Published in final edited form as:
J Multivar Anal. 2009 October 1; 100(9): 2018–2030. doi:10.1016/j.jmva.2009.03.013.

Maximum Likelihood Inference for the Cox Regression Model with Applications to Missing Covariates

Ming-Hui Chen, **Joseph G. Ibrahim**, and **Qi-Man Shao**
Ming-Hui Chen is Professor, Department of Statistics, University of Connecticut, 215 Glenbrook Road, U-4120, Storrs, CT 06269-4120, Email: mhchen@merlot.stat.uconn.edu. Joseph G. Ibrahim is Professor, Department of Biostatistics, University of North Carolina, McGavran-Greenberg Hall, Chapel Hill, NC 27599, Email: ibrahim@bios.unc.edu. Qi-Man Shao is Professor, Department of Mathematics, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, Email: maqmshao@ust.hk

Abstract

In this paper, we carry out an in-depth theoretical investigation for existence of maximum likelihood estimates for the Cox model (Cox, 1972, 1975) both in the full data setting as well as in the presence of missing covariate data. The main motivation for this work arises from missing data problems, where models can easily become difficult to estimate with certain missing data configurations or large missing data fractions. We establish necessary and sufficient conditions for existence of the maximum partial likelihood estimate (MPLE) for completely observed data (i.e., no missing data) settings as well as sufficient conditions for existence of the maximum likelihood estimate (MLE) for survival data with missing covariates via a profile likelihood method. Several theorems are given to establish these conditions. A real dataset from a cancer clinical trial is presented to further illustrate the proposed methodology.

Keywords

Missing at random (MAR); Monte Carlo EM algorithm; Existence of partial maximum likelihood estimate; Necessary and sufficient conditions; Partial likelihood; Proportional hazards model

1 Introduction

There is a vast literature on parameter estimation in the Cox model in presence of missing covariates, including Schluchter and Jackson (1989), Lin and Ying (1993), Lipsitz and Ibrahim (1996, 1998, 2000), Paik (1997), Paik and Tsai (1997), Chen and Little (1999), Herring and Ibrahim (2001), Leong, Lipsitz, and Ibrahim (2001), Chen (2002), Pons (2002), Herring, Ibrahim, and Lipsitz (2002, 2004), and Chen, Ibrahim, and Shao (2006). However, there is very little literature addressing specific theoretical conditions for the existence of MLE's of the Cox model in either the full data case or in the presence of missing covariate data. We are not aware of specific literature that establishes specific theoretical results for existence of such estimates. This is what we set out to do in this paper. Specifically, we provide necessary and sufficient conditions for existence of the *Maximum Partial Likelihood Estimate* (MPLE) with no missing data as well as sufficient conditions for existence of the Maximum Likelihood Estimate (MLE) with Missing at Random (MAR) covariate data via the profile likelihood method. The methodology proposed here is quite new and will shed light on the

AMS 2000 subject classifications. Primary 62N02, 62F15; secondary 62N99, 65C05.

characterizations of existence of the MPLE or MLE for the Cox model with complete data as well as with missing covariate data. The profile likelihood method for obtaining the MLE in the presence of MAR covariates is quite different from the other parametric and semiparametric approaches seen in the literature. The profile likelihood method is genuinely non-parametric in estimating the cumulative baseline hazard and does not require a semi-parametric estimate of the baseline hazard as is required in Lipsitz and Ibrahim (1998) and Herring and Ibrahim (2001).

We mention that Jacobsen (1989) establishes a necessary and sufficient condition for existence of the MPLE for the Cox model without missing covariate data, Chen, Ibrahim, and Shao (2004) consider issues in posterior propriety and characterize conditions for existence of the MLE in generalized linear models with MAR covariate data, and Huang, Chen, and Ibrahim (2005) carry out a detailed investigation of posterior propriety in generalized linear models with nonignorably missing covariate data. The methods and models considered in those papers are quite different from the Cox model setting. In the Cox model, i) we no longer have independence between the observations in the construction of the partial likelihood, that is, the complete data log-likelihood is not a sum of n independent observations, ii) the Cox regression model, and in particular, Cox's partial likelihood, is an inherently semiparametric model, and thus a profile likelihood method considered here is quite different than the fully parametric models considered in Chen, Ibrahim, and Shao (2004) and Huang, Chen, and Ibrahim (2005), and iii) right censoring and tied observations require new theory not developed in Chen, Ibrahim, and Shao (2004) and Huang, Chen, and Ibrahim (2005). Thus, i) – iii) will require new theory for characterizing conditions for existence of the MPLE and MLE of the regression coefficients in the Cox model allowing for tied observations.

The significance of this work thus has two aspects. First, the proposed methodology will allow the data analyst to determine, for a given dataset, whether the MPLE or MLE exists *before* carrying out the analysis. Such a methodology is critical since it is not always clear from the computer output in an analysis whether the MPLE or MLE exists or not. Second, such conditions will be useful for determining suitable starting values for EM-type algorithms when fitting these models. Thus, the practical consequences of the proposed methodology is that we provide valuable tools for checking existence of the MPLE or MLE as well as inferential and computational tools for maximum likelihood based inference for the Cox model with or without MAR covariates.

The rest of this article is organized as follows. Section 2 presents several motivating examples. We give necessary and sufficient conditions for the existence of the MPLE with no missing data in Section 3 and give sufficient conditions for existence of the MLE in the presence of MAR covariate data in Section 4. The computational development involving the Monte Carlo EM (MCEM) algorithm is given in Section 5. Section 6 presents a detailed analysis of a lung cancer dataset to further illustrate the proposed methodology. Proofs of all theorems are given in the Appendix.

2 Motivating Examples

To fix ideas, 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 i^{th} subject. Denote by $\boldsymbol{\beta} = (\beta_1,\ldots,\beta_p)'$ the $p \times 1$ vector of regression coefficients. Also, $\delta_i = 1\{T_i = y_i\}$ is the indicator for the event for $i = 1, 2, \ldots, n$, 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

$$L_p(\beta|D_{obs})=\prod_{i=1}^{n}\left[\frac{\exp(\boldsymbol{x}_i'\beta)}{\sum_{j\in\mathcal{R}(y_i)}\exp(\boldsymbol{x}_j'\beta)}\right]^{\delta_i}, \tag{2.1}$$

where $D_{obs} = \{(y_i, \delta_i, \boldsymbol{x}_i) : i = 1, 2, \ldots, n\}$ is the observed univariate right censored survival data. As usual, we assume throughout that $\boldsymbol{x}_i$ does not include an intercept, since the interceptis not estimable in the Cox partial likelihood, and that given $\boldsymbol{x}_i$, T_i and C_i are independent. For the completely observed data D_{obs}, the maximum partial likelihood estimate (MPLE) is defined as $\hat{\boldsymbol{\beta}}$ = arg max $L_p(\boldsymbol{\beta}|D_{obs})$. The asymptotic properties of $\hat{\boldsymbol{\beta}}$ have been well studied in the literature, and in fact, the MPLE can be computed via standard statistical software, such as the SAS procedure, PROC PHREG. However, it remains unclear when the MPLE exists and when it does not for a given dataset. To motivate the proposed methodology, we consider the following two examples.

Example 1: A Simple Illustration

Suppose $n = 3$, y_1 and y_2 are two failure times, y_3 is a right censored survival time, and we have one binary covariate x. Let x_1, x_2, and x_3 denote the three observed values of x. Assuming $y_1 < y_2 < y_3$, the partial likelihood of Cox (1975) is then given by

$$L_p(\beta_1|D_{obs})=\frac{\exp(x_1\beta_1)}{\exp(x_1\beta_1)+\exp(x_2\beta_1)+\exp(x_3\beta_1)}\times\frac{\exp(x_2\beta_1)}{\exp(x_2\beta_1)+\exp(x_3\beta_1)},$$

where $D_{obs} = \{(y_i, x_i), i = 1, 2, 3\}$. We consider two special cases.

Case 1—$x_1 = x_2 = 0$ and $x_3 = 1$. In this case, we have $L_p(\beta_1|D)=\frac{1}{2+\exp(\beta_1)}\times\frac{1}{2+\exp(\beta_1)}$. Then, we can see that the maximum value of $L_p(\beta_1|D)$ is attained at $\beta_1 = -\infty$. Thus, the MPLE does not exist.

Case 2—$x_1 = 0$, $x_2 = 1$, and $x_3 = 0$. In this case, we have $L_p(\beta_1|D_{obs})=\frac{1}{2+\exp(\beta_1)}\times\frac{\exp(\beta_1)}{\exp(\beta_1)+1}$. Then, the MPLE does exist. In fact, the MPLE of β_1 is $\frac{1}{2}\log(2)$.

In Example 1, the partial likelihood function behaves quite differently by simply switching two observed values of the covariate: one leads to the existence of the MPLE and the other does not. Thus, a natural question is what are general *if and only if conditions* for the existence of the MPLE in the Cox model? From this illustrative example, we can see that this is not an easy problem to solve, as it requires an in-depth theoretical investigation to find such conditions.

Example 2: Prostate Cancer Data

We consider data, which consist of $n = 550$ men who were treated with radiation therapy following with six months of with short-course androgen suppression therapy for localized prostate cancer with at least one adverse risk factor (prostate-specific antigen [PSA] > 10 ng/mL, biopsy Gleason score 7 to 10, or 2002 American Joint Commission on Cancer (AJCC) clinical tumor category T2b or T2c) between 1989 and 2002. The outcome variable (y_i) in years was time to prostate cancer death, which is continuous and subject to right censoring, and δ_i = pfail denotes the censoring indicator which equals 1 if the i^{th} subject died due to prostate cancer, and 0 otherwise. The goal of this study was to determine whether the number of risk factors present was associated with time to prostate cancer death (Tsai et al., 2006).

Define $A = I$ {PSA > 10}, $B = I$ {Gleason ≥ 7}, and $C = I$ {T2b or T2c}. We consider five covariates: AB, AC, BC, ABC, and age. There are no missing values in this data set. A Cox proportional hazards model was fitted to this data set. The following outputs were produced by SAS Procedure PHREG:

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq
AB	1	0.39759	1.23355	0.1039	0.7472
AC	1	−14.30314	2107	0.0000	0.9946
BC	1	0.59060	1.22714	0.2316	0.6303
ABC	1	2.22155	0.80450	7.6253	0.0058
age	1	0.02262	0.04821	0.2201	0.6390

From the above results, we see that although SAS Procedure PHREG does produce the estimates for all five covariates, clearly there is some identifiability problem with the covariate, AC, as it has a large value of the estimate along with a huge standard error compared to all other covariates. Now, the question is: are the MPLEs are really existent in this Cox model?

Example 3: Small Cell Lung Cancer Data

We consider data from a phase III advanced non-small-cell lung cancer (SCLC) clinical trial conducted by the University of North Carolina at Chapel Hill (LCCC 9719). The results of this study have been published in Socinski et al. (2002). The goal of this trial was to compare a defined duration of therapy (A) to continuous therapy followed by second line therapy (B) in order to determine optimal duration of therapy in SCLC patients. LCCC 9719 had $n = 230$ patients. We consider here five prognostic factors: x_1 = treatment (2 arms: A and B, coded as 1 and 0), x_2 = gender (female and male, coded as 0 and 1), x_3 = age in years, x_4 = highest grade toxicity (recorded by cycle) (2 levels: 0 versus > 0, coded as 0 and 1), and x_5 = quality of life (QOL) FACTG score. For these five prognostic factors, x_4 and x_5 had missing information and x_1, x_2, and x_3 were completely observed for all cases. In this dataset, there is a total missing covariate data fraction of 52.74% on these two covariates. The outcome variable (y_i in months) is time to progression, which is continuous and subject to right censoring, and δ_i denotes the censoring indicator which equals 1 if the i^{th} subject had disease progression, and 0 otherwise. The median follow up time is 3.94 months and the range of the follow up time is (0.10, 12.26) months. There are $d = 102$ distinct progression times and ties are present in the dataset. A summary of the dataset is given in Table 1. In the presence of missing covariates, a joint probability distribution must be specified for the progression time and the missing covariates, and a profile likelihood method is hence proposed for obtaining the MLE in Section 4, as a partial likelihood approach in this context may not be as desirable.

3 Existence of the MPLE With No Missing Data

In this section, we characterize very general conditions for the existence of the MPLE of $\boldsymbol{\beta}$ for a given dataset D_{obs} under the Cox model with no missing covariate data. Define X^* to be

$$X^* = (\delta_i(\boldsymbol{x}_j - \boldsymbol{x}_i),\ j \in \mathcal{R}(y_i),\ 1 \le i \le n)'. \tag{3.1}$$

Let k_i denote the number of subjects in $\mathcal{R}(y_i)$ for $i = 1, 2, \ldots, n$. Also let $K = \sum_{i=1}^{n} k_i$. Then, X^* is a $K \times p$ matrix. Using X^*, we are led to the following theorem.

Theorem 3.1

The MPLE of **β** in (2.1) exists if the following conditions are satisfied:

(C1) X^* is of full rank p; and

(C2) There exists a positive vector **v**, i.e., each component of **v** is positive, such that

$$X^{*'} \boldsymbol{v}=0. \tag{3.2}$$

In addition, if (C1) is satisfied, then (C2) is a necessary condition for the existence of MPLE for **β**.

The proof of Theorem 3.1 is given in the Appendix.

Remark 3.1—In X^* defined by (3.1), the rows corresponding to $\delta_i = 0$ or $\boldsymbol{x}_j = \boldsymbol{x}_i$ can be excluded. Thus, the effective numbers of rows in X^* can be reduced substantially. Specifically, let $k_i^*=\sum_{j\in\mathcal{R}(y_i)} 1\{\boldsymbol{x}_j \neq \boldsymbol{x}_i\}$, where the indicator function $1\{\mathbf{x}_j \neq \mathbf{x}_i\} = 1$ if $\mathbf{x}_j \neq \mathbf{x}_i$ and 0 otherwise. Then, the effective numbers of rows in X^* is given by $K^*=\sum_{i=1}^n \delta_i k_i^*$.

Remark 3.2—When ties are present, as discussed in Klein and Moeschberger (2003, Chapter 8), the partial likelihood may be defined as

$$L_{pt}(\beta|D_{obs})=\prod_{i=1}^{d} \frac{\exp(\boldsymbol{z}_i'\beta)}{\left[\sum_{j\in\mathcal{R}(y_i)} \exp(\boldsymbol{x}_j'\beta)\right]^{d_i}}, \tag{3.3}$$

where $d=\sum_{i=1}^n \delta_i$, $\boldsymbol{z}_i = \Sigma_{j\in\mathcal{D}_i} \mathbf{x}_j$, d_i = the number of events at y_i, and $\mathcal{D}_i$ is the set of all individuals who have the event at time y_i. We can thus rewrite (3.3) as

$$L_{pt}(\beta|D_{obs})=\prod_{i=1}^{n} \frac{\exp(\delta_i \boldsymbol{x}_i'\beta)}{\left[\sum_{j\in\mathcal{R}(y_i)} \exp(\boldsymbol{x}_j'\beta)\right]^{\delta_i}},$$

and Theorem 3.1 can be easily extended to the cases when ties are present. Note that the partial likelihood given by (3.3) is the likelihood of Breslow (1974), and the Breslow likelihood is the default choice in SAS to handle ties in the failure times.

Remark 3.3—Suppose $y_1 \leq y_2 \leq \cdots \leq y_n$. Then, from condition ($C$2), it is easy to observe that if there exists a j such that $x_{1j} \leq x_{2j} \leq \cdots x_{nj}$, the MPLE of $\boldsymbol{\beta}$ does not exist. Also, when one of the components of $\boldsymbol{x}_i$, say, x_{ij}, is binary and the x_{ij}'s take the same value for $\delta_i = 1$ or the the x_{ij}'s take the same value for $\delta_i = 0$, then the MPLE of β does not exist.

Remark 3.4—When conditions (C1) and (C2) are satisfied for a subset of the data, the MPLE still does exist. To see this, we assume that the subset consists of the first n^* observations. Then we have

$$L_p(\beta|D_{obs}) \le \prod_{i=1}^{n^*}\left[\frac{\exp(\boldsymbol{x}_i'\beta)}{\sum_{j\in\mathcal{R}(y_i),j\le n^*}\exp(\boldsymbol{x}_j'\beta)}\right]^{\delta_i}.$$

The existence of the MPLE can obtain by simply applying Theorem 3.1 to the above upper bound. These subset conditions are only sufficient but not necessary. However, this result is particularly useful for large datasets, for which checking conditions ($C1$) and ($C2$) may not be computationally feasible.

Remark 3.5—Jacobsen (1989) also characterizes a necessary and sufficient condition for the existence of the MPLE. His condition can be stated as follows: there is no $\boldsymbol{a} \in R^p$ such that $\boldsymbol{a}'\delta_i(\boldsymbol{x}_j - \boldsymbol{x}_i) \ge 0$ for $j \in \mathcal{R}(y_i)$ and $1 \le i \le n$. According to Lemma A.1, we can see that Jacobsen's condition implies ($C2$). Thus, ($C2$) is necessary for existence and for uniqueness. We note that the conditions stated in Theorem 3.1 are sufficient. However, compared to Jacobsen's condition, the conditions ($C1$) and ($C2$) given in Theorem 3.1 are easier to check. First, it is straightforward to check condition ($C1$) that X^*has full column rank. As discussed in Appendix A of Roy and Hobert (2007), condition ($C2$) can be checked with a simple linear program using the 'simplex' function from the 'boot' library in the R programming language.

Example 1: A Simple Illustration (revisited): Recall that in Example 1, we have $n = 3$, $y_1 < y_2 < y_3$, $\delta_1 = \delta_2 = 1$, and $\delta_3 = 0$. For **Case 1** in which $x_1 = x_2 = 0$ and $x_3 = 1$, we have $k_1 = 3$, $k_2 = 2$, $k_3 = 1$, and $K = 6$. Thus, using (3.1), $(X^*) = (0, 0, 1, 0, 1, 0)'$, which is a 6×1 matrix. After excluding the rows corresponding to $\delta_i = 0$ or $\boldsymbol{x}_j = \boldsymbol{x}_i$, the effective number of rows in X^* is $K^* = 2$. It is easy to see that X^* is of full rank, which is 1. Also, for any $\boldsymbol{v} = (v_1, v_2, v_3, v_4, v_5, v_6)'$ such that $v_i > 0$, $(X^*)'\boldsymbol{v} = v_3 + v_5 > 0$. Thus, by Theorem 3.1, the MPLE does not exist.

For **Case 2**, where $x_1 = 0$, $x_2 = 1$, and $x_3 = 0$, using (3.1), we have $(X^*)' = (0, 1, 0, 0, -1, 0)$. Let $\boldsymbol{v} = (v_1, v_2, v_3, v_4, v_2, v_5)'$ for $v_j > 0$, $j = 1, 2, \ldots, 5$, $(X^*)'\boldsymbol{v} \equiv 0$. Obviously, X^* is of full rank. Thus, the MPLE does exist by Theorem 3.1.

In general, we have $X^* = (0, x_2 - x_1, x_3 - x_1, 0, x_3 - x_2, 0)'$ and (3.2) reduces to

$$\begin{aligned}&(0, x_2 - x_1, x_3 - x_1, 0, x_3 - x_2, 0)(v_1, v_2, v_3, v_4, v_5, v_6)'\\&=v_2(x_2 - x_1)+v_3(x_3 - x_1)+v_5(x_3 - x_2)=0.\end{aligned} \tag{3.4}$$

Condition ($C1$) requires that at least two of x_1, x_2, and x_3 are different. If $x_1 = x_2$ and condition ($C1$) holds, then there is no positive solution $\boldsymbol{v}$ to (3.4) regardless of the value of x_3. Thus, the MPLE always does not exist when $x_1 = x_2$. However, if $x_1 < x_2$, then the MPLE exists if $x_3 < x_2$ and does not exist if $x_3 \ge x_2$. Similarly, if $x_2 < x_1$, the MPLE exists if $x_3 > x_2$ and does not exist if $x_3 \le x_2$. One interesting observation is that even if $x_3 < x_1 < x_2$, the MPLE still exists although x_3, for which $\delta_3 = 0$, is distinct from $\{x_1, x_2\}$ in the sense that $\delta_1 = \delta_2 = 1$. Thus, the condition for existence of the MPLE cannot be characterized by the value of δ_i alone by fitting, for example, a binary regression model to δ_i while treating $(1, x_i)'$ as a vector of covariates.

Example 2: Prostate Cancer Data (revisited): After we further examined the data, we found that

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Variable	**δ= 0**	**δ= 1**	**Total**
Only one of A, B, C	253	2	255
Only AB not C	116	1	117
Only AC not B	35	0	35
Only BC not A	64	1	65
ABC	71	7	78
Total	539	11	550

From the above table, "only AC not B" is the only group, in which there are no events. This explains why we obtained the unusual estimate and standard error for the regression coefficient corresponding to AC. From Remark 3.3, it becomes apparent that the MPLEs do not exist for this dataset if we fit the five covariates in the Cox model. One way to fix this problem is to combine AB, AC, and BC as one variable, which was called the two-factors only variable in Tsai et al. (2006).

4 Profile Maximum Likelihood Estimation in the Presence of Missing Covariates

When there are missing covariates, we assume that the distribution of the censoring time C_i does not depend on the missing covariates and the missingness is MAR. In this case, we cannot directly use the Cox partial likelihood since we need to model the failure time and the covariates jointly. Thus, instead of the partial likelihood approach, we use a profile likelihood approach when we have MAR covariates.

For notational simplicity, we assume that all failure times are distinct and let $y_1, y_2, \ldots, y_d$ be d distinct failure times. Let $h_0(y) \geq 0$ denote the baseline hazard function and also let $H_0(y)=\int_0^y h_0(u)du$ denote the baseline cumulative hazard function. Let $\boldsymbol{x}_i=(\boldsymbol{x}'_{i,mis}, \boldsymbol{x}'_{i,obs})'$ and $D_{obs} = (y_i, \delta_i, \boldsymbol{x}_{i,obs}, i = 1,2, \ldots, n)$. Also let $D = (y_i, \delta_i, \boldsymbol{x}_{i,obs}, \boldsymbol{x}_{i,mis}, i = 1, 2,\ldots, n)$ denote the complete data. In addition, let $\boldsymbol{r}_i = (r_{i1}, r_{i2}, \ldots, r_{ip})'$ to be the vector of the p missing covariate indicators such that $r_{il} = 0$ when x_{il} is missing and $r_{il} = 1$ when x_{il} is observed for $i = 1, 2, \cdots, n$ and $l = 1, 2, \cdots, p$. Since we assume ignorable missingness in the covariates (i.e., MAR covariates and the parameters of the missing data mechanism are distinct from the sampling model), we do not need to model $\boldsymbol{r}_i$. Also, we assume that the parameters of the distributions for the censoring times C_i's are distinct from the sampling model. Thus, for ignorably missing covariates, ignoring the parts adhering to censoring and the missing data mechanism, the observed data likelihood function based on the Cox model (Cox, 1972) is given by

$$L(\beta, h_0, \alpha | D_{obs}) = \int \left[\prod_{i=1}^{d} h_0(y_i) \exp(\boldsymbol{x}'_i \beta) \right] \exp\{-\sum_{j=1}^{n} H_0(y_j) \exp(\boldsymbol{x}'_j \beta)\} \left[\prod_{i=1}^{n} f(\boldsymbol{x}_{i,mis}, \boldsymbol{x}_{i,obs} | \alpha) \right] d\boldsymbol{x}_{mis}, \quad (4.1)$$

where $\boldsymbol{x}_{mis} = (\boldsymbol{x}_{i,mis}, i = 1, 2, \ldots, n)$, $f(\boldsymbol{x}_{i,mis}, \boldsymbol{x}_{i,obs}|\boldsymbol{\alpha})$ denotes the joint distribution of $\boldsymbol{x}_i$, and $\boldsymbol{\alpha}$ is the vector of parameters for the covariate distribution.

It is well known that the partial likelihood can be expressed as a profile likelihood (Johansen, 1983) by substituting a nonparametric maximum likelihood estimator for the cumulative baseline hazard function $H_0(y)$, which is a function of the fixed coefficients $\boldsymbol{\beta}$, and that this nonparametric maximum likelihood estimator is necessarily a pure-jump estimator with jumps precisely at the observed event times. Following the profile likelihood approach (see, for example, Klein and Moeschberger (2003, Chapter 8)), we have

$$\begin{aligned}
&\sup_{h_0} L(\beta, h_0, \alpha|D_{obs}) \\
&\leq \int \sup_{h_0} \left[\prod_{i=1}^{d} h_0(y_i)\exp(\boldsymbol{x}_i'\beta)\right] \exp\{-\sum_{j=1}^{n} H_0(y_j)\exp(\boldsymbol{x}_j'\beta)\} \left[\prod_{i=1}^{n} f(\boldsymbol{x}_{i,mis}, \boldsymbol{x}_{i,obs}|\alpha)\right] d\boldsymbol{x}_{mis} \\
&= \int \sup_{h_0} \left[\prod_{i=1}^{d} h_0(y_i)\exp(\boldsymbol{x}_i'\beta)\right] \exp\{-\sum_{j=1}^{n} (\sum_{y_k \leq y_j} h_0(y_k))\exp(\boldsymbol{x}_j'\beta)\} \left[\prod_{i=1}^{n} f(\boldsymbol{x}_{i,mis}, \boldsymbol{x}_{i,obs}|\alpha)\right] d\boldsymbol{x}_{mis} \\
&= \int \sup_{h_0} \prod_{i=1}^{d} \left[h_0(y_i)\exp(\boldsymbol{x}_i'\beta)\exp\{-h_0(y_i) \sum_{j\in\mathcal{R}(y_i)} \exp(\boldsymbol{x}_j'\beta)\}\right] \left[\prod_{i=1}^{n} f(\boldsymbol{x}_{i,mis}, \boldsymbol{x}_{i,obs}|\alpha)\right] d\boldsymbol{x}_{mis} \\
&= \int \left\{\prod_{i=1}^{d} \exp(\boldsymbol{x}_i'\beta) \left[\sum_{j\in\mathcal{R}(y_i)} \exp(\boldsymbol{x}_j'\beta)\right]^{-1}\right\} \left[\prod_{i=1}^{n} f(\boldsymbol{x}_{i,mis}, \boldsymbol{x}_{i,obs}|\alpha)\right] d\boldsymbol{x}_{mis}.
\end{aligned} \tag{4.2}$$

We note that in (4.2), the function

$$\left[\prod_{i=1}^{d} h_0(y_i)\exp(\boldsymbol{x}_i'\beta)\right] \exp\{-\sum_{j=1}^{n} H_0(y_j)\exp(\boldsymbol{x}_j'\beta)\}$$

is maximized when $h_0(y_j) = 0$ except for the times at which events occur. Thus, the MLE of $(\boldsymbol{\beta},\boldsymbol{\alpha})$ exists if the upper bound in the right-hand side of (4.2) goes to zero when $\| \beta \| + \| \alpha \| = \sqrt{\beta'\beta} + \sqrt{\alpha'\alpha} \to \infty$. Write

$$L(\alpha|D_{obs}) = \int \prod_{i=1}^{n} f(\boldsymbol{x}_{i,mis}, \boldsymbol{x}_{i,obs}|\alpha) d\boldsymbol{x}_{mis}. \tag{4.3}$$

The following theorem characterizes the conditions for existence of the MLE of $(\beta, h_0, \boldsymbol{\alpha})$ when the x_{ij}'s are bounded.

Theorem 4.1

If the x_{ij}'s are bounded, i.e., $a_i \leq x_{ij} \leq b_i$, define X^{**} to be $X^{**} = (\delta_i(\boldsymbol{x}_j^* - \boldsymbol{x}_i^*), j \in \mathcal{R}(y_i), \delta_j = 0, 1 \leq i \leq n)'$, *where* $x_i^* = ((\boldsymbol{x}_{i,mis}^R)', \boldsymbol{x}_{i,obs}')'$ and each component of $\boldsymbol{x}_{i,mis}^R$ is equal to either $a_i^* = \delta_i a_i + (1-\delta_i) b_i$ or $b_i^* = (1-\delta_i) a_i + \delta_i b_i$ for all i. Then, the MLE of $(\boldsymbol{\beta}, h_0, \boldsymbol{\alpha})$ in (4.1) exists if the following conditions are satisfied: (C1*) $\lim_{\|\boldsymbol{\alpha}\| \to \infty} L(\boldsymbol{\alpha}|D_{obs}) = 0$; (C2*) X^{**} is of full rank; and (C3*) there exists a positive vector $\mathbf{v}$ such that $\mathbf{X}^{**\prime}\mathbf{v} = 0$.

The proof of Theorem 4.1 is given in the Appendix. The main intuition behind Theorem 4.1 is that when the MLE exists under conditions ($C2^*$) and ($C3^*$) for the most extreme possible values of the missing covariates, then the MLE also exists for any intermediate values of the missing covariates, and averaging over the missing values will not affect the existence of the MLE. In Theorem 4.1, the elements of the matrix X^* corresponding to the missing covariates are "filled-in" by either $a_i^* = \delta_i a_i + (1-\delta_i) b_i$ or $b_i^* = (1-\delta_i) a_i + \delta_i b_i$, where a_i^* and b_i^* are in fact the two possible extreme values of the missing covariates when the x_{ij}'s are bounded.

The next theorem gives the sufficient conditions for existence of the MLE of (β, h_0, $\boldsymbol{\alpha}$) when the x_{ij}'s unbounded.

Theorem 4.2

If the x_{ij}'s are unbounded, the MLE of ($\boldsymbol{\beta}$, h_0, $\boldsymbol{\alpha}$) in (4.1) exists if condition ($C1^*$) in Theorem 4.1 and conditions (C1) and (C2) in Theorem 3.1 are satisfied for the completely observed cases.

The proof of Theorem 4.2 is given in the Appendix. Theorem 4.2 is practically useful as the conditions stated in this theorem are easy to check than those given in Theorem 4.1. We note that in Theorem 4.2, we are not doing a complete case analysis. Instead, we use a subset of the data with the completely observed cases to establish the sufficient conditions for the existence of the MLE when the missing covariates are unbounded.

Remark 4.1—Assume that the maximum number of missing components of $\boldsymbol{x}_i$, $i = 1, \ldots, n$, is p_i. Then, to verify the conditions given in Theorem 4.1, we need to check only the conditions ($C2^*$) and ($C3^*$) for at most 2^{p_i} possible X^{**}'s.

Remark 4.2—When there are no missing covariates, it is easy to observe that the profile maximum likelihood estimate of $\boldsymbol{\beta}$ reduces to the MPLE, while the profile maximum likelihood estimate of $\boldsymbol{\alpha}$ is the MLE.

Remark 4.3—Ibrahim, Lipsitz and Chen (1999) and Chen and Ibrahim (2001) provide a comprehensive set of guidelines for specifying the joint distribution of the covariate vector $\boldsymbol{x}_i$ through a series of one dimensional conditional distributions. Condition ($C1^*$) stated in Theorem 4.1 holds for many covariate distributions considered in Ibrahim, Lipsitz and Chen (1999) and Chen and Ibrahim (2001).

Remark 4.4—When there are ties in the event times, similar to Remark 3.2, the upper bound given in (4.2) can be modified as

$$K\int\left\{\prod_{i=1}^{d}\exp(\boldsymbol{z}_i'\beta)\left[\sum_{j\in\mathcal{R}(y_i)}\exp(\boldsymbol{x}_j'\beta)\right]^{-d_i}\right\}\left[\prod_{i=1}^{n}f(\boldsymbol{x}_{i,mis},\boldsymbol{x}_{i,obs}|\alpha)\right]d\boldsymbol{x}_{mis},$$

where $K > 0$ is independent of $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$, and $\boldsymbol{x}_i$, and $\boldsymbol{z}_i$ and d_i are defined in (3.3). Thus, all the theory developed in this subsection is still valid in the presence of ties.

Next, we consider an interesting special case where each missing component of $\boldsymbol{x}_i$ is discrete and bounded.

Corollary 4.1

Assume that each missing component of $\mathbf{x}_i$ is discrete and bounded. Then condition ($C3^*$) given in Theorem 4.1 is also necessary for the existence of the MLE for ($\boldsymbol{\beta}$, h_0) if condition ($C2^*$) is satisfied.

The proof of Corollary 4.1 directly follows from the fact that when each missing component of $\boldsymbol{x}_i$ is discrete and bounded, we have

$$\sup_{h_0} L(\beta, h_0, \alpha|D_{obs}) = \sum_{\boldsymbol{x}_{mis}} \sup_{h_0} \left[\prod_{i=1}^{d} h_0(y_i)\exp(\boldsymbol{x}_i'\beta)\right] \exp\{-\sum_{j=1}^{n} H_0(y_j)\exp(\boldsymbol{x}_j'\beta)\} \times \left[\prod_{i=1}^{n} f(\boldsymbol{x}_{i,mis}, \boldsymbol{x}_{i,obs}|\alpha)\right].$$

Thus, details of the proof are omitted for brevity.

5 Computational Development

When there are no missing covariates, computing the MPLE of $\boldsymbol{\beta}$ is straightforward and, in fact, the MPLE can be computed via standard statistical software, such as the SAS procedure, PROC PHREG. In the presence of missing covariates, the EM algorithm is required. Martinussen (1999) proposes an efficient EM algorithm for computing the MLE and its standard error in the presence of discrete missing covariates. When $\boldsymbol{x}_{i,mis}$ is continuous or mixed continuous and categorical, we need to develop a Monte Carlo EM (MCEM) algorithm, which is an extension of Martinussen's algorithm for computing the MLE's of $\boldsymbol{\beta}$, h_0, and $\boldsymbol{\alpha}$ as well as their standard errors.

To implement the MCEM algorithm, let $\boldsymbol{\gamma} = (\boldsymbol{\beta}, h_0, \boldsymbol{\alpha})$. Let $\boldsymbol{\gamma}^{(t)}$ denote the parameter estimate of $\boldsymbol{\gamma}$ at the t^{th} EM iteration. In the E-step, we take an MCMC sample of size $m_i^{(t)}, \boldsymbol{x}_{i,mis}^{(t1)}, \boldsymbol{x}_{i,mis}^{(t2)}, \ldots, \boldsymbol{x}_{i,mis}^{(tm_i^{(t)})}$, from

$$f(x_{i,mis}|x_{i,obs}, y_i, \gamma^{(t)}) \propto \exp(\delta_i \boldsymbol{x}_i'\beta^{(t)})\exp\{-H_0^{(t)}(y_i)\exp(\boldsymbol{x}_i'\beta^{(t)})\} f(\boldsymbol{x}_{i,mis}, \boldsymbol{x}_{i,obs}|\alpha^{(t)})$$

for $i = 1, 2, \ldots, n$. Note that this conditional distribution is log-concave as long as $f(\boldsymbol{x}_{i,mis}, \boldsymbol{x}_{i,obs}|\boldsymbol{\alpha}^{(t)})$ is log-concave in each component of $\boldsymbol{x}_{i,mis}$. We then compute

$$Q(\gamma|\gamma^{(t)}) = \sum_{i=1}^{d}\left[\log h_0(y_i) + \frac{1}{m_i^{(t)}}\sum_{k=1}^{m_i^{(t)}} \boldsymbol{x}_i^{(tk)'}\beta\right] - \sum_{j=1}^{n} H_0(y_j)\left[\frac{1}{m_i^{(t)}}\sum_{k=1}^{m_j^{(t)}} \exp(\boldsymbol{x}_j^{(tk)'}\beta)\right] + \sum_{i=1}^{n}\frac{1}{m_i^{(t)}}\sum_{k=1}^{m_i^{(t)}} \log f(\boldsymbol{x}_{i,mis}^{(tk)}, \boldsymbol{x}_{i,obs}|\alpha), \tag{5.1}$$

where $\boldsymbol{x}_i^{(tk)} = (\boldsymbol{x}_{i,mis}^{(tk)'}, \boldsymbol{x}_{i,obs}')'$ and $H_0(y_j) = \Sigma_{yl \le y_j, \delta_l = 1} h_0(yl)$. In the M-step, we compute

$$\beta^{(t+1)} = \arg\max_{\beta} \sum_{i=1}^{d}\left\{\left[\frac{1}{m_i^{(t)}}\sum_{k=1}^{m_i^{(t)}} \boldsymbol{x}_i^{(tk)'}\beta\right] - \log\left[\sum_{j \in \mathcal{R}(y_i)} \frac{1}{m_j^{(t)}}\sum_{k=1}^{m_j^{(t)}} \exp(\boldsymbol{x}_j^{(tk)'}\beta)\right]\right\}, \tag{5.2}$$

$$h_0^{(t+1)}(y_i) = \left[\sum_{j \in \mathcal{R}(y_i)} \frac{1}{m_j^{(t)}}\sum_{k=1}^{m_j^{(t)}} \exp(\boldsymbol{x}_j^{(tk)'}\beta^{(t+1)})\right]^{-1}, H_0^{(t+1)}(y_i) = \sum_{y_j \le y_i, \delta_j = 1} h_0^{(t+1)}(y_j), \tag{5.3}$$

NIH-PA Author Manuscript

NIH-PA Author Manuscript

and

$$\alpha^{(t+1)} = \arg\max_{\alpha} \sum_{i=1}^{n} \frac{1}{m_i^{(t)}} \sum_{k=1}^{m_i^{(t)}} \log f(\boldsymbol{x}_{i,mis}^{(tk)}, \boldsymbol{x}_{i,obs}|\alpha).$$

Following Booth and Hobert (1999), in the MCEM algorithm, we take $m^{(t+1)} = m^{(t)} + \Delta m$, where $\Delta m > 0$. With this dynamic MCMC sample size $m^{(t)}$, the MCEM algorithm requires much less computational time. Also a large $m^{(t)}$ is not needed in early iterations of the algorithm since $\boldsymbol{\gamma}^{(t)}$ is still far from the MLE $\hat{\boldsymbol{\gamma}}$ and the algorithm is not near convergence. As t increases $m^{(t)}$ increases, and a more computationally accurate estimate of $Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{(t)})$ is obtained in the E-step.

When $\boldsymbol{x}_{i,mis}$ is categorical, the E-step at the $(t + 1)^{st}$ iteration reduces to the *EM by the Method of Weights* (Ibrahim, 1990). With the EM by the Method of Weights, a similar M-step can be developed. We refer to Ibrahim (1990) and Martinussen (1999) for the detailed development of the EM algorithm in this case. It is easy to see from (5.2) that when there are no missing covariates, $\boldsymbol{\beta}^{(t+1)}$ is the MPLE of $\boldsymbol{\beta}$, which is consistent with Remark 4.2.

Let $\hat{\boldsymbol{\gamma}}$ denote the estimate of $\boldsymbol{\gamma}$ at EM convergence. Using Louis's method (Louis, 1982), the estimated observed information matrix of $\boldsymbol{\gamma}$ based on the observed data is not difficult to compute. Note that the complete-data likelihood function can be written as

$$L(\beta, h_0, \alpha|D) = \prod_{i=1}^{n} \left[(h_0(y_i)\exp(\boldsymbol{x}_i'\beta))^{\delta_i} \exp\{-H_0(y_i)\exp(\boldsymbol{x}_i'\beta)\} f(\boldsymbol{x}_i|\alpha) \right]. \tag{5.4}$$

Thus, the log-likelihood function for the i^{th} observation is given by

$$l(\gamma|\boldsymbol{x}_i, y_i, \delta_i) = \delta_i[\log h_0(y_i) + \boldsymbol{x}_i'\beta] - H_0(y_i)\exp(\boldsymbol{x}_i'\beta) + \log(f(\boldsymbol{x}_i|\alpha)). \tag{5.5}$$

Write the gradient vector of $Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{(t)})$ as

$$\dot{Q} = (\gamma|\gamma^{(t)}) = \sum_{i=1}^{n} \dot{Q}_i(\gamma|\gamma^{(t)}) = \sum_{i=1}^{n} \frac{1}{m_i^{(t)}} \sum_{k=1}^{m_i^{(t)}} \frac{\partial l(\gamma|\boldsymbol{x}_i^{(tk)}, y_i, \delta_i)}{\partial \gamma},$$

and write the matrix of second derivatives of $Q(\boldsymbol{\gamma}|\boldsymbol{\gamma}^{(t)})$ as

$$\ddot{Q} = (\gamma|\gamma^{(t)}) = \sum_{i=1}^{n} \frac{1}{m_i^{(t)}} \sum_{k=1}^{m_i^{(t)}} \frac{\partial^2 l(\gamma|\boldsymbol{x}_i^{(tk)}, y_i, \delta_i)}{\partial \gamma\, \partial \gamma'}.$$

In addition, write the complete data score vector as

$$S(\gamma|D)=\sum_{i=1}^{n}S_i(\gamma|\boldsymbol{x}_i,y_i,\delta_i)=\sum_{i=1}^{n}\frac{\partial l(\gamma|\boldsymbol{x}_i,y_i,\delta_i)}{\partial\gamma}.$$

Then, the estimated observed information matrix of $\hat{\gamma}$ is given by

$$\mathcal{I}(\hat{\gamma})=-\ddot{Q}(\hat{\gamma}|\hat{\gamma})-\left\{\left[\sum_{i=1}^{n}\frac{1}{m_i^{(t)}}\sum_{k=1}^{m_i^{(t)}}S_i(\hat{\gamma}|\boldsymbol{x}_i^{(k)},y_i)S_i(\hat{\gamma}|\boldsymbol{x}_i^{(k)},y_i)'\right]-\sum_{i=1}^{n}\dot{Q}_i(\hat{\gamma}|\hat{\gamma})\dot{Q}_i(\hat{\gamma}|\hat{\gamma})'\right\},\tag{5.6}$$

where $\boldsymbol{x}_{i,mis}^{(1)},\boldsymbol{x}_{i,mis}^{(2)},\ldots,\boldsymbol{x}_{i,mis}^{(m_i^{(t)})}$ is an MCMC sample of size $m_i^{(t)}$, from $f(x_{i,mis}|x_{i,obs},\hat{\gamma})$, and $\boldsymbol{x}_i^{(k)}=(\boldsymbol{x}_{i,mis}^{(k)\prime},\boldsymbol{x}_{i,obs}')'$. Thus, the estimate of the asymptotic covariance matrix of $\hat{\gamma}$ is $[\mathcal{I}(\hat{\gamma})]^{-1}$.

Finally, we note that when there are ties in the failure times, (5.7), (5.2), and (5.3) can be modified as

$$Q(\gamma|\gamma^{(t)})=\sum_{i=1}^{d}\left[d_i\log h_0(y_i)+\frac{1}{m_i^{(t)}}\sum_{k=1}^{m_i^{(t)}}\boldsymbol{z}_i^{(tk)\prime}\beta\right]-\sum_{i=1}^{d}h_0(y_i)\sum_{j\in\mathcal{R}(y_i)}\left[\frac{1}{m_j^{(t)}}\sum_{k=1}^{m_j^{(t)}}\exp(\boldsymbol{x}_j^{(tk)\prime}\beta)\right]$$
$$+\sum_{i=1}^{n}\frac{1}{m_i^{(t)}}\sum_{k=1}^{m_i^{(t)}}\log f(\boldsymbol{x}_{i,mis}^{(tk)},\boldsymbol{x}_{i,obs}|\alpha),\tag{5.7}$$

where $\boldsymbol{z}_i^{(tk)}=\sum_{j\in\mathcal{D}_i}(\boldsymbol{x}_{j,mis}^{(tk)\prime},\boldsymbol{x}_{j,obs}')'$,

$$\beta^{(t+1)}=\arg\max_{\beta}\sum_{i=1}^{d}\left\{\left[\frac{1}{m_i^{(t)}}\sum_{k=1}^{m_i^{(t)}}\boldsymbol{z}_i^{(tk)\prime}\beta\right]-d_i\log\left[\sum_{j\in\mathcal{R}(y_i)}\frac{1}{m_j^{(t)}}\sum_{k=1}^{m_j^{(t)}}\exp(\boldsymbol{x}_j^{(tk)\prime}\beta)\right]\right\},\tag{5.8}$$

and

$$h_0^{(t+1)}(y_i)=d_i\left[\sum_{j\in\mathcal{R}(y_i)}\frac{1}{m_j^{(t)}}\sum_{k=1}^{m_j^{(t)}}\exp(\boldsymbol{x}_j^{(tk)\prime}\beta^{(t+1)})\right]^{-1},\ H_0^{(t+1)}(y_i)=\sum_{y_j\le y_i,\delta_j=1}h_0^{(t+1)}(y_j).\tag{5.9}$$

The calculation of $\mathcal{I}(\hat{\gamma})$ needs to be modified accordingly in the presence of ties. Again, the above formulation can be easily extended to the case where $\boldsymbol{x}_{i,mis}$ is categorical.

6 Analysis of Small Cell Lung Cancer Data

For the LCCC 9719 data discussed in Section 2, we use the proposed methods to estimate the regression coeffients assuming the missing covariates are MAR. We consider a Cox regression model for $[y_i|\boldsymbol{x}_i,\boldsymbol{\beta},h_0]$ allowing for right censoring. Thus, we have

NIH-PA Author Manuscript

NIH-PA Author Manuscript

$$f(y_i|\delta_i, \boldsymbol{x}_i, \beta, h_0) = \left[h_0(y_i)\exp\{\boldsymbol{x}_i'\beta)\}\right]^{\delta_i}\exp\{-H_0(y_i)\exp(\boldsymbol{x}_i'\beta)\},$$

where $\boldsymbol{x}_i = (x_{i1},\ldots, x_{i5})'$ is a 5 × 1 vector of covariates, $i = 1, 2, \ldots, n$, $\boldsymbol{\beta} = (\beta_1,\ldots,\beta_5)'$ is the vector of the corresponding regression coeffcients, $h_0(y_i)$ and $H_0(y_i)$ denote the baseline hazard function and the cumulative baseline hazard function, respectively. Since x_{i1}, x_{i2}, and x_{i3} are always observed, they do not need to be modeled. Thus, we only need to model two missing covariates (x_4, x_5) conditioning on the completely observed covariates throughout. We consider two models: $[x_4|x_1, x_2, x_3][x_5|x_1, x_2, x_3, x_4]$ and $[x_4|x_1, x_2, x_3, x_5][x_5|x_1, x_2, x_3]$ for (x_4, x_5). We use a logistic regression model for x_{i4} and a normal linear regression model for x_{i5}. Specifically, for example, for $[x_4|x_1, x_2, x_3][x_5|x_1, x_2, x_3, x_4]$, we have

$$f(x_{i4}|x_{i1}, x_{i2}, x_{i3}, \alpha_4) = \frac{\exp\{x_{i4}(\alpha_{40}+\alpha_{41}x_{i1}+\alpha_{42}x_{i2}+\alpha_{43}x_{i3})\}}{1+\exp(\alpha_{40}+\alpha_{41}x_{i1}+\alpha_{42}x_{i2}+\alpha_{43}x_{i3})},$$

where $\boldsymbol{\alpha}_4 = (\alpha_{40}, \alpha_{41}, \alpha_{42}, \alpha_{43})'$, and

$$f(x_{i5}|x_{i1}, x_{i2}, x_{i3}, x_{i4}, \alpha_5) = \frac{1}{\sqrt{2\pi\alpha_{55}}}\exp\{-[x_{i5} - (\alpha_{50}+\alpha_{51}x_{i1}+\cdots+\alpha_{54}x_{i4})]^2/(2\alpha_{55})\},$$

where $\boldsymbol{\alpha}_5 = (\alpha_{50}, \alpha_{51}, \ldots, \alpha_{55})'$.

To illustrate how to apply the Theorems presented in Sections 2 and 3, we consider a subset of the LCCC 9719 data, which is given in Table 2. Since all of the covariates are observed in this subset, using (3.1) after excluding the rows corresponding to $\delta_i = 0$ or $\boldsymbol{x}_j = \boldsymbol{x}_i$, X^* is a 35× 5 matrix. The first 8 rows are given by $\boldsymbol{x}_i' - \boldsymbol{x}_1'$ for $i = 1, 2, \ldots, 8$, and the last row is given by $\boldsymbol{x}_9' - \boldsymbol{x}_7'$. Using Maple (Version 8) *linsolve*, with Maple code "linsolve($\boldsymbol{X}^*$, $\boldsymbol{v}$);", after loading a *linalg* package, we obtain a closed form solution for $X^{*\prime}\boldsymbol{v} = 0$ and find that there indeed exists a positive vector $\boldsymbol{v} > 0$ satisfying $X^{*\prime}\boldsymbol{v} = 0$. Also, $|X^{*\prime}X^*| = 9.2344 \times 10^{10} > 0$. Thus, conditions $(C1)$ and $(C2)$ given in Theorem 3.1 are met for this subset. As discussed in Remark 3.4, when the conditions $(C1)$ and $(C2)$ are satisfied for a subset of the data, these two conditions hold for the entire set of completely observed cases. In addition, we can show that $\lim_{\|\boldsymbol{\alpha}\|\to\infty} L(\boldsymbol{\alpha}|D_{obs}) = 0$, where $L(\boldsymbol{\alpha}|D_{obs})$ is defined by (4.3), using the results established in Chen and Shao (2001) and hence, details are omitted here for brevity. Thus, based on Theorem 4.1, the MLE does exist for the entire dataset.

Since the MPLE of $\boldsymbol{\beta}$ and the MLE of $(\boldsymbol{\beta}, h_0, \boldsymbol{\alpha})$ exist for this dataset, we can compute various estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}_4$ and $\boldsymbol{\alpha}_5$. We standardized age and QOL score in order to make the numerical computations more stable. We used the SAS procedure PHREG to obtain the MPLE of $\boldsymbol{\beta}$ for the complete case (CC) analysis (i.e, an analysis deleting all of the missing values). The MCEM algorithm discussed in Section 5 was implemented using FORTRAN 77 with IMSL, the estimated observed information matrix $\mathcal{Q}(\hat{\boldsymbol{\gamma}})$ given by (5.6) is of dimension (102 +15)×(102+15), and its inverse was computed via the IMSL subroutine DLINDS. The Gibbs sampling algorithm was used to generate the Monte Carlo sample with 500 "burn-in" iterations at each MCEM iteration. In the MCEM, we took $m^{(0)} = 500$ and $\Delta m = 50$. The convergence criterion for the MCEM algorithm for obtaining the MLE was that the squared distance between the t^{th} and $(t + 20)^{th}$ iterations was less than 10^{-3}. The MCEM algorithm for obtaining the MLE of $(\boldsymbol{\beta}, h_0, \boldsymbol{\alpha})$ required only 25 iterations using $m(t) = 1750$ at convergence.

The resulting MPLEs and MLEs are shown in Tables 3, 4, and 5 for the complete case (CC) analysis as well as analyses incorporating all of the cases with two different models for (x_4, x_5). In the tables, standard errors (SEs), Z-statistics, p-values, and 95% confidence intervals for $\boldsymbol{\beta}$ are also reported. We can see some differences between the estimates in Tables 3 and 4. In the CC analysis, the 95% confidence interval for β_1 is (−0.024, 0.967) while the 95% confidence interval is (0.133, 0.820) in the analysis incorporating all of the cases, which indicates that the regression coefficient for treatment is not significant at the 0.05 level in the CC analysis, but significant in the analysis incorporating all of the cases. This indicates that continuous therapy followed by second line therapy may have a strong effect (i.e., more beneficial) compared to defined duration of therapy with respect to time to progression. Also, the SEs from the analysis incorporating all of the cases are consistently smaller than those from the CC analysis for all of the β_j's. This is expected since more information is used in the all case analysis.

The reason why we considered two models for (x_4, x_5) is that there are two possibilities in modeling the joint covariate distribution as a sequence of one dimensional conditional distributions. As Ibrahim, Lipsitz, and Chen (1999) point out, it is important to conduct a sensitivity analysis to examine whether inference about the parameters of primary interest, which are the β_j's in this case, is robust with respect to the order of conditioning in the covariate distributions. From Tables 4 and 5, both estimates and SEs for all the β_j's are very close for these two joint covariate distributions. Thus, inference about $\boldsymbol{\beta}$ is quite robust with respect to these two different orders of conditioning.

Finally, the estimated baseline hazard functions $h_0(y)$ are plotted in Figure 1 for the complete case analysis as well as the analysis incorporating all of the cases, labeled Complete Cases and All Cases, respectively. In the all case analysis, the model $[x_4|x_1, x_2, x_3][x_5|x_1, x_2, x_3, x_4]$ for (x_4, x_5) was used since an almost identical estimated baseline hazard function was obtained under the model $[x_4|x_1, x_2, x_3, x_5][x_5|x_1, x_2, x_3]$. Strikingly, the CC analysis resulted in a much different (larger) estimate of the baseline hazard than the all case analysis, which further demonstrates the importance of incorporating all of the cases into the analysis.

Acknowledgments

The authors wish to thank the Editor, the Associate Editor, and two referees for helpful comments and suggestions which have improved the paper. Dr. Chen and Dr. Ibrahim's research was partially supported by NIH grants #GM 70335 and #CA 74015. Dr. Shao's research was partially supported by HKUST DAG05/06.SC27 and RGC 602206.

References

Booth JG, Hobert JP. Maximizing Generalized Linear Mixed Model likelihoods with an Automated Monte Carlo EM Algorithm. Journal of the Royal Statistical Society, Series B 1999;61:265–285.

Breslow NE. Covariance Analysis of Censored Survival Data. Biometrics 1974;30:89–99. [PubMed: 4813387]

Chen HY. Double Nonparametric Likelihood Method for the Cox regression model with missing covariates. Journal of the American Statistical Association 2002;97:565–576.

Chen HY, Little RJA. Proportional Hazards Regression with Missing Covariates. Journal of the American Statistical Association 1999;94:896–908.

Chen MH, Ibrahim JG. Maximum Likelihood Methods for Cure Rate Models with Missing Covariates. Biometrics 2001;57:43–52. [PubMed: 11252617]

Chen MH, Ibrahim JG, Shao QM. On Propriety of the Posterior Distribution and Existence of the Maximum Likelihood Estimator for Regression models with Covariates Missing at Random. Journal of the American Statistical Association 2004;99:421–438.

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, Shao QM. Propriety of Posterior Distribution for Dichotomous Quantal Response Models With General Link Functions. Proceedings of the American Mathematical Society 2001;129:293–302.

Cox DR. Regression Models and Life Tables (with Discussion). Journal of the Royal Statistical Society, Series B 1972;34:187–220.

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

Herring AH, Ibrahim JG. Likelihood-based Methods for Missing Covariates in the Cox Proportional Hazards Model. Journal of the American Statistical Association 2001;96:292–302.

Herring AH, Ibrahim JG, Lipsitz SR. Frailty Models With Missing Covariates. Biometrics 2002;58:98–109. [PubMed: 11890332]

Herring AH, Ibrahim JG, Lipsitz SR. Nonignorably Missing Covariate Data in Survival Analysis: A Case Study of an International Breast Cancer Study Group Trial. Applied Statistics 2004;53:293–310.

Huang L, Chen M-H, Ibrahim JG. Bayesian Analysis for Generalized Linear Models with Nonignorably Missing Covariates. Biometrics 2005;61:729–737. [PubMed: 16135024]

Ibrahim JG. Incomplete Data in Generalized Linear Models. Journal of the American Statistical Association 1990;85:765–769.

Ibrahim JG, Lipsitz SR, Chen MH. Missing Covariates in Generalized Linear Models When the Missing Data Mechanism is Nonignorable. Journal of the Royal Statistical Society, Series B 1999;61:173–190.

Jacobsen M. Existence and Unicity of MLEs in Discrete Exponential Family Distributions. Scandinavian Journal of Statistics 1989;16:335–349.

Johansen S. An Extension of Cox's Regression Model. International Statistical Review 1983;51:258–262.

Klein, JP.; Moeschberger, ML. Survival Analysis. Vol. 2. New York: Springer-Verlag; 2003.

Lin DY, Ying Z. Cox Regression With Incomplete Covariate Measurements. Journal of the American Statistical Association 1993;88:1341–1349.

Lipsitz SR, Ibrahim JG. Using the EM Algorithm for Survival Data with Incomplete Categorical Covariates. Lifetime Data Analysis 1996;2:5–14. [PubMed: 9384645]

Lipsitz SR, Ibrahim JG. Estimating Equations with Incomplete Categorical Covariates in the Cox Model. Biometrics 1998;54:1002–1013. [PubMed: 9750248]

Lipsitz SR, Ibrahim JG. Estimation with Correlated Censored Survival Data with Missing Covariates. Biostatistics 2000;1:315–327. [PubMed: 12933512]

Louis T. Finding the Observed Information Matrix When Using the EM Algorithm. Journal of the Royal Statistical Society, Series B 1982;44:226–233.

Martinussen T. Cox Regression with Incomplete Covariate Measurements Using the EM-Algorithm. Scandinavian Journal of Statistics 1999;26:479–491.

Paik MC. Multiple Imputation for the Cox Proportional Hazards Model with Missing Covariates. Lifetime Data Analysis 1997;3:289–298. [PubMed: 9384657]

Paik MC, Tsai WY. On Using the Cox Proportional Hazards Model with Missing Covariates. Biometrika 1997;84:579–593.

Pons O. Estimation in the Cox Model with Missing Covariate Data. Journal of Nonparametric Statistics 2002;14:223–247.

Rockafellar, RT. Convex Analysis. Princeton, N.J: Princeton University Press; 1970.

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, Series B 2007;69:607–623.

Schluchter M, Jackson K. Log-linear Analysis of Censored Survival Data with Partially Observed Covariates. Journal of the American Statistical Association 1989;84:42–52.

Socinski MA, Schell MJ, Peterman A, Bakri K, Yates S, Gitten R, Unger P, Lee J, Lee Ji, Tynan M, Moore M, Kies M. Phase III Trial Comparing Defined Duration of Therapy Versus Continuous Therapy Followed by Second-Line Therapy in Advanced-Stage IIIB/IV Non-Small-Cell Lung Cancer. Journal of Clinical Oncology 2002;20:1335–1343. [PubMed: 11870177]

Tsai HK, Chen M-H, McLeod DG, Carroll PR, Richie JP, D'Amico AV. Cancer-Specific Mortality Following Radical Prostatectomy or Radiation Therapy with Short-Course Hormonal Therapy in Men with Localized, Unfavorable-Risk Prostate Cancer. 2006Submitted

Appendix: Proofs of Theorems

We first establish a useful result, which is formally stated in the following lemma.

Lamma A.1

Let X^* be an $n^* \times p$ matrix ($p < n^*$). Also let R^{n^*} denote the n^*-dimensional Euclidean space. If there is no positive vector $\mathbf{v} = (v_1, v_2, \ldots, v_{n^*})' \in R^{n^*}$ (denoted by $\mathbf{v} > 0$, i.e., $v_i > 0$ for i = 1, 2, …, n^*) such that

$$X^{*\prime}\boldsymbol{v}=0, \tag{A.1}$$

then there exists a non-zero vector $\mathbf{b} \in R^p$ such that

$$\boldsymbol{b}'\boldsymbol{x}_i^* \le 0, \tag{A.2}$$

where $\boldsymbol{x}_i^*$ is the i^{th} row of X^*.

Proof

Let $\mathcal{V} = \{X^{*\prime}\boldsymbol{v}: \boldsymbol{v} > 0, \boldsymbol{v} \in R^{n^*}\}$. Then $\mathcal{V}$ is a convex cone in R^p (see Theorem 2.6 in Rockafellar (1970)). Since (A.1) does not hold, by Corollary 11.7.3 of Rockafellar (1970), there exists some non-zero vector $\boldsymbol{b}$ such that $\forall\ \boldsymbol{v} > 0$, $\boldsymbol{b}'X^{*\prime}\boldsymbol{v} \le 0$ and hence $\forall\ \boldsymbol{v} \ge 0$, $\boldsymbol{b}'X^{*\prime}\boldsymbol{v} \le 0$. In particular, (A.2) holds.

Proof of Theorem 3.1

Observe that for $\delta = 0$ or 1 and $x > -1$

$$\left(\frac{1}{1+x}\right)^{\delta}=\int_0^{\infty} e^{-t(1+\delta x)}dt. \tag{A.3}$$

Without loss of generality (WLOG), assume $y_1 \le y_2 \le \ldots \le y_n$. Then

$$\begin{aligned}
L_p(\beta|D_{obs})&=\prod_{i=1}^{n}\left(\frac{\exp(\boldsymbol{x}_i'\beta)}{\sum_{j\in\mathcal{R}(y_i)}\exp(\boldsymbol{x}_j'\beta)}\right)^{\delta_i}=\prod_{i=1}^{n}\left(\frac{1}{1+\sum_{j>i}\exp((\boldsymbol{x}_j-\boldsymbol{x}_i)'\beta)}\right)^{\delta_i}\\
&=\prod_{i=1}^{n}\int_0^{\infty}\exp\left(-t_i(1+\delta_i\sum_{j>t}\exp((\boldsymbol{x}_j-\boldsymbol{x}_i)'\beta))\right)dt_i\\
&=\int_{R^{+n}}\exp(-\sum_{i=1}^{n}t_i)\prod_{1\le i\le n, j>i}\left(\exp\left(-\exp(-(\boldsymbol{x}_i-\boldsymbol{x}_j)'\beta)\right)\right)^{t_i\delta_i}d\boldsymbol{t}\\
&=\int_{R^{+n}}\exp(-\sum_{i=1}^{n}t_i)\prod_{1\le i\le n, j>i}F((\boldsymbol{x}_i-\boldsymbol{x}_j)'\beta)^{t_i\delta_i}d\boldsymbol{t},
\end{aligned} \tag{A.4}$$

where $\boldsymbol{t} = (t_1, t_2, \ldots, t_n)'$, $R^{+n} = R^+ \times \ldots \times R^+$ with $R^+ = (0, \infty)$, and $F(u) = \exp(-\exp(-u))$.

Sufficiency—WLOG, we assume that $L_p(\boldsymbol{\beta}|D_{obs}) \not\equiv 0$. Then, there exists a $\boldsymbol{\beta}_0$ such that $L_p(\boldsymbol{\beta}_0|D_{obs}) > 0$. Let $M > 1$ such that

$$\frac{1}{1 - \log F(-M)} < L_p(\beta_0|D_{obs}).$$

For $\boldsymbol{\beta}$ satisfying $\max_{1\le i\le n, j>i} \delta_i(\boldsymbol{x}_j - \boldsymbol{x}_i)'\boldsymbol{\beta} > M$, there exist i_0 and j_0 achieving the maximum such that $\delta_{i0}(\boldsymbol{x}_{i0} - \boldsymbol{x}_{j0})'\boldsymbol{\beta} < -M$. Since F is a nondecreasing distribution function, we have

$$\begin{aligned} L_p(\beta|D_{obs}) &\le \int_{R^{+n}} \exp(-\sum_{i=1}^n t_i) F(\delta_{i_0}(\boldsymbol{x}_{i_0} - \boldsymbol{x}_{j_0})'\beta)^{t_{i_0}} d\boldsymbol{t} \le \int_{R^{+n}} \exp(-\sum_{i=1}^n t_i) F(-M)^{t_{i_0}} d\boldsymbol{t} \\ &= \int_0^\infty \exp\{-[1 - \log F(-M)] t_{i_0}\} dt_{i_0} = \frac{1}{1 - \log F(-M)} \le L_p(\beta_0|D_{obs}). \end{aligned} \quad (A.5)$$

When $\max_{1\le i\le n, j>i} \delta_i(\boldsymbol{x}_j - \boldsymbol{x}_i)'\boldsymbol{\beta} \le M$, following Lemma 4.1 in Chen and Shao (2001), conditions ($C1$) and ($C2$) imply that

$$\| \beta \| = \sqrt{\beta'\beta} \le D \ \text{ for some } 0 < D < \infty. \quad (A.6)$$

Combining (A.5) and (A.6) leads to $\sup_{\boldsymbol{\beta}} L_p(\boldsymbol{\beta}|D_{obs}) = \sup_{\boldsymbol{\beta}:\|\boldsymbol{\beta}\|\le D} L_p(\boldsymbol{\beta}|D_{obs})$. Since $L_p(\boldsymbol{\beta}|D_{obs})$ is a continuous and bounded function, there exists a $\hat{\boldsymbol{\beta}}$ such that

$$L_p(\widehat{\beta}|D_{obs}) = \sup_{\beta:\|\beta\|\le D} L_p(\beta|D_{obs})$$

and hence the MPLE exists.

Necessity—Assume that the MPLE of $\boldsymbol{\beta}$ exists. Then, there is a $\boldsymbol{\beta}^*$ such that

$$L_p(\beta^*|D_{obs}) = \sup_\beta L_p(\beta|D_{obs}) \ \text{ and } \ \| \beta^* \| < \infty.$$

Assume that condition ($C2$) does not hold. Then, by Lemma A.1, there exists a non-zero vector $\boldsymbol{b}$ such that $\delta_i(\boldsymbol{x}_j - \boldsymbol{x}_i)'\boldsymbol{b} \le 0$ for all $1 \le i \le n$ and $j > i$. Thus,

$$\begin{aligned} L_p(\beta^* + s\boldsymbol{b}|D_{obs}) &= \int_{R^{+n}} \exp(-\sum_{i=1}^n t_i) \prod_{1\le i\le n, j>i} F(\delta_i(\boldsymbol{x}_i - \boldsymbol{x}_j)'\beta^* + \delta_i s(\boldsymbol{x}_i - \boldsymbol{x}_j)'\boldsymbol{b})^{t_i\delta_i} d\boldsymbol{t} \\ &= \int_{R^{+n}} \exp(-\sum_{i=1}^n t_i) \prod_{1\le i\le n, j>i} F(\delta_i(\boldsymbol{x}_i - \boldsymbol{x}_j)'\beta^* - s\delta_i(\boldsymbol{x}_j - \boldsymbol{x}_i)'\boldsymbol{b})^{t_i\delta_i} d\boldsymbol{t} \end{aligned}$$

which is an increasing function of s when condition ($C1$) holds. This is a contradiction. This shows that condition ($C2$) is necessary for the existence of the MPLE for $\boldsymbol{\beta}$ if condition ($C1$) is satisfied.

Proof of Theorem 4.1

Write

$$L^*(\beta,\alpha|D_{obs})=\int\left\{\prod_{i=1}^{d}\exp(\boldsymbol{x}_i'\beta)\left[\sum_{j\in\mathcal{R}(y_i)}\exp(\boldsymbol{x}_j'\beta)\right]^{-1}\right\}\left[\prod_{i=1}^{n}f(\boldsymbol{x}_{i,mis},\boldsymbol{x}_{i,obs}|\alpha)\right]d\boldsymbol{x}_{mis} \tag{A.7}$$

It is sufficient to prove that

$$\lim_{\|\beta\|+\|\alpha\|\to\infty}L^*(\beta,\alpha|D_{obs})=0. \tag{A.8}$$

Observe that

$$L_p(\beta|D)=\prod_{i=1}^{d}\exp(\boldsymbol{x}_i'\beta)\left[\sum_{j\in\mathcal{R}(y_i)}\exp(\boldsymbol{x}_j'\beta)\right]^{-1}$$

is an increasing function in $\boldsymbol{x}_j'\beta$ for $\delta_j = 1$ and a decreasing function in $\boldsymbol{x}_j'\beta$ for $\delta_j = 0$. For $1 \le l \le p$, let $x_{il}^R=b_i^*=\delta_i a_i+(1-\delta_i)b_i$ if $\beta_l \ge 0$ and $x_{il}^R=a_i^*=(1-\delta_i)a_i+\delta_i b_i$ if $\beta_l < 0$. Write $x_i^*=((\boldsymbol{x}_{i,mis}^R)',\boldsymbol{x}_{i,obs}')'$ and $\boldsymbol{x}_{i,mis}^R=(x_{il}^R, r_{il}=0, 1\le l\le p)'$. Let $\mathcal{R}_i = \mathcal{R}(y_i) - \{i\}$. Then we have

$$L_p(\beta|D)\le L_p^*(\beta|D)\equiv\prod_{i=1}^{d}\exp((\boldsymbol{x}_i^*)'\beta)\left[\exp((\boldsymbol{x}_i^*)'\beta)+\sum_{\delta_j=0,\,j\in\mathcal{R}_i}\exp((\boldsymbol{x}_j^*)'\beta)\right]^{-1}. \tag{A.9}$$

It directly follows from (A.7), (A.9) and (4.3) that

$$L^*(\beta,\alpha|D_{obs})\le L_p^*(\beta|D)L(\alpha|D_{obs}).$$

Following the proof of Theorem 3.1, $\lim_{\|\beta\|\to\infty}L_p^*(\beta|D)=0$ if conditions ($C2^*$) and ($C3^*$) are satisfied. Consequently, we obtain (A.8) under condition ($C1^*$).

Proof of Theorem 4.2

Let $\mathbf{1} = (1, 1, \ldots, 1)'$. Then we have

$$L_p(\beta|D) \le \prod_{\boldsymbol{r}_i=\mathbf{1},\ 1\le i\le d} \exp(\boldsymbol{x}_i'\beta)\left[\exp(\boldsymbol{x}_i'\beta)+\sum_{\delta_j=0,\boldsymbol{r}_j=\mathbf{1},\ j\in\mathcal{R}_i} \exp(\boldsymbol{x}_j'\beta)\right]^{-1}.$$

Therefore, the above inequality, condition ($C1^*$) in Theorem 4.1, and conditions ($C1$) and ($C2$) stated in Theorem 3.1 directly yield the existence of the MLE of ($\boldsymbol{\beta}$, h_0, $\boldsymbol{\alpha}$).

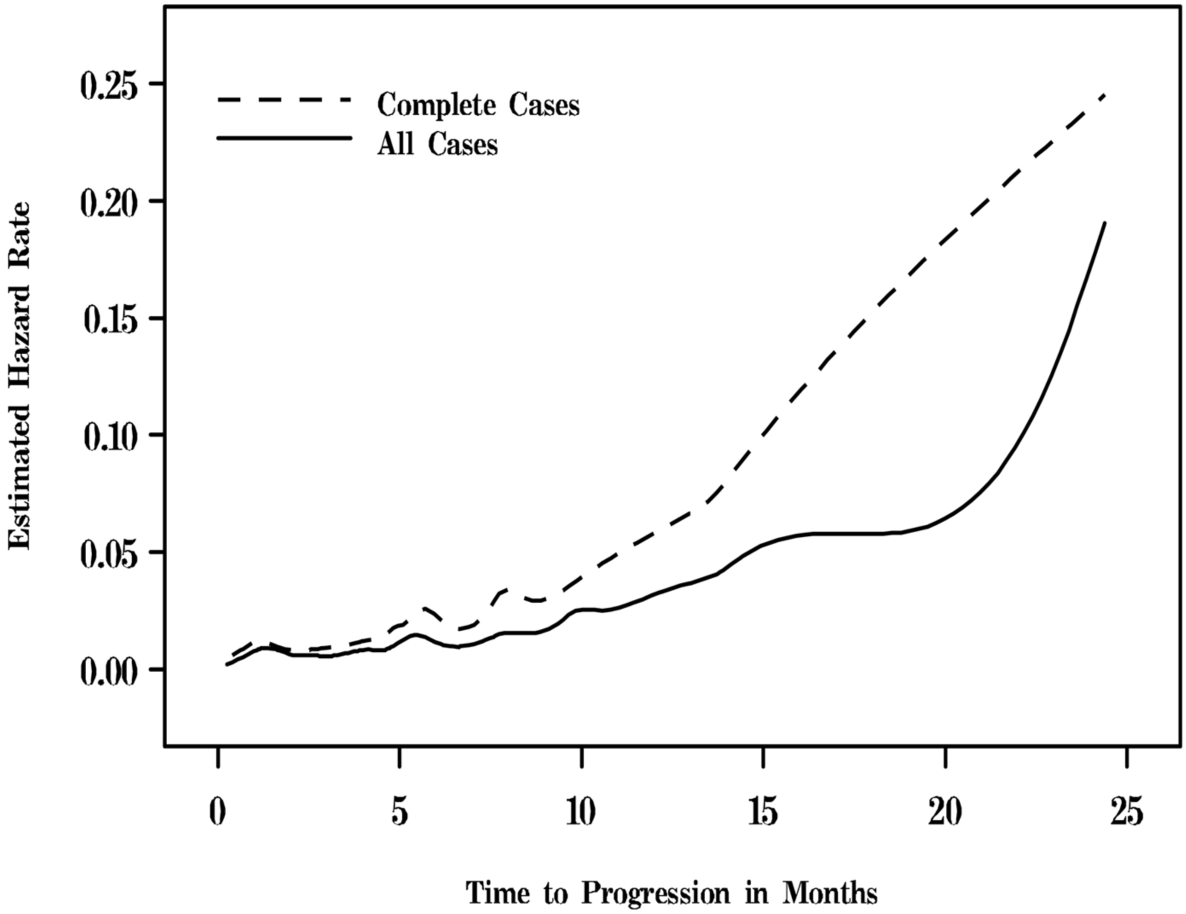

Figure 1.
Estimated baseline hazard function ($h_0(y)$) for CC and all cases analyses.

Table 1

Summary of LCCC 9719 Data

completely observed variables		
x_1 (frequency)	A	114
	B	116
x_2 (frequency)	Male	144
	Female	86
x_3 (years)	mean	62.24
	std dev	10.17
y (frequency)	censored	83
	relapsed	147
missing covariates		
x_4 (frequency)	0	155
	1	10
	missing	65
x_5 (QOL score)	mean	78.14
	std dev	15.31
	missing	81
both x_4 and x_5	missing	27
one of x_4 or x_5	missing	119

Table 2

A Subset of LCCC 9719 Data

Obs (*i*)	y_i	δ_i	x_{i1}	x_{i2}	x_{i3}	x_{i4}	x_{i5}
1	0.394	1	1	0	68	0	54
2	1.083	1	0	0	81	0	79
3	1.116	1	1	1	82	0	64
4	1.149	1	0	1	58	1	86
5	1.313	1	1	1	52	1	54
6	3.973	1	1	0	69	1	92
7	6.665	1	0	0	54	1	83
8	9.521	0	1	0	62	0	67
9	14.380	0	0	1	81	0	80

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 3

Maximum Partial Likelihood Estimates of β for Complete Case Analysis

Parameter	**MPLE**	**SE**	**Z-statistic**	***p*-value**	**95% CI**
β_1	0.471	0.253	1.864	0.062	(−0.024, 0.967)
β_2	0.068	0.243	0.280	0.780	(−0.409, 0.545)
β_3	−0.020	0.130	−0.154	0.878	(−0.275, 0.235)
β_4	0.878	0.411	2.140	0.032	(0.074, 1.684)
β_5	−0.138	0.119	−1.158	0.247	(−0.372, 0.096)

Table 4

Maximum Likelihood Estimates of β Based On All Observed Data and Model $[x_4|x_1, x_2, x_3][x_5|x_1, x_2, x_3, x_4]$ for Missing Covariates

Parameter	MLE	SE	*Z*-statistic	*p*-value	95% CI
β_1	0.477	0.175	2.723	0.006	(0.133, 0.820)
β_2	0.174	0.180	0.966	0.334	(−0.179, 0.528)
β_3	−0.021	0.090	−0.238	0.812	(−0.198, 0.155)
β_4	0.914	0.381	2.400	0.016	(0.168, 1.661)
β_5	−0.052	0.105	−0.490	0.624	(−0.258, 0.155)

NIH-PA Author Manuscript
NIH-PA Author Manuscript

Table 5

Maximum Likelihood Estimates of β Based On All Observed Data and Model $[x_4|x_1, x_2, x_3, x_5][x_5|x_1, x_2, x_3]$ for Missing Covariates

Parameter	MLE	SE	*Z*-statistic	*p*-value	95% CI
β_1	0.477	0.175	2.722	0.006	(0.133, 0.820)
β_2	0.173	0.180	0.959	0.338	(−0.181, 0.527)
β_3	−0.021	0.090	−0.233	0.816	(−0.197, 0.155)
β_4	0.914	0.388	2.356	0.018	(0.154, 1.674)
β_5	−0.053	0.106	−0.501	0.616	(−0.261, 0.155)