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

Jing Wu

Department of Computer Science & Statistics, University of Rhode Island,
Kingston, RI 02881, USA
E-Mail: jing_wu@uri.edu

Joint work with Mário de Castro, Elizabeth D. Schifano, and Ming-Hui Chen

Outline

Model Identifiability in Practice

- In clinical trials involving time-to-event data, it is often the case that patients within at least one arm of the trial will experience very few events.
- This could be due to the length of the trial, but it could also be due to the nature of the trial itself.

Model Identifiability in Practice

- Among the particular subset of SEER prostate cancer data we are interested in, no patients receiving surgery treatment encountered the event.

Model Identifiability in Practice

- Among the particular subset of SEER prostate cancer data we are interested in, no patients receiving surgery treatment encountered the event.

- If one wishes to analyze the surgery treatment effect of the time-to-event data in one of our motivating prostate cancer studies, then the zero event in surgery treatment group will lead to the model identifiability issue.

The Monotone Likelihood Problem

- Standard analysis of the Cox proportional hazards model (Cox, 1972) involves parameter estimation through maximization of the logarithm of the partial likelihood function.

The Monotone Likelihood Problem

- Standard analysis of the Cox proportional hazards model (Cox, 1972) 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 $+\infty$.

The Monotone Likelihood Problem

- Standard analysis of the Cox proportional hazards model (Cox, 1972) 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.

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.

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.

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.
- Many issues are still not well understood, and need to be further studied from both practical and theoretical points of view.

I. The Prostate Cancer (PC) SEER Data

- We consider 1,840 men subjects from the SEER prostate cancer data between 1973 to 2013, 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.

I. The Prostate Cancer (PC) SEER Data

- We consider 1,840 men subjects from the SEER prostate cancer data between 1973 to 2013, 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.
- The outcome variable in years was the time to death due to prostate cancer or other causes, which is right censored and continuous.
- 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).

I. A Preliminary Analysis of the PC SEER Data

- The cause-specific proportional hazards competing risk model was fit and the MPLEs produced by SAS procedure PHREG are given in Table 1.
- The results in Table 1 indicate that RP covariate is not identifiable for the death caused by prostate cancer, which is due to the 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.

Table 1: Maximum Partial Likelihood Estimates (MPLEs) for the PC SEER Data

Prostate Cancer Death					
Variable	Number of censored	Number of death	Est	SE	*p*-value
PSA	1840		0.253	0.468	0.5881
RP	842	0	-17.745	1680	0.9916
RT	576	3	-1.150	0.742	0.1210
Black	279	1	-0.539	1.125	0.6318
Year_diag	1840		-0.377	0.743	0.6118
Age	1840		-0.372	0.416	0.3712
Other Cause Death					
Variable	Number of censored	Number of death	Est	SE	*p*-value
PSA	1840		0.074	0.173	0.6657
RP	842	13	-1.082	0.388	0.0053
RT	576	19	-0.785	0.300	0.0089
Black	279	8	0.198	0.391	0.6132
Year_diag	1840		-0.204	0.197	0.3009
Age	1840		0.575	0.166	0.0006

II. The Prostate Cancer Data

- The study cohort in Chen et al. (2009) included 550 men suffering from localized prostate cancer with at least 1 risk factor:
  - clinical tumor stage T2b or T2c,
  - gleason score 7 to 10, and
  - prostate-specific antigen (PSA) level > 10 ng/mL.
- They were treated with radiation therapy combined with short-term androgen suppression therapy between 1989 and 2002.

II. The Prostate Cancer Data

- The study cohort in Chen et al. (2009) included 550 men suffering from localized prostate cancer with at least 1 risk factor:
  - clinical tumor stage T2b or T2c,
  - gleason score 7 to 10, and
  - prostate-specific antigen (PSA) level > 10 ng/mL.
- They were treated with radiation therapy combined with short-term androgen suppression therapy between 1989 and 2002.
- The outcome variable in years was the time to death due to prostate cancer or other causes, which is right censored and continuous.
- Define $A = 1\{\text{PSA} > 10\}$, $B = 1\{\text{Gleason} > 7\}$, and $C = 1\{\text{T2b or T2c}\}$.
- In Chen et al. (2009), they considered five covariates: AB, AC, BC, ABC, and age.

II. A Preliminary Analysis of the PC Data

- The cause-specific proportional hazards competing risk model was fit and the MPLEs produced by SAS procedure PHREG are given in Table 2.
- The results in Table 2 indicate that the AC covariate is not identifiable for the death caused by prostate cancer, which is due to the absence of events (prostate cancer death) in the "only AC not B" group.
- This case study motivates us to carry out further examination of monotone partial likelihoods.

Table 2: Maximum Partial Likelihood Estimates (MPLEs) for the PC Data

Prostate Cancer Death					
Variable	Number of censored	Number of death	Est	SE	p-value
Only one of A, B, C	233	1			
Only AB not C	109	2	0.398	1.234	0.7472
Only AC not B	34	0	-14.303	2107	0.9946
Only BC not A	59	1	0.591	1.227	0.6303
ABC	64	7	2.222	0.805	0.0058
Age	499	11	0.023	0.048	0.6390
Other Cause Death					
Variable	Number of censored	Number of death	Est	SE	p-value
Only one of A, B, C	233	20			
Only AB not C	109	7	-0.163	0.439	0.7103
Only AC not B	34	1	-1.198	1.025	0.2426
Only BC not A	59	5	-0.167	0.500	0.7392
ABC	64	7	0.079	0.441	0.8585
Age	499	40	0.060	0.029	0.0360

Cox Model

- Let y_i denote the right-censored failure time and let δ_i denote the failure indicator such that $\delta_i = 1$ if y_i is a failure time and 0 otherwise.
- Let $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip})'$ be the $p \times 1$ vector of covariates and $\boldsymbol{\beta}$ the $p \times 1$ vector of regression coefficients.
- In addition, n denotes the total number of observations and $\mathcal{R}(t) = \{i : y_i \geq t\}$ is the set of subjects at risk at time t.

Cox Model

- Let y_i denote the right-censored failure time and let δ_i denote the failure indicator such that $\delta_i = 1$ if y_i is a failure time and 0 otherwise.
- Let $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip})'$ be the $p \times 1$ vector of covariates and $\boldsymbol{\beta}$ the $p \times 1$ vector of regression coefficients.
- In addition, n denotes 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

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

where $\mathcal{D}_{\text{obs}} = \{(y_i, \delta_i, \boldsymbol{x}_i) : i = 1, \ldots, n\}$ denotes the observed data.

The Cause-Specific Hazards Model

- One generalization of the Cox model is the cause-specific hazards model which was discussed in Gaynor et al. (1993) and Ge and Chen (2012).

The Cause-Specific Hazards Model

- One 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 for cause j is defined by
$$h_{Cj}(t|\boldsymbol{x}_i) = \lim_{\Delta t \to 0} \frac{\Pr(t \le T < t + \Delta t,\ \delta = j | T \ge t, \boldsymbol{x}_i)}{\Delta t}.$$
- Assume the Cox proportional hazards structure for $h_{Cj}(t|\boldsymbol{x}_i)$, i.e.,
$$h_{Cj}(t|\boldsymbol{x}_i) = h_{Cj0}(t) \exp(\boldsymbol{x}'\boldsymbol{\beta}_j).$$
- The likelihood function given $\mathcal{D}_{\text{obs}}$ is
$$\begin{aligned} &L_C(\boldsymbol{\beta}, \boldsymbol{h}_{C0}|\mathcal{D}_{\text{obs}}) \\ &= \prod_{j=1}^{J} \prod_{i=1}^{n} \{h_{Cj0}(y_i) \exp(\boldsymbol{x}_i'\boldsymbol{\beta}_j)\}^{1\{\delta_i = j\}} \exp\{-H_{Cj0}(y_i) \exp(\boldsymbol{x}_i'\boldsymbol{\beta}_j)\}, \end{aligned}$$
where $H_{Cj0}(y_i) = \int_0^{y_i} h_{Cj0}(u) du$ for $j = 1, \ldots, J$, $\boldsymbol{h}_{C0} = (h_{C10}, \ldots, h_{CJ0})'$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \ldots, \boldsymbol{\beta}_J')'$, and $1\{\delta_i = j\}$ is the indicator function taking a value of 1 if $\delta_i = j$ and 0 otherwise, with $\delta_i = 0$ denoting a right-censored observation.

Partial Likelihood under the Cause-Specific Hazards Model

- Using the conditional probability argument or the profile likelihood approach (e.g., Klein and Moeschberger, 2003), it is easy to show that the partial likelihood is given by

$$L_{Cp}(\boldsymbol{\beta}|\mathcal{D}_{\text{obs}}) = \prod_{j=1}^{J}\prod_{i=1}^{n}\left\{\frac{\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_j)}{\sum_{\ell\in\mathcal{R}(y_i)}\exp(\boldsymbol{x}_\ell'\boldsymbol{\beta}_j)}\right\}^{1\{\delta_i=j\}}.$$

Partial Likelihood under the Cause-Specific Hazards Model

- Using the conditional probability argument or the profile likelihood approach (e.g., Klein and Moeschberger, 2003), it is easy to show that the partial likelihood is given by

$$L_{Cp}(\boldsymbol{\beta}|\mathcal{D}_{\text{obs}}) = \prod_{j=1}^{J}\prod_{i=1}^{n}\left\{\frac{\exp(\boldsymbol{x}_i'\boldsymbol{\beta}_j)}{\sum_{\ell\in\mathcal{R}(y_i)}\exp(\boldsymbol{x}_\ell'\boldsymbol{\beta}_j)}\right\}^{1\{\delta_i=j\}}.$$

- Due to similarity of the partial likelihoods under these two models, we only focus on the conditions for the existence of MPLE and posterior propriety under the Cox model throughout.

Conditions for the Existence of MPLE and Posterior Propriety

- Define $\boldsymbol{X}^*$ to be

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

- Let k_i denote the number of subjects in $\mathcal{R}(y_i)$ and $K = \sum_{i=1}^n k_i$. Then, $\boldsymbol{X}^*$ is a $K \times p$ matrix.

Conditions for the Existence of MPLE and Posterior Propriety

- Define $\boldsymbol{X}^*$ to be

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

- Let k_i denote the number of subjects in $\mathcal{R}(y_i)$ 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 $\boldsymbol{\beta}$ are given by

C1. (Multicolliearity) The matrix $\boldsymbol{X}^*$ is of full column rank and

C2. (Separation) There exists a positive vector $\boldsymbol{v}$ such that $\boldsymbol{X}^{*'}\boldsymbol{v} = \boldsymbol{0}$.

Conditions for the Existence of MPLE and Posterior Propriety

- Define $\boldsymbol{X}^*$ to be

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

- Let k_i denote the number of subjects in $\mathcal{R}(y_i)$ 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 $\boldsymbol{\beta}$ are given by

C1. (Multicolliearity) The matrix $\boldsymbol{X}^*$ is of full column rank and

C2. (Separation) There exists a positive vector $\boldsymbol{v}$ such that $\boldsymbol{X}^{*'}\boldsymbol{v} = \boldsymbol{0}$.

- If C1 holds, C2 is the "iff" condition for the existence of MPLE.

Conditions for the Existence of MPLE and Posterior Propriety

- 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.

  C1$'$. The matrix $\boldsymbol{X}_s^*$ is of full column rank and

  C2$'$. There exists a positive vector $\boldsymbol{v}$ such that $\boldsymbol{X}_s^{*'}\boldsymbol{v} = \mathbf{0}$.

- C1' and C2' are sufficient for C1 and C2.

The Conditions for Survival Data with Binary Covariates

- Let $\{y_{(i)}\}$ denote the rearranged $\{y_i\}$ in the ascending order associated with $\{\boldsymbol{x}_{(i)}\}$ and $\{\delta_{(i)}\}$, where each x_{ij} takes a value of 0 or 1.
- Let $i_0 = \operatorname*{argmin}_i \{\delta_{(i)} = 1\}$ denote the index for the first event.

The Conditions for Survival Data with Binary Covariates

- Let $\{y_{(i)}\}$ denote the rearranged $\{y_i\}$ in the ascending order associated with $\{\boldsymbol{x}_{(i)}\}$ and $\{\delta_{(i)}\}$, where each x_{ij} takes a value of 0 or 1.
- Let $i_0 = \operatorname*{argmin}_i \{\delta_{(i)} = 1\}$ denote the index for the first event.
- We must have $n \geq i_0 + p$ in order to satisfy C1.

The Conditions for Survival Data with Binary Covariates

- Let $\{y_{(i)}\}$ denote the rearranged $\{y_i\}$ in the ascending order associated with $\{\boldsymbol{x}_{(i)}\}$ and $\{\delta_{(i)}\}$, where each x_{ij} takes a value of 0 or 1.
- Let $i_0 = \underset{i}{\operatorname{argmin}}\{\delta_{(i)} = 1\}$ denote the index for the first event.
- We must have $n \geq i_0 + p$ in order to satisfy C1.
- For binary covaraite data, the minimum sample size required for both C1 and C2 to hold is $n = i_0 + p + 1$ with at least 2 events.
- In addition, $\{\boldsymbol{x}_{(i)}\}$ should not take monotone values, e.g., $1, \ldots, 1, 0, \ldots, 0$.

How Many Events are required?

- When $p = 1$, we should have two events with one in each arm and then followed by one observation whose covariate takes the same value as the first event.
- For example, if the first two events take values $x_1 = 1$ and $x_2 = 0$, the following observation (can be either censored or died) should have $x_3 = x_1 = 1$. Thus, we have $n = i_0 + 2$ and both C1 and C2 hold.

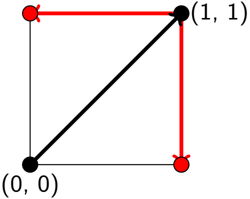
How Many Events are required?

- When $p = 1$, we should have two events with one in each arm and then followed by one observation whose covariate takes the same value as the first event.
- For example, if the first two events take values $x_1 = 1$ and $x_2 = 0$, the following observation (can be either censored or died) 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 and then followed by some observations without monotone covariate pattern such that $n = i_0 + p + 1$.

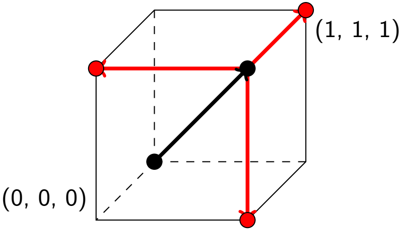
How Many Events are required?

- When $p = 1$, we should have two events with one in each arm and then followed by one observation whose covariate takes the same value as the first event.
- For example, if the first two events take values $x_1 = 1$ and $x_2 = 0$, the following observation (can be either censored or died) 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 and then followed by some observations without monotone covariate pattern such that $n = i_0 + p + 1$.
- Furthermore, if the number of events is exactly two, the two events should be in two completely opposite arms. For example in Figure 1 ($p = 3$), if $x_1 = (0, 0, 0)$ and there are only two events, the other event must occur in the arm $x_2 = (1, 1, 1)$ in order to satisfy both C1 and C2 conditions.

Figure 1: Diagrams for p=2 and $p = 3$



(a) $p = 2$



(b) $p = 3$

- The black vertices refer to the covariates of the two events while the red vertices stand for the covariates of the subsequent observations (can be either censored or not).

Penalized Maximum Partial Likelihood Estimation

- 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 partial likelihood function.
- The estimate $\widehat{\boldsymbol{\beta}}^*$ is obtained as the solution of the penalized score equation

$$U^*(\boldsymbol{\beta}) = \frac{\partial L_p(\boldsymbol{\beta}|\mathcal{D}_{\text{obs}})}{\partial \boldsymbol{\beta}} + \frac{1}{2}\,\text{trace}\left\{ \boldsymbol{I}(\boldsymbol{\beta})^{-1} \frac{\partial I(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \right\} = 0,$$

where $I(\boldsymbol{\beta})$ denotes the information matrix.

The Information Matrix

$$
\begin{aligned}
\boldsymbol{I}(\boldsymbol{\beta}) &= -\frac{\partial^2 \log\{L_p(\boldsymbol{\beta}|\mathcal{D}_{\text{obs}})\}}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'} \\
&= \sum_{i=1}^{n} \delta_i \left\{ \sum_{j_i \in \mathcal{R}(y_i)} w_{ij_i} \boldsymbol{x}_{j_i} \boldsymbol{x}_{j_i}' - \left(\sum_{j_i \in \mathcal{R}(y_i)} w_{ij_i} \boldsymbol{x}_{j_i} \right) \left(\sum_{m_i \in \mathcal{R}(y_i)} w_{im_i} \boldsymbol{x}_{m_i} \right)' \right\} \\
&= \sum_{i=1}^{n} \delta_i \boldsymbol{A}_i,
\end{aligned}
$$

where $w_{ij} = \exp(\boldsymbol{x}_j'\boldsymbol{\beta}) / \sum_{l \in \mathcal{R}(y_i)} \exp(\boldsymbol{x}_l'\boldsymbol{\beta})$ and

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

Bayesian Formulation and Properties of Jeffreys-type prior

- Jeffreys-type prior is given by

$$\pi(\boldsymbol{\beta}) \propto |I(\boldsymbol{\beta})|^{\frac{1}{2}}.$$

- The posterior is of the form

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

Bayesian Formulation and Properties of Jeffreys-type prior

- Jeffreys-type prior is given by

$$\pi(\boldsymbol{\beta}) \propto |I(\boldsymbol{\beta})|^{\frac{1}{2}}.$$

- The posterior is of the form

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

- Theorem 1: *If Condition C1 holds, then the Jeffreys-type exists and is proper. Otherwise,* $\pi(\boldsymbol{\beta})$ *is not defined.*
- Corollary 1: *The Jeffreys-type has finite modes. Since the profile likelihood function is bounded above, the posterior distribution has finite modes too.*
- Proposition 1: *The Jeffreys-type has lighter tails than a p-dimensional multivariate t distribution with v degrees of freedom for* $v > 0$*, and heavier tails than a p-dimensional multivariate normal distribution.*

Remark 4.1: Heinze and Schemper's Conditions

- Heinze and Schemper (2001, p. 116) presented two sufficient conditions for the existence of finite estimates of $\boldsymbol{\beta}$ using the penalized partial likelihood.
- One of these conditions requires at least p distinct failure times.
- Our condition C1 is much weaker and is illustrated in Example 4.1 below.

Example 4.1: An Illustration

- Consider a dataset with $n = 3$ observations, $p = 2$ covariates, $\boldsymbol{x}_1 = (0, 1)'$, $\boldsymbol{x}_2 = (1, 0)'$, $\boldsymbol{x}_3 = (1, 1)'$, $\boldsymbol{\delta} = (1, 0, 0)'$, $y_1 < y_2$, and $y_1 < y_3$.
- Since there is only one failure time, Heinze and Schemper's conditions do not hold.
- In this case, $\mathcal{R}(y_1) = \{1, 2, 3\}$, and $\boldsymbol{X}^*$ has lines $(0, 0)$, $(1, -1)$, and $(1, 0)$, so that $\mathrm{rank}(\boldsymbol{X}^*) = 2$ and condition C1 holds.
- 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 $w_3 = e^{\beta_1+\beta_2}/(e^{\beta_1} + e^{\beta_2} + e^{\beta_1+\beta_2})$.
- Since $L_p(\boldsymbol{\beta}|\mathcal{D}_{\mathrm{obs}}) = w_1$, Condition C2 is not satisfied and, thus, the partial likelihood function is monotone as shown in Figure 2(a).

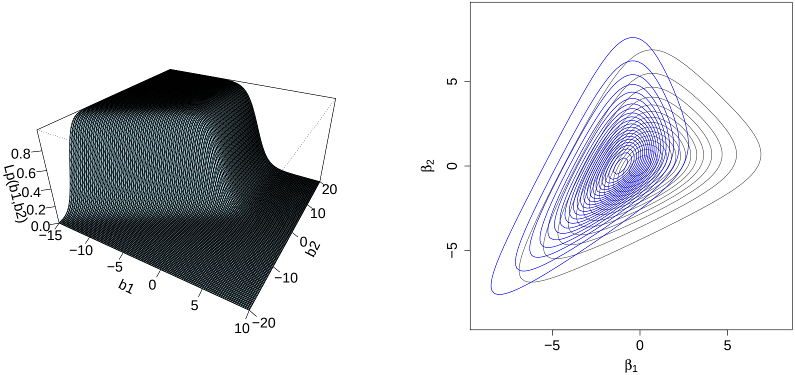
Example 4.1: An Illustration

- Consider a dataset with $n = 3$ observations, $p = 2$ covariates, $\boldsymbol{x}_1 = (0, 1)'$, $\boldsymbol{x}_2 = (1, 0)'$, $\boldsymbol{x}_3 = (1, 1)'$, $\boldsymbol{\delta} = (1, 0, 0)'$, $y_1 < y_2$, and $y_1 < y_3$.
- Since there is only one failure time, Heinze and Schemper's conditions do not hold.
- In this case, $\mathcal{R}(y_1) = \{1, 2, 3\}$, and $\boldsymbol{X}^*$ has lines $(0, 0)$, $(1, -1)$, and $(1, 0)$, so that $\text{rank}(\boldsymbol{X}^*) = 2$ and condition C1 holds.
- 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 $w_3 = e^{\beta_1+\beta_2}/(e^{\beta_1} + e^{\beta_2} + e^{\beta_1+\beta_2})$.
- Since $L_p(\boldsymbol{\beta}|\mathcal{D}_{\text{obs}}) = w_1$, Condition C2 is not satisfied and, thus, the partial likelihood function is monotone as shown in Figure 2(a).
- After some algebra, we obtain

$$\boldsymbol{I}(\boldsymbol{\beta}) = \boldsymbol{A}_1 = \begin{bmatrix} w_1(1 - w_1) & -w_1 w_2 \\ -w_1 w_2 & w_2(1 - w_2) \end{bmatrix}.$$

- Hence, Jeffreys-type is $\pi(\boldsymbol{\beta}) \propto |\boldsymbol{I}(\boldsymbol{\beta})|^{1/2} = e^{\beta_1+\beta_2}/(e^{\beta_1} + e^{\beta_2} + e^{\beta_1+\beta_2})^{3/2}$, which is proper.
- Figure 2(b) shows the contour plots of the prior and posterior distributions of $(\beta_1, \beta_2)'$.

Example 4.1: (a) Partial likelihood; (b) Contour plots of the prior (grey) and posterior (blue) distributions for $(\beta_1, \beta_2)'$



(a) (b)

Shifted Jeffreys-type

- The shifting idea is motivated from the variable selection point of view.
- Tibshirani (1996) proposed the least absolute shrinkage and selection operator (lasso), which aims to improve the overall prediction accuracy as well as selecting the most significant covariates by shrinking to 0 for some covariates.

Shifted Jeffreys-type

- The shifting idea is motivated from the variable selection point of view.
- Tibshirani (1996) proposed the least absolute shrinkage and selection operator (lasso), which aims to improve the overall prediction accuracy as well as selecting the most significant covariates by shrinking to 0 for some covariates.
- Let $\boldsymbol{\beta}_M$ be a mode of the Jeffreys-type.
- A shifted Jeffreys-type is given by

$$\pi_s(\boldsymbol{\beta}) \propto \left|\boldsymbol{I}(\boldsymbol{\beta} + \boldsymbol{\beta}_M)\right|^{1/2},$$

so that the mode is shifted to $\boldsymbol{\beta} = \mathbf{0}$.

Jeffreys-type Based on the First Risk Set

- Recall that $i_0 = \underset{i}{\text{argmin}}\{\delta_{(i)} = 1\}$. Let $\boldsymbol{X}^*_{(\boldsymbol{i_0})}$ be the submatrix of $\boldsymbol{X}^*$ corresponding to the first risk set.

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

- Theorem 2: *Condition C1 holds if and only if C1$''$ holds.*

Jeffreys-type Based on the First Risk Set

- Recall that $i_0 = \operatorname*{argmin}_i \{\delta_{(i)} = 1\}$. Let $\boldsymbol{X}^*_{(\boldsymbol{i_0})}$ be the submatrix of $\boldsymbol{X}^*$ corresponding to the first risk set.

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

- Theorem 2: *Condition C1 holds if and only if C1$''$ holds.*
- WLOG, we assume $i_0 = 1$.
- We propose a new variation of Jeffreys-type which only depends on the first risk set, given by

$$\pi_f(\boldsymbol{\beta}) \propto |A_1|^{1/2},$$

where $A_1 =$
$\sum_{j_1 \in \mathcal{R}(y_1)} w_{1j_1} \Big\{ \sum_{\ell_1 \in \mathcal{R}(y_1)} w_{1\ell_1} (\boldsymbol{x}_{j_1} - \boldsymbol{x}_{\ell_1}) \Big\} \Big\{ \sum_{m_1 \in \mathcal{R}(y_1)} w_{1m_1} (\boldsymbol{x}_{j_1} - \boldsymbol{x}_{m_1}) \Big\}'$.

Jeffreys-type Based on the First Risk Set

- Recall that $i_0 = \underset{i}{\text{argmin}}\{\delta_{(i)} = 1\}$. Let $\boldsymbol{X}^*_{(\boldsymbol{i_0})}$ be the submatrix of $\boldsymbol{X}^*$ corresponding to the first risk set.

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

- Theorem 2: *Condition C1 holds if and only if C1$''$ holds.*
- WLOG, we assume $i_0 = 1$.
- We propose a new variation of Jeffreys-type which only depends on the first risk set, given by
$$\pi_f(\boldsymbol{\beta}) \propto |A_1|^{1/2},$$
where $A_1 =$
$\sum_{j_1 \in \mathcal{R}(y_1)} w_{1j_1} \left\{ \sum_{\ell_1 \in \mathcal{R}(y_1)} w_{1\ell_1} (\boldsymbol{x}_{j_1} - \boldsymbol{x}_{\ell_1}) \right\} \left\{ \sum_{m_1 \in \mathcal{R}(y_1)} w_{1m_1} (\boldsymbol{x}_{j_1} - \boldsymbol{x}_{m_1}) \right\}'$.
- Proposition 2: *The Jeffreys-type based on the whole dataset exists and is proper iff $\pi_f(\boldsymbol{\beta})$ exists and is proper.*

Jeffreys-type Based on the First Risk Set

- Recall that $i_0 = \underset{i}{\text{argmin}}\{\delta_{(i)} = 1\}$. Let $\boldsymbol{X}^*_{(\boldsymbol{i_0})}$ be the submatrix of $\boldsymbol{X}^*$ corresponding to the first risk set.

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

- Theorem 2: *Condition C1 holds if and only if C1$''$ holds.*
- WLOG, we assume $i_0 = 1$.
- We propose a new variation of Jeffreys-type which only depends on the first risk set, given by

$$\pi_f(\boldsymbol{\beta}) \propto |A_1|^{1/2},$$

where $A_1 =$
$\sum_{j_1 \in \mathcal{R}(y_1)} w_{1j_1} \left\{ \sum_{\ell_1 \in \mathcal{R}(y_1)} w_{1\ell_1} (\boldsymbol{x}_{j_1} - \boldsymbol{x}_{\ell_1}) \right\} \left\{ \sum_{m_1 \in \mathcal{R}(y_1)} w_{1m_1} (\boldsymbol{x}_{j_1} - \boldsymbol{x}_{m_1}) \right\}'$.

- Proposition 2: *The Jeffreys-type based on the whole dataset exists and is proper iff $\pi_f(\boldsymbol{\beta})$ exists and is proper.*
- The first risk set Jeffreys-type leads to a substantial gain in computation.

A Simple Localized Metropolis Algorithm

- Let $\pi(\beta_j|\boldsymbol{\beta}_{(-j)}, \mathcal{D}_{\text{obs}})$ denote the full conditional distribution of β_j given $\boldsymbol{\beta}_{(-j)}$ and $\mathcal{D}_{\text{obs}}$, where $\boldsymbol{\beta}_{(-j)}$ is $\boldsymbol{\beta}$ with the j^{th} component deleted.
- The Localized Metropolis Algorithm to sample β_j operates as follows:

Step 1. Let β_j be the current value.

Step 2. Compute $\hat{\beta}_j = \arg\max_{\beta_j} \log \pi(\beta_j|\boldsymbol{\beta}_{(-j)}, \mathcal{D}_{\text{obs}})$ and use a quadratic regression to compute a by approximating $y = ax^2 + bx + c$ to $\log \pi(\beta_j|\boldsymbol{\beta}_{(-j)}, \mathcal{D}_{\text{obs}})$ in the neighborhood of $\hat{\beta}_j$.

Step 3. Draw β_j^* from $N(\hat{\beta}_j, -\frac{1}{2a})$.

Step 4. A move from β_j to β_j^* is made with probability $\alpha = \min\left\{\frac{\pi(\beta_j^*|\boldsymbol{\beta}_{(-j)}, \mathcal{D}_{\text{obs}})q(\beta_j)}{\pi(\beta_j|\boldsymbol{\beta}_{(-j)}, \mathcal{D}_{\text{obs}})q(\beta_j^*)}, 1\right\}$, where $q(\cdot)$ is the density function of $N(\hat{\beta}_j, -\frac{1}{2a})$.

Simulation Design

- In the data generation, we first generate $n = 100$ independent $x_{i1} \sim \text{Bernoulli}(0.9)$ and $x_{i2} \sim \text{Bernoulli}(0.5)$.
- 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.

Simulation Design

- The failure times are subject to administrative censoring with duration set to 5.0, 30.0 in order to achieve average censoring rates around 90%, and 50%.
- The percentages of zero events corresponding to $x_1 = 0$ amount to 86.6%, and 43.8%, respectively.

Simulation Design

- The failure times are subject to administrative censoring with duration set to 5.0, 30.0 in order to achieve average censoring rates around 90%, and 50%.
- The percentages of zero events corresponding to $x_1 = 0$ amount to 86.6%, and 43.8%, respectively.
- For censoring percentage 86.6%, the total computation time of the four approaches was 1577.3 minutes using all the data and only 936.9 minutes if we just use the first risk set.

Simulation Design

- The failure times are subject to administrative censoring with duration set to 5.0, 30.0 in order to achieve average censoring rates around 90%, and 50%.
- The percentages of zero events corresponding to $x_1 = 0$ amount to 86.6%, and 43.8%, respectively.
- For censoring percentage 86.6%, the total computation time of the four approaches was 1577.3 minutes using all the data and only 936.9 minutes if we just use the first risk set.
- For censoring percentage 43.8%, the total computation time of the four approaches was 4794.5 minutes using all the data and only 2585.0 minutes if we just use the first risk set.

Table 3: MPLEs Based on 500 Replications with Censoring Percentage 86.6%

Variable	True	Est	SE	SimE	RMSE	CP
MPLE/ Shifted MPLE			All data			
β_1	2.000	0.647	1.436	0.531	1.453	0.870
		1.045	1.724	0.523	1.088	0.926
β_2	-0.800	-0.861	0.695	0.670	0.672	0.988
		-0.833	0.694	0.673	0.673	0.988
MPLE/ Shifted MPLE			First risk set			
β_1	2.000	0.642	1.432	0.529	1.458	0.870
		1.019	1.702	0.514	1.107	0.922
β_2	-0.800	-0.855	0.694	0.666	0.668	0.988
		-0.830	0.693	0.671	0.671	0.988

- True: true value; Est: estimate; SE: average of thestandard errors;
- SimE: simulation standard error; RMSE: root mean squared error; and CP: coverage probability of the 95% confidence interval.

Table 4: Posterior Summaries Based on 500 Replications with Censoring Percentage 86.6%

Variable	True	Est	SD	SimE	RMSE	CP
Jeffreys-type/ Shifted Jeffreys-type				All data		
β_1	2.000	1.294	1.706	0.854	1.107	0.926
		1.315	1.728	0.764	1.026	0.932
β_2	-0.800	-0.924	0.716	0.764	0.773	0.968
		-0.898	0.721	0.772	0.778	0.962
Jeffreys-type/ Shifted Jeffreys-type				First risk set		
β_1	2.000	1.299	1.707	0.901	1.141	0.928
		1.186	1.736	0.766	1.118	0.928
β_2	-0.800	-0.919	0.717	0.760	0.769	0.966
		-1.106	0.730	0.792	0.849	0.956

- True: true value; Est: posterior estimate; SE: average of the posterior standard deviations;
- SimE: simulation standard error; RMSE: root mean squared error; and CP: coverage probability of the 95% HPD interval.

Table 5: MPLEs Based on 500 Replications with Censoring Percentage 43.8%

Variable	True	Est	SE	SimE	RMSE	CP
MPLE/ Shifted MPLE			All data			
β_1	2.000	1.795	1.052	0.801	0.826	0.894
		1.893	1.103	0.807	0.814	0.924
β_2	-0.800	-0.826	0.300	0.304	0.305	0.948
		-0.816	0.300	0.304	0.304	0.944
MPLE/ Shifted MPLE			First risk set			
β_1	2.000	1.788	1.048	0.800	0.827	0.892
		1.875	1.093	0.800	0.809	0.924
β_2	-0.800	-0.818	0.300	0.301	0.302	0.946
		-0.812	0.300	0.302	0.302	0.944

- True: true value; Est: estimate; SE: average of thestandard errors;
- SimE: simulation standard error; RMSE: root mean squared error; and CP: coverage probability of the 95% confidence interval.

Table 6: Posterior Summaries Based on 500 Replications with Censoring Percentage 43.8%

Variable	True	Est	SD	SimE	RMSE	CP
Jeffreys-type/			All data			
Shifted Jeffreys-type						
β_1	2.000	2.236	1.250	1.162	1.185	0.934
		2.294	1.258	1.073	1.112	0.936
β_2	-0.800	-0.832	0.301	0.308	0.309	0.958
		-0.806	0.302	0.309	0.309	0.948
Jeffreys-type/			First risk set			
Shifted Jeffreys-type						
β_1	2.000	2.230	1.250	1.163	1.185	0.932
		2.193	1.254	1.070	1.087	0.934
β_2	-0.800	-0.824	0.301	0.305	0.305	0.958
		-0.865	0.302	0.307	0.313	0.952

- True: true value; Est: posterior estimate; SE: average of the posterior standard deviations;
- SimE: simulation standard error; RMSE: root mean squared error; and CP: coverage probability of the 95% HPD interval.

Table 1: Maximum Partial Likelihood Estimates (MPLEs) for the PC SEER Data

Prostate Cancer Death					
Variable	Number of censored	Number of death	Est	SE	*p*-value
PSA	1840		0.253	0.468	0.5881
RP	842	0	-17.745	1680	0.9916
RT	576	3	-1.150	0.742	0.1210
Black	279	1	-0.539	1.125	0.6318
Year_diag	1840		-0.377	0.743	0.6118
Age	1840		-0.372	0.416	0.3712

Table 7: MPLEs for PC Death for the PC SEER Data

Variable	EST	SE	HR	95% CI
MPLE/Shifted MPLE		All data		
PSA	0.277	0.467	1.319	(0.528, 3.295)
	0.227	0.470	1.255	(0.500, 3.152)
RP	-3.675	1.687	0.025	(0.001, 0.692)
	-3.484	1.611	0.031	(0.001, 0.721)
RT	-1.098	0.755	0.334	(0.076, 1.464)
	-0.981	0.752	0.375	(0.086, 1.638)
Black	-0.173	1.012	0.841	(0.116, 6.112)
	-0.432	1.107	0.649	(0.074, 5.680)
Year_diag	-0.210	0.740	0.811	(0.190, 3.459)
	-0.245	0.742	0.782	(0.183, 3.351)
Age	-0.361	0.417	0.697	(0.308, 1.579)
	-0.319	0.420	0.727	(0.319, 1.655)
MPLE/Shifted MPLE		First risk set		
PSA	0.295	0.467	1.343	(0.538, 3.351)
	0.222	0.470	1.249	(0.497, 3.139)
RP	-3.644	1.677	0.026	(0.001, 0.700)
	-3.493	1.612	0.030	(0.001, 0.716)
RT	-1.086	0.754	0.338	(0.077, 1.479)
	-0.984	0.752	0.374	(0.086, 1.633)
Black	-0.152	1.007	0.859	(0.119, 6.186)
	-0.464	1.117	0.629	(0.070, 5.618)
Year_diag	-0.199	0.739	0.820	(0.193, 3.488)
	-0.240	0.741	0.787	(0.184, 3.364)
Age	-0.353	0.417	0.702	(0.310, 1.591)
	-0.324	0.420	0.723	(0.318, 1.647)

Table 8: Posterior Estimates for PC Death for the PC SEER Data

Variable	EST	SE	HR	95% HPD
Jeffreys-type/ Shifted Jeffreys-type prior		All data		
PSA	0.262	0.462	1.299	(0.407, 2.800)
	0.211	0.465	1.234	(0.362, 2.707)
RP	-4.513	2.134	0.011	(0.000, 0.134)
	-4.156	1.957	0.016	(0.000, 0.171)
RT	-1.134	0.718	0.322	(0.028, 1.015)
	-0.996	0.735	0.369	(0.024, 1.171)
Black	-0.410	1.061	0.664	(0.035, 3.040)
	-0.683	1.163	0.505	(0.008, 2.500)
Year_diag	-0.280	0.744	0.756	(0.050, 2.403)
	-0.323	0.765	0.724	(0.035, 2.275)
Age	-0.374	0.416	0.688	(0.240, 1.399)
	-0.329	0.416	0.720	(0.268, 1.461)
Jeffreys-type/ Shifted Jeffreys-type prior		First risk set		
PSA	0.282	0.465	1.326	(0.417, 2.866)
	0.204	0.468	1.226	(0.384, 2.726)
RP	-4.232	1.748	0.015	(0.000, 0.152)
	-4.314	2.196	0.013	(0.000, 0.175)
RT	-1.119	0.749	0.327	(0.027, 1.044)
	-1.016	0.740	0.362	(0.021, 1.177)
Black	-0.374	1.034	0.688	(0.034, 3.071)
	-0.676	1.070	0.509	(0.013, 2.397)
Year_diag	-0.261	0.718	0.771	(0.071, 2.364)
	-0.311	0.790	0.733	(0.039, 2.172)
Age	-0.356	0.415	0.700	(0.247, 1.424)
	-0.337	0.417	0.714	(0.241, 1.424)

Concluding Remarks

- We have thoroughly investigated the conditions of the monotone partial likelihood and developed equivalent sufficient/necessary conditions based on the first risk set.
- We have showed that Jeffreys-type has finite modes and lighter tail than a multivariate t distribution and heavier tail than a multivariate normal distribution.

Concluding Remarks

- We have thoroughly investigated the conditions of the monotone partial likelihood and developed equivalent sufficient/necessary conditions based on the first risk set.
- We have showed that Jeffreys-type has finite modes and lighter tail than a multivariate t distribution and heavier tail than a multivariate normal distribution.
- Based on the idea of the first risk set and variable selection, we have proposed two variations of Jeffreys-type, i.e., Jeffreys-type based on the first risk set, which has led to huge computational saving, and shifted Jeffreys-type.

Concluding Remarks

- We have thoroughly investigated the conditions of the monotone partial likelihood and developed equivalent sufficient/necessary conditions based on the first risk set.
- We have showed that Jeffreys-type has finite modes and lighter tail than a multivariate t distribution and heavier tail than a multivariate normal distribution.
- Based on the idea of the first risk set and variable selection, we have proposed two variations of Jeffreys-type, i.e., Jeffreys-type based on the first risk set, which has led to huge computational saving, and shifted Jeffreys-type.
- Future work includes a theoretically investigation of the unimodality of Jeffreys-type and more proprieties of the first risk set.
- We also need to further investigate additional properties of the two variations of Jeffreys-type.
- We conjecture that $\lim_{\|\boldsymbol{\beta}\|\to\infty} \frac{|I(\boldsymbol{\beta})|}{|I_f(\boldsymbol{\beta})|} \to c$, where c is a constant.
- We also expect that the shifted prior, which is inspired by "lasso", may potentially work well for variable selection.

Thank you !

Selected References

- Chen, M.-H., Ibrahim, J.G., Lam, P., Yu, A., and Zhang, Y. (2011). Bayesian Design of Non-Inferiority Trials for Medical Devices Using Historical Data. *Biometrics, 67*, 1163-1170.
- Ibrahim, J. G., Chen, M.-H., and Sinha, D. (2001). *Bayesian Survival Analysis*, New York: Springer.
- Ibrahim, J.G., Chen, M.-H., Xia, H.A., and Liu, T. (2012). Bayesian Meta-Experimental Design: Evaluating Cardiovascular Risk in New Antidiabetic Therapies to Treat Type 2 Diabetes. *Biometrics* **68**, 578-586.
- Song, R., Kosorok, M.R., and Cai, J. (2008). Robust Covariate-Adjusted Log-Rank Statistics and Corresponding Sample Size Formula for Recurrent Events Data. *Biometrics* **64**, 741-750.
- Wang, F., and Gelfand, A. E. (2002). A simulation-based approach to Bayesian sample size determination for performance under a given model and for separating models. *Statistical Science* **17**, 193-208.

Appendices

Table 2: Maximum Partial Likelihood Estimates (MPLEs) for the PC Data

			Prostate Cancer Death		
Variable	Number of censored	Number of death	Est	SE	p-value
Only one of A, B, C	233	1			
Only AB not C	109	2	0.398	1.234	0.7472
Only AC not B	34	0	-14.303	2107	0.9946
Only BC not A	59	1	0.591	1.227	0.6303
ABC	64	7	2.222	0.805	0.0058
Age	499	11	0.023	0.048	0.6390
			Other Cause Death		
Variable	Number of censored	Number of death	Est	SE	p-value
Only one of A, B, C	233	20			
Only AB not C	109	7	-0.163	0.439	0.7103
Only AC not B	34	1	-1.198	1.025	0.2426
Only BC not A	59	5	-0.167	0.500	0.7392
ABC	64	7	0.079	0.441	0.8585
Age	499	40	0.060	0.029	0.0360

Table 9: MPLEs for PC Death for the PC Data

Variable	EST	SE	HR	95% CI
MPLE/Shifted MPLE		All data		
AB	0.564	1.162	1.758	(0.180, 17.14)
	0.419	1.143	1.521	(0.162, 14.29)
AC	0.078	1.727	1.081	(0.037, 31.92)
	-0.262	1.873	0.770	(0.020, 30.26)
BC	0.768	1.150	2.156	(0.226, 20.52)
	0.561	1.154	1.753	(0.183, 16.83)
ABC	2.063	0.810	7.872	(1.608, 38.54)
	1.926	0.775	6.865	(1.502, 31.38)
Age	0.018	0.048	1.018	(0.926, 1.118)
	0.020	0.048	1.020	(0.928, 1.122)
MPLE/Shifted MPLE		First risk set		
AB	0.557	1.175	1.745	(0.174, 17.47)
	0.343	1.172	1.409	(0.142, 14.02)
AC	0.127	1.710	1.135	(0.040, 32.38)
	-0.220	1.842	0.802	(0.022, 29.68)
BC	0.778	1.157	2.176	(0.225, 21.03)
	0.557	1.156	1.745	(0.181, 16.83)
ABC	2.094	0.815	8.116	(1.642, 40.11)
	1.947	0.775	7.009	(1.535, 32.00)
Age	0.017	0.048	1.017	(0.926, 1.118)
	0.022	0.049	1.023	(0.930, 1.125)

Table 10: Posterior Estimates for PC Death for the PC Data

Variable	EST	SD	HR	95% HPD
Jeffreys-type/		All data		
Shifted Jeffreys-type prior				
AB	0.490	1.172	1.632	(0.022, 9.825)
	0.264	1.195	1.302	(0.022, 7.890)
AC	-0.467	2.139	0.627	(0.001, 7.800)
	-0.786	1.972	0.455	(0.001, 5.584)
BC	0.718	1.155	2.050	(0.038, 12.20)
	0.474	1.200	1.606	(0.036, 9.553)
ABC	2.206	0.795	9.078	(0.901, 37.82)
	2.057	0.799	7.820	(0.904, 32.26)
Age	0.021	0.048	1.021	(0.928, 1.122)
	0.026	0.048	1.026	(0.935, 1.130)
Jeffreys-type/		First risk set		
Shifted Jeffreys-type prior				
AB	0.429	1.155	1.536	(0.030, 9.227)
	0.331	1.189	1.392	(0.026, 8.728)
AC	-1.521	3.415	0.219	(0.001, 5.753)
	-0.853	1.985	0.426	(0.001, 5.494)
BC	0.637	1.198	1.890	(0.027, 11.03)
	0.501	1.186	1.651	(0.016, 9.971)
ABC	2.108	0.779	8.234	(1.007, 33.25)
	2.044	0.803	7.718	(0.884, 32.15)
Age	0.021	0.048	1.022	(0.930, 1.124)
	0.024	0.048	1.024	(0.935, 1.127)