.000	0.000	4	0.077	0.915	0.000	0.000	0.000	10	0.003	0.777	0.000	0.000	0.000	5
	RIDGE	0.007	0.822	100.000	0.000	80.000	20	0.033	0.916	100.000	0.000	50.000	20	0.034	0.768	100.000	0.000	75.000	20
	ENET	0.006	0.824	77.125	0.000	75.059	16	0.010	0.917	92.100	0.000	47.825	19	0.013	0.772	85.600	0.000	71.735	18
	LASSO	0.005	0.825	54.812	0.000	66.856	13	0.010	0.917	79.300	0.000	43.831	18	0.013	0.773	62.800	0.000	64.416	14
	MCP	0.002	0.827	3.000	0.000	6.611	4	0.006	0.918	3.800	0.000	2.548	10	0.004	0.776	3.333	0.000	6.175	6
	SCAD	0.002	0.827	4.375	0.000	9.176	5	0.005	0.918	4.300	0.000	2.974	10	0.004	0.776	7.067	0.000	13.510	6
	DLASSO	0.005	0.823	100.000	0.000	80.000	20	0.107	0.870	100.000	0.000	50.000	20	0.013	0.776	100.000	0.000	75.000	20
	BLASSO	0.003	0.825	58.125	0.000	68.582	13	0.009	0.917	84.700	0.000	45.664	18	0.008	0.773	66.000	0.000	65.570	15
	piMOM	0.002	0.829	0.000	0.000	0.000	4	0.006	0.918	0.000	0.000	0.000	10	0.005	0.776	0.533	1.600	1.333	5
	GD	0.002	0.827	0.875	0.000	2.733	4	0.006	0.918	0.000	0.000	0.000	10	0.004	0.776	0.867	1.400	2.042	5
100	ORACLE	0.000	0.829	0.000	0.000	0.000	4	0.015	0.915	0.000	0.000	0.000	10	0.001	0.779	0.000	0.000	0.000	5
	RIDGE	0.007	0.790	100.000	0.000	96.000	100	0.012	0.902	100.000	0.000	90.000	100	0.028	0.719	100.000	0.000	95.000	100
	ENET	0.004	0.819	38.115	0.000	89.608	41	0.009	0.910	55.711	0.000	83.121	60	0.017	0.756	44.232	0.000	89.036	47
	LASSO	0.003	0.824	23.031	0.000	82.837	26	0.007	0.913	39.844	0.000	77.587	46	0.010	0.768	27.484	0.000	83.019	31
	MCP	0.000	0.829	0.688	0.000	7.479	5	0.001	0.917	0.522	0.000	3.261	10	0.001	0.778	0.411	0.000	4.909	5
	SCAD	0.000	0.829	1.927	0.000	14.263	6	0.001	0.917	1.456	0.000	7.652	11	0.001	0.777	1.863	0.000	18.844	7
	DLASSO	0.005	0.803	100.000	0.000	96.000	100	0.028	0.857	100.000	0.000	90.000	100	0.086	0.755	100.000	0.000	95.000	100
	BLASSO	0.001	0.826	24.104	0.000	83.863	27	0.002	0.914	41.756	0.000	78.480	48	0.002	0.772	28.621	0.000	83.527	32
	piMOM	0.001	0.829	0.000	0.000	0.000	4	0.001	0.918	0.011	0.000	0.091	10	0.001	0.779	0.032	0.600	0.500	5
	GD	0.001	0.827	0.792	0.000	11.930	5	0.001	0.917	0.078	0.000	0.636	10	0.001	0.776	0.621	2.200	8.763	5
200	ORACLE	0.000	0.829	0.000	0.000	0.000	4	0.008	0.916	0.000	0.000	0.000	10	0.000	0.779	0.000	0.000	0.000	5
	RIDGE	0.009	0.737	100.000	0.000	98.000	200	0.014	0.880	100.000	0.000	95.000	200	0.029	0.657	100.000	0.000	97.500	200
	ENET	0.003	0.816	24.719	0.000	91.777	52	0.007	0.908	39.062	0.000	87.858	84	0.013	0.747	27.055	0.000	90.920	58
	LASSO	0.002	0.822	15.730	0.000	87.126	35	0.005	0.912	26.941	0.000	83.151	61	0.007	0.762	18.073	0.000	86.582	40
	MCP	0.000	0.829	0.622	0.000	12.401	5	0.000	0.919	0.740	0.000	7.932	11	0.000	0.777	0.351	0.211	7.829	6
	SCAD	0.000	0.829	1.378	0.000	20.283	7	0.000	0.919	1.442	0.000	13.799	13	0.000	0.777	1.301	0.000	25.579	8
	DLASSO	0.005	0.778	100.000	0.000	98.000	200	0.012	0.915	100.000	0.000	95.000	200	0.041	0.764	100.000	0.000	97.500	200
	BLASSO	0.000	0.825	16.622	0.000	88.071	37	0.001	0.915	29.002	0.000	84.184	65	0.001	0.771	18.829	0.000	87.160	42
	piMOM	0.000	0.829	0.000	0.000	0.000	4	0.001	0.919	0.000	0.000	0.000	10	0.001	0.770	0.876	0.000	2.014	6
	GD	0.000	0.826	1.122	0.000	28.832	6	0.001	0.918	0.208	0.000	3.289	10	0.001	0.771	0.934	3.789	20.762	7

Table 4 Comparison of the Average Computing Time (Seconds) to Attain MAP Estimates for Various Bayesian Methods when $n = 100, 1, 250$.

Method	n = 100			n = 1250		
	Scenario 1	Scenario 2	Scenario 3	Scenario 1	Scenario 2	Scenario 3
BLASSO	9	13	13	78	79	80
pMOM	8	10	8	65	67	70
piMOM	6	9	8	53	59	64
GD	7	12	11	56	59	66

mass index (kg/m^2), oxygen saturation, body temperature, respiratory rate, diastolic and systolic blood pressure, and heart rate. In total, $p = 51$ demographic and clinical predictors were considered candidate features.

We first applied our method to the entire dataset to present a summary of a posterior distribution. As in simulations, the hyperparameter κ of the GD prior was selected over $[0.1, 3]$. For selection and prediction performance measures, we used 10-fold cross-validation by randomly partitioning the data into training and test sets, using the training set to fit the model and the testing set to evaluate the model performance. Observations in each fold were randomly chosen to maintain a similar censoring rate across folds. In each training set, we generated 3,000 times posterior samples after discarding 500 burn-in samples. By using 10-fold cross-validation, the average C-index and integrated Brier score [IBS; 39] across folds were measured for prediction performance; in addition, we defined the selection probability of each feature using the empirical percentage of selecting this feature across folds.

Table 6 provides the MAP estimates and overall posterior summaries of the regression coefficients of the GD prior approach, compared with the DLASSO method; the other competing methods did not draw an inference and, therefore, are not reported in this table. The predictors chosen by the GD prior approach are listed in the top panel above the separation line. The Bayesian point estimates agreed with the DLASSO estimates; however, the GD prior approach was able to select variables and, in the meantime, quantify the uncertainty for all the variables, whereas DLASSO did not select variables. That is, fundamentally different from the DLASSO method, our proposed GD prior method performed variable selection and inference simultaneously.

The covariates selected by the GD prior method included age, blood loss anemia, diastolic blood pressure (non-invasive), an indicator of fluid electrolyte disorders, heart rate, oxygen saturation, psychoses, respiratory rate, and solid cancer with and without metastasis. Of the non-selected variables, the 95% HPD intervals obtained from the proposed method were narrow, containing zeroes. We found that age (0.83; 0.70 to 0.98), blood loss anemia (0.30; 0.04 to 0.58), fluid electrolyte disorders (1.40; 0.82 to 1.97), heart rate (0.23; 0.10 to 0.33), psychoses (0.3; 0.08 to 0.55), respiratory rate (0.21; 0.11 to 0.34), and solid tumor without metastasis (0.33; 0.08 to 0.56) were significant with the 95% HPD intervals not containing 0, and were associated with increased risk of in-hospital death. On the other hand, high oxygen saturation (-0.41; -0.52 to -0.29) and high diastolic blood pressure (-0.13; -0.22 to -0.03) were associated with improved survival. Figure 3 illustrates how the tuning parameter works internally

Table 5 Over 100 repetitions with $n = 100$ and $p = 20$, comparisons of MAP estimates obtained by the GD prior approach with the estimates by the MCP, BLASSO, and DLASSO methods. All noise variables with true values of 0 are averaged and grouped as "Noise."

	COEF	TRUE	MCP		BLASSO		DLASSO		GD	
			MEAN(SE)	MSE	MEAN(SE)	MSE	MEAN(SE)	MSE	MEAN(SE)	MSE
Scenario 1	β_1	1	1.03(0.16)	0.03	0.95(0.15)	0.03	1.02(0.18)	0.03	0.93(0.15)	0.03
	β_2	1	1.05(0.16)	0.03	0.96(0.16)	0.03	1.03(0.2)	0.04	0.96(0.16)	0.03
	β_5	1	1.06(0.15)	0.03	0.97(0.15)	0.02	1.04(0.17)	0.03	0.96(0.14)	0.02
	β_6	1	1.04(0.2)	0.04	0.95(0.18)	0.04	1.02(0.2)	0.04	0.94(0.17)	0.03
	Noise	0	0(0.01)	0.00	0(0.01)	0.00	0(0.03)	0.00	0(0.01)	0.00
Scenario 2	β_1	1	1.13(0.25)	0.08	0.88(0.19)	0.05	0.96(0.25)	0.06	0.92(0.19)	0.04
	β_2	1	1.15(0.32)	0.12	0.91(0.22)	0.05	1.02(0.31)	0.09	0.97(0.26)	0.07
	β_3	1	1.11(0.29)	0.10	0.9(0.2)	0.05	0.99(0.27)	0.07	0.93(0.23)	0.06
	β_4	1	1.14(0.3)	0.11	0.9(0.21)	0.05	1(0.28)	0.08	0.95(0.24)	0.06
	β_5	1	1.12(0.32)	0.12	0.88(0.21)	0.06	0.98(0.27)	0.07	0.94(0.21)	0.05
	β_{11}	1	1.13(0.31)	0.11	0.88(0.22)	0.06	0.97(0.29)	0.08	0.94(0.24)	0.06
	β_{12}	1	1.1(0.32)	0.11	0.89(0.22)	0.06	1.01(0.3)	0.09	0.93(0.24)	0.06
	β_{13}	1	1.11(0.31)	0.11	0.89(0.19)	0.05	0.99(0.27)	0.07	0.94(0.23)	0.05
	β_{14}	1	1.14(0.33)	0.13	0.91(0.22)	0.06	1.01(0.27)	0.07	0.96(0.23)	0.06
	β_{15}	1	1.17(0.3)	0.12	0.91(0.2)	0.05	1(0.29)	0.08	0.96(0.23)	0.05
	Noise	0	0(0.02)	0.00	0(0.03)	0.00	0(0.03)	0.00	0(0.01)	0.00
Scenario 3	β_1	1	1.05(0.2)	0.04	0.87(0.19)	0.05	0.9(0.18)	0.04	0.95(0.17)	0.03
	β_2	1	1.07(0.22)	0.05	0.87(0.21)	0.06	0.89(0.23)	0.07	0.96(0.22)	0.05
	β_3	1	1.04(0.2)	0.04	0.86(0.19)	0.06	0.89(0.2)	0.05	0.94(0.19)	0.04
	β_4	-2.121	-2.25(0.31)	0.11	-1.74(0.33)	0.25	-1.83(0.39)	0.24	-1.98(0.27)	0.09
	β_5	0.333	0.26(0.18)	0.04	0.26(0.12)	0.02	0.29(0.13)	0.02	0.21(0.21)	0.06
	Noise	0	0(0.01)	0.00	-0.01(0.02)	0.00	-0.01(0.03)	0.00	0(0.01)	0.00

Table 6 For COVID-19 patients admitted to ICU, the MAP estimates, posterior means, and 95% highest posterior density intervals obtained by the GD prior approach, along with the results obtained by the DLASSO method.

Variable	DLASSO		GD			piMOM
	MEAN(SE)	INT[a]	MAP	MEAN (SD)	INT[b]	COEF
Age	0.84(0.02)	(0.81,0.87)	0.83	0.84(0.07)	(0.7,0.98)	0.89
Blood Loss Anemia	0.22(0.02)	(0.18,0.27)	0.31	0.30(0.14)	(0.04,0.58)	0.30
Diastolic Blood Pressure (Non-Invasive)	-0.11(0.01)	(-0.13,-0.08)	-0.14	-0.13(0.06)	(-0.22,0.03)	-0.13
Fluid Electrolyte Disorders	0.92(0.05)	(0.81,1.02)	1.4	1.4(0.29)	(0.82,1.97)	1.12
Heart Rate	0.19(0.01)	(0.17,0.22)	0.23	0.22(0.06)	(0.1,0.33)	0.22
Oxygen Saturation (SpO2)	-0.39(0.01)	(-0.41,-0.37)	-0.41	-0.41(0.06)	(-0.52,-0.29)	-0.43
Psychoses	0.19(0.03)	(0.14,0.24)	0.3	0.3(0.13)	(0.08,0.55)	0.29
Respiratory Rate	0.22(0.01)	(0.2,0.25)	0.2	0.21(0.06)	(0.11,0.34)	0.23
Solid Tumor Without Metastasis	0.23(0.02)	(0.19,0.28)	0.33	0.33(0.12)	(0.08,0.56)	0.33
Gender Male	0.09(0.02)	(0.04,0.14)	0	0(3.5e-06)	(-6.9e-06,6.5e-06)	0.16
Race Black	1.20e-06(0.03)	(-0.06,0.06)	0	0(5.9e-06)	(-1.11e-05,1.16e-05)	0
Race Asian	2.37e-06(0.07)	(-0.14,0.14)	0	0(1.81e-05)	(-3.38e-05,3.49e-05)	0
Race Other	0.13(0.05)	(0.03,0.22)	0	0(1.38e-05)	(-2.72e-05,2.56e-05)	0.81
Ethnicity Hispanic or Latino	0.01(0.06)	(-0.11,0.13)	0	0(1.5e-05)	(-3.21e-05,2.55e-05)	0
Alcohol Abuse	-0.01(0.03)	(-0.06,0.05)	0	0(7.2e-06)	(-1.44e-05,1.32e-05)	-0.25
Cardiac Arrhythmias	0.08(0.03)	(0.01,0.14)	0	03.2e-06)	(-5.8e-06,6.4e-06)	0.14
Chronic Pulmonary Disease	3.30e-04(0.02)	(-0.05,0.05)	0	0(4.2e-06)	(-8.9e-06,7.4e-06)	0
Coagulopathy	0.15(0.02)	(0.1,0.2)	0	0(4.6e-06)	(-9.1e-06,8.9e-06)	0.18
Congestive Heart Failure	0(0.02)	(-0.05,0.05)	0	0(4.4e-06)	(-7.5e-06,8.6e-06)	
Deficiency Anemia	-0.05(0.02)	(-0.1,0)	0	0(5.1e-06)	(-1.03e-05,9.4e-06)	-0.20
Depression	7.50e-04(0.02)	(-0.05,0.05)	0	0(4.2e-06)	(-7.9e-06,8.2e-06)	0
Diabetes Complicated	4.30e-05(0.02)	(-0.05,0.04)	0	0(4.4e-06)	(-9.4e-06,7.4e-06)	-0.07
Diabetes Uncomplicated	5.80e-06(0.02)	(-0.05,0.05)	0	0(4.3e-06)	(-7.7e-06,8.7e-06)	0
Drug Abuse	2.10e-05(0.03)	(-0.06,0.05)	0	0(6.6e-06)	(-1.25e-05,1.33e-05)	0
Hypertension Complicated	1.60e-05(0.02)	(-0.05,0.05)	0	0(4.1e-06)	(-7.2e-06,8.6e-06)	0
Hypertension Uncomplicated	-0.01(0.03)	(-0.07,0.05)	0	0(3.3e-06)	(-6e-06,7.2e-06)	-0.14
Hypothyroidism	0.06(0.02)	(0.02,0.11)	0	0(5.7e-06)	(-9.9e-06,1.18e-05)	0.13
Liver Disease	0.10(0.02)	(0.05,0.15)	0	0(5.2e-06)	(-1.02e-05,9.7e-06)	0
Lymphoma	0.01(0.03)	(-0.05,0.06)	0	0(8.8e-06)	(-1.58e-05,1.85e-05)	0
Metastatic Cancer	0.08(0.02)	(0.03,0.12)	0	0(5.9e-06)	(-1.26e-05,1.07e-05)	0.07
Obesity	9.30e-04(0.02)	(-0.05,0.05)	0	0(4.1e-06)	(-7.9e-06,8.3e-06)	0
Other Neurological Disorders	0.15(0.02)	(0.1,0.2)	0	0(4.8e-06)	(-9.3e-06,9.5e-06)	0
Paralysis	5.60e-04(0.03)	(-0.06,0.06)	0	0(7.2e-06)	(-1.39e-05,1.36e-05)	0
Peptic Ulcer Disease Excluding Bleeding	2.32e-04(0.03)	(-0.06,0.05)	0	0(7.9e-06)	(-1.43e-05,1.53e-05)	0
Peripheral Vascular Disorders	0.08(0.02)	(0.03,0.13)	0	0(4.6e-06)	(-8.4e-06,9.3e-06)	-0.02
Pulmonary Circulation Disorders	9.52e-04(0.02)	(-0.05,0.05)	0	0(5e-06)	(-9.5e-06,1.03e-05)	0
Renal Failure	3.74e-04(0.02)	(-0.05,0.05)	0	0(4e-06)	(-8e-06,7.6e-06)	0
Rheumatoid Arthritis Collagen Vascular Diseases	-0.01(0.03)	(-0.07,0.04)	0	0(6.6e-06)	(-1.18e-05,1.28e-05)	0
Valvular Disease	-0.12(0.02)	(-0.17,-0.08)	0	0(5e-06)	(-1.1e-05,8.4e-06)	0
Weight Loss	2.31e-04(0.02)	(-0.05,0.05)	0	0(4.7e-06)	(-8.4e-06,9.2e-06)	0
Total Score	0.01(0.01)	(-0.01,0.04)	0	0(2.8e-06)	(-5.7e-06,5.4e-06)	0
BMI	-0.06(0.01)	(-0.09,-0.03)	0	0(2.6e-06)	(-5.3e-06,4.8e-06)	0
Temperature	0.08(0.01)	(0.06,0.1)	0	0(2.8e-06)	(-5.2e-06,5.7e-06)	0.10
Mean Blood Pressure (Non-Invasive)	0.11(0.01)	(0.1,0.12)	0	0(2.9e-06)	(-5.6e-06,5.8e-06)	0
Systolic Blood Pressure (Non-Invasive)	-0.1(0.01)	(-0.13,-0.08)	0	0(2.9e-06)	(-5.5e-06,5.6e-06)	0
Affluence	1.36e-05(0.01)	(-0.02,0.03)	0	0(2.8e-06)	(-5.3e-06,5.6e-06)	0.01
Socially Disadvantage	2.81e-05(0.01)	(-0.03,0.02)	0	0(2.8e-06)	(-5e-06,5.9e-06)	0
Immigrant	-0.02(0.01)	(-0.04,0.01)	0	0(2.8e-06)	(-4.9e-06,5.7e-06)	0
Pediatric	0.06(0.01)	(0.04,0.08)	0	0(2.8e-06)	(-4.9e-06,5.6e-06)	0

[a] 95% confidence interval

[b] 95% highest posterior density interval

Table 7 Comparisons of C-index, IBS and the number of selected variables across different methods applied to the COVID-19 ICU dataset.

Method	C-index	IBS	Average No. of Variables
RIDGE	0.762	0.015	51
ENET	0.765	0.015	34
LASSO	0.758	0.015	25
MCP	0.754	0.016	18
SCAD	0.753	0.016	20
BLASSO	0.759	0.016	30
DLASSO	0.754	0.016	51
pMOM	0.727	0.016	7
piMOM	0.739	0.016	10
GD	0.756	0.016	17

in the algorithm and performs selection. The convergence of the model is verified by the trace plots and autocorrelation plots (Web Figures S5 and S6).

With 10-fold cross-validation, Figure 4 shows the variables sorted by selection probability and, in particular, those with probability over 0.8, which were highlighted in red. A similar set of predictors were chosen with high probability as applied to the entire data. Table 7 presents the average C-index and integrated Brier score for prediction evaluation in the testing datasets and the average number of selected variables in the training datasets. The proposed method had a reasonable C-index and IBS on par with the competing methods, while producing the sparsest model.

Our results agree with various studies on patients admitted to ICU for severe coronavirus illness. For example, studies have found that older age was associated with COVID-19 outcomes [40, 41], lower oxygen saturation and fluid and electrolyte imbalance were associated with higher COVID-19 mortality [42], increase in respiratory rate was associated with a higher hazard of COVID-19 death [43], and pneumonia and acute respiratory distress caused high mortality rates among COVID-19 patients [44]. Moreover, [45] and [46] found that more advanced tumors elevated mortality among patients with COVID-19, and in particular, [47] linked the increase in COVID-19 mortality to lung cancer patients.

To recapitulate, by leveraging an EHR dataset on a large-scale COVID-19 population admitted to ICU, we examined the prognostic utility of demographic and clinical features on predicting in-hospital mortality and obtained results consistent with the literature. Compared to the competing methods that performed either statistical inference or variable selection, our proposed GD prior method performed variable selection and inference simultaneously.

7 Concluding Remarks

We have developed a new Bayesian method for analyzing Electronic Health Record (EHR) data with survival endpoints in response to ongoing concerns about future resurgences of COVID-19 and the need to improve treatment and management of

critically infected individuals. Our findings can inform critical care providers of treatment priorities, empower healthcare stakeholders with effective disease management strategies, and aid health policymakers in optimizing the allocation of medical resources.

From a statistical perspective, our approach provides a broad framework for performing variable selection, model estimation, and inference simultaneously. To control sparsity, we recommend using the marginal likelihood to determine the optimal tuning parameter. Although our method was primarily motivated by the analysis of ICU-admitted COVID-19 patients, it can be applied to other clinical or longer-term care settings. Through simulation studies, we have demonstrated that our Bayesian approach has the dual benefits of fast MCMC convergence and high inferential precision.

Several future research directions are possible. First, we could expand the GD prior approach to address other practical needs, such as extending it to the Accelerated Failure Time (AFT) model widely used for survival analysis. Furthermore, the GD prior could be used for competing risk models or frailty models. We anticipate that the GD prior may outperform rival techniques in sparsity and accuracy as in Cox regression settings. Second, we may integrate longitudinal measures in a time-dependent model to predict patient death using COVID-19 data. Finally, as with most EHR studies, there may be an inherent selection bias among patients seen at Michigan Medicine and subsequently admitted for COVID-19-related complications. Causal inference approaches, combined with the GD prior, could be explored to address both observed and unmeasured confounders.

Acknowledgements. The work is supported by a grant from the NIH. We thank the Editor, the AE, and two referees for their insightful comments and suggestions that have substantially improved the quality and presentation of the submission.

Supplementary information. Tables and Figures referenced in Section 5 are presented. The computing code is available at https://github.com/js0314/bayes-cox.

References

[1] Cox, D.R.: Regression models and life-tables. Journal of the Royal Statistical Society: Series B (Methodological) **34**(2), 187–202 (1972)

[2] Akaike, H.: A new look at the statistical model identification. IEEE Transactions on Automatic Control **19**(6), 716–723 (1974)

[3] Tibshirani, R.: Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Statistical Methodology) **58**(1), 267–288 (1996)

[4] Tibshirani, R.: The lasso method for variable selection in the Cox model. Statistics in Medicine **16**(4), 385–395 (1997)

[5] Gui, J., Li, H.: Penalized Cox regression analysis in the high-dimensional and low-sample size settings, with applications to microarray gene expression data. Bioinformatics **21**(13), 3001–3008 (2005)

[6] Efron, B., Hastie, T., Johnstone, I., Tibshirani, R.: Least angle regression. Annals of Statistics **32**(2), 407–499 (2004)

[7] Fan, J., Li, R.: Variable selection for Cox's proportional hazards model and frailty model. The Annals of Statistics **30**(1), 74–99 (2002)

[8] Zhang, C.-H., *et al.*: Nearly unbiased variable selection under minimax concave penalty. The Annals of Statistics **38**(2), 894–942 (2010)

[9] Zou, H., Hastie, T.: Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology) **67**(2), 301–320 (2005)

[10] Xia, L., Nan, B., Li, Y.: Statistical inference for Cox proportional hazards models with a diverging number of covariates. Scandinavian Journal of Statistics **in press** (2022)

[11] Fang, E.X., Ning, Y., Liu, H.: Testing and confidence intervals for high dimensional proportional hazards models. Journal of the Royal Statistical Society: Series B (Statistical Methodology) **79**(5), 1415–1437 (2017)

[12] Mitchell, T.J., Beauchamp, J.J.: Bayesian variable selection in linear regression. Journal of the American Statistical Association **83**(404), 1023–1032 (1988)

[13] Sha, N., Tadesse, M.G., Vannucci, M.: Bayesian variable selection for the analysis of microarray data with censored outcomes. Bioinformatics **22**(18), 2262–2268 (2006)

[14] Tang, Z., Shen, Y., Zhang, X., Yi, N.: The spike-and-slab lasso cox model for survival prediction and associated genes detection. Bioinformatics **33**(18), 2799–2807 (2017)

[15] Griffin, J.E., Brown, P.J., *et al.*: Inference with normal-gamma prior distributions in regression problems. Bayesian Analysis **5**(1), 171–188 (2010)

[16] Tipping, M.: The relevance vector machine. Advances in neural information processing systems **12** (1999)

[17] Carvalho, C.M., Polson, N.G., Scott, J.G.: Handling sparsity via the horseshoe. In: Artificial Intelligence and Statistics, pp. 73–80 (2009). PMLR

[18] Bhattacharya, A., Pati, D., Pillai, N.S., Dunson, D.B.: Dirichlet-Laplace priors for optimal shrinkage. Journal of the American Statistical Association **110**(512), 1479–1490 (2015)

[19] Ibrahim, J.G., Chen, M.-H., MacEachern, S.N.: Bayesian variable selection for proportional hazards models. Canadian Journal of Statistics **27**(4), 701–717 (1999)

[20] Lee, K.H., Chakraborty, S., Sun, J.: Bayesian variable selection in semiparametric proportional hazards model for high dimensional survival data. The International Journal of Biostatistics **7**(1), 21 (2011)

[21] Nikooienejad, A., Wang, W., Johnson, V.E.: Bayesian variable selection for survival data using inverse moment priors. The Annals of Applied Statistics **14**(2), 809 (2020)

[22] Mu, J., Liu, Q., Kuo, L., Hu, G.: Bayesian variable selection for the Cox regression model with spatially varying coefficients with applications to Louisiana respiratory cancer data. Biometrical Journal **63**(8), 1607–1622 (2021)

[23] Ročková, V., George, E.I.: Emvs: The em approach to bayesian variable selection. Journal of the American Statistical Association **109**(506), 828–846 (2014)

[24] Goh, G., Dey, D.K.: Bayesian MAP estimation using Gaussian and diffused-gamma prior. Canadian Journal of Statistics **46**(3), 399–415 (2018)

[25] Prentice, R.L., Kalbfleisch, J.D.: Hazard rate models with covariates. Biometrics **35**(1), 25–39 (1979)

[26] Cox, D.R.: Partial likelihood. Biometrika **62**(2), 269–276 (1975)

[27] Breslow, N.: Covariance analysis of censored survival data. Biometrics, 89–99 (1974)

[28] Hertz-Picciotto, I., Rockhill, B.: Validity and efficiency of approximation methods for tied survival times in cox regression. Biometrics, 1151–1156 (1997)

[29] DeLong, D.M., Guirguis, G.H., So, Y.C.: Efficient computation of subset selection probabilities with application to cox regression. Biometrika **81**(3), 607–611 (1994)

[30] Efron, B.: The efficiency of cox's likelihood function for censored data. Journal of the American statistical Association **72**(359), 557–565 (1977)

[31] Kalbfleisch, J.D., Prentice, R.L.: Marginal likelihoods based on Cox's regression and life model. Biometrika **60**(2), 267–278 (1973)

[32] Ibrahim, J.G., Chen, M.-H., Sinha, D., Ibrahim, J., Chen, M.: Bayesian Survival Analysis vol. 2. Springer, Berlin (2001)

[33] Yi, N., Tang, Z., Zhang, X., Guo, B.: Bhglm: Bayesian hierarchical GLMs and survival models, with applications to genomics and epidemiology. Bioinformatics **35**(8), 1419–1421 (2019)

[34] Friedman, J., Hastie, T., Tibshirani, R.: Regularization paths for generalized linear models via coordinate descent. Journal of statistical software **33**(1), 1 (2010)

[35] Simon, N., Friedman, J., Hastie, T., Tibshirani, R.: Regularization paths for cox's proportional hazards model via coordinate descent. Journal of statistical software **39**(5), 1 (2011)

[36] Breheny, P., Huang, J.: Coordinate descent algorithms for nonconvex penalized regression, with applications to biological feature selection. The annals of applied statistics **5**(1), 232 (2011)

[37] Fan, J., Feng, Y., Wu, Y.: High-dimensional variable selection for cox's proportional hazards model. In: Borrowing Strength: Theory Powering Applications–a Festschrift for Lawrence D. Brown vol. 6, pp. 70–87. Institute of Mathematical Statistics, New York (2010)

[38] Harrell, F.E., Califf, R.M., Pryor, D.B., Lee, K.L., Rosati, R.A.: Evaluating the yield of medical tests. JAMA **247**(18), 2543–2546 (1982)

[39] Brier, G.W., *et al.*: Verification of forecasts expressed in terms of probability. Monthly Weather Review **78**(1), 1–3 (1950)

[40] Wu, Z., McGoogan, J.M.: Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention. JAMA **323**(13), 1239–1242 (2020)

[41] Shahid, Z., Kalayanamitra, R., McClafferty, B., Kepko, D., Ramgobin, D., Patel, R., Aggarwal, C.S., Vunnam, R., Sahu, N., Bhatt, D., *et al.*: COVID-19 and older adults: what we know. Journal of the American Geriatrics Society **68**(5), 926–929 (2020)

[42] Nahkuri, S., Becker, T., Schueller, V., Massberg, S., Bauer-Mehren, A.: Prior fluid and electrolyte imbalance is associated with COVID-19 mortality. Communications Medicine **1**(1), 1–10 (2021)

[43] Miller, D.J., Capodilupo, J.V., Lastella, M., Sargent, C., Roach, G.D., Lee, V.H., Capodilupo, E.R.: Analyzing changes in respiratory rate to predict the risk of COVID-19 infection. PloS one **15**(12), 0243693 (2020)

[44] Fan, B.E., Ong, K.H., Chan, S.S.W., Young, B.E., Chong, V.C.L., Chen, S.P.C., Lim, S.P., Lim, G.P., Kuperan, P.: Blood and blood product use during COVID-19 infection. American Journal of Hematology **95**(7), 158–160 (2020)

[45] Mohiuddin, M., Kasahara, K.: Cancer metastasis may increase COVID-19 mortality: Suitable targets required to impede cancer metastasis. Journal of Infection

and Public Health **15**(2), 153 (2022)

[46] Liu, Y., Lu, H., Wang, W., Liu, Q., Zhu, C.: Clinical risk factors for mortality in patients with cancer and COVID-19: a systematic review and meta-analysis of recent observational studies. Expert Review of Anticancer Therapy **21**(1), 107–119 (2021)

[47] Lei, H., Yang, Y., Zhou, W., Zhang, M., Shen, Y., Tao, D., Wang, L., Lei, Q., Wang, Y., Wu, Y.: Higher mortality in lung cancer patients with COVID-19? a systematic review and meta-analysis. Lung Cancer **157**, 60–65 (2021)

Fig. 2 MC coverage probability for Simulation Scenario 1 with $p = 100$ (top) and $p = 200$ (bottom). The legend True represents 1 if the variable is nonzero or 0 if the variable is zero. In addition, the x-axis indicates each covariate.

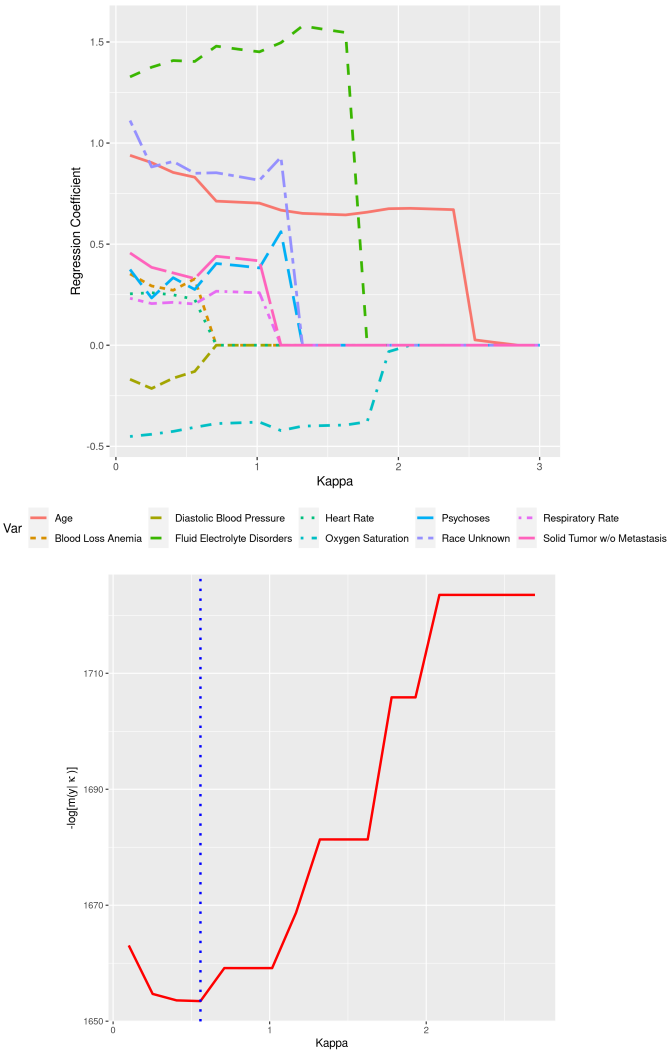

Fig. 3 Solution path (top) and marginal likelihood estimate (bottom) for selecting tuning parameter κ in the COVID-19 data.

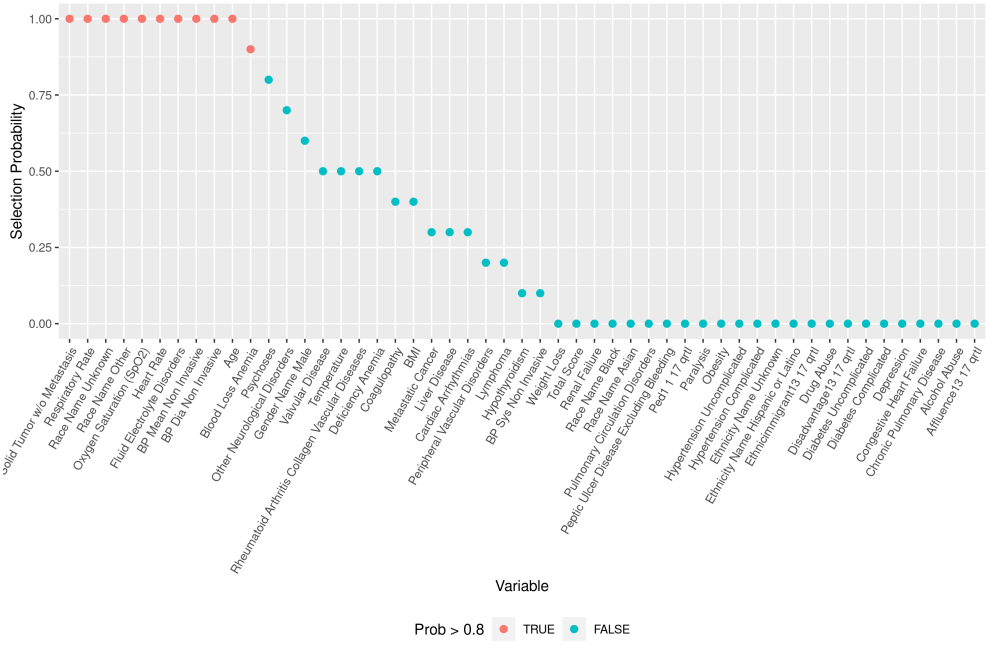

Fig. 4 Selection probability of covariates calculated by 10-fold cross-validation. Selection probabilities greater than 0.8 are flagged in red.