HHS Public Access
Author manuscript
J Stat Theory Pract. Author manuscript; available in PMC 2018 May 23.

Published in final edited form as:
J Stat Theory Pract. 2018 ; 12(1): 23–41. doi:10.1080/15598608.2017.1299058.

Assessing covariate effects using Jeffreys-type prior in the Cox model in the presence of a monotone partial likelihood

Jing Wu[a], **Mário de Castro**[b], **Elizabeth D. Schifano**[a], and **Ming-Hui Chen**[a]

[a]Department of Statistics, University of Connecticut, Storrs, Connecticut, USA

[b]Universidade de São Paulo, Instituto de Ciências Matemáticas e de Computação, São Carlos, SP, Brazil

Abstract

In medical studies, the monotone partial likelihood is frequently encountered in the analysis of time-to-event data using the Cox model. For example, with a binary covariate, the subjects can be classified into two groups. If the event of interest does not occur (zero event) for all the subjects in one of the groups, the resulting partial likelihood is monotone and consequently, the covariate effects are difficult to estimate. In this article, we develop both Bayesian and frequentist approaches using a data-dependent Jeffreys-type prior to handle the monotone partial likelihood problem. We first carry out an in-depth examination of the conditions of the monotone partial likelihood and then characterize sufficient and necessary conditions for the propriety of the Jeffreys-type prior. We further study several theoretical properties of the Jeffreys-type prior for the Cox model. In addition, we propose two variations of the Jeffreys-type prior: the shifted Jeffreys-type prior and the Jeffreys-type prior based on the first risk set. An efficient Markov-chain Monte Carlo algorithm is developed to carry out posterior computation. We perform extensive simulations to examine the performance of parameter estimates and demonstrate the applicability of the proposed method by analyzing real data from the SEER prostate cancer study.

Keywords

Bayesian estimates; cause-specific hazards model; first risk set; penalized maximum likelihood; shifted Jeffreys-type prior; zero events

AMS SUBJECT CLASSIFICATION

62N02

1. Introduction

In medical studies involving time-to-event data, it is often the case that patients within at least one arm of the study will experience very few events. This could be due to the length of

CONTACT: Ming-Hui Chen, ming-hui.chen@uconn.edu, Department of Statistics, University of Connecticut, 215 Genbrook Road, U-4120, Storrs, CT 06269, USA.

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/ujsp.

the study, or due to the nature of the study itself. For instance, among the particular subset of SEER prostate cancer data we are interested in, no patients receive surgery treatment. If one wishes to analyze the surgery treatment effect of the time-to-event data in our motivating prostate cancer study, then the zero event in surgery treatment group will lead to the model identifiability issue. Such an identifiability issue is a well-known problem in the analysis of time-to-event data using the semiparametric proportional hazards model of Cox (1972). Standard analysis of the Cox proportional hazards model involves parameter estimation through maximization of the logarithm of the partial likelihood function. However, it is not uncommon for the partial likelihood to converge to a finite value while at least one parameter estimate goes to $-\infty$ or $+\infty$. This phenomenon is known as the monotone likelihood problem. Bryson and Johnson (1981) state that when estimating covariate parameters for the Cox proportional hazards model, there is a nonzero probability for any finite sample that the maximum partial likelihood estimate will be infinite. Heinze and Ploner (2002) further remark that the probability of monotone likelihood is “too high to be negligible,” thus necessitating solutions to the monotone likelihood problem. In the example of the SEER prostate cancer study, one might consider removing the surgery treatment covariate to eliminate the monotone likelihood problem. However, this may not be desirable since the surgery treatment effect is of great clinical interest. Based on a procedure by Firth (1993), Heinze and Schemper (2001) proposed a solution to the monotone partial likelihood problem by means of penalized maximum likelihood estimation. For a recent discussion on the role of penalization in logistic and survival regression, see Greenland and Mansournia (2015). However, many issues are still not well understood and need to be further studied from both practical and theoretical points of view. Although we mainly focus on the Cox proportional hazards model, the monotone likelihood problem may also exist under alternative models such as accelerated failure time models (Kalbfleisch and Prentice 2011), parametric hazards models, and so on.

In this article, we first examine the conditions that lead to the monotone partial likelihood problem and establish easy-to-check sufficient conditions for the survival data with binary baseline covariates. We then characterize the sufficient and necessary conditions for the existence and propriety of a data-dependent Jeffreys-type prior. Our conditions are certainly much weaker than those of Heinze and Schemper (2001, 116). In addition, we show that the Jeffreys-type prior has finite modes and thus the maximum partial likelihood estimate (MPLE) exists. We also compare the tail behavior between the Jeffreys-type prior and the multivariate *t* and normal distributions. We further develop two variations of the Jeffreys-type prior: the shifted Jeffreys-type prior and the Jeffreys-type prior based on the first risk set. The shifted Jeffreys-type prior leads to less biased estimates (when compared to MPLE). The Jeffreys-type prior based on the first risk set has the same propriety conditions as and similar parameter estimates to the Jeffreys-type prior based on the whole data set. As demonstrated in both the simulation study and the real data analysis, the Jeffreys-type prior based on the first risk set leads to a substantial decrease in computing time over the prior based on the whole data set, especially when the sample size is large. The approach for constructing Jeffreys-type prior based on the first risk set is not only useful for survival data but also applicable for other types of data such as count data with excessive zeros and missing data. Finally, we propose a computationally less expensive but efficient localized

Metropolis algorithm, which avoids direct computation of the second derivative of the Jeffreys-type prior.

This article unfolds as follows. Section 2 presents a motivating case study where the MPLE is not identifiable. In section 3, we provide an in-depth investigation and characterization of monotone partial likelihoods as well as posterior propriety under improper uniform priors. In section 4, we obtain sufficient and necessary conditions for the propriety of the Jeffreys-type prior and propose the two variations of the Jeffreys-type prior. The computational development involving the localized Metropolis algorithm is given in section 5. An extensive simulation is carried out in section 6. Section 7 presents a detailed analysis of the motivating SEER prostate cancer data. We conclude the article with some discussion in section 8. Proofs are given in the appendix.

2. A motivating prostate cancer case study

We consider 1840 men who were subjects in the SEER prostate cancer data between 1973 to 2013, and who have all of the three intermediate risk factors: clinical tumor stage is T2b or T2c, Gleason score equals 7, and prostate-specific antigen (PSA) level between 10 and 20 ng/mL. Among those 1840 subjects, the total number of events due to prostate cancer is 8, and the total number of events due to other causes is 63. The covariates considered in our analysis are PSA, surgery treatment indicator (RP), radiation treatment only indicator (RT), African-American indicator (Black), year of diagnosed (Year_diag), and age (Age). The covariates RP, RT, and Black are binary covariates, taking value 0 or 1. We fit the cause-specific hazards model described in section 3.1 to the SEER prostate cancer data, in which there are two causes of death, namely, prostate cancer and other causes. The resulting maximum partial likelihood estimates (MPLEs) are shown in Table 1. We see from Table 1 that for RP, the MPLE (Est) and the standard error (SE) were −17.745 and 1680, respectively, for death due to prostate cancer. These results indicate that RP is not identifiable for the death caused by prostate cancer, which is due to the absence of events (prostate cancer death) in the “surgery treatment” group of patients. This case study motivates us to carry out further examination of monotone partial likelihoods. Note that in Table 1, N_c is the number of censored, N_{pc} the number of prostate cancer deaths, and N_{oc} the number of other cause deaths.

3. Monotone partial likelihood and posterior propriety

In this section, we first introduce the Cox model and the cause-specific hazards model. We then provide the conditions under which the maximum partial likelihood estimator of the regression coefficients exists. Two examples are given to demonstrate the necessity of those two conditions.

3.1. The Cox proportional hazards model and the cause-specific hazards model

Let y_i denote the minimum of the censoring time C_i and the survival time T_i, and let $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip})'$ be the $p \times 1$ vector of covariates associated with y_i for the ith subject. Denote by $\beta = (\beta_1, \ldots, \beta_p)'$ the $p \times 1$ vector of regression coefficients. Also, $\delta_i = 1\{T_i = y_i\}$ is the failure indicator for $i = 1, \ldots, n$, where n is the total number of observations and $\mathcal{R}(t) = \{i : y_i$

$\geq t\}$ is the set of subjects at risk at time t. Then the partial likelihood of Cox (1975) is given by

$$L_p(\beta \mid \mathscr{D}_{\text{obs}}) = \prod_{i=1}^{n} \left\{ \frac{\exp\left(\boldsymbol{x}'_i \beta\right)}{\sum_{j \in \mathscr{R}(y_i)} \exp\left(\boldsymbol{x}'_j \beta\right)} \right\}^{\delta_i}, \quad (1)$$

where $\mathscr{D}_{\text{obs}}=\{(y_i, \delta_i, \boldsymbol{x}_i) : i = 1, \ldots, n\}$ is the observed right censored data. As usual, we assume throughout this article that β does not include an intercept, since the intercept cannot be estimated in the Cox partial likelihood, and that, given $\boldsymbol{x}_i$, T_i and C_i are independent. Maximization of the partial likelihood function leads to the MPLE of β.

A generalization of the Cox model is the cause-specific hazards model, which was discussed in Gaynor et al. (1993) and Ge and Chen (2012). For $j = 1, \ldots, J$, the cause-specific hazard function for cause j is defined by $h_{Cj}(t) = \lim_{\Delta t \to 0} \Pr(t \leq T < t + \Delta t, \delta = j \mid T \geq t)/\Delta t$. The overall survival function is $S(t) = \Pr(T > t) = \exp\left\{-\sum_{j=1}^{J} \int_0^t h_{Cj}(u)du\right\}$. Let β_j and $h_{Cj0}(t)$ be the vector of regression coefficients without an intercept and the cause-specific baseline hazard function at time t for cause j, respectively. Assume the Cox proportional hazards structure for $h_{Cj}(t)$, that is, $h_{Cj}(t|\boldsymbol{x}) = h_{Cj0}(t) \exp(\boldsymbol{x}'\beta_j)$. The likelihood function is

$$L_C(\beta, \boldsymbol{h}_{C0} \mid \mathscr{D}_{\text{obs}}) = \prod_{j=1}^{J} \prod_{i=1}^{n} \{h_{Cj0}(y_i) \quad \exp(\boldsymbol{x}_{i'}\beta_j)\}^{1\{\delta_i = j\}} \quad \exp\{-H_{Cj0}(y_i) \exp(\boldsymbol{x}_{i'}\beta_j)\},$$

where $\beta = (\beta_1{}', \ldots, \beta_J{}')'$, $\boldsymbol{h}_{C0} = (h_{C10}, \ldots, h_{CJ0})'$, $H_{Cj0}(y_i) = \int_0^{y_i} h_{Cj0}(u)du$ for $j = 1, \ldots, J$ and $1\{\delta_i = j\}$ is the indicator function for j in $\{0, 1, \ldots, J\}$, with 0 denoting a censored observation. Assume there are no ties among the event times. The conditional probability that an individual dies from cause j at time y_i given one death in the risk set $R(y_i)$ is given by

$$\Pr(\text{individual dies at } y_i \text{ due to cause j} \mid \text{one death at } y_i) = 1\{\delta_\text{i} = \text{j}\}$$

$$\frac{\Pr(\text{individual dies at } y_i \text{ due to cause } j \mid \text{survival prior to } y_i)}{\Pr(\textit{one} \text{ death at } y_i \text{ due to cause j} \mid \text{survival prior to } y_i)} = \frac{\left\{h_{Cj0}(y_i) \exp(\boldsymbol{x}'_i\beta_j)\right\}^{1\left\{\delta_i = j\right\}}}{\left\{\sum_{l \in \mathscr{R}(y_i)} h_{cj0}(y_i) \exp(\boldsymbol{x}'_l\beta_j)\right\}^{1\{\delta_i = j\}}}$$

$$= \left\{ \frac{\exp(\boldsymbol{x}'_i\beta_j)}{\sum_{l \in \mathscr{R}(y_i)} \exp(\boldsymbol{x}'_l\beta_j)} \right\}^{1\{\delta_i = j\}}.$$

The partial likelihood function is thus given by multiplying the conditional probabilities over all deaths and causes, resulting in

$$L_p(\beta \mid \mathscr{D}_{\text{obs}}) = \prod_{j=1}^{J} \prod_{i=1}^{n} \left\{ \frac{\exp(\boldsymbol{x}'_i \beta_j)}{\sum_{l \in \mathscr{R}(y_i)} \exp(\boldsymbol{x}'_l \beta_j)} \right\}^{1\{\delta_i = j\}}. \quad (2)$$

3.2. Conditions for the existence of MPLE and posterior propriety

Here we only focus on the conditions of the existence of MPLE for the Cox model in Eq. (1), since the generalization to the cause-specific model is straightforward. Define $\boldsymbol{X}^*$ to be

$$\boldsymbol{X}^* = [\delta_i(\boldsymbol{x}_j - \boldsymbol{x}_i) : j \in \mathscr{R}(y_i),\ i = 1, \ldots, n]'. \quad (3)$$

Let k_i denote the number of subjects in $R(y_i)$ and $K = \sum_{i=1}^{n} k_i$. Then $\boldsymbol{X}^*$ is a $K \times p$ matrix. The necessary and sufficient conditions established in Chen et al. (2006) for propriety of the posterior when an improper uniform prior is assumed for β are given by

$$\begin{array}{ll} \text{C1.} & \text{The matrix } \boldsymbol{X}^* \text{ is of full column rank and} \\ \text{C2.} & \text{There exists a positive vector } \boldsymbol{v} \text{ such that } \boldsymbol{X}^{*'}\boldsymbol{v} = \boldsymbol{0}. \end{array} \quad (4)$$

A positive vector $\boldsymbol{v}$ means that each component of $\boldsymbol{v}$ is positive. Condition C2 can be checked by solving a linear programming problem (Roy and Hobert 2007, Appendix A).

Under the frequentist point of view, Chen et al. (2009) established that the MPLE of β exists if conditions C1 and C2 are satisfied. Moreover, if C1 is satisfied, then C2 is a necessary condition for the existence of the MPLE for β. We can also consider the identifiability of the MPLE problem from a Bayesian point of view. Kalbfleisch (1978) and Sinha et al. (2003) showed that the partial likelihood in Eq. (1) can be obtained as a limiting case of the marginal posterior of β with continuous time survival data under a gamma process prior for the cumulative baseline hazard function $H_0(\cdot)$ using the likelihood function

$$L(\beta, h_0 \mid \mathscr{D}_{\text{obs}}) = \prod_{i=1}^{n} \{h_0(y_i) \exp(\boldsymbol{x}_{i'}\beta)\}^{\delta_i} \exp\{-\textstyle\sum_{j=1}^{n} H_0(y_j) \exp(\boldsymbol{x}_{j'}\beta)\}.$$

If we treat the partial likelihood $L_p(\beta|\mathscr{D}_{\text{obs}})$ in Eq. (1) as the likelihood function, the posterior distribution for β is then given by

$$\pi(\beta \mid \mathscr{D}_{\text{obs}}) \propto \pi(\beta) L_p(\beta \mid \mathscr{D}_{\text{obs}}) = \pi(\beta) \prod_{i=1}^{n} \left\{ \frac{\exp(\boldsymbol{x}'_i \beta)}{\sum_{j \in \mathscr{R}(y_i)} \exp(\boldsymbol{x}'_j \beta)} \right\}^{\delta_i}, \quad (5)$$

where $\pi(\beta)$ denotes the prior distribution for β. Taking $\pi(\beta) \propto 1$, the existence of the MPLE is thus equivalent to the propriety of the posterior distribution (5). Chen et al. (2006) proved that with $\pi(\beta) \propto 1$, the posterior distribution (5) is proper if and only if C1 and C2 are satisfied. Actually, when C1 and C2 are satisfied for a subset of the data, conditions C1 and C2 in Eq. (4) will automatically be satisfied. Assume $\boldsymbol{X}_s^*$ is a $K_s \times p$ submatrix of $\boldsymbol{X}^*$, where $p < K_s < K$. The conditions for $\boldsymbol{X}_s^*$ are stated as follows:

$$\begin{array}{ll} \mathrm{C1}'. & \text{The matrix } \boldsymbol{X}_s^* \text{ is of full column rank and} \\ \mathrm{C2}'. & \text{There exists a positive vector } \boldsymbol{v} \text{ such that } \boldsymbol{X}_s^{*\prime}\boldsymbol{v} = \boldsymbol{0}. \end{array} \quad (6)$$

Theorem 3.2.1—C1′ and C2′ are sufficient for C1 and C2.

Here we consider a simple example that does not satisfy C1 and C2, and thus the MPLE of β does not exist.

Example 1—Take a data set with $n = 3$ observations, $p = 2$ covariates, $\boldsymbol{x}_1 = (0, 1)'$, $\boldsymbol{x}_2 = (1, 0)'$, $\boldsymbol{x}_3 = (0, 1)'$, $\delta = (1, 0, 0)'$, $y_1 < y_2$, and $y_1 < y_3$. In this case, $\mathcal{R}(y_1) = \{1, 2, 3\}$ and $\boldsymbol{X}^*$ in Eq. (3) has rows (0, 0), (1,−1), and (0, 0), so that rank($\boldsymbol{X}^*$) = 1 and condition C1 in Eq. (4) breaks down. The partial likelihood in Eq. (1) is given by $L_p(\beta|\mathcal{D}_{\text{obs}}) = \exp(\beta_2)/\{\exp(\beta_1) + 2\exp(\beta_2)\}$. The maximum of $L_p(\beta|\mathcal{D}_{\text{obs}})$ is attained when $\beta_1 - \beta_2 \to -\infty$.

3.3. Survival data with binary covariates

Binary covariates are quite common in survival analysis. Therefore, it is of great interest to study the minimum requirement for binary covariates data such that both C1 and C2 hold. Let $y_{(i)}$ denote the rearranged y_i in ascending order associated with $\boldsymbol{x}_{(i)}$ and $\delta_{(i)}$, for $i = 1, \ldots, n$. Denote the index for the first event i_0 as $\min(i \in \{1, \ldots, n\}: \delta_{(i)} = 1)$. We know that $n \geq i_0 + p$ in order to satisfy C1 (see the proof of Theorem 4.3.1 in the appendix). Specifically, for binary covariates data, the minimum sample size required for both C1 and C2 is $n = i_0 + p + 1$ with at least two events. In addition, each covariate in $\boldsymbol{x}_{(i)}$ should not take monotone values across the n observations, for example, 1, …, 1, 0, …, 0.

To be specific, when $p = 1$, we should have at least two events with one in each arm (i.e., $x_{(i)} = 0$ or $x_{(i)} = 1$) and an additional observation after the second event time in which the covariate takes the same value as in the observation corresponding to the first event. Otherwise, $x_{(i)}$ takes monotone values. For example, if for the first two events the values are $x_{(1)} = 1$ and $x_{(2)} = 0$, the following observation (can be either censored or an event) should have $x_{(3)} = x_{(1)} = 1$. Thus, we have $n = i_0 + 2$ and both C1 and C2 hold. Similarly, for $p \geq 2$, we should have at least two events followed by some observations without a monotone pattern for each covariate and $n \geq i_0 + p + 1$. Furthermore, if the number of events is exactly two, the two events should be in two completely opposite arms (i.e., $\boldsymbol{x}$ and $\boldsymbol{1} - \boldsymbol{x}$, where $\boldsymbol{1} = (1, \ldots, 1)'$), as shown in Figures 1a and 1b, where the black vertices refer to the covariates of the two events and the red vertices indicate the covariates of the subsequent observations (can be either censored or not). For example, in Figure 1b, if $\boldsymbol{x}_{(1)} = (0, 0, 0)$ corresponds to an event and there are only two events, the other event must occur in the arm $\boldsymbol{x}_{(2)} = (1, 1, 1)$ in order to satisfy both conditions C1 and C2.

4. Characterization and variation of Jeffreys-type prior

We begin presenting the Jeffreys-type prior in the context of the penalized maximum likelihood estimation method. Based on a procedure by Firth (1993), Heinze and Schemper (2001) proposed a solution to the monotone likelihood problem by means of penalized maximum partial likelihood estimation. Heinze and Ploner (2002) and Ploner and Heinze (2010) developed SAS, SPLUS, and R programs for inference in the Cox model using the penalized likelihood function. The MPLE of β is a solution of the score equations $\boldsymbol{U}(\beta) = \mathbf{0}$, where $\boldsymbol{U}(\beta) = \partial L_p(\beta|\mathscr{D}_{\text{obs}})/\partial\beta$. Heinze and Schemper (2001) suggested a modification of the score function. The estimate $\hat{\beta}^*$ is obtained as the solution of the equations $\mathbf{U}^*(\beta) = \mathbf{0}$, where

$$\boldsymbol{U}^*(\beta) = \frac{\partial}{\partial\beta} L_p(\beta \mid \mathscr{D}_{\text{obs}}) + \frac{1}{2}\text{trace}\left\{\boldsymbol{I}(\beta)^{-1}\frac{\partial}{\partial\beta}\boldsymbol{I}(\beta)\right\}, \quad (7)$$

with $\boldsymbol{I}(\beta)$ denoting the negated Hessian matrix. From Eq. (1) we compute

$$\boldsymbol{I}(\beta) = -\frac{\partial^2}{\partial\beta\partial\beta'}\log\{L_p(\beta \mid \mathscr{D}_{\text{obs}})\} \quad (8)$$

$$= \sum_{i=1}^{n}\delta_i\left\{\sum_{j_i \in \mathscr{R}(y_i)} w_{ij_i}\boldsymbol{x}_{j_i}\boldsymbol{x}_{j_i}{}' - \left(\sum_{j_i \in \mathscr{R}(y_i)} w_{ij_i}\boldsymbol{x}_{j_i}\right)\left(\sum_{m_i \in \mathscr{R}(y_i)} w_{im_i}\boldsymbol{x}_{m_i}\right)'\right\} = \sum_{i=1}^{n}\delta_i\boldsymbol{A}_i,$$

where

$$w_{ij_i} = \exp(\boldsymbol{x}_{j_i}{}'\beta)/\sum_{l \in \mathscr{R}(y_i)}\exp(\boldsymbol{x}_{l'}\beta)$$

and

$$\boldsymbol{A}_i = \boldsymbol{A}_i(\beta) = \sum_{j_i \in \mathscr{R}(y_i)} w_{ij_i}\left\{\sum_{l_i \in \mathscr{R}(y_i)} w_{il_i}(\boldsymbol{x}_{j_i} - \boldsymbol{x}_{l_i})\right\}\left\{\sum_{m_i \in \mathscr{R}(y_i)} w_{im_i}(\boldsymbol{x}_{j_i} - \boldsymbol{x}_{m_i})\right\}'. \quad (9)$$

The modified score function in Eq. (1) arises from the penalized likelihood function $L_p^*(\beta \mid \mathscr{D}_{\text{obs}}) = L_p(\beta \mid \mathscr{D}_{\text{obs}}) \mid \boldsymbol{I}(\beta) \mid^{1/2}$, where $|\cdot|$ denotes determinant. While Heinze and Schemper (2001) did not study the case of the cause-specific hazards model, we may similarly consider a modified score function arising from

$L_p^*(\beta \mid \mathscr{D}_{\text{obs}}) = L_p(\beta \mid \mathscr{D}_{\text{obs}}) \mid \boldsymbol{I}(\beta) \mid^{1/2}$, where from Eq. (2), for $j = 1, \ldots, J$, we compute

$\boldsymbol{I}(\beta_j) = -\frac{\partial^2}{\partial \beta_j \partial \beta_{j'}} \log \{L_p(\beta \mid \mathscr{D}_{\text{obs}})\} = \sum_{i=1}^{n} 1\{\delta_i = j\}\boldsymbol{A}_{ij}$, where

$\boldsymbol{A}_{ij} = \sum_{k_i \in \mathscr{R}(y_i)} w_{ik_ij} \left\{ \sum_{l_i \in R(y_i)} w_{il_ij}(\boldsymbol{x}_{k_i} - \boldsymbol{x}_{l_i}) \right\} \{ \sum_{m_i \in \mathscr{R}(y_i)} w_{im_ij}(\boldsymbol{x}_{k_i} - \boldsymbol{x}_{m_i})\}'$, with

$w_{ik_ij} = \exp(\boldsymbol{x}_{k_i}{}'\beta_j) / \sum_{l \in \mathscr{R}(y_i)} \exp(\boldsymbol{x}_{l'}\beta_j)$. It can be seen that the negated Hessian matrix $\boldsymbol{I}(\beta)$ is block diagonal with blocks $\boldsymbol{I}(\beta_1), \ldots, \boldsymbol{I}(\beta_J)$. We discuss the Bayesian formulation of this problem with the Jeffreys-type prior in the next section. For the sake of space, our presentation covers only the Coxmodel.

4.1. Bayesian formulation

Recall that the posterior distribution for β is given by Eq. (5). If we take

$$\pi(\beta) \propto \mid \boldsymbol{I}(\beta) \mid^{1/2} \quad (10)$$

in Eq. (5), the proposal by Heinze and Schemper (2001) has a Bayesian interpretation under the Jeffreys-type prior for β. They also noted that the penalty function $|\boldsymbol{I}(\beta)|^{1/2}$ in the penalized likelihood function of the Cox model, which is $L_p^*(\beta \mid \mathscr{D}_{\text{obs}}) = L_p(\beta \mid \mathscr{D}_{\text{obs}}) \mid \boldsymbol{I}(\beta) \mid^{1/2}$, is exactly the Jeffreys-type prior. Since the negated Hessian matrix $I(\beta)$ depends on the survival time y_i via the risk set $\mathcal{R}(y_i)$, the prior in Eq. (10) is data-dependent. Thus, the construction of the Jeffreys-type prior is based on a heuristic rule rather than a formal one. Similarly, the penalty function of the cause-specific hazards model is also the Jeffreys-type prior $\pi(\beta) \propto |\boldsymbol{I}(\beta)|^{1/2}$, with $\mid \boldsymbol{I}(\beta) \mid \; = \prod_{j=1}^{J} \left| \boldsymbol{I}(\beta_j) \right|$. In Theorem 4.1.1, we characterize when this choice of prior will exist and be proper.

Theorem 4.1.1—*Consider the prior distribution $\pi(\beta)$ in* Eq. (10) *with $\mathbf{I}(\beta)$ as in* Eq. (8). *If condition C1 in* Eq. (4) *holds, then $\pi(\beta)$ exists and is proper. Otherwise, $\pi(\beta)$ does not exist —that is, the negated Hessian matrix $\mathbf{I}(\beta)$ is singular for all β.*

Example 1 (Revisited)—Omitting the first index, we compute $w_1 = w_3 = e^{\beta_2}/(e^{\beta_1} + 2e^{\beta_2})$ and $w_2 = e^{\beta_1}/(e^{\beta_1} + 2e^{\beta_2}$. From Eqs. (8) and (9), $\boldsymbol{I}(\beta) = \boldsymbol{A}_1 = \kappa \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$, where $\kappa = 2w_1w_2$. Because of that,$|\boldsymbol{I}(\beta)|$=0, and $\pi(\beta)$ does not exist.

Remark 4.1.1—*Due to the result in Proposition A.2 (see* appendix), *each summand in* Eq. (17) *with $|\tilde{\boldsymbol{X}}_h| > 0$ is bounded above by a unimodal and symmetric function around $\beta = \mathbf{0}$. Moreover, the upper bound is an integrable function. Hence, the prior distribution in* Eq. (10) *has finite modes.*

Corollary 4.1.1—*If the Jeffreys-type prior in* Eq. (10) *has finite modes, then the partial posterior distribution in* Eq. (5) *also has finite modes, since the profile likelihood function in* Eq. (1) *is bounded above.*

Theorem 4.2.1—*Assume that the Jeffreys-type prior exists. Then the Jeffreys-type prior has lighter tails than a p-dimensional multivariate t distribution with ν degrees of freedom for $\nu > 0$, and heavier tails than a p-dimensional multivariate normal distribution.*

Remark 4.2.1—Heinze and Schemper (2001, *116) presented two sufficient conditions for the existence of finite estimates of β using* Eq. (7). *One of these conditions requires at least p distinct failure times. Our condition C2 in* Eq. (4) *is weaker, and is illustrated in Example 2.*

Example 2—Take a data set with n =3 observations, p = 2 covariates, $\boldsymbol{x}_1 = (0, 1)'$, $\boldsymbol{x}_2 = (1, 0)'$, $\boldsymbol{x}_3 = (1, 1)'$, $\delta = (1, 0, 0)'$, $y_1 < y_2$, and $y_1 < y_3$. In this case, $R(y_1) = \{1, 2, 3\}$ and $\boldsymbol{X}^*$ in Eq. (3) has rows (0, 0), (1, −1), and (1, 0), so that rank($\boldsymbol{X}^*$) = 2 and condition C1 in Eq. (4) holds true. Omitting the first index, we compute $w_1 = e^{\beta_2}/(e^{\beta_1} + e^{\beta_2} + e^{\beta_1+\beta_2})$, $w_2 = e^{\beta_1}/(e^{\beta_1} + e^{\beta_2} + e^{\beta_1+\beta_2})$, and $w_3 = e^{\beta_1+\beta_2}/(e^{\beta_1} + e^{\beta_2} + e^{\beta_1+\beta_2})$. From Eqs. (8) and (9), after some algebraic manipulations we obtain $\boldsymbol{I}(\beta) = \boldsymbol{A}_1 = \begin{bmatrix} w_1(1-w_1) & -w_1w_2 \\ -w_1w_2 & w_2(1-w_2) \end{bmatrix}$. Hence, $|\boldsymbol{I}(\beta)| = w_1w_2w_3 = e^{2(\beta_1+\beta_2)}/(e^{\beta_1} + e^{\beta_2} + e^{\beta_1+\beta_2})^3$ and $\pi(\beta) \propto |\boldsymbol{I}(\beta)|^{1/2} = e^{\beta_1+\beta_2}/(e^{\beta_1} + e^{\beta_2} + e^{\beta_1+\beta_2})^{3/2}$. Using Eq. (1), we obtain $L_p(\beta|\mathscr{D}_{obs}) = w_1$ and it is easy to check that condition C2 in Eq. (4) is not satisfied in this case. The partial likelihood function has the monotone behavior portrayed in Figure 2a. The posterior distribution is such that $\pi(\beta|\mathscr{D}_{obs}) \propto e^{\beta_1+2\beta_2}/(e^{\beta_1} + e^{\beta_2} + e^{\beta_1+\beta_2})^{5/2}$. Figure 2b shows the contour plots of the prior and posterior distributions of $(\beta_1, \beta_2)'$.

4.2. Shifted Jeffreys-type prior

It is very common to specify a prior distribution for regression coefficients centered at $\beta = \mathbf{0}$. Based on this idea, we introduce the shifted Jeffreys-type prior. Let β_M be a mode of the prior in Eq. (10). By adding β_M to β we get a shifted Jeffreys-type prior given by

$$\pi_s(\beta) \propto |\boldsymbol{I}(\beta + \beta_M)|^{1/2}, \quad (11)$$

so that its mode is shifted to $\beta = \mathbf{0}$. Using $\pi_s(\beta)$, a different posterior $\pi_s(\beta|\mathscr{D}_{obs})$ is obtained from Eq. (5). Our simulation study in section 6 empirically suggests that the shifted Jeffreys-type prior may potentially reduce biases in MPLEs and posterior estimates of the regression coefficients.

4.3. Jeffreys-type prior based on the first risk set

In section 3.2, we mention that C1′ in Eq. (6) is a sufficient condition of C1 in Eq. (4). However, C1 does not imply C1′ unless the subset of the data corresponds to the first risk set, which is the risk set corresponding to the first event defined in section 3.3. As in section 3.3, the data set is rearranged and i_0 denotes the index of the first event, that is, $\delta_i = 0$, $i = 1$,

…, $i_0 - 1$, and $\boldsymbol{X}^*_{(i_0)}$ is the submatrix of $\boldsymbol{X}^*$ corresponding to the first risk set. The following is the condition for C1′:

$$\text{C1}''. \quad \text{The matrix } \boldsymbol{X}^*_{(i_0)} \text{ is of full column rank.} \quad (12)$$

Theorem 4.3.1—*Condition C1 holds if and only if C1″ holds.*

Based on this finding, we propose a new variation of the Jeffreys-type prior that only depends on the first risk set,

$$\pi_f(\beta) \propto |\boldsymbol{A}_{i_0}|^{1/2}, \quad (13)$$

where

$$\boldsymbol{A}_{i_0} = \sum_{j \in \mathscr{R}(y_{i_0})} w_{i_0 j} \left\{ \sum_{l \in \mathscr{R}(y_{i_0})} w_{i_0 l} (\boldsymbol{x}_j - \boldsymbol{x}_l) \right\} \left\{ \sum_{m \in \mathscr{R}(y_{i_0})} w_{i_0 m} (\boldsymbol{x}_j - \boldsymbol{x}_m) \right\}'.$$

Proposition 4.3.1—*The Jeffreys-type prior based on the whole dataset exists and is proper if and only if $\pi_f(\beta)$ in* Eq. (13) *exists and is proper.*

One benefit of the prior in Eq. (13) is the computing time. Using the first risk set to build the prior saves computation time, especially for data sets with a large number of observations and covariates. Furthermore, constructing the prior based on the first risk set will not lose much information, which is verified by both the simulation study and the real data analysis in sections 6 and 7.

5. Bayesian computation

According to Heinze and Schemper (2001, 116), when estimating the standard errors of the estimators obtained by solving $U^*(\beta_r) = 0$ in Eq. (8), $r = 1, \ldots, p$, the negated Hessian matrix in (4.2) and the second derivative of the logarithm of the penalized likelihood function $L_p^*(\beta \mid \mathscr{D}_{\text{obs}})$ can be used, both evaluated at $\beta = \hat{\beta}^*$, where $\hat{\beta}^*$ maximizes $L_p^*(\beta \mid \mathscr{D}_{\text{obs}})$. From their experience, the differences in the estimates are negligible. Based on this finding, in order to sample from the posterior distribution in Eq. (5) with $\pi(\beta)$ in Eq. (10), $\pi_s(\beta)$ in Eq. (11), or $\pi_f(\beta)$ in Eq. (12), one may use the Metropolis–Hastings algorithm (Tierney 1994) to jointly sample β. This algorithm would work if a global proposal density based on $\hat{\beta}^*$ and the available negated Hessian matrix and second derivative of the logarithm of the penalized likelihood function can be constructed so that it mimics the posterior based on the Jeffreys-type prior. However, finding such a good proposal is very difficult. Another possible solution is to adapt the variance of a normal proposal in a Metropolis within Gibbs sampler by controlling the acceptance rate of batches of the simulation (Roberts and Rosenthal 2009). Here we consider an adaptive localized Metropolis algorithm discussed in Chen et al. (2000) to sample each component of β in turn within the Gibbs sampling framework. The localized Metropolis algorithm requires the computation of the second derivative of the logarithm of the full conditional posterior distributions, which is a challenging task due to the highly demanding computation of the Jeffreys-type prior.

Let $\pi(\beta_j|\beta_{(-j)},\mathscr{D}_{\mathrm{obs}})$ denote the full conditional distribution of β_j given $\beta_{(-j)}$, where $\beta_{(-j)}$ denotes β with the jth component deleted. To avoid direct computation of the second derivative of log $\pi(\beta_j|\beta_{(-j)},\mathscr{D}_{\mathrm{obs}})$, we first find the mode of $\pi(\beta_j|\beta_{(-j)},\mathscr{D}_{\mathrm{obs}})$ and then use a quadratic curve $y = ax^2 + bx + c$ to approximate log $\pi(\beta_j|\beta_{(-j)},\mathscr{D}_{\mathrm{obs}})$ around this mode. Since the localized Metropolis algorithm uses a normal proposal, the variance of this proposal can then be approximated by $-1/(2a)$. This approach requires the evaluation of log $\pi(\beta_j|\beta_{(-j)},\mathscr{D}_{\mathrm{obs}})$ at a few values of β_j around its mode. This algorithm operates as follows: Step 1. Let β_j be the current value; Step 2. Compute $\hat{\beta}_j = \mathrm{argmax}_{\beta_j} \log \pi(\beta_j|\beta_{(-j)},\mathscr{D}_{\mathrm{obs}})$ and use a quadratic regression to compute a by approximating $y = ax^2 + bx + c$ to log $\pi(\beta_j|\beta_{(-j)},\mathscr{D}_{\mathrm{obs}})$ in the neighborhood of $\hat{\beta}_j$; Step 3. Draw β_j^* from $N(\hat{\beta}_j, -\frac{1}{2a})$; and Step 4. A move from β_j to β_j^* is made with probability $\alpha = \min\left\{\frac{\pi(\beta_j^* \mid \beta_{(-j)}, \mathscr{D}_{\mathrm{obs}})q(\beta_j)}{\pi(\beta_j \mid \beta_{(-j)}, \mathscr{D}_{\mathrm{obs}})q(\beta_j^*)}, 1\right\}$, where $q(\cdot)$ is the density function of $N(\hat{\beta}_j, -\frac{1}{2a})$.

6. Simulation studies

In this section, we conduct a simulation study to assess the properties of estimators under different approaches. In the data generation, we first generate $n = 100$ independent x_{i1}~Bernoulli(0.9) and x_{i2}~Bernoulli(0.5). The failure times follow an exponential distribution with hazards $0.005 \exp(\beta_1 x_{i1} + \beta_2 x_{i2})$, $i = 1, \ldots, n$, where the true values of β_1 and β_2 are 2.0 and −0.8. These values remain fixed throughout the 500 replications of the simulations. The failure times are subject to administrative censoring with duration set to 5.0, and 30.0 in order to reach an average censoring rate around 90%, and 50%. The percentage of zero events corresponding to $x_1 = 0$ amounts to 86.6%, and 43.8%, respectively, which lead to a monotone partial likelihood, and therefore the corresponding parameter β_1 is not identifiable. We use the localized Metropolis algorithm to generate samples from the posterior distribution. After discarding the first 2000 iterations of the sampler, we used the next 10,000 iterations for each parameter. We compare the simulation results of the MPLE, shifted MPLE, Jeffreys-type prior, and shifted Jeffreys-type prior approaches using all data as well as only the first risk set to build our prior. The code was written in the FORTRAN language using IMSL subroutines with double precision.

In Table 2 we report the true value of the parameter (True), the average of the MPLEs or of the posterior medians (Est), the average of the standard errors or of the posterior standard deviations (SE), the standard deviation of the estimates (SD), the root of the mean squared error of the MPLEs or of the posterior medians (RMSE), and the coverage probability (CP) of the Wald 95% asymptotic confidence interval (CI) or of the 95% highest posterior density

(HPD) interval for each parameter. First, we note that SD, SE, and RMSE for β_2 are very close to each other since it does not have the "zero event" issue, but they are different from each other for the problematic parameter, that is, β_1. The estimates are biased as was expected when duration equals 5.0, with a high percentage of zero events (86.6%). As the censoring duration becomes larger (duration sets to 30.0), the percentage of zero events becomes smaller (43.8%), which makes the estimates more accurate. Moreover, under both cases (duration sets to 5.0 or 30.0), Bayesian approaches perform better than the frequentist approaches in terms of the coverage probabilities (closer to 95%). As exhibited in Table 2, shifted approaches are more likely to provide more accurate estimates (closer to the true values) for both β_1 and β_2 than the corresponding non-shifted approaches, and provide different SE, and RMSE for the problematic parameter (β_1) only.

Another interesting finding in Table 2 is that for the same approach, the results based on the original Jeffreys type prior, which uses all data, and the first risk set prior were quite similar, indicating that using the specific partial data will not lose much information. The first risk set also has the advantage of computation time. For the censoring duration equal to 5.0, the total computation time for all the four approaches using entire data was 1577.3 minutes on an Intel i7-4770 processor machine with 16 GB of RAM memory using a GNU/Linux operating system, while the computation time was almost half (936.91 minutes) if we just used the first risk set, noticing that the difference in Table 2 increases when the number of events increases.

7. Analysis of the SEER prostate cancer data

We can easily check that the SEER prostate cancer data satisfies condition C1″ (consequently condition C1). According to Theorem 4.1.1, Jeffreys-type prior has finite modes and thus the posterior mode under the Bayesian formulation exists. Table 3 shows the MPLEs or the posterior medians (Est), the standard errors or the posterior standard deviations for the regression coefficients, the MPLEs or the posterior medians of the hazard ratios (HR), and the 95% confidence intervals (CI) or the 95% HPD intervals for HR under MPLE, shifted MPLE, Jeffreys-type prior, and shifted Jeffreys-type prior approaches using all the data as well as the first risk set.

Recall that there was no monotone likelihood issue for death due to other causes. Therefore, the estimates computed by the shifted MPLE approach and the SAS procedure PHREG, as shown in Table 1, were almost identical. For example, under the shifted MPLE using the all data approach, $\hat{\beta} = (0.074, 1.049, -0.764, 0.207, -0.170, 0.569)'$. If we used the SAS procedure PHREG, $\hat{\beta} = (0.074, -1.082, -0.785, 0.198, -0.204, 0.575)'$. The estimates given by the other approaches in Table 3 were also similar to the results in Table 1, which further empirically confirms that the Jeffreys-type prior is noninformative and does not introduce bias if the data do not have a monotone partial likelihood issue. However, the Jeffreys-type prior does improve the estimates if the monotone problem exists. For death due to prostate cancer, the covariate RP (Table 1) had a huge standard error, namely, SE = 1680, and was thus not identifiable. The problematic covariate (RP) is now identifiable, and even significant under all the four approaches in Table 3.

Similar to the simulation study, the estimates obtained by using Jeffreys-type priors based on the entire data set as well as the first risk set ($n - i_0$ = 1203 for prostate cancer death and $n - i_0$ = 1825 for other causes death) were very similar, especially for nonproblematic covariates. For example, the estimates for prostate cancer death under shifted the Jeffreys-type prior and first risk set shifted Jeffreystype prior were $\hat{\beta} = (0.211, -4.156, -0.996, -0.683, -0.323, -0.329)'$ and $\hat{\beta} = (0.204, -4.314, -1.016, -0.676, -0.311, -0.337)'$, respectively. For the prostate cancer data, the computation times for all the four approaches were 93.0 minutes if we used the entire data set for the prior and 39.3 minutes if we used the first risk set for the prior. The computation gains are more obvious for the other causes death data (830.0 minutes using entire dataset and 242.1 minutes using the first risk set), which have more events than the prostate cancer death data. Thus, the first risk set approach was more computationally efficient. SE for covariates with rare events (RP and Black) under shifted approaches are different from SE under the corresponding nonshifted approaches, while SE for other covariates are quite similar among all approaches, which is consistent with the simulation results. For death due to the prostate cancer data, the coefficient of RP, which is originally not identifiable, is now significant based on the 95% CI/HPD for HR under all the four approaches. None of the other coefficients are significant within this group of data. For death due to other causes, surgical treatment (RP), radiation only treatment (RT), and age (Age) are all significant effects under the four approaches.

8. Discussion

In this article, we have thoroughly investigated the conditions of the monotone partial likelihood and developed equivalent sufficient and necessary conditions based on the first risk set. Under mild conditions, we have shown that the Jeffreys-type prior is proper and has finite modes. Moreover, it has lighter tails than a multivariate t distribution and heavier tails than a multivariate normal distribution. We have proposed two variations of the Jeffreys-type prior, namely, the Jeffreys-type prior based on the first risk set and the shifted Jeffreys-type prior. We emphasize that C1″ in (4.6) is an easy-to-check condition and plays a key role in the solution of the monotone likelihood problem. Future work includes a theoretical investigation of the unimodality of the Jeffreys-type prior and more proprieties of the prior based on the first risk set—for example, in the presence of time dependent covariates (Heinze and Dunkler 2008). We also need to further investigate additional properties of the two variations of Jeffreys-type prior. These two priors would lead to similar posterior estimates if $\lim_{\|\beta\| \to \infty} |\boldsymbol{I}(\beta)|/|\boldsymbol{A}_{i_0}| = c$, where c is a constant. We also envision extending the Jeffreys-type prior to other models for competing risks data (for recent contributions and a literature overview, see Ge and Chen 2012; Beyersmann and Scheike 2013; Chen et al. 2013; Fine and Lindqvist 2014).

Acknowledgments

We would like to thank the editor in chief, the guest editor, the associate editor, and the two anonymous reviewers for their very helpful comments and suggestions, which have led to a much improved version of the article.

Funding

Dr. de Castro's research was partially supported by CNPq, Brazil, grant 306579/2014-6. Dr. Chen's research was partially supported by US Nationsal Institutes of Health grants GM 70335 and P01 CA142538.

References

Beyersmann, J., Scheike, T. Classical regression models for competing risks. In: Klein, JP.van Houwelingen, HC.Ibrahim, JG., Scheike, TH., editors. Handbook of survival analysis. Boca Raton, FL: Chapman & Hall/CRC; 2013. p. 157-77.

Bryson MC, Johnson ME. The incidence of monotone likelihood in the Cox model. Technometrics. 1981; 23:381–83.

Chen, M-H., Shao, Q-M., Ibrahim, JG. Monte Carlo methods in Bayesian computation. New York, NY: Springer; 2000.

Chen MH, Ibrahim JG, Shao QM. Posterior propriety and computation for the Cox regression model with applications to missing covariates. Biometrika. 2006; 93:791–807.

Chen MH, Ibrahim JG, Shao QM. Maximum likelihood inference for the Cox regression model with applications to missing covariates. Journal of Multivariate Analysis. 2009; 100:2018–30. [PubMed: 19802375]

Chen, M-H., de Castro, M., Ge, M., Zhang, Y. Bayesian regression models for competing risks. In: Klein, JP.van Houwelingen, HC.Ibrahim, JG., Scheike, TH., editors. Handbook of survival analysis. Boca Raton, FL: Chapman & Hall/CRC; 2013. p. 179-98.

Cox DR. Regression models and life-tables. Journal of the Royal Statistical Society B. 1972; 34:187–220.

Cox DR. Partial likelihood. Biometrika. 1975; 62:269–76.

Fine J, Lindqvist BH. Competing risks. Lifetime Data Analysis. 2014; 20:159–60. [PubMed: 24633917]

Firth D. Bias reduction of maximum likelihood estimates. Biometrika. 1993; 80:27–38.

Gaynor JJ, Feuer EJ, Tan CC, Wu DH, Little CR, Straus DJ, Clarkson BD, Brennan MF. On the use of cause-specific failure and conditional failure probabilities: Examples from clinical oncology data. Journal of the American Statistical Association. 1993; 88:400–9.

Ge M, Chen MH. Bayesian inference of the fully specified subdistribution model for survival data with competing risks. Lifetime Data Analysis. 2012; 18:339–63. [PubMed: 22484596]

Greenland S, Mansournia MA. Penalization, bias reduction, and default priors in logistic and related categorical and survival regressions. Statistics in Medicine. 2015; 34:3133–43. [PubMed: 26011599]

Heinze G, Dunkler D. Avoiding infinite estimates of time-dependent effects in small-sample survival studies. Statistics in Medicine. 2008; 27:6455–69. [PubMed: 18816502]

Heinze G, Ploner M. SAS and SPLUS programs to perform Cox regression without convergence problems. Computer Methods and Programs in Biomedicine. 2002; 67:217–23. [PubMed: 11853948]

Heinze G, Schemper M. A solution to the problem of monotone likelihood in Cox regression. Biometrics. 2001; 57:114–19. [PubMed: 11252585]

Ibrahim JG, Laud PW. On Bayesian analysis of generalized linear models using Jeffreys's prior. Journal of the American Statistical Association. 1991; 86:981–86.

Kalbfleisch JD. Non-parametric Bayesian analysis of survival time data. Journal of the Royal Statistical Society B. 1978; 40:214–21.

Kalbfleisch, JD., Prentice, RL. The statistical analysis of failure time data. 2. Hoboken, NJ: Wiley; 2011.

Ploner, M., Heinze, G. coxphf: Cox regression with Firth's penalized likelihood. R package version 1.05. 2010. http://CRAN.R-project.org/package=coxphf

Roberts GO, Rosenthal JS. Examples of adaptive MCMC. Journal of Computational and Graphical Statistics. 2009; 18:349–67.

Roy V, Hobert JP. Convergence rates and asymptotic standard errors for Markov chain Monte Carlo algorithms for Bayesian probit regression. Journal of the Royal Statistical Society B. 2007; 69:607–623.

Sinha D, Ibrahim JG, Chen MH. A Bayesian justification of Cox's partial likelihood. Biometrika. 2003; 90:629–41.

Tierney L. Markov chains for exploring posterior distributions. Annals of Statistics. 1994; 22:1701–62.
Zhang, F. Matrix theory. Basic results and techniques. New York, NY: Springer; 1999.

Appendix: Proofs

Proof of Theorem 3.2.1—It is easy to prove that C1′ implies C1, since the column rank of a matrix is always greater than or equal to the column rank of any of its submatrices. Let $\boldsymbol{X}_s^*$ be a $K_s \times p$ submatrix of $\boldsymbol{X}^*$ and let $\boldsymbol{x}_{s_i}^{*\prime}$ be a row vector of $\boldsymbol{X}_s^*$, $i = 1, \ldots, K_s$. According to C2′, there exists a positive vector **v** such that $\sum_{i=1}^{K_s} \boldsymbol{x}_{s_i}^* v_i = 0$. Since the row vectors in $\boldsymbol{X}^*$ are linear combinations of those in $\boldsymbol{X}_s^*$, according to condition C1′, $\boldsymbol{X}^{*\prime}\boldsymbol{v}^*$ can be expressed as $\sum_{i=1}^{K_s} \boldsymbol{x}_{s_i}^* v_i^* + \sum_{j=K_s+1}^{K} \sum_{i=1}^{K_s} k_{i_j} \boldsymbol{x}_{s_i}^* v_j^*$. Let $v_i^* = v_i - \sum_{j=K_s+1}^{K} k_{i_j} v_j^*$, for $i = 1, \ldots, K_s$, where $0 < v_j^* < \min_{i \in \{1, \ldots, K_s\}} v_i / \sum_{j=K_s+1}^{K} | k_{i_j} |$, for $j = K_s$) $1, \ldots, K$. Thus, we have $\boldsymbol{X}^{*\prime}\mathbf{v}^* = \mathbf{0}$ and $\mathbf{v}^*$ is a positive vector. Therefore, C2′ implies C2. In order to prove Theorem 4.1.1, we need to establish the following two propositions.

Proposition A.1

For $\boldsymbol{A}_i$ *given in* Eq. (9) *and*

$$\boldsymbol{B}_i = \boldsymbol{B}_i(\beta) = \sum_{j_i \in \mathscr{R}(y_i)} w_{ij_i} \sum_{l_i \in \mathscr{R}(y_i)} w_{il_i} (\boldsymbol{x}_{j_i} - \boldsymbol{x}_{l_i})(\boldsymbol{x}_{j_i} - \boldsymbol{x}_{l_i})', \quad (14)$$

we have that $\boldsymbol{A}_i \preceq \boldsymbol{B}_i$, meaning that $\boldsymbol{B}_i - \boldsymbol{A}_i$ is a nonnegative definite matrix, $i = 1, \ldots, n$.

Proof

For all $\boldsymbol{a} \in \mathbb{R}^p$ we can write

$$\begin{aligned}
\mathbf{a}'\boldsymbol{A}_i\boldsymbol{a} = \mathbf{a}' & \left[\sum_{j_i} w_{ij_i} \left\{ \sum_{l_i} w_{il_i}^2 (\boldsymbol{x}_{j_i} - \boldsymbol{x}_{l_i})(\boldsymbol{x}_{j_i} - \boldsymbol{x}_{l_i})' + \sum_{l_i < m_i} w_{il_i} w_{im_i} (\boldsymbol{x}_{j_i} - \boldsymbol{x}_{l_i}) \right.\right. \\
& \left.\left. (\boldsymbol{x}_{j_i} - \boldsymbol{x}_{m_i})' + \sum_{l_i > m_i} w_{il_i} w_{im_i} (\boldsymbol{x}_{j_i} - \boldsymbol{x}_{l_i})(\boldsymbol{x}_{j_i} - \boldsymbol{x}_{m_i})' \right\} \right] \boldsymbol{a} \\
\le & \sum_{j_i} w_{ij_i} \left[\sum_{l_i} w_{il_i}^2 \mathbf{a}'(\boldsymbol{x}_{j_i} - \boldsymbol{x}_{l_i})(\boldsymbol{x}_{j_i} - \boldsymbol{x}_{l_i})'\boldsymbol{a} + \sum_{l_i < m_i} w_{il_i} w_{im_i} \right. \\
& \left. \{\mathbf{a}'(\boldsymbol{x}_{j_i} - \boldsymbol{x}_{l_i})(\boldsymbol{x}_{j_i} - \boldsymbol{x}_{l_i})'\boldsymbol{a} + \mathbf{a}'(\boldsymbol{x}_{j_i} - \boldsymbol{x}_{m_i})(\boldsymbol{x}_{j_i} - \boldsymbol{x}_{m_i})'\boldsymbol{a}\} \right] \\
= & \sum_{j_i} w_{ij_i} \left\{ \sum_{l_i} w_{il_i} \mathbf{a}'(\boldsymbol{x}_{j_i} - \boldsymbol{x}_{l_i})(\boldsymbol{x}_{j_i} - \boldsymbol{x}_{l_i})'\boldsymbol{a} \right\} = \boldsymbol{a}'\boldsymbol{B}_i\boldsymbol{a}.
\end{aligned}$$

Therefore, $\boldsymbol{a}'(\boldsymbol{A}_i - \boldsymbol{B}_i)\boldsymbol{a} \le 0$, concluding the proof.

Proposition A.2

For j_i and $l_i \in \mathcal{R}(y_i)$, we have $w_{ij_i} w_{il_i} \le w^*_{i,j_i,l_i}$, where w^*_{i,j_i,l_i} is given by

$$w^*_{i,j_i,l_i} = \frac{\exp\{(\boldsymbol{x}_{l_i} - \boldsymbol{x}_{j_i})'\beta\}}{[1 + \exp\{(\boldsymbol{x}_{l_i} - \boldsymbol{x}_{j_i})'\beta\}]^2}. \quad (15)$$

Proof

From the expression after Eq. (8),
$w_{ij_i} w_{il_i} = \frac{\exp(\boldsymbol{x}_{j_i}'\beta)}{\sum_{m_i \in \mathcal{R}(y_i)} \exp(\boldsymbol{x}_{m_i}'\beta)} \frac{\exp(\boldsymbol{x}_{l_i}'\beta)}{\sum_{m_i \in \mathcal{R}(y_i)} \exp(\boldsymbol{x}_{m_i}'\beta)} \le \frac{\exp\{(\boldsymbol{x}_{l_i} - \boldsymbol{x}_{j_i})'\beta\}}{[1 + \exp\{(\boldsymbol{x}_{l_i} - \boldsymbol{x}_{j_i})'\beta\}]^2} = w^*_{i,j_i,l_i}$, as claimed in the proposition. We are now ready to prove Theorem 4.1.1.

Proof of Theorem 4.1.1—Let $\boldsymbol{X}_i$ and $\mathbf{w}_i$ denote, respectively, the $k_i^2 \times p$ matrix with row $\boldsymbol{x}_{j_i}' - \boldsymbol{x}_{l_i}'$ and the $k_i^2 \times 1$ vector with element $w_{ij_i} w_{il_i}$, for j_i and $l_i \in \mathcal{R}(y_i)$. The matrix in Eq. (14) can be written as $\boldsymbol{B}_i = \boldsymbol{X}_i' \mathbf{D}(\mathbf{w}_i)\boldsymbol{X}_i$, where $\mathbf{D}(\cdot)$ denotes a diagonal matrix.

The number of failures is denoted by n_1. Without loss of generality, we assume $\delta_i = 1$ for $i = 1, \ldots, n_1 \le n$, so that $\boldsymbol{I}(\beta) = \sum_{i=1}^{n_1} \boldsymbol{A}_i$. Let $\tilde{\mathbf{X}}$ and $\tilde{\boldsymbol{w}}$ be, respectively, the $K_2 \times p$ matrix and the $K_2 \times 1$ vector given by $\tilde{\mathbf{X}} = [\boldsymbol{X}_1', \ldots, \boldsymbol{X}_{n_1}']'$ and $\tilde{\boldsymbol{w}} = (w_{ij_i} w_{il_i}: j_i$ and $l_i \in \mathcal{R}(y_i), i = 1, \ldots, n_1)'$, where $K_2 = \sum_{i=1}^{n_1} k_i^2$. We define $\tilde{\boldsymbol{B}} = \tilde{\mathbf{X}}' \mathbf{D}(\tilde{\boldsymbol{w}})\, \tilde{\mathbf{X}}$, noting that $\tilde{\boldsymbol{B}} = \sum_{i=1}^{n_1} \boldsymbol{B}_i$. Since $\boldsymbol{B}_i$, $i = 1, \ldots, n_1$, are symmetric nonnegative definite matrices, then from Proposition A.1, we have that $\boldsymbol{I}(\beta) \preccurlyeq \sum_{i=1}^{n_1} \boldsymbol{B}_i = \tilde{\boldsymbol{B}}$ and according to Zhang (1999, Theorem 6.8), $|\boldsymbol{I}(\beta)| \le |\tilde{\boldsymbol{B}}|$. Let $\tilde{\boldsymbol{w}}^*$ be the $K_2 \times 1$ vector with elements w^*_{i,j_i,l_i} in Eq. (15). Using a determinant expansion (see, e.g., Ibrahim and Laud 1991) and Proposition A.2, we obtain

$$|\boldsymbol{I}(\beta)| \le \sum_{h \in \mathcal{H}} |\tilde{\mathbf{X}}_h|^2 \prod_{k=1}^{p} \tilde{w}_{i_k} \le \sum_{h \in \mathcal{H}} |\tilde{\mathbf{X}}_h|^2 \prod_{k=1}^{p} \tilde{w}^*_{i_k}, \quad (16)$$

where $\mathcal{H} = \{(i_1, \ldots, i_p): 1 \le i_1 < \cdots < i_p \le K_2\}$ and $\tilde{\mathbf{X}}_h$ is a $\boldsymbol{p} \times \boldsymbol{p}$ matrix with columns $\tilde{\mathbf{x}}_{i_1}, \ldots, \tilde{\mathbf{x}}_{i_p}$. Hence,

$$\pi(\beta) \propto \ |\boldsymbol{I}(\beta)|^{1/2} \leq \sum_{h \in \mathscr{H}} |\tilde{\mathbf{X}}_h| \prod_{k=1}^{p} (\tilde{w}^*_{i_k})^{1/2} \quad (17)$$

and if $|\tilde{\mathbf{X}}_h| = 0$ for all $h \in H$, then $\pi(\beta)$ does not exist

Under C1 in Eq. (4), there is an $h \in H$ such that $\varphi_h = \tilde{\mathbf{X}}_h{}'\beta$ is a one-to-one transformation with Jacobian $1/|\tilde{\mathbf{X}}_h|$. Taking into account Eq. (15), we get $\int_{\mathbb{R}^p} |\boldsymbol{I}(\beta)|^{1/2} d\beta \leq \sum_{h \in \mathscr{H}:\, |\tilde{\mathbf{X}}_h| > 0} \int_{\mathbb{R}^p} \prod_{k=1}^{p} \frac{\exp(\varphi_{hk}/2)}{1 + \exp(\varphi_{hk})} d\varphi_h$. Since $\int_{-\infty}^{\infty} \exp(\varphi_{hk}/2)/\{1 + \exp(\varphi_{hk})\} d\varphi_{hk} = 2\int_0^{\infty} 1/(1+u^2) du = \pi$, for $k = 1, \ldots, p$, we conclude that the prior distribution for β is proper, that is, $\int_{\mathbb{R}^p} \pi(\beta)\, d\beta < \infty$.

Proof of Theorem 4.2.1—Let $t_\nu(\beta|\beta_0, \Sigma_0)$ denote the density function of a p-dimensional multivariate t distribution with location vector β_0, scale matrix Σ_0, and ν degrees of freedom, which is given by

$$t_\nu(\beta \mid \beta_0, \boldsymbol{\Sigma}_0) = \frac{\Gamma(\frac{\nu + p}{2})}{\Gamma(\nu/2)(\nu\pi)^{p/2}} \left\{1 + \frac{1}{\nu}(\beta - \beta_0)' \boldsymbol{\Sigma}_0^{-1} (\beta - \beta_0)\right\}^{-(\nu + p)/2}.$$

The tail condition in the proposition is represented by $\lim_{\|\beta\| \to \infty} |\boldsymbol{I}(\beta)|^{1/2}/t_\nu(\beta|\beta_0, \Sigma_0) = 0$, for all $\nu > 0$. It is equivalent to $\lim_{\|\beta\| \to \infty} |\boldsymbol{I}(\beta)|^{1/2}\left\{1 + (\beta - \beta_0)'\boldsymbol{\Sigma}_0^{-1}(\beta - \beta_0)/\nu\right\} = 0$, for all $\nu > 0$. Recall from Eq. (16) that $|\boldsymbol{I}(\beta)| \leq \sum_{h \in H} |\tilde{\mathbf{X}}_h|^2 \prod_{k=1}^{p} \tilde{w}^*_{i_k}$, and from Eq. (15), $\tilde{w}^*_{i_k} \leq \exp(-\tilde{\mathbf{x}}_{hk}{}'\beta)$, where $\tilde{\mathbf{x}}_{hk}{}'$ is the kth row of $\tilde{\mathbf{X}}_h$. Therefore,

$$\begin{aligned} 0 &\leq \lim_{\|\beta\| \to \infty} |\boldsymbol{I}(\beta)|^{1/2}\{1 + \frac{1}{\nu}(\beta - \beta_0)'\boldsymbol{\Sigma}_0^{-1}(\beta - \beta_0)\} \\ &\leq \lim_{\|\beta\| \to \infty} \left\{\sum_{h \in H} |\tilde{\mathbf{X}}_h|^2 \prod_{k=1}^{p} \exp(-\tilde{\mathbf{x}}_{hk}{}'\beta)\right\}^{1/2} \{1 + \frac{1}{\nu}(\beta - \beta_0)'\boldsymbol{\Sigma}_0^{-1}(\beta - \beta_0)\} = 0, \end{aligned}$$

and consequently we have $\lim_{\|\beta\| \to \infty} \boldsymbol{I}(\beta)|^{1/2}/t_\nu(\beta|\beta_0, \Sigma_0) = 0$.

Let $\varphi(\beta|\beta_0, \Sigma_0)$ be the probability density function of the p-dimensional normal distribution, where Σ_0 is a $p \times p$ positive definite matrix, which is given by $\varphi(\beta \big| \beta_0, \boldsymbol{\Sigma}_0) = (2\pi)^{-p/2} \big|\boldsymbol{\Sigma}_0\big|^{-1/2} \exp\left\{-(\beta - \beta_0)'\boldsymbol{\Sigma}_0^{-1}(\beta - \beta_0)/2\right\}$. The tail condition in the proposition is represented by $\lim_{\|\beta\| \to \infty} /\boldsymbol{I}(\beta)|^{1/2}/\varphi(\beta|\beta_0, \Sigma_0) = \infty$. It is equivalent to $\lim_{\|\beta\| \to \infty} |\boldsymbol{I}(\beta)| \exp\left\{(\beta - \beta_0)'\boldsymbol{\Sigma}_0^{-1}(\beta - \beta_0)\right\} = \infty$.

Without loss of generality, we assume there are no ties for each component of $\boldsymbol{x}_{j_i}$, where $j_i \in R(y_i)$. Since $|\boldsymbol{I}(\beta)| \ge |\boldsymbol{A}_{i_0}|$, with $\boldsymbol{A}_{i_0}$ as in Eq. (13), we need to prove only for $\boldsymbol{A}_{i_0}$. We further take $i_0 = 1$ for simplicity. For $p = 1$, $A_1 = \Sigma_{j_1 \in R(y_1)} w_{1j_1} \{\Sigma_{l_1 \in R(y_1)} w_{1l_1} (x_{j_1} - x_{l_1})\}^2$ and $\boldsymbol{\Sigma}_0 = \sigma_0^2$. Recall that $w_{1j_1} = \exp(x_{j_1}\beta)/\Sigma_{l \in R(y_1)} \exp(x_{l_1}\beta)$. Note that

$$\lim_{\beta \to +\infty} w_{1j_1} = \begin{cases} 1, & \text{if } x_{j_1} = \max\{x_{l_1}: l_1 \in R(y_1)\}, \\ 0, & \text{otherwise}. \end{cases} \tag{18}$$

Let j_0 be such that $w_{1j_0} = \min\{w_{1j_1}: j_1 \in R(y_1)\}$ and j^* be such that $x_{j*} = \max\{x_{j_1}: j_1 \in R(y_1)\}$. Thus, we have

$$\begin{aligned} &\lim_{\beta \to +\infty} A_1 \exp\{(\beta - \beta_0)^2/\sigma_0^2\} \\ &\ge \lim_{\beta \to +\infty} w_{1j_0} \exp\{(\beta - \beta_0)^2/\sigma_0^2\} \lim_{\beta \to +\infty} \sum_{j_1 \in \mathscr{R}(y_1)} \left\{ \sum_{l_1 \in \mathscr{R}(y_1)} w_{1l_1}(x_{j_1} - x_{l_1}) \right\}^2 \\ &= \lim_{\beta \to +\infty} \frac{\exp\{(\beta - \beta_0)^2/\sigma_0^2\}}{\Sigma_{l_1 \in \mathscr{R}(y_1)} \exp\{(x_{l_1} - x_{j_0})\beta\}} \sum_{j_1 \in \mathscr{R}(y_1)} (x_{j_1} - x_{j^*})^2 = \infty . \end{aligned}$$

Similarly, we can prove that $\lim_{\beta \to -\infty} A_1 \exp\{(\beta - \beta_0)^2/\sigma_0^2\} = \infty$ using $-\beta$ and $-x_{j_1}$ to replace β and x_{j_1} in A_1

For $p > 1$, let $\beta = r\mathbf{d}_k$, where $r \in \mathbb{R}$ and $\mathbf{d}_k$ is the kth unit vector in the canonical basis of $\mathbb{R}^p$. It suffices to show that for any $\mathbf{d}_k$, $\lim_{r \to \infty} |\boldsymbol{A}_1| \exp\{(r\boldsymbol{d}_k - \beta_0)'\boldsymbol{\Sigma}_0^{-1}(r\boldsymbol{d}_k - \beta_0)\} = \infty$. Let j_k^* be such that $x_{j_k^*} = \max\{x_{j_1,k}: j_1 \in R(y_1)\}$. Similar to the case when $p = 1$,

$$\lim_{r \to +\infty} |\boldsymbol{A}_1| \exp\{(r\boldsymbol{d}_k - \beta_0)'\boldsymbol{\Sigma}_0^{-1}(r\boldsymbol{d}_k - \beta_0)\} \ge \lim_{r \to +\infty} w_{1j_0} \exp\{(r\boldsymbol{d}_k - \beta_0)'\boldsymbol{\Sigma}_0^{-1}(r\boldsymbol{d}_k - \beta_0)\} \left| \Sigma_{j_1 \in \mathscr{R}(y_1)} (\boldsymbol{x}_{j_1} - \boldsymbol{x}_{j_k^*})(\boldsymbol{x}_{j_1} - \boldsymbol{x}_{j_k^*})' \right| = \infty$$

Therefore, $\lim_{\|\beta\| \to \infty} \boldsymbol{A}_1|^{1/2}/\varphi(\beta|\beta_0, \Sigma_0) = \infty$.

Proof of Theorem 4.3.1—The matrix $\boldsymbol{X}^*$ is given by $\boldsymbol{X}^*\left[\mathbf{0}, \cdots, \mathbf{0}, (X_{(i_0)}^*)', \cdots, \delta_{(n-1)}(\boldsymbol{x}_{(n)} - \boldsymbol{x}_{(n-1)}), \mathbf{0}\right]'$, where $\boldsymbol{X}_{(i_0)}^*$ is the $(n - i_0) \times p$ submatrix generated by the first risk set with rows $\mathbf{0}'$, $(\boldsymbol{x}_{(i_0+1)} - \boldsymbol{x}_{(i_0)})'$, …, $(\boldsymbol{x}_{(n)} - \boldsymbol{x}_{(i_0)})'$. Notice that all the rows above $\boldsymbol{X}_{(i_0)}^*$ are $\mathbf{0}'$ since $\delta_{(i)} = 0$, $i = 1, \ldots, i_0 - 1$. The remaining row vectors are simply linear combinations of the row vectors in $\boldsymbol{X}^*_{(i_0)}$. Thus, the matrix $\boldsymbol{X}^*$ is of full column rank p if and only if $\boldsymbol{X}^*_{(i_0)}$ is of full column rank. This also implies that $n \geq i_0 + p$.

Proof of Proposition 4.3.1—Using only the risk set $R(y_{(i_0)})$, the proof follows steps similar to those found in the proof of Theorem 4.1.1 after replacing $\boldsymbol{I}(\beta)$ by $\boldsymbol{A}_{i_0}$ given in Eq. (13).

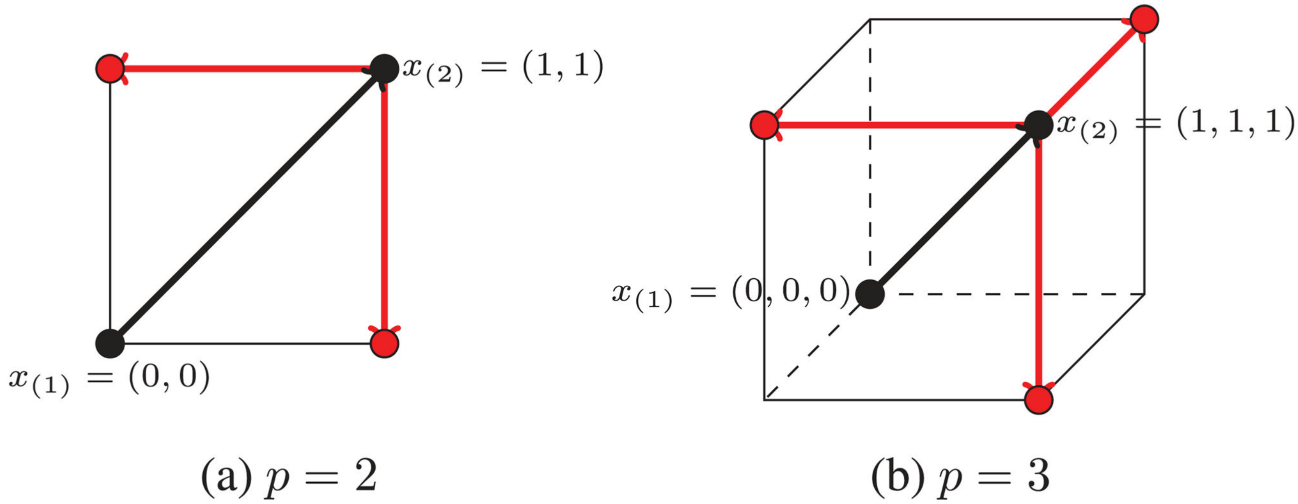
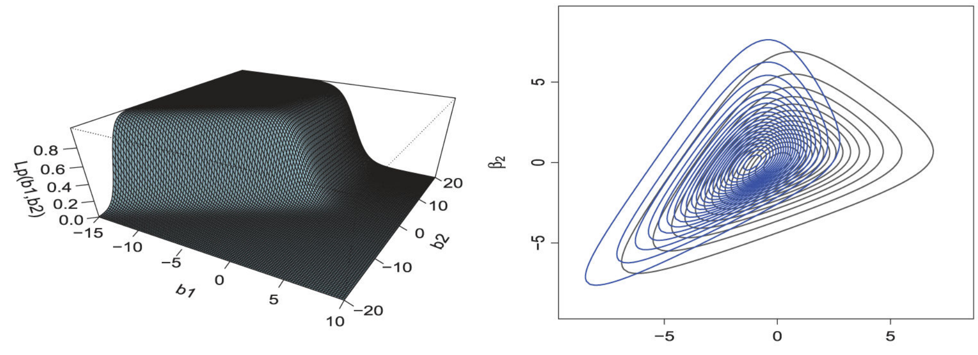


Figure 1.
Illustration of C1 and C2 when there are exactly two events.



(a) (b)

Figure 2.

Example 2: (a) partial likelihood function for $(\beta_1, \beta_2)'$; (b) contour plots of the prior (gray) and posterior (blue) distributions for $(\beta_1, \beta_2)'$.

Table 1

Maximum partial likelihood estimates for the prostate cancer data.

		Death due to prostate cancer				Death due to other causes			
Variable	N_c	N_{pc}	**Est**	**SE**	***p* value**	N_{oc}	**Est**	**SE**	***p* value**
PSA	1840		0.253	0.468	0.5881		0.074	0.173	0.6657
RP	842	0	−17.745	1680	0.9916	13	−1.082	0.388	0.0053
RT	576	3	−1.150	0.742	0.1210	19	−0.785	0.300	0.0089
Black	279	1	−0.539	1.125	0.6318	8	0.198	0.391	0.6132
Year_diag	1840		−0.377	0.743	0.6118		−0.204	0.197	0.3009
Age	1840		−0.372	0.416	0.3712		0.575	0.166	0.0006

Table 2

Estimates from 500 replications and computation times in minutes.

	True	All data					First risk set				
		Est	SE	SD	RMSE	CP	Est	SE	SD	RMSE	CP
MPLE/Shifted MPLE Zero event percentage = 86.6%											
β_1	2.0	0.647	1.436	0.531	1.453	0.870	0.642	1.432	0.529	1.458	0.870
		1.045	1.724	0.523	1.088	0.926	1.037	1.717	0.520	1.095	0.926
β_2	−0.8	−0.861	0.695	0.670	0.672	0.988	−0.855	0.694	0.666	0.668	0.988
		−0.833	0.694	0.673	0.673	0.988	−0.832	0.693	0.672	0.672	0.988
Jeffreys-type/Shifted Jeffreys-type prior											
β_1	2.000	1.294	1.706	0.854	1.107	0.926	1.299	1.707	0.901	1.141	0.928
		1.315	1.728	0.764	1.026	0.932	1.186	1.736	0.766	1.118	0.928
β_2	−0.800	−0.924	0.716	0.764	0.773	0.968	−0.919	0.717	0.760	0.769	0.966
		−0.898	0.721	0.772	0.778	0.962	−1.106	0.730	0.792	0.849	0.956
Computation time				1577.3					936.91		
MPLE/Shifted MPLE Zero event percentage = 43.8%											
β_1	2.0	1.795	1.052	0.801	0.826	0.894	1.788	1.048	0.800	0.827	0.892
		1.893	1.103	0.807	0.814	0.924	1.875	1.093	0.800	0.809	0.924
β_2	−0.8	−0.826	0.300	0.304	0.305	0.948	−0.818	0.300	0.301	0.302	0.946
		−0.816	0.300	0.304	0.304	0.944	−0.812	0.300	0.302	0.302	0.944
Jeffreys-type/Shifted Jeffreys-type prior											
β_1	2.0	2.236	1.250	1.162	1.185	0.934	2.230	1.250	1.163	1.185	0.932
		2.294	1.258	1.073	1.112	0.936	2.193	1.254	1.070	1.087	0.934
β_2	−0.8	−0.832	0.301	0.308	0.309	0.958	−0.824	0.301	0.305	0.305	0.958
		−0.806	0.302	0.309	0.309	0.948	−0.865	0.302	0.307	0.313	0.952
Computation time				4794.5					2585.0		

Table 3

Parameter estimates for SEER prostate cancer death and death due to other causes.

	Death due to prostate cancer				Death due to other causes			
Variable	**Est**	**SE**	**HR**	**95% CI/HPD**	**Est**	**SE**	**HR**	**95% CI/HPD**
MPLE/shifted MPLE								
PSA	0.277	0.467	1.319	(0.528, 3.295)	0.083	0.173	1.087	(0.775, 1.524)
	0.227	0.470	1.255	(0.500, 3.152)	0.074	0.173	1.077	(0.767, 1.512)
RP	−3.675	1.687	0.025	(0.001, 0.692)	−1.067	0.386	0.344	(0.162, 0.734)
	−3.484	1.611	0.031	(0.001, 0.721)	−1.049	0.386	0.350	(0.165, 0.746)
RT	−1.098	0.755	0.334	(0.076, 1.464)	−−0.778	0.300	0.459	(0.255, 0.827)
	−0.981	0.752	0.375	(0.086, 1.638)	−0.764	0.300	0.466	(0.259, 0.839)
Black	−0.173	1.012	0.841	(0.116, 6.112)	0.248	0.383	1.282	(0.605, 2.716)
	−0.432	1.107	0.649	(0.074, 5.680)	0.207	0.388	1.230	(0.574, 2.632)
Year diag	−0.210	0.740	0.811	(0.190, 3.459)	−0.166	0.199	0.847	(0.574, 1.251)
	−0.245	0.742	0.782	(0.183, 3.351)	−0.170	0.199	0.843	(0.571, 1.246)
Age	−0.361	0.417	0.697	(0.308, 1.579)	0.570	0.166	1.769	(1.277, 2.450)
	−0.319	0.420	0.727	(0.319, 1.655)	0.569	0.167	1.767	(1.275, 2.449)
Jeffreys-type/shifted Jeffreys-type prior								
PSA	0.262	0.462	1.299	(0.407, 2.800)	0.076	0.175	1.079	(0.733, 1.468)
	0.211	0.465	1.234	(0.362, 2.707)	0.068	0.175	1.070	(0.730, 1.465)
RP	−4.513	2.134	0.011	(0.000, 0.134)	−1.078	0.387	0.340	(0.130, 0.659)
	−4.156	1.957	0.016	(0.000, 0.171)	−1.068	0.386	0.344	(0.133, 0.666)
RT	−1.134	0.718	0.322	(0.028, 1.015)	−0.786	0.297	0.456	(0.229, 0.763)
	−0.996	0.735	0.369	(0.024, 1.171)	−0.775	0.301	0.461	(0.239, 0.779)
Black	−0.410	1.061	0.664	(0.035, 3.040)	0.219	0.380	1.245	(0.448, 2.243)
	−0.683	1.163	0.505	(0.008, 2.500)	0.172	0.398	1.188	(0.425, 2.242)
Year diag	−0.280	0.744	0.756	(0.050, 2.403)	−0.171	0.198	0.843	(0.550, 1.206)
	−0.323	0.765	0.724	(0.035, 2.275)	−0.174	0.197	0.840	(0.547, 1.187)
Age	−0.374	0.416	0.688	(0.240, 1.399)	0.572	0.166	1.772	(1.247, 2.393)
	−0.329	0.416	0.720	(0.268, 1.461)	0.570	0.166	1.768	(1.228, 2.379)
Computation time			93.0				830.0	

Variable	Death due to prostate cancer				Death due to other causes			
	Est	SE	HR	95% CI/HPD	Est	SE	HR	95% CI/HPD
First risk set MPLE/shifted MPLE								
PSA	0.295	0.467	1.343	(0.538, 3.351)	0.085	0.173	1.088	(0.776, 1.526)
	0.222	0.470	1.249	(0.497, 3.139)	0.075	0.173	1.078	(0.768, 1.512)
RP	−3.644	1.677	0.026	(0.001, 0.700)	−1.064	0.386	0.345	(0.162, 0.736)
	−3.493	1.612	0.030	(0.001, 0.716)	−1.046	0.386	0.351	(0.165, 0.748)
RT	−1.086	0.754	0.338	(0.077, 1.479)	−0.778	0.300	0.460	(0.255, 0.827)
	−0.984	0.752	0.374	(0.086, 1.633)	−0.765	0.300	0.465	(0.259, 0.838)
Black	−0.152	1.007	0.859	(0.119, 6.186)	0.248	0.383	1.281	(0.605, 2.715)
	−0.464	1.117	0.629	(0.070, 5.618)	0.214	0.388	1.238	(0.579, 2.647)
Year diag	−0.199	0.739	0.820	(0.193, 3.488)	−0.164	0.199	0.849	(0.575, 1.253)
	−0.240	0.741	0.787	(0.184, 3.364)	−0.166	0.199	0.847	(0.574, 1.251)
Age	−0.353	0.417	0.702	(0.310, 1.591)	0.571	0.166	1.769	(1.277, 2.451)
	−0.324	0.420	0.723	(0.318, 1.647)	0.573	0.167	1.773	(1.279, 2.457)
First risk set Jeffreys-type/shifted Jeffreys-type prior								
PSA	0.282	0.465	1.326	(0.417, 2.866)	0.077	0.174	1.080	(0.744, 1.473)
	0.204	0.468	1.226	(0.384, 2.726)	0.066	0.173	1.068	(0.735, 1.464)
RP	−4.232	1.748	0.015	(0.000, 0.152)	−1.084	0.388	0.338	(0.137, 0.662)
	−4.314	2.196	0.013	(0.000, 0.175)	−1.064	0.384	0.345	(0.134, 0.668)
RT	−1.119	0.749	0.327	(0.027, 1.044)	−0.780	0.297	0.458	(0.220, 0.756)
	−1.016	0.740	0.362	(0.021, 1.177)	−0.771	0.301	0.463	(0.230, 0.774)
Black	−0.374	1.034	0.688	(0.034, 3.071)	0.216	0.380	1.241	(0.469, 2.268)
	−0.676	1.070	0.509	(0.013, 2.397)	0.184	0.389	1.202	(0.420, 2.212)
Year diag	−0.261	0.718	0.771	(0.071, 2.364)	−0.171	0.197	0.843	(0.557, 1.207)
	−0.311	0.790	0.733	(0.039, 2.172)	−0.171	0.197	0.843	(0.550, 1.193)
Age	−0.356	0.415	0.700	(0.247, 1.424)	0.572	0.166	1.771	(1.253, 2.403)
	−0.337	0.417	0.714	(0.241, 1.424)	0.575	0.165	1.776	(1.254, 2.407)
Computation time			39.3				242.1	