Sankhyā: The Indian Journal of Statistics
Special Issue on Biostatistics
2000, Volume 62, Series B, Pt. 1, pp. 25–42

MISSPECIFICATION OF MARGINAL HAZARDS MODELS IN MULTIVARIATE FAILURE TIME DATA

By LIMIN X. CLEGG*
National Cancer Institute, Bethesda
JIANWEN CAI**
PRANAB KUMAR SEN
and
LAWRENCE L. KUPPER
University of North Carolina, Chapel Hill

SUMMARY. Hazard functions for correlated censored data are usually formulated through the Cox regression model in a marginal regression framework. We investigate properties of the maximum pseudo partial likelihood estimator vector under a possibly misspecified marginal Cox regression model. The estimator vector is shown to be consistent for an implicitly defined parameter vector and is asymptotically Gaussian as well, with a covariance matrix that can be consistently estimated. The general results are applied to some special cases, including the case of misspecifying the type of baseline hazards function for the Cox model when the regression functional form is correctly specified. Simulation results confirm that the asymptotic results are applicable for sample sizes seen in practice.

1. Introduction

Multivariate failure time data, also referred to as correlated or clustered failure time data, arise when more than one failure outcome is observed for an individual, or when group randomization or cluster sampling is used, or both. For example, in the well known Framingham Heart Study (Dawber (1980)), times to myocardial infarction, cerebrovascular accident, cancer, and so on, were observed for each individual. Also, some individuals in the study were related because the sampling unit was family. In the data used by Klein (1992), for instance, about 54% were married couples and 25% were siblings. The Cox proportional hazards model (Cox (1972, 1975)) has been extensively used to assess covariate effects on failure times. The

---

AMS (1991) *subject classification.* 62H99, 62P10.
Key words and phrases. Censoring, Cox regression, correlated failure times, marginal hazard rate, martingale, model misspecification, multivariate failure times, partial likelihood, semiparametric model, survival data.
*Work partially supported by the grant R29 HL57444
**Research supported by the grants T32 ES07018 and R29 HL57444 from the (U.S.) National Institute of Health.

Cox model, however, may not be applied directly to multivariate failure time data because of its basic assumption of independence among failure times, conditional on the covariates in the model.

Much recent research effort has been devoted to generalizing the Cox regression model to deal with multivariate failure time data via two different approaches. In the copula or frailty model approach, the dependency in multivariate failure time data is modeled parametrically (Vaupel, Manton and Stallard (1979), Clayton and Cuzick (1985), Hougaard (1987), Oakes (1989), Klein (1992), Nielsen, Gill, Andersen and Sørensen (1992), Andersen, Borgan, Gill and Keiding (1993), Murphy (1994, 1995), and Bandeen-Roche and Liang (1996)). In contrast, the marginal regression modeling approach focuses on "population-averaged" (Zeger, Liang and Albert (1988)) covariate effects and the dependence structure is left unspecified (Wei, Lin and Weissfeld (1989), Huster, Brookmeyer and Self (1989), Lee, Wei and Amato (1992), Liang, Self and Chang (1993), Lin (1994), Cai and Prentice (1995, 1997), Therneau (1996), Clegg, Cai and Sen (1996, 1999), and Prentice and Hsu (1997)).

Let $\boldsymbol{T}_i' = (T_{i11}, \cdots, T_{i1K}, \cdots, T_{iJK}), i = 1, \cdots, n$, denote independent failure time response vectors, where T_{ijk} is the failure time for the kth type of failure on subject j in cluster i. For example, in the Framingham Heart Study, $(T_{i11}, T_{i12}, T_{i21}, T_{i22})$ may denote the times from entering into the study until the first occurrence of coronary heart disease or the first occurrence of cerebrovascular accident for the husband and wife from the ith family. Wei et al. (1989) and Cai and Prentice (1995) postulated that the marginal hazards model for T_{ijk} vary functionally with the p-dimensional covariate vector $\boldsymbol{Z}_{ijk}(t)$ as

$$\lambda_{ijk}(t; \boldsymbol{Z}_{ijk}(t)) = \lambda_{0jk}(t) \exp\{\boldsymbol{\beta}' \boldsymbol{Z}_{ijk}(t)\}; \tag{1}$$

in contrast, Lee et al. (1992), Liang et al. (1993) and Cai and Prentice (1997) modeled the marginal hazard as

$$\lambda_{ijk}(t; \boldsymbol{Z}_{ijk}(t)) = \lambda_0(t) \exp\{\boldsymbol{\beta}' \boldsymbol{Z}_{ijk}(t)\}. \tag{2}$$

In both of these formulations, $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)'$ is a p-dimensional vector of the unknown regression coefficients to be estimated, and $\lambda_{0jk}(t)$ and $\lambda_0(t)$ are unspecified baseline hazard functions. Note that we use $\boldsymbol{\beta}$ to denote the true value of the regression parameter and a general argument as well. Where it is necessary to make the distinction, $\boldsymbol{\beta}_0$ is used to denote the true parameter value. Notice that the only difference between models (1) and (2) is that model (1) uses a different baseline hazard function, $\lambda_{0jk}(t)$, for each type of failure and each subject in a cluster and model (2) assumes an identical baseline hazard, $\lambda_0(t)$, for all types of failures from all subjects. Hence, we refer to (1) as a distinct baseline hazards model and to (2) as a common baseline hazard model, respectively.

Instead of using either a different baseline hazard function for each type of failure and for each subject in a cluster, or an identical baseline function for all types of failure from all subjects, Clegg et al. (1996, 1999) proposed a marginal mixed baseline hazards model. A model is called a mixed baseline hazards model if the baseline hazard function is identical for some combinations of subjects and failure types but different for others. The mixed baseline hazards model is useful when, for instance, the baseline hazards are the same for all subjects, but are heterogeneous for different failure types, or vice versa. More generally, the mixed baseline hazards model may have an identical baseline for some subjects (or some failure types) in a cluster, but different baselines for the rest of the subjects (or failure types). For example, a mixed baseline hazards model with a different baseline for each subject and an identical baseline for all types of failures for a subject would be written as:

$$\lambda_{ijk}(t; \boldsymbol{Z}_{ijk}(t)) = \lambda_{0j}(t) \exp\{\boldsymbol{\beta}' \boldsymbol{Z}_{ijk}(t)\}. \tag{3}$$

The estimator, $\hat{\boldsymbol{\beta}}$, for the regression parameter is obtained via maximizing the pseudo partial likelihood. In the proof of the consistency and asymptotic Gaussian distribution of $\sqrt{n}\hat{\boldsymbol{\beta}}$, it is assumed that the marginal hazards functions $\lambda_{ijk}(t; \boldsymbol{Z}_{ijk}(t))$ are correctly specified. However, misspecification of a marginal hazards model may occur in a variety of ways. First, the model family may be misspecified. For example, the Cox regression model is assumed for analysis when, in fact, the accelerated failure time model holds. Second, if the choice of the Cox model is indeed correct, the wrong functional form for the regression portion, $\exp\{\boldsymbol{\beta}' \boldsymbol{Z}_{ijk}(t)\}$, may be assumed, which includes omitting important covariates from the model and/or choosing the wrong functional form for a covariate. Third, the baseline types may also be misspecified even when the Cox model is applicable and the correct functional form of $\exp\{\boldsymbol{\beta}' \boldsymbol{Z}_{ijk}(t)\}$ is used; for example, a common baseline hazard model could be assumed when a mixed baseline hazards model should actually be used.

Although the marginal regression modeling approach has become popular in analysis of multivariate failure time data, virtually no prior research on marginal hazards model misspecification has been done. In this paper we study the consequences of misspecification of $\lambda_{ijk}(t; \boldsymbol{Z}_{ijk}(t))$ in a marginal regression modeling framework. We shall confine our attention hereafter to the mixed baseline hazards model (3) since the technical development for the mixed baseline hazards model combines the features of both the distinct baseline hazards model and the common baseline hazard model. In the next section, we derive the asymptotic distributions of the estimator under a working marginal hazards model which is possibly misspecified. In Section 3, we apply the general results developed in Section 2 to some special cases, including the case of misspecifying the type of baseline hazards function for the Cox model when the functional form of the regression portion, $\exp\{\boldsymbol{\beta}' \boldsymbol{Z}_{ijk}(t)\}$, is correct. Simulation studies are conducted in Section 4 to obtain information on the consequences of misspecifying marginal hazards models for finite sample sizes applicable in practice. Finally, some concluding remarks are made in Section 5.

2. Asymptotic Distributions of Parameter Estimators under Possibly Misspecified Marginal Hazards Models

2.1 *Notation and definitions.* Suppose that there are n independent clusters. In each cluster, there are J subjects. For each subject, K types of failure may occur. We use (i,j,k) to denote the kth type of failure on subject j in cluster i. Let C_{ijk} denote the potential censoring time for T_{ijk}. Denote $X_{ijk} = \min(T_{ijk}, C_{ijk})$ and $\delta_{ijk} = I(T_{ijk} = X_{ijk})$, where $I(.)$ is the indicator function. We observe $(X_{ijk}, \delta_{ijk}, \boldsymbol{Z}_{ijk})$ for $(i,j,k)(i = 1, \cdots, I, j = 1, \cdots, J,$ and $k = 1, \cdots, K)$. If (i,j,k) does not exist, we set $C_{ijk} = 0$, which allows us to have varying cluster size and varying number of failures. Consequently, the data are assumed implicitly to be missing completely at random (MCAR) in the sense of Rubin (1976) as in any marginal approaches. In terms of counting process and martingale formulation, we let $Y_{ijk}(t) = I(X_{ijk} \geq t)$ denote the at risk indicator process for (i,j,k), $N_{ijk}(t) = I(X_{ijk} \leq t, \delta_{ijk} = 1)$ the counting process which registers failure for (i,j,k), and $\lambda_{ijk}(t)$ and $M_{ijk}(t) = N_{ijk}(t) - \int_0^t Y_{ijk}(u)\lambda_{ijk}(u)du$ the corresponding marginal hazards function and marginal martingale, respectively, with respect to the filtration $\mathcal{F}_{jk}(t^-)$, where $\mathcal{F}_{jk}(t) = \sigma\{N_{ijk}(u), Y_{i11}(u), \cdots,$ $Y_{iJK}(u), \boldsymbol{Z}_{i11}(u), \cdots, \boldsymbol{Z}_{iJK}(u); 0 \leq u \leq t, i = 1, \cdots, n\}$. Note that in the filtration the time-varying covariates $\boldsymbol{Z}_{ifg}(t)$, for $f \neq j$ and $g \neq k$, are external to the failure process for the kth type of failure on subject j whereas $\boldsymbol{Z}_{ijk}(t)$ can be internal time-varying covariates (Kalbfleisch and Prentice (1980), pp.123-125).

To avoid imposing extra asymptotic stability conditions, we assume that $(X_{ijk}, \delta_{ijk}, \boldsymbol{Z}_{ijk})$ $(i = 1, \cdots, n)$ are n independent and identically distributed replicates of $(X_{1jk}, \delta_{1jk}, \boldsymbol{Z}_{1jk})$ for $j = 1, \cdots, J$ and $k = 1, \cdots, K$. We also make the following common assumptions: time is absolutely continuous on the interval $\mathcal{T} = [0, \tau]$, where τ is the terminal time of the study; censoring is independent and noninformative; J and K are finite and they are not functions of n.

Denote the true marginal hazards function for (i,j,k) by $\lambda_{ijk}(t)$. Note that the true marginal hazards model $\lambda_{ijk}(t)$ may not even belong to the Cox regression model family. Let the assumed marginal hazards function be $\lambda_{ijk}(t; \boldsymbol{Z}_{ijk}(t))$, which is in the Cox regression form. Hereafter, the following notation will be used for the true marginal hazards function $\lambda_{ijk}(t)$ and for the assumed marginal hazards function $\lambda_{ijk}(t; \boldsymbol{Z}_{ijk}(t))$:

$$\boldsymbol{S}_{jk}^{(d)}(t) = n^{-1}\sum_{i=1}^{n} Y_{ijk}(t)\lambda_{ijk}(t)\boldsymbol{Z}_{ijk}(t)^{\otimes d}, \quad \boldsymbol{S}_{j.}^{(d)}(t) = \sum_{k=1}^{K}\boldsymbol{S}_{jk}^{(d)}(t), \quad \boldsymbol{s}_{jk}^{(d)}(t) = E(\boldsymbol{S}_{jk}^{(d)}(t)),$$

$$\boldsymbol{s}_{j.}^{(d)}(t) = \sum_{k=1}^{K}\boldsymbol{s}_{jk}^{(d)}(t), \quad \boldsymbol{s}_{..}^{(d)}(t) = \sum_{j=1}^{J}\sum_{k=1}^{K}\boldsymbol{s}_{jk}^{(d)}(t),$$

$$\boldsymbol{S}_{jk}^{(d)}(\boldsymbol{\beta}, t) = n^{-1}\sum_{i=1}^{n} Y_{ijk}(t)\boldsymbol{Z}_{ijk}(t)^{\otimes d}\exp\{\boldsymbol{\beta}'\boldsymbol{Z}_{ijk}(t)\},$$

$$\boldsymbol{S}_{j.}^{(d)}(\boldsymbol{\beta},t)=\sum_{k=1}^{K}\boldsymbol{S}_{jk}^{(d)}(\boldsymbol{\beta},t),\quad \boldsymbol{S}_{..}^{(d)}(\boldsymbol{\beta},t)=\sum_{j=1}^{J}\sum_{k=1}^{K}\boldsymbol{S}_{jk}^{(d)}(\boldsymbol{\beta},t),\quad \boldsymbol{s}_{jk}^{(d)}(\boldsymbol{\beta},t)=E(\boldsymbol{S}_{jk}^{(d)}(\boldsymbol{\beta},t)),$$

$$\boldsymbol{s}_{j.}^{(d)}(\boldsymbol{\beta},t)=\sum_{k=1}^{K}\boldsymbol{s}_{jk}^{(d)}(\boldsymbol{\beta},t),\quad s_{..}^{(d)}(\boldsymbol{\beta},t)=\sum_{j=1}^{J}\sum_{k=1}^{K}\boldsymbol{s}_{jk}^{(d)}(\boldsymbol{\beta},t)$$

for $d=0,1,2$, where, for a column vector $\boldsymbol{a}$, $\boldsymbol{a}^{\otimes 2}$ refers to the matrix $\boldsymbol{aa}'$, $\boldsymbol{a}^{\otimes 1}$ to the vector $\boldsymbol{a}$, and $\boldsymbol{a}^{\otimes 0}$ to the scalar 1 and the expectations are taken with respect to the true model of (i,j,k).

2.2 *Asymptotic distributions of estimators.* Let the mixed baseline hazards model (3) be the assumed marginal hazards model for (i,j,k). Under a working independence assumption, i.e., ignoring the dependence among failure times in a cluster, the pseudo partial likelihood ($PL(\boldsymbol{\beta})$) for the mixed baseline model is:

$$PL(\boldsymbol{\beta})=\prod_{k=1}^{K}\prod_{j=1}^{J}\prod_{i=1}^{n}\left[\frac{\exp\{\boldsymbol{\beta}'\boldsymbol{Z}_{ijk}(X_{ijk})\}}{\sum_{l,g\in R_j(X_{ijk})}\exp\{\boldsymbol{\beta}'\boldsymbol{Z}_{ljg}(X_{ijk})\}}\right]^{\delta_{ijk}},$$

where $R_j(t)=\{l,g : X_{ljg}\geq t\}$, that is, the set of clusters and failure types for which subject j is at risk just prior to time t. The estimator $\hat{\boldsymbol{\beta}}$ for $\boldsymbol{\beta}$ is defined as the solution to the pseudo score equations

$$\boldsymbol{U}(\boldsymbol{\beta})=\partial\log PL(\boldsymbol{\beta})/\partial\boldsymbol{\beta}=\boldsymbol{0}, \tag{4}$$

Formulated by means of counting processes, the pseudo partial likelihood scores

$$\boldsymbol{U}(\boldsymbol{\beta})=\sum_{k=1}^{K}\sum_{j=1}^{J}\sum_{i=1}^{n}\int_0^{\tau}\left[\boldsymbol{Z}_{ijk}(t)-\frac{\sum_{g=1}^{K}\sum_{l=1}^{n}Y_{ljg}(t)\boldsymbol{Z}_{ljg}(t)\exp\{\boldsymbol{\beta}'\boldsymbol{Z}_{ljg}(t)\}}{\sum_{g=1}^{K}\sum_{l=1}^{n}Y_{ljg}(t)\exp\{\boldsymbol{\beta}'\boldsymbol{Z}_{ljg}(t)\}}\right]dN_{ijk}(t).$$

When the assumed marginal hazards model (3) is correct, under some regularity conditions (Clegg et al. (1996, 1999)), $\hat{\boldsymbol{\beta}}$ is a consistent estimator for $\boldsymbol{\beta}_0$ and is asymptotical Gaussian with a "sandwich" type robust covariance matrix which can be consistently estimated. Here, we derive the asymptotic distribution of $\sqrt{n}\hat{\boldsymbol{\beta}}$ when the marginal model (3) may be misspecified.

THEOREM 1. *Let $\boldsymbol{\beta}^*$ be the unique solution to the system of p equations $\boldsymbol{q}(\boldsymbol{\beta})=\boldsymbol{0}$, where*

$$\boldsymbol{q}(\boldsymbol{\beta})=\sum_{j=1}^{J}\int_0^{\tau}\left\{\boldsymbol{s}_{j.}^{(1)}(t)-\boldsymbol{s}_{j.}^{(1)}(\boldsymbol{\beta},t)s_{j.}^{(0)}(t)/s_{j.}^{(0)}(\boldsymbol{\beta},t)\right\}dt. \tag{5}$$

Suppose that the following conditions hold:

Condition 1. For $j=1,\cdots,J$ and $k=1,\cdots,K$,

$$E\{\sup_{t\in[0,\tau]}Y_{1jk}(t)|Z_{1jk}(t)|^2\lambda_{ijk}(t)\}<\infty$$

and there exists a neighborhood $\mathcal{B}$ of $\boldsymbol{\beta}^*$ such that

$$E\{\sup_{t\in[0,\tau],\boldsymbol{\beta}\in\mathcal{B}} Y_{1jk}(t)|Z_{1jk}(t)|^2 \exp(\boldsymbol{\beta}' Z_{1jk}(t))\} < \infty,$$

$\lambda_{0j}(t) \geq 0$, *and* $\int_0^\tau \lambda_{0j}(t)dt < \infty$.

Condition 2. $\Pr\{Y_{1jk}(t) = 1 \;\; \forall t \in [0,\tau]\} > 0$.

Condition 3. The matrices $\boldsymbol{I}_j(\boldsymbol{\beta}^*) = \sum_{k=1}^K \int_0^\tau \boldsymbol{v}_j(\boldsymbol{\beta}^*, t) s_{jk}^{(0)}(t)dt = \int_0^\tau \boldsymbol{v}_j(\boldsymbol{\beta}^*, t) s_{j.}^{(0)}(t)dt$ *are positive definite,* $\forall j = 1, \cdots, J$, *where* $\boldsymbol{v}_j(\boldsymbol{\beta}, t) = \boldsymbol{s}_{j.}^{(2)}(\boldsymbol{\beta}, t)/s_{j.}^{(0)}(\boldsymbol{\beta}, t) - \left\{\boldsymbol{s}_{j.}^{(1)}(\boldsymbol{\beta}, t)/s_{j.}^{(0)}(\boldsymbol{\beta}, t)\right\}^{\otimes 2}$.

Then the estimator $\hat{\boldsymbol{\beta}}$ is a consistent estimator of $\boldsymbol{\beta}^*$.

The proof of Theorem 1 is given in the appendix. The conditions assumed in the theorem are analogous to those used by Andersen and Gill (1982) for univariate maximum partial likelihood estimation when the hazard model is correctly specified. The result in the theorem is a multivariate failure time generalization of Theorem 2.1 of Struthers and Kalbfleisch (1986) and (2.1) of Lin and Wei (1989). In the next section, we shall discuss some special cases of Theorem 1. Since $\boldsymbol{\beta}^* \neq \boldsymbol{\beta}_0$ in general, $\hat{\boldsymbol{\beta}}$ is not consistent for $\boldsymbol{\beta}_0$. As shown in the next section, however, $\hat{\boldsymbol{\beta}}$ or a subset of the p components of $\hat{\boldsymbol{\beta}}$ is consistent for $\boldsymbol{\beta}_0$ or for a subset of the elements of $\boldsymbol{\beta}_0$ in some special cases. Hence, it is useful to know the asymptotic distribution of $\hat{\boldsymbol{\beta}}$. The asymptotic distribution of $\hat{\boldsymbol{\beta}}$ under a possibly misspecified mixed baseline hazards model is given in the following theorem.

THEOREM 2. *Assume that Condition 1 - Condition 3 given in Theorem 1 are satisfied. Let*

$$F_{jk:n}(t) = n^{-1}\sum_{i=1}^n N_{ijk}(t), \quad F_{jk}(t) = E\{F_{jk:n}(t)\},$$

$$F_{j.:n}(t) = \sum_{k=1}^K F_{jk:n}(t) \text{ and } F_{j.}(t) = \sum_{k=1}^K F_{jk}(t).$$

Suppose that $\sqrt{n}\{F_{j.:n}(t) - F_{j.}(t)\}$ *and* $\sqrt{n}\left\{S_{j.}^{(0)}(\boldsymbol{\beta}^*, t) - s_{j.}^{(0)}(\boldsymbol{\beta}^*, t)\right\}$ *are Gaussian processes with mean* 0. *Then* $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) \xrightarrow{\mathrm{w}} N_p(\boldsymbol{0}, \boldsymbol{\Sigma}(\boldsymbol{\beta}^*))$ *as* $n \longrightarrow \infty$, *where*

$$\boldsymbol{\Sigma}(\boldsymbol{\beta}) = \boldsymbol{I}^{-1}(\boldsymbol{\beta})\boldsymbol{A}(\boldsymbol{\beta})\boldsymbol{I}^{-1}(\boldsymbol{\beta}), \quad \boldsymbol{I}(\boldsymbol{\beta}) = \sum_{j=1}^J \boldsymbol{I}_j(\boldsymbol{\beta}),$$

$$\boldsymbol{A}(\boldsymbol{\beta}) = E\{\sum_{j=1}^J\sum_{k=1}^K\sum_{f=1}^J\sum_{g=1}^K \boldsymbol{w}_{1jk}(\boldsymbol{\beta})\boldsymbol{w}'_{1fg}(\boldsymbol{\beta})\},$$

$$\begin{aligned}\boldsymbol{w}_{ijk}(\boldsymbol{\beta}) &= \int_0^\tau \left\{\boldsymbol{Z}_{ijk}(t) - \boldsymbol{s}_{j.}^{(1)}(\boldsymbol{\beta},t)/s_{j.}^{(0)}(\boldsymbol{\beta},t)\right\} \\ &\quad \times \left\{dN_{ijk}(t) - Y_{ijk}(t)\exp\{\boldsymbol{\beta}'\boldsymbol{Z}_{ijk}(t)\}/s_{j.}^{(0)}(\boldsymbol{\beta},t)dF_{j.}(t)\right\},\end{aligned}$$

and $\boldsymbol{I}_j(\boldsymbol{\beta})$ *is defined as in Condition 3 in Theorem 1.*

The proof of Theorem 2 is also shown in Appendix. Note that Theorem 2.1 of Lin and Wei (1989) is a special (univariate) case of Theorem 2 where $J = 1$ and $K = 1$. The covariance matrix of $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)$, $\boldsymbol{\Sigma}(\boldsymbol{\beta}^*)$, can be estimated from the data by

$$\hat{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\beta}}) = \hat{\boldsymbol{I}}^{-1}(\hat{\boldsymbol{\beta}})\hat{\boldsymbol{A}}(\hat{\boldsymbol{\beta}})\hat{\boldsymbol{I}}^{-1}(\hat{\boldsymbol{\beta}}), \tag{6}$$

where

$$\hat{\boldsymbol{I}}(\boldsymbol{\beta}) = n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{J}\sum_{k=1}^{K}\delta_{ijk}\boldsymbol{V}_j(\boldsymbol{\beta}, X_{ijk}),$$

$$\boldsymbol{V}_j(\boldsymbol{\beta},t) = \boldsymbol{S}_{j.}^{(2)}(\boldsymbol{\beta},t)/S_{j.}^{(0)}(\boldsymbol{\beta},t) - \left\{\boldsymbol{S}_{j.}^{(1)}(\boldsymbol{\beta},t)/S_{j.}^{(0)}(\boldsymbol{\beta},t)\right\}^{\otimes 2},$$

$$\hat{\boldsymbol{A}}(\boldsymbol{\beta}) = n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{f=1}^{J}\sum_{g=1}^{K}\boldsymbol{W}_{ijk}(\boldsymbol{\beta})\boldsymbol{W}'_{ifg}(\boldsymbol{\beta}),$$

and

$$\begin{aligned}\boldsymbol{W}_{ijk}(\boldsymbol{\beta}) &= \delta_{ijk}\left\{\boldsymbol{Z}_{ijk}(X_{ijk}) - \boldsymbol{S}_{j.}^{(1)}(\boldsymbol{\beta},X_{ijk})/S_{j.}^{(0)}(\boldsymbol{\beta},X_{ijk})\right\} \\ &\quad -\sum_{l=1}^{n}\sum_{g=1}^{K}\delta_{ljg}Y_{ijk}(X_{ljg})\exp\{\boldsymbol{\beta}'\boldsymbol{Z}_{ijk}(X_{ljg})\}\{nS_{j.}^{(0)}(\boldsymbol{\beta},X_{ljg})\}^{-1} \\ &\quad \times\left\{\boldsymbol{Z}_{ijk}(X_{ljg}) - \boldsymbol{S}_{j.}^{(1)}(\boldsymbol{\beta},X_{ljg})/S_{j.}^{(0)}(\boldsymbol{\beta},X_{ljg})\right\}.\end{aligned}$$

Note that $\hat{\boldsymbol{\Sigma}}$ is obtained from $\boldsymbol{\Sigma}$ by replacing $\boldsymbol{\beta}^*$, $\boldsymbol{s}_{j.}^{(1)}(\boldsymbol{\beta}^*,t)$, $s_{j.}^{(0)}(\boldsymbol{\beta}^*,t)$, and $F_{j.}(t)$ with $\hat{\boldsymbol{\beta}}$, $\boldsymbol{S}_{j.}^{(1)}(\hat{\boldsymbol{\beta}},t)$, $S_{j.}^{(0)}(\hat{\boldsymbol{\beta}},t)$, and $F_{j.:n}(t)$, respectively. It can be shown that $\hat{\boldsymbol{\Sigma}}$ is a consistent estimator for $\boldsymbol{\Sigma}$. The proof requires some tedious work to establish asymptotic equivalence of corresponding terms by applying empirical distribution theory. Note that $\boldsymbol{I}^{-1}(\boldsymbol{\beta})$ and $\boldsymbol{\Sigma}(\boldsymbol{\beta})$ are asymptotically equivalent if the assumed marginal hazards model is correct and if the failure times for each subject and for each failure in a cluster are indeed independent of each other.

Similar asymptotic distributions of $\hat{\boldsymbol{\beta}}$ hold under an assumed marginal distinct baseline hazards model or under an assumed marginal common baseline hazard model. Note that, under the assumed distinct baseline hazards model (1), $\hat{\boldsymbol{\beta}}$ is defined as the solution to the system of pseudo partial likelihood score equations $\sum_{k=1}^{K}\sum_{j=1}^{J}\sum_{i=1}^{n}\int_0^\tau\left\{\boldsymbol{Z}_{ijk}(t) - \boldsymbol{S}_{jk}^{(1)}(\boldsymbol{\beta},t)/S_{jk}^{(0)}(\boldsymbol{\beta},t)\right\}dN_{ijk}(t) = \boldsymbol{0}$, and $\boldsymbol{\beta}^*$ is the unique solution to the system of equations

$$\boldsymbol{q}(\boldsymbol{\beta}) = \sum_{j=1}^{J}\sum_{k=1}^{K}\int_0^\tau\left\{\boldsymbol{s}_{jk}^{(1)}(t) - \boldsymbol{s}_{jk}^{(1)}(\boldsymbol{\beta},t)s_{jk}^{(0)}(t)/s_{jk}^{(0)}(\boldsymbol{\beta},t)\right\}dt = \boldsymbol{0}. \tag{7}$$

Similarly, under an assumed common baseline hazard model (2), $\hat{\boldsymbol{\beta}}$ is defined as the solution to the pseudo partial likelihood score equations

$$\sum_{k=1}^{K}\sum_{j=1}^{J}\sum_{i=1}^{n}\int_0^{\tau}\left\{\boldsymbol{Z}_{ijk}(t)-\boldsymbol{S}_{..}^{(1)}(\boldsymbol{\beta},t)/S_{..}^{(0)}(\boldsymbol{\beta},t)\right\}dN_{ijk}(t)=\mathbf{0},$$

and $\boldsymbol{\beta}^*$ is the unique solution to the system of equations $\boldsymbol{q}(\boldsymbol{\beta})=\mathbf{0}$, where

$$\boldsymbol{q}(\boldsymbol{\beta})=\int_0^{\tau}\left\{\boldsymbol{s}_{..}^{(1)}(t)-\boldsymbol{s}_{..}^{(1)}(\boldsymbol{\beta},t)s_{..}^{(0)}(t)/s_{..}^{(0)}(\boldsymbol{\beta},t)\right\}dt.$$

3. Special Cases

The asymptotic results derived in Section 2 are very general in the sense that they apply to any kind of marginal hazards model misspecification. In this section we consider some special cases of model misspecification. When the assumed mixed baseline hazards model (3) is correct, i.e., $\lambda_{ijk}(t)=\lambda_{0j}(t)\exp\{\boldsymbol{\beta}_0'\boldsymbol{Z}_{ijk}(t)\}$, we then have $\boldsymbol{s}_{jk}^{(d)}(t)=\lambda_{0j}(t)\boldsymbol{s}_{jk}^{(d)}(\boldsymbol{\beta}_0,t)$ for $d=0,1$. Hence, $\boldsymbol{\beta}^*=\boldsymbol{\beta}_0$ is the solution to $\boldsymbol{q}(\boldsymbol{\beta})=\mathbf{0}$, a trivial consequence of Theorem 1.

Now, we examine the consequences of baseline type misspecification when the true marginal hazards model is in the Cox model family and the regression part of the hazards model, $\exp\{\boldsymbol{\beta}'\boldsymbol{Z}_{ijk}(t)\}$, is specified correctly. The results are summarized in Corollary 1.

Corollary 1. *If the true marginal hazards model is a common baseline hazard model, the estimator $\hat{\boldsymbol{\beta}}$ under either a mixed baseline hazards model or a distinct baseline hazards model is consistent for $\boldsymbol{\beta}_0$.*

Proof. If the common baseline hazard model is true but we fit as a mixed baseline hazards model, then $\boldsymbol{s}_{j.}^{(d)}(t)=\sum_{k=1}^{K}E\{\lambda_0(t)Y_{1jk}(t)\boldsymbol{Z}_{1jk}(t)^{\otimes d}\exp[\boldsymbol{\beta}_0'\boldsymbol{Z}_{1jk}(t)]\}=\lambda_0(t)\boldsymbol{s}_{j.}^{(d)}(\boldsymbol{\beta}_0,t)$, $d=0,1,2$. Hence, from (5), $\boldsymbol{q}(\boldsymbol{\beta})=\sum_{j=1}^{J}\int_0^{\tau}\left\{\boldsymbol{s}_{j.}^{(1)}(\boldsymbol{\beta}_0,t)-\boldsymbol{s}_{j.}^{(1)}(\boldsymbol{\beta},t)\;s_{j.}^{(0)}(\boldsymbol{\beta}_0,t)/s_{j.}^{(0)}(\boldsymbol{\beta},t)\right\}\lambda_0(t)dt$. Obviously, $\boldsymbol{\beta}^*=\boldsymbol{\beta}_0$ is the solution to the system of equations. Therefore, by Theorem 1, the estimator $\hat{\boldsymbol{\beta}}$ under the assumed mixed baseline hazards model is consistent for $\boldsymbol{\beta}_0$ when the true marginal hazards model is actually a common baseline hazard model. By the same argument, it is easy to show that the estimator $\hat{\boldsymbol{\beta}}$ from the assumed distinct baseline hazards model is consistent for $\boldsymbol{\beta}_0$ when the true marginal hazards model is, in fact, a common baseline hazard model. □

The results in Corollary 1 guarantee the consistency of the estimator $\hat{\boldsymbol{\beta}}$ obtained from mixed baseline hazards model or distinct baseline hazards model when the true model is a common baseline hazard model. However, if the true marginal hazards model is a distinct baseline hazards model and $\boldsymbol{\beta}_0\neq\mathbf{0}$, it can be easily shown that the estimator $\hat{\boldsymbol{\beta}}$ from either a mixed baseline hazards model or a common baseline hazard model is not consistent for $\boldsymbol{\beta}_0$. Similarly, the estimator $\hat{\boldsymbol{\beta}}$ under a common baseline hazard model is not consistent for $\boldsymbol{\beta}_0 (\boldsymbol{\beta}_0 \neq \mathbf{0})$ in a mixed baseline hazards model.

When $\boldsymbol{\beta}^*$ is not equal to $\boldsymbol{\beta}_0$, the estimator $\hat{\boldsymbol{\beta}}$ will not be consistent for $\boldsymbol{\beta}_0$. In some special cases, however, part of components of $\hat{\boldsymbol{\beta}}$ could be consistent for the corresponding part of $\boldsymbol{\beta}_0$. Without loss of generality, we focus our attention here on $\beta_{0,1}$, the first component of the true parameter vector, $\boldsymbol{\beta}_0$.

COROLLARY 2. *Let $\beta_{0,1}$, $Z_{ijk,1}(t)$, and $\hat{\beta}_1$ be the first component of the parameter vector $\boldsymbol{\beta}_0$, the covariate vector $\boldsymbol{Z}_{ijk}(t)$, and the estimator vector $\hat{\boldsymbol{\beta}}$, respectively. Suppose that $\beta_{0,1} = 0$ and that $Z_{ijk,1}(t)$ is independent of censoring and the other $(p-1)$ components of the covariate vector $Z_{ijk,2}(t), \cdots, Z_{ijk,p}(t)$, some of which may even be mistakenly omitted from the model. Then $\hat{\beta}_1$ under (1) is consistent, that is, $\hat{\beta}_1 \longrightarrow 0$ in probability as $n \longrightarrow \infty$. In addition, if $Z_{ijk,1}(t)$ is identically distributed across k (or across j and k) and $\hat{\boldsymbol{\beta}}$ is estimated under (3) (or under (2)), then $\hat{\beta}_1$ is consistent.*

Note that, in the corollary, the true underlying marginal hazards model $\lambda_{ijk}(t)$ may not even be in the Cox model family.

PROOF. Let $s^{(1)}_{jk,1}(t)$ and $s^{(1)}_{jk,1}(\boldsymbol{\beta}^*, t)$ be the first components of $\boldsymbol{s}^{(1)}_{jk}(t)$ and $\boldsymbol{s}^{(1)}_{jk}(\boldsymbol{\beta}^*, t)$, respectively. Let $\hat{\boldsymbol{\beta}}$ be the estimator obtained from (1). To show that $\hat{\beta}_1 \longrightarrow 0$ in probability when $\beta_{0,1} = 0$, it suffices to show that $\beta^*_1 = 0$ when $\beta_{0,1} = 0$ in view of (7). In other words, it is sufficient to verify that when $\beta_{0,1} = 0$, $\beta^*_1 = 0$ satisfies the following condition regardless of the values of the other $p-1$ components of $\boldsymbol{\beta}^*$:

$$\sum_{j=1}^{J}\sum_{k=1}^{K}\int_0^\tau \left\{ s^{(1)}_{jk,1}(t) - s^{(1)}_{jk,1}(\boldsymbol{\beta}^*, t) s^{(0)}_{jk}(t) / s^{(0)}_{jk}(\boldsymbol{\beta}^*, t) \right\} dt = 0. \tag{8}$$

Notice that when $\beta_{0,1} = 0$, $s^{(1)}_{jk,1}(t) = s^{(0)}_{jk}(t) E\{Z_{1jk,1}(t)\}$. Also, when $\beta^*_1 = 0$, $s^{(1)}_{jk,1}(\boldsymbol{\beta}^*, t) = s^{(0)}_{jk}(\boldsymbol{\beta}^*, t) E\{Z_{1jk,1}(t)\}$, since $Z_{ijk,1}(t)$ is independent of censoring and other $(p-1)$ components of $\boldsymbol{Z}_{ijk}(t)$. Therefore, (8) is satisfied. Hence, the estimator $\hat{\beta}_1$ obtained from a distinct baseline hazards model is consistent.

In addition, if $Z_{ijk,1}(t)$ is identically distributed for $k = 1, \cdots, K$, we have

$$s^{(1)}_{jk,1}(t) = s^{(0)}_{jk}(t) E\{Z_{1j1,1}(t)\} \quad \text{and} \quad s^{(1)}_{jk,1}(\boldsymbol{\beta}^*, t) = s^{(0)}_{jk}(\boldsymbol{\beta}^*, t) E\{Z_{1j1,1}(t)\},$$

so that $\sum_{j=1}^{J} \int_0^\tau \left\{ s^{(1)}_{j\cdot,1}(t) - s^{(1)}_{j\cdot,1}(\boldsymbol{\beta}^*, t) s^{(0)}_{j\cdot}(t) / s^{(0)}_{j\cdot}(\boldsymbol{\beta}^*, t) \right\} dt = 0$ is satisfied, where $s^{(1)}_{j\cdot,1}(t)$ and $s^{(1)}_{j\cdot,1}(\boldsymbol{\beta}^*, t)$ are the first component of $\boldsymbol{s}^{(1)}_{j\cdot}(t)$ and $\boldsymbol{s}^{(1)}_{j\cdot}(\boldsymbol{\beta}^*, t)$, respectively. It follows from Theorem 1 that $\hat{\beta}_1$ obtained under a mixed baseline hazards model is consistent for $\beta_{0,1}$. Using the same argument, we can easily show that $\hat{\beta}_1$ from a common baseline hazard model is consistent for $\beta_{0,1}$, if, additionally, $Z_{ijk,1}(t)$ is identically distributed for $j = 1, \cdots, J$ and $k = 1, \cdots, K$. □

What is the practical implication of Corollary 2? Suppose that $Z_{ijk,1}(t)$ is the indicator of the treatment assigned in a randomized trial. The random assignment of the treatment should guarantee that $Z_{ijk,1}(t)$ is independent of censoring and the other baseline covariates of interest. The results in Corollary 2 imply that, if the treatment has no effect on failure times, the treatment effect will be consistently estimated even if the marginal hazards model is misspecified.

4. Simulations

Simulation studies were conducted to obtain information on the consequences of misspecification of marginal hazards models in finite sample situations. Tables 1 and 2 present results when the true marginal hazards model is the mixed baseline hazards model (3). In these tables, the failure times $T_{i11}, T_{i12}, T_{i21}$, and T_{i22} for the ith cluster are generated from the multivariate Clayton-Oakes distribution (Clayton and Cuzick (1985)and Oakes (1989)) with a marginal exponential distribution for two types of failure and two subjects in a cluster:

$$pr(T_{i11} > t_{i11}, T_{i12} > t_{i12}, T_{i21} > t_{i21}, T_{i22} > t_{i22} | z_{i11}, z_{i12}, z_{i21}, z_{i22})$$
$$= \left[\sum_{j=1}^{2} \sum_{k=1}^{2} \exp\{t_{ijk} \lambda_{0j} \theta^{-1} \exp(\beta_0 z_{ijk})\} - 3 \right]^{-\theta}.$$

Table 1. Simulation results (based on 500 simulation runs) for the true mixed baseline hazards model with N(0,1) covariate and uniform(0, 1) censoring distribution (42% censoring).

β_0	θ	n	mean $\hat{\beta}_m$	mean $\hat{\beta}_d$	mean $\hat{\beta}_c$	mean $\hat{V}_m$	mean $\hat{V}_d$	mean $\hat{V}_c$
0	.25	100	.001	.000	-.001	.0043	.0044	.0043
		200	-.000	-.000	-.001	.0021	.0022	.0021
	.80	100	.000	-.000	-.002	.0043	.0044	.0043
		200	-.000	-.000	-.001	.0021	.0022	.0021
	1.50	100	.000	.001	-.002	.0044	.0044	.0044
		200	.000	.000	-.001	.0021	.0021	.0021
	3.00	100	-.000	-.000	-.003	.0044	.0044	.0043
		200	-.000	-.000	-.001	.0021	.0021	.0021
.7	.25	100	.706	.703	.541	.0068	.0068	.0051
		200	.699	.698	.538	.0034	.0034	.0025
	.80	100	.705	.704	.542	.0060	.0060	.0050
		200	.700	.699	.539	.0030	.0030	.0025
	1.50	100	.703	.703	.541	.0057	.0057	.0050
		200	.700	.699	.539	.0028	.0028	.0025
	3.00	100	.703	.703	.541	.0055	.0056	.0050
		200	.699	.699	.539	.0027	.0028	.0025

Table 2. Simulation results (based on 500 simulation runs) for the true mixed baseline hazards model with $\beta_0 = 0.7$, Bernoulli(0.5) or N(0,1) (truncated at ± 5) covariate, uniform(0, 5) censoring distribution (9% censoring for Bernoulli(0.5) covariate and 14% censoring for N(0,1)).

Covariate Z	θ	n	mean $\hat{\beta}_m$	mean $\hat{\beta}_d$	mean $\hat{\beta}_c$	mean $\hat{V}_m$	mean $\hat{V}_d$	mean $\hat{V}_c$
Bern(0.5)	.25	100	.704	.697	.515	.0139	.0142	.0113
		200	.701	.697	.513	.0071	.0072	.0057
	.80	100	.700	.695	.512	.0129	.0132	.0111
		200	.699	.696	.511	.0066	.0067	.0056
	1.50	100	.698	.696	.511	.0124	.0126	.0111
		200	.698	.696	.511	.0063	.0064	.0056
	3.00	100	.698	.697	.511	.0121	.0122	.0111
		200	.697	.696	.510	.0061	.0061	.0056
N(0,1)	.25	100	.708	.704	.521	.0056	.0056	.0036
		200	.701	.699	.518	.0029	.0029	.0018
	.80	100	.706	.704	.521	.0048	.0048	.0034
		200	.701	.699	.519	.0024	.0024	.0017
	1.50	100	.704	.703	.520	.0044	.0044	.0034
		200	.700	.699	.518	.0022	.0022	.0017
	3.00	100	.704	.703	.520	.0041	.0041	.0034
		200	.700	.700	.518	.0020	.0021	.0017

Here, we assume a different baseline hazard function, λ_{0j}, for each subject in a cluster and an identical baseline for the two failure times from the same subject. This would be the case, for example, in a vision loss study involving husbands and wives where we treat vision loss from each eye as one type of failure. It should be reasonable to assume a mixed baseline hazards model with different baseline hazards for the husband and wife from the same family to account for their different susceptibilities to vision loss and an identical baseline hazard for the left and right eye vision loss, because there are no biological differences to support that one eye is superior or inferior to the other eye. In the simulation study, $\lambda_{01} = 1$ and $\lambda_{02} = 5$ were used for subject 1 and subject 2 in each cluster, respectively. The true regression coefficient β_0 was 0 or 0.7, which corresponds to a relative risk of 1 or approximately 2. Distributions for the scalar covariate Z_{ijk} were taken to be independent Bernoulli(0.5) or independent standard Gaussian distribution N(0, 1), truncated at ± 5. The parameter θ represents the degree of pairwise dependence of failure times in a cluster. When $\beta_0 = 0$, $\theta \longrightarrow 0$ gives the maximal positive correlation of 1. Independence is the limiting case of $\theta \longrightarrow \infty$. In simulations the values for the dependence parameter θ were 0.25, 0.80, 1.50, and 3.00, which resulted in pairwise correlation coefficients between failure times being 0.94, 0.71, 0.51, and 0.30 for $\beta = 0$, and 0.42, 0.32, 0.23, and 0.14 for $\beta = 0.70$ with N(0,1) covariate truncated at ± 5, respectively, when there is no censoring. Censoring times C_{ijk} were from either uniform (0,1) or uniform (0, 5) distributions and they were generated independent of each other and of T_{ijk} and Z_{ijk}. Five hundred simulation, with the number of clusters (n) being 100 and 200, were carried out for each configuration. A Newton-Raphson iterative procedure was used in each run to obtain the estimates. In the simulations, the only part of model misspecification was the type of baseline hazards; the $\exp(\beta Z)$ part was correct.

The simulation results are summarized within each configuration by mean estimates (mean $\hat{\beta}_m$, mean $\hat{\beta}_d$, and mean $\hat{\beta}_c$) and mean variance estimates (mean $\hat{V}_m$, mean $\hat{V}_d$, and mean $\hat{V}_c$). The subscripts m, d, and c indicate that the estimates are obtained under a marginal mixed (m) baseline hazards model (i.e., the correct model in this simulation study), under a distinct (d) baseline hazards model (1), and under a common (c) baseline hazard model (2), respectively. The robust variance estimator in (6) is used in calculating $\hat{V}_m$. Similar robust variance estimators for the distinct baseline hazards model and for the common baseline hazard model are used for $\hat{V}_d$ and $\hat{V}_c$.

When $\beta_0 = 0.7$, these simulation results suggest that the estimate $\hat{\beta}_d$ under the distinct baseline hazards model is approximately unbiased. In contrast, the estimate $\hat{\beta}_c$ obtained from the common baseline hazard model is severely biased towards zero for the given simulation configurations if the underlying marginal hazards model is the mixed baseline hazards model. All estimates, including $\hat{\beta}_c$, are essentially unbiased for each simulation configuration when $\beta_0 = 0$. This is expected as a result of Corollary 3 since the covariate Z is identically distributed for j and k in the simulations.

The variance estimate under the assumed distinct baseline hazards model agrees with that obtained from the true mixed baseline hazards model, as judged by the values of mean $\hat{V}_d$ and mean $\hat{V}_m$, although mean $\hat{V}_d$ tends to be consistently little larger than mean $\hat{V}_m$. This result agrees with the conclusion for univariate failure time data that the loss of efficiency in estimating $\boldsymbol{\beta}$ is generally not severe when the stratification is used unnecessarily [c.f., Page 88 of Kalbfleisch and Prentice (1980)]. This might be explained by the small number of strata. The loss of efficiency could be severe when the number of strata is large. Because $\hat{\beta}_c$ is biased towards zero, the variance estimate, $\hat{V}_c$, calculated under the common baseline hazard model underestimates the variance V_m except when $\beta_0 = 0$, in which case $\hat{\beta}_c$ is consistent for β_0. The degree of underestimation of variance, however, lessens as a result of either decreasing the dependence or increasing the censoring percentage.

The Wald-type empirical coverage probabilities (data not shown) of the mixed baseline hazards model and the distinct baseline hazards model are close to the nominal levels of 0.90 and 0.95, indicating that the proposed Gaussian approximation for the estimator distribution is quite good. Because $\hat{\beta}_c$ is biased, the coverage obtained from the common baseline hazard model range from 0.060 to 0.280 for the nominal level of 0.90 and from 0.114 to 0.380 for the nominal level of 0.95 when $\beta_0 = 0.7$ for the simulation configurations in Table 1; as expected, the coverage is near their nominal levels of 0.90 (range from 0.894 to 0.906) and 0.95 (from 0.940 to 0.954) when $\beta_0 = 0$.

In brief, these simulation results indicate that the asymptotic results derived in Section 3 are quite adequate for practical sample sizes.

5. Concluding Remarks

Consequences of misspecification of the Cox regression model for univariate (i.e., independently distributed) failure time data on the regression parameter $\boldsymbol{\beta}$ have been investigated by several authors, including Gail, Wieand and Piantadosi (1984), Lagakos and Schoenfeld (1984), Solomon (1984), Morgan (1986), Struthers and Kalbfleisch (1986), Lagakos (1988), Lin and Wei (1989), and Anderson and Fleming (1995). We studied the consequences of hazards model misspecification in a marginal regression modeling framework for multivariate failure time data, where virtually no research has been done.

Our results indicate that misspecification of the marginal hazards model in a marginal regression framework will lead to inconsistent estimators in general. This conclusion is not unique to correlated failure time data. In a marginal regression modeling framework for longitudinal data, such as the generalized estimating equation (GEE) approach (Liang and Zeger (1986)), the consistency of the regression parameter estimator also depends on the correct specification of the underlying marginal model.

In classical multiple linear regression, it is already known that over-fitting a model (i.e., including terms with truly zero regression coefficients) will not bias the other estimated regression coefficients in the model, but it will lead to larger standard errors for the estimated regression coefficients. However, under-fitting a model (i.e, leaving out terms with truly non-zero regression coefficients) may lead to biases in estimated regression coefficients. Corollary 1 confirms this in the marginal hazards model setting. More specifically, if the baseline hazards are heterogeneous and we assume a common baseline hazard (under-fitting), then we get bias; on the other hand, if we allow for heterogeneous baseline hazards but homogeneity is the true state of nature (over-fitting), then the estimates are consistent, though we may lose some precision. So, when in doubt, it is probably better to overfit to avoid bias.

Appendix

Theoretical Details. We provide here a sketch of the proof that the estimator $\hat{\boldsymbol{\beta}}$ from solving (4) is consistent for $\boldsymbol{\beta}^*$ and that $\sqrt{n}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}^*)$ is asymptotically Gaussian.

Consider first the consistency of the estimator $\hat{\boldsymbol{\beta}}$ for $\boldsymbol{\beta}^*$ (Theorem 1). Following the arguments of Anderson and Gill (1982, p.1111) we can show that Condition 1 and Condition 2 imply for $j = 1, \cdots, J$, $k = 1, \cdots, K$, and $d = 0, 1, 2$, that

$$\sup_{t \in [0,\tau]} \|\boldsymbol{S}_{jk}^{(d)}(t) - \boldsymbol{s}_{jk}^{(d)}(t)\| \xrightarrow{\text{p}} 0$$

and that there exists a neighborhood $\mathcal{B}$ of $\boldsymbol{\beta}^*$ such that

$$\sup_{t \in [0,\tau], \boldsymbol{\beta} \in \mathcal{B}} \|\boldsymbol{S}_{jk}^{(d)}(\boldsymbol{\beta}, t) - \boldsymbol{s}_{jk}^{(d)}(\boldsymbol{\beta}, t)\| \xrightarrow{\text{p}} 0 \tag{9}$$

as $n \to \infty$; $\boldsymbol{s}_{jk}^{(d)}(t)$ is bounded on $(0, \tau)$ and $\boldsymbol{s}_{jk}^{(d)}(\boldsymbol{\beta}, t)$ is bounded on $\mathcal{B} \times (0, \tau)$; and $s_{jk}^{(0)}(t)$ and $s_{jk}^{(0)}(\boldsymbol{\beta}, t)$ are bounded away uniformly from zero on $(0, \tau)$ and $\mathcal{B} \times (0, \tau)$, respectively. Let $l(\boldsymbol{\beta}, t) = \log PL(\boldsymbol{\beta}, t)$. Define the process

$$\begin{aligned}
\tilde{\boldsymbol{G}}(\boldsymbol{\beta}, t) &= \frac{1}{n}\left(l(\boldsymbol{\beta}, t) - l(\boldsymbol{\beta}^*, t)\right) \\
&= \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{J} \sum_{k=1}^{K} \int_0^t [(\boldsymbol{\beta} - \boldsymbol{\beta}^*)' \boldsymbol{Z}_{ijk}(u) \\
&\quad - \log\left\{S_{j.}^{(0)}(\boldsymbol{\beta}, u) / S_{j.}^{(0)}(\boldsymbol{\beta}^*, u)\right\}\Big] \, dN_{ijk}(u) \\
&= \sum_{j=1}^{J} \sum_{k=1}^{K} \tilde{\boldsymbol{G}}_{jk}(\boldsymbol{\beta}, t)
\end{aligned}$$

and

$$\tilde{\boldsymbol{G}}_{jk}(\boldsymbol{\beta}, t) = \frac{1}{n} \sum_{i=1}^{n} \int_0^t \left[(\boldsymbol{\beta} - \boldsymbol{\beta}^*)' \boldsymbol{Z}_{ijk}(u) - \log\left\{S_{j.}^{(0)}(\boldsymbol{\beta}, u) / S_{j.}^{(0)}(\boldsymbol{\beta}^*, u)\right\}\right] dN_{ijk}(u).$$

Applying Conditions 1 and 2 and the Lenglart's inequality (Fleming and Harrington (1991), Lemma 8. 2. 1) to $\tilde{\boldsymbol{G}}_{jk}(\boldsymbol{\beta}, \tau) - \boldsymbol{G}_{jk}(\boldsymbol{\beta}, \tau)$, one can easily show that $\tilde{\boldsymbol{G}}_{jk}(\boldsymbol{\beta}, \tau)$ is asymptotically equivalent to $\boldsymbol{G}_{jk}(\boldsymbol{\beta}, \tau) = n^{-1} \sum_{i=1}^{n} \int_0^\tau [(\boldsymbol{\beta} - \boldsymbol{\beta}^*)' \boldsymbol{Z}_{ijk}(u)$ $- \log\{s_{j.}^{(0)}(\boldsymbol{\beta}, u) / s_{j.}^{(0)}(\boldsymbol{\beta}^*, u)\}\Big] \, dN_{ijk}(u)$. Hence, $\tilde{\boldsymbol{G}}(\boldsymbol{\beta}, \tau)$ is asymptotically equivalent to $\boldsymbol{G}(\boldsymbol{\beta}, \tau) = \sum_{j=1}^{J} \sum_{k=1}^{K} \boldsymbol{G}_{jk}(\boldsymbol{\beta}, \tau)$. The compensator of $\boldsymbol{G}_{jk}(\boldsymbol{\beta}, t)$ is $\boldsymbol{A}_{jk}(\boldsymbol{\beta}, t)$ $= \int_0^t \left[(\boldsymbol{\beta} - \boldsymbol{\beta}^*)' \boldsymbol{S}_{jk}^{(1)}(u) - S_{jk}^{(0)}(u) \ \log\{s_{j.}^{(0)}(\boldsymbol{\beta}, u) \ / \ s_{j.}^{(0)}(\boldsymbol{\beta}^*, u)\}\right] du$ and $\boldsymbol{G}_{jk}(\boldsymbol{\beta}, t) -$ $\boldsymbol{A}_{jk}(\boldsymbol{\beta}, t)$ is a local square integrable martingale with respect to the filtration $\mathcal{F}_{jk}(t)$ for each j and k. Following the arguments of Anderson and Gill (1982, p.1105-6), one can show that $\boldsymbol{G}_{jk}(\boldsymbol{\beta}, \tau)$ and $\boldsymbol{A}_{jk}(\boldsymbol{\beta}, \tau)$ have the same limit, and hence $\boldsymbol{G}(\boldsymbol{\beta}, \tau)$ and $\boldsymbol{A}(\boldsymbol{\beta}, \tau)$ have the same limit $\boldsymbol{g}(\boldsymbol{\beta}) = \sum_{j=1}^{J} \sum_{k=1}^{K} \int_0^\tau [(\boldsymbol{\beta} - \boldsymbol{\beta}^*)'$ $\boldsymbol{s}_{jk}^{(1)}(u) - s_{jk}^{(0)}(u) \log\{s_{j.}^{(0)}(\boldsymbol{\beta}, u) / s_{j.}^{(0)}(\boldsymbol{\beta}^*, u)\}\Big] \, du$. Under the assumed conditions and the definition of $\boldsymbol{\beta}^*$, it is straightforward to show that the first derivative vector of

$\boldsymbol{g}(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$ is zero at $\boldsymbol{\beta}^*$ and the second derivative matrix of $\boldsymbol{g}(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$ is negative definite at $\boldsymbol{\beta}^*$. Consequently, the random concave function $\tilde{\boldsymbol{G}}(\boldsymbol{\beta})$ converges to $\boldsymbol{g}(\boldsymbol{\beta})$ in probability uniformly over B with a unique maximum at $\boldsymbol{\beta}^*$ by Theorem II. 1 of Andersen and Gill (1982). It follows that $\hat{\boldsymbol{\beta}}$ is a consistent estimator of $\boldsymbol{\beta}^*$.

Now consider the asymptotic normality of $\hat{\boldsymbol{\beta}}$ (Theorem 2). Define

$$\boldsymbol{U}_{ijk}(\boldsymbol{\beta}) = \int_0^\tau \left\{ \boldsymbol{Z}_{ijk}(u) - \boldsymbol{S}_{j.}^{(1)}(\boldsymbol{\beta}, u)/S_{j.}^{(0)}(\boldsymbol{\beta}, u) \right\} dN_{ijk}(u)$$

and

$$\boldsymbol{U}_j(\boldsymbol{\beta}) = \sum_{i=1}^n \sum_{k=1}^K \boldsymbol{U}_{ijk}(\boldsymbol{\beta}).$$

From (4) $\boldsymbol{U}(\boldsymbol{\beta}) = \sum_{i=1}^n \sum_{j=1}^J \sum_{k=1}^K \boldsymbol{U}_{ijk}(\boldsymbol{\beta}) = \sum_{j=1}^J \boldsymbol{U}_j(\boldsymbol{\beta})$. Define $\boldsymbol{w}_j(\boldsymbol{\beta}) = \sum_{i=1}^n \sum_{k=1}^K \boldsymbol{w}_{ijk}(\boldsymbol{\beta})$ and $\boldsymbol{w}(\boldsymbol{\beta}) = \sum_{j=1}^J \boldsymbol{w}_j(\boldsymbol{\beta})$, where $\boldsymbol{w}_{ijk}(\boldsymbol{\beta})$ is defined as in Theorem 2. Then

$$\begin{aligned} & -n^{-1/2} \left\{ \boldsymbol{U}_j(\boldsymbol{\beta}^*) - \boldsymbol{w}_j(\boldsymbol{\beta}^*) \right\} \\ = \quad & \sqrt{n} \int_0^\tau \left\{ \frac{\boldsymbol{S}_{j.}^{(1)}(\boldsymbol{\beta}^*, u)}{S_{j.}^{(0)}(\boldsymbol{\beta}^*, u)} - \frac{\boldsymbol{s}_{j.}^{(1)}(\boldsymbol{\beta}^*, u)}{s_{j.}^{(0)}(\boldsymbol{\beta}^*, u)} \right\} d\{F_{j.:n}(u) - F_{j.}(u)\} \qquad (10) \\ - \sqrt{n} & \int_0^\tau \left\{ \frac{\boldsymbol{S}_{j.}^{(1)}(\boldsymbol{\beta}^*, u)}{S_{j.}^{(0)}(\boldsymbol{\beta}^*, u)} - \frac{\boldsymbol{s}_{j.}^{(1)}(\boldsymbol{\beta}^*, u)}{s_{j.}^{(0)}(\boldsymbol{\beta}^*, u)} - \frac{\boldsymbol{S}_{j.}^{(1)}(\boldsymbol{\beta}^*, u)}{s_{j.}^{(0)}(\boldsymbol{\beta}^*, u)} + \frac{\boldsymbol{s}_{j.}^{(1)}(\boldsymbol{\beta}^*, u)}{s_{j.}^{(0)}(\boldsymbol{\beta}^*, u)} \frac{S_{j.}^{(0)}(\boldsymbol{\beta}^*, u)}{s_{j.}^{(0)}(\boldsymbol{\beta}^*, u)} \right\} dF_{j.}(u). \qquad (11) \end{aligned}$$

Now, $\sqrt{n}\{F_{j.:n}(t) - F_{j.}(t)\}$ converges weakly to a zero-mean Gaussian process by assumption. As the consequence of the tightness of the Gaussian process $\sqrt{n}\{F_{j.:n}(t) - F_{j.}(t)\}$, the convergence of $\boldsymbol{S}_{jk}^{(d)}(\boldsymbol{\beta}, t)$ established in (9), along with the strict positivity of $s_{jk}^{(0)}(\boldsymbol{\beta}, t)$ and the boundedness of $\boldsymbol{s}_{jk}^{(d)}(\boldsymbol{\beta}, t)$, (10) vanishes as $n \longrightarrow \infty$. We also have that

$$\begin{aligned} \|(11)\| \quad \leq \quad & \sup_{u \in [0,\tau]} \left\{ \left\| \frac{\boldsymbol{S}_{j.}^{(1)}(\boldsymbol{\beta}^*, u)}{S_{j.}^{(0)}(\boldsymbol{\beta}^*, u)} - \frac{\boldsymbol{s}_{j.}^{(1)}(\boldsymbol{\beta}^*, u)}{s_{j.}^{(0)}(\boldsymbol{\beta}^*, u)} \right| s_{j.}^{(0)}(\boldsymbol{\beta}^*, u)^{-1} \right\} \\ & \int_0^\tau \sqrt{n} \left| S_{j.}^{(0)}(\boldsymbol{\beta}^*, u) - s_{j.}^{(0)}(\boldsymbol{\beta}^*, u) \right| dF_{j.}(u) \\ = \quad & o_p(1), \end{aligned}$$

because

$$\sup_{u \in [0,\tau]} \left\{ \left\| \frac{\boldsymbol{S}_{j.}^{(1)}(\boldsymbol{\beta}^*, u)}{S_{j.}^{(0)}(\boldsymbol{\beta}^*, u)} - \frac{\boldsymbol{s}_{j.}^{(1)}(\boldsymbol{\beta}^*, u)}{s_{j.}^{(0)}(\boldsymbol{\beta}^*, u)} \right\| s_{j.}^{(0)}(\boldsymbol{\beta}^*, u)^{-1} \right\} = o_p(1)$$

in view of (9) along with the strict positivity of $s_{jk}^{(0)}(\boldsymbol{\beta}, t)$ and the boundedness of $s_{j.}^{(0)}(\boldsymbol{\beta}, t)$, plus $\sqrt{n}\left|S_{j.}^{(0)}(\boldsymbol{\beta}^*, u) - s_{j.}^{(0)}(\boldsymbol{\beta}^*, u)\right| = O_p(1)$ from the assumption of the Gaussian nature of $\sqrt{n}\{S_{j.}^{(0)}(\boldsymbol{\beta}^*, u) - s_{j.}^{(0)}(\boldsymbol{\beta}^*, u)\}$. Hence, $n^{-1/2}\boldsymbol{U}(\boldsymbol{\beta}^*)$ is asymptotically equivalent to $n^{-1/2}\boldsymbol{w}(\boldsymbol{\beta}^*)$. Notice that $n^{-1/2}\boldsymbol{w}(\boldsymbol{\beta}^*)$ is a sum of n independently and identically distributed $p-$component random vectors with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{A}(\boldsymbol{\beta}^*) = E\{\sum_{j=1}^J \sum_{k=1}^K \sum_{f=1}^J \sum_{g=1}^K \boldsymbol{w}_{1jk}(\boldsymbol{\beta}^*)\boldsymbol{w}'_{1fg}(\boldsymbol{\beta}^*)\}$. It follows from the multivariate central limit theorem [c.f., Page 123 of Sen and Singer (1993) that $n^{-1/2}\boldsymbol{U}(\boldsymbol{\beta}^*) \xrightarrow{\text{w}} N_p\left(\mathbf{0}, \boldsymbol{A}(\boldsymbol{\beta}^*)\right)$.

The first order Taylor expansion for $\boldsymbol{U}(\boldsymbol{\beta})$ centered at $\boldsymbol{\beta}^*$ of $\boldsymbol{\beta}$ gives

$$n^{-1/2}\boldsymbol{U}(\boldsymbol{\beta}^*) = \sum_{j=1}^J \sum_{k=1}^K \left[n^{-1}\left\{-\partial \boldsymbol{U}_{jk}(\boldsymbol{\beta})/\partial \boldsymbol{\beta}\right\}\big|_{\boldsymbol{\beta}=\tilde{\boldsymbol{\beta}}}\right] \sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)$$

where $\tilde{\boldsymbol{\beta}}$ is on a line segment between $\hat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}^*$ and $\boldsymbol{U}_{jk}(\boldsymbol{\beta}) = \sum_{i=1}^n \boldsymbol{U}_{ijk}(\boldsymbol{\beta})$. Denote

$$\boldsymbol{I}_{jk}(\boldsymbol{\beta}^*) = \int_0^\tau \boldsymbol{v}_j(\boldsymbol{\beta}^*, t) s_{jk}^{(0)}(t) dt,$$

then $\boldsymbol{I}(\boldsymbol{\beta}^*) = \sum_{j=1}^J \sum_{k=1}^K \boldsymbol{I}_{jk}(\boldsymbol{\beta}^*)$. Following along the lines of the last part of the proof of Theorem 3.2 in Andersen and Gill (1982), one can show easily that

$$n^{-1}\left\{-\partial \boldsymbol{U}_{jk}(\boldsymbol{\beta})/\partial \boldsymbol{\beta}\right\}\big|_{\boldsymbol{\beta}=\tilde{\boldsymbol{\beta}}} \xrightarrow{\text{p}} \boldsymbol{I}_{jk}(\boldsymbol{\beta}^*),$$

for any random $\tilde{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}(n)$ such that $\tilde{\boldsymbol{\beta}} \xrightarrow{\text{p}} \boldsymbol{\beta}^*$ as $n \longrightarrow \infty$. Therefore,

$$n^{-1}\left\{-\partial \boldsymbol{U}(\boldsymbol{\beta})/\partial \boldsymbol{\beta}\right\}\big|_{\boldsymbol{\beta}=\tilde{\boldsymbol{\beta}}} \xrightarrow{\text{p}} \boldsymbol{I}(\boldsymbol{\beta}^*).$$

Hence, $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*) \xrightarrow{\text{w}} N_p\left(\mathbf{0}, \boldsymbol{\Sigma}(\boldsymbol{\beta}^*)\right)$.

Acknowledgment. The authors are grateful to Dr. B. K. Sinha for helpful comments.

References

ANDERSEN, P. K. AND GILL, R. D. (1982). Cox's regression model for counting processes: A large sample study. *Ann. Statist.* **10:** 1100-1120.

ANDERSEN, P. K., BORGAN, Ø., GILL, R. D. AND KEIDING, N. (1993). *Statistical Models Based on Counting Processes*. New York: Springer.

ANDERSON, G. L. AND FLEMING, T. R. (1995). Model misspecification in proportional hazards regression. *Biometrika* **82:** 527-541.

BANDEEN-ROCHE, K. J. AND LIANG, K. Y. (1996). Modeling failure-time associations in data with multiple levels of clustering. *Biometrika* **83:** 29-39.

CAI, J. AND PRENTICE, R. L. (1995). Estimating equations for hazard ratio parameters based on correlated failure time data. *Biometrika* **82:** 151-164.

–––– (1997). Regression estimation using multivariate failure time data and a common baseline hazard function model. *Lifetime Data Analysis* **3:** 197-213.

Clayton, D. and Cuzick J. (1985). Multivariate generalizations of the proportional hazards model. *J. Roy. Statist. Soc., Ser. A* **148:** 82-117.

Clegg, L. X., Cai, J. and Sen, P. K. (1996). Marginal models for failure time data with generalized dependence structure. *Amer. Statist. Assoc. 1996 Proceedings of the Biometrics Section* 280-285.

–––– (1999). A marginal mixed baseline hazards model for multivariate failure time data. *Biometrics* **55:** 805-812.

Cox, D. R. (1972). Regression models and life-tables (with discussion). *J. Roy. Statist. Soc., Ser. B* **34:**187-220.

–––– (1975). Partial likelihood. *Biometrika* **62:** 269-276.

Dawber, T. R. (1980). *The Framingham Study, The Epidemiology of Atherosclerotic Disease.* Harvard University Press.

Fleming, T. R. and Harrington, D. P. (1991). *Counting Processes and Survival Analysis.* New York: Wiley.

Gail, M. H., Wieand, S. and Piantadosi, S. (1984). Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates. *Biometrika* **71:** 431-444.

Hougaard, P. (1987). Modeling multivariate survival. *Scandinavian Journal of Statistics* **14:** 291-304.

Huster, W. J., Brookmeyer, R. and Self, S. G. (1989). Modeling paired survival data with covariates. *Biometrics* **45:** 145-156.

Kalbfleisch, J. D. and Prentice, R. L. (1980). *The Statistical Analysis of Failure Time Data.* New York: Wiley.

Klein, J. P. (1992). Semiparametric estimation of random effects using Cox model based on the EM algorithm. *Biometrics* **48:** 795-806.

Lagakos, S. W. (1988). The loss in efficiency from misspecifying covariates in proportional hazards regression models. *Biometrika* **75:** 156-160.

Lagakos, S. W. and Schoenfeld, D. A. (1984). Properties of proportional hazards score tests under misspecified regression models. *Biometrics* **40:** 1037-1048.

Lee, E. W., Wei, L. J. and Amato, D. A. (1992). Cox-type regression analysis for large numbers of small groups of correlated failure time observations. In J. P. Klein and P. K. Goel, eds., *Survival Analysis: State of the Art* (237-247). Kluwer Academic Publishers: Dordrecht.

Liang, K. Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73:** 13-22.

Liang, K. Y., Self , S. G. and Chang, Y. (1993). Modeling marginal hazards in multivariate failure time data. *J. Roy. Statist. Soc., Ser. B* **55:** 441-453.

Lin, D. Y. (1994). Cox regression analysis of multivariate failure time data: The marginal approach. *Statistics in Medicine* **13:** 2233-2247.

Lin, D. Y. and Wei, L. J. (1989). The robust inference for the Cox proportional hazards model. *Jour. Amer. Statist. Assoc.* **84:** 1074-1078.

Morgan, T. M. (1980). Omitting covariates from the proportional hazards model. *Biometrics* **42:** 993-995.

Murphy, S. A. (1994). Consistency in a proportional hazards model incorporating a random effect. *Ann. Statist.* **22:** 712-731.

–––– (1995). Asymptotic theory for the frailty model. *Ann. Statist.* **23:** 182-198.

Nielsen, G. G., Gill, R. D., Andersen, P. K. and Sørensen, T. I. A. (1992). A counting process approach to maximum likelihood estimation in frailty models. *Scandinavian Journal of Statistics* **19:** 25-43.

Oakes, D. (1982). A model for association in bivariate survival data. *J. Roy. Statist. Soc., Ser. B* **44:** 414-422.

–––– (1989). Bivariate survival models induced by frailties. *Jour. Amer. Statist. Assoc.* **84:** 487-493.

Prentice, R. L. and Hsu, L. (1997). Regression on hazard ratios and cross ratios in multivariate failure time analysis. *Biometrika* **84:** 349-363.

Sen, P. K. and Singer, J. M. (1993). *Large Sample Methods in Statistics.* New York: Wiley.

SOLOMON, P. J. (1984). Effect of misspecification of regression models in the analysis of survival data. *Biometrika* **71:** 291-298. Amendment in *Biometrika* **73:** 245 (1986).

STRUTHERS, C. A. AND KALBFLEISCH, J. D. (1986). Misspecified proportional hazard models. *Biometrika* **73:** 363-369.

THERNEAU, T. M. (1996). Extending the Cox model. *Technical Report no. 58.* Mayo Foundation.

VAUPEL, J. W., MANTON, K. G. AND STALLARD, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* **16:** 439-454.

WEI, L. J., LIN, D. Y. AND WEISSFELD, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distribution. *Jour. Amer. Statist. Assoc.* **84:** 1065-1073.

ZEGER, S. L., LIANG, K. Y. AND ALBERT, P. S. (1988). Models for longitudinal data: A generalized estimating equation approach. *Biometrics* **44:** 1049-1060.

LIMIN CLEGG
NATIONAL CANCER INSTITUTE, NIH
BETHESDA, MD
U.S.A.
E-mail: lin_clegg@nih.gov

JIANWEN CAI, PRANAB KUMAR SEN
AND LAWRANCE L. KUPPER
UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL
CHAPEL HILL, NC
U.S.A.
E-mails: cai@bios.unc.edu
pksen@bios.unc.edu
kupper@bios.unc.edu