Marginal Regression Models for Multivariate Failure Time Data

C. F. SPIEKERMAN and D. Y. LIN

In this article we propose a general Cox-type regression model to formulate the marginal distributions of multivariate failure time data. This model has a nested structure in that it allows different baseline hazard functions among distinct failure types and imposes a common baseline hazard function on the failure times of the same type. We prove that the maximum "quasi-partial-likelihood" estimator for the vector of regression parameters under the independence working assumption is consistent and asymptotically normal with a covariance matrix for which a consistent estimator is provided. Furthermore, we establish the uniform consistency and joint weak convergence of the Aalen-Breslow type estimators for the cumulative baseline hazard functions, and develop a resampling technique to approximate the joint distribution of these processes, which enables one to make simultaneous inference about the survival functions over the time axis and across failure types. Finally, we assess the small-sample properties of the proposed methods through Monte Carlo simulation, and present an application to a real dental study.

KEY WORDS: Censoring; Clustered data; Confidence band; Correlated survival times; Cox regression; Marginal model; Proportional hazards; Survival function.

1. INTRODUCTION

There is much current interest in studying and utilizing statistical models and methods for handling multivariate or clustered failure time data. These data are commonly encountered in scientific investigations because each study subject may experience multiple events (e.g., recurrences of a given disease or occurrences of several diseases) or because the study involves several members from each group (e.g., twins, classmates, etc.). Lin (1994) provided a detailed description of multivariate failure time data along with some real biomedical examples. Statistical analysis of such data needs to account for the intracluster dependence. To this end, two classes of models—proportional hazards frailty models and marginal proportional hazards models—have been proposed. The former approach formulates the dependence explicitly; the latter does not specify the dependence structure in the model formulation but adjusts for it in the inference. This article is concerned with the latter approach.

If each cluster consists of K failure times $T_1, \ldots, T_K$ with corresponding, possibly time-dependent, covariate vectors $\mathbf{Z}_1(t), \ldots, \mathbf{Z}_K(t)$, then the marginal proportional hazards model specifies that the marginal hazard functions for T_k $(k = 1, \ldots, K)$ are

$$\lambda_k(t; \mathbf{Z}_k) = \lambda_0(t) e^{\beta_0^T \mathbf{Z}_k(t)} \tag{1}$$

or

$$\lambda_k(t; \mathbf{Z}_k) = \lambda_{0k}(t) e^{\beta_0^T \mathbf{Z}_k(t)}, \tag{2}$$

depending on whether the K baseline hazard functions are identical or different (Lin 1994). In this article we propose a general model which includes (1) and (2) as special cases. Specifically, suppose that there are K distinct failure types, each of which consists of L exchangeable failure times. Then it is natural to postulate that the marginal hazard function for the lth component of the kth type of failure is related to the corresponding covariate vector $\mathbf{Z}_{kl}(t)$ by

$$\lambda_{kl}(t; \mathbf{Z}_{kl}) = \lambda_{0k}(t) e^{\beta_0^T \mathbf{Z}_{kl}(t)}, \tag{3}$$

where $\lambda_{0k}(t)$ $(k = 1, \ldots, K)$ are unspecified positive functions and β_0 is a $p \times 1$ vector of unknown regression parameters. Note that $\mathbf{Z}_k(t)$ and β_0 in (2) as well as $\mathbf{Z}_{kl}(t)$ and β_0 in (3) may be specified in a manner that allows distinct regression parameter vectors for $k = 1, \ldots, K$.

Model (3) resembles the familiar stratified Cox model for univariate failure time data (Kalbfleisch and Prentice 1980, p. 33), but here the strata are correlated and there is clustering of failure times within each stratum. Stratified models may be used to accommodate nonproportional hazards or stratified random sampling. Obviously, model (3) reduces to (1) if $K = 1$ and to (2) if $L = 1$. The general formulation given here not only provides additional modelling flexibility, but also allows one to present theoretical results for models (1) and (2) in a compact form.

One motivating example for model (3) is the genetic application such as the schizophrenia study described by Liang, Self, and Chang (1993) and Lin (1994). In such a study, the data are collected on the age at onset of a genetic disease from several relatives of each proband, and it is reasonable to assume that the male relatives share a common baseline hazard function, which is different from that shared by the female relatives. A second example arises in dentistry. Each person has about 30 teeth. The teeth in different positions, such as molars versus anteriors, have different distributions with respect to tooth survival time, whereas teeth in similar positions, such as contralateral ones, tend to have similar survival times. In a real dental example given in Section 3.2, model (3) with $K = 6$ is shown to be a good choice.

As in the case of univariate failure time data, the main statistical issues surrounding models (1)–(3) are the estimation of the regression parameters and the estimation of

C. F. Spiekerman is Senior Fellow, Department of Dental Public Health Sciences, D. Y. Lin is Professor, Department of Biostatistics, University of Washington, Seattle, WA 98195. Please address all correspondence to the second author. This research was supported by National Institutes of Health grants AI29168, GM47845, and T32 DE07227. The authors thank Zhiliang Ying and referees for useful comments.

the hazard and survival functions. The estimation of β_0 under models (1) and (2) was studied by Lee, Wei, and Amato (1992) and Wei, Lin, and Weissfeld (1989), using what might be termed "quasi-partial likelihood" estimating equations with an independence working assumption. In this article we extend their ideas to model (3). We develop a large-sample theory for the resulting estimator of β_0 in a more rigorous fashion than was done by Lee et al. and Wei et al., filling several critical gaps in the existing proofs. More importantly we establish the uniform consistency and joint weak convergence of the Aalen–Breslow type estimators for the cumulative hazard functions $\Lambda_{0k}(t) = \int_0^t \lambda_{0k}(u)\,du$ $(k = 1, \ldots, K)$ under model (3) and construct confidence bands for these functions and related quantities. Such results have not been available even for the special cases of models (1) and (2). Thus the main methodological contributions of this article are the inference procedures for the cumulative hazard functions and survival functions. Figures 1 and 2 illustrate the applications of these inference procedures to the dental study mentioned earlier.

The rest of this article is organized as follows. In the next section we present the main theoretical results. In Section 3 we report some simulation results along with the dental example. In Section 4 we give a few concluding remarks. Appendixes A and B contain some technical material.

2. THEORETICAL RESULTS

2.1 Notation and Assumptions

For $i = 1, \ldots, n, k = 1, \ldots, K$, and $l = 1, \ldots, L$, let T_{ikl} and C_{ikl} be the failure and censoring times with respect to the lth component of the kth failure type in the ith cluster, and let $\mathbf{Z}_{ikl} = (Z_{1ikl}, \ldots, Z_{pikl})^T$ be the corresponding (possibly time-varying) covariate vector. The marginal distribution of T_{ikl} is related to $\mathbf{Z}_{ikl}$ through model (3). Define $\mathbf{T}_i = \{T_{ikl}; k = 1, \ldots, K, l = 1, \ldots, L\}$, with $\mathbf{C}_i$ and $\mathbf{Z}_i$ defined similarly. Suppose that $(\mathbf{T}_i, \mathbf{C}_i, \mathbf{Z}_i)$ $(i = 1, \ldots, n)$ are iid and that $\mathbf{T}_i$ is independent of $\mathbf{C}_i$ conditional on $\mathbf{Z}_i$. Write $X_{ikl} = T_{ikl} \wedge C_{ikl}$ and $\Delta_{ikl} = 1(T_{ikl} \le C_{ikl})$, where $a \wedge b = \min(a, b)$ and $1(\cdot)$ is the indicator function. Assume that K and L are fixed constants. The clusters are allowed to have different sizes, which is achieved by setting C_{ikl} to 0 whenever T_{ikl} is missing. We assume that the number of nonmissing observations per failure type tends to ∞ as the number of clusters goes to ∞. Let

$$Y_{ikl}(t) = 1(X_{ikl} \ge t), \qquad N_{ikl}(t) = \Delta_{ikl} 1(X_{ikl} \le t),$$

and

$$M_{ikl}(t) = N_{ikl}(t) - \int_0^t Y_{ikl}(u)\lambda_{0k}(u)e^{\beta_0^T \mathbf{Z}_{ikl}(u)}\,du. \quad (4)$$

One may modify $Y_{ikl}(\cdot)$ to allow left truncation and other general at-risk processes. Note that $M_{ikl}(t)$ is a martingale with respect to the marginal filtration $\mathcal{F}_{t,ikl} =$

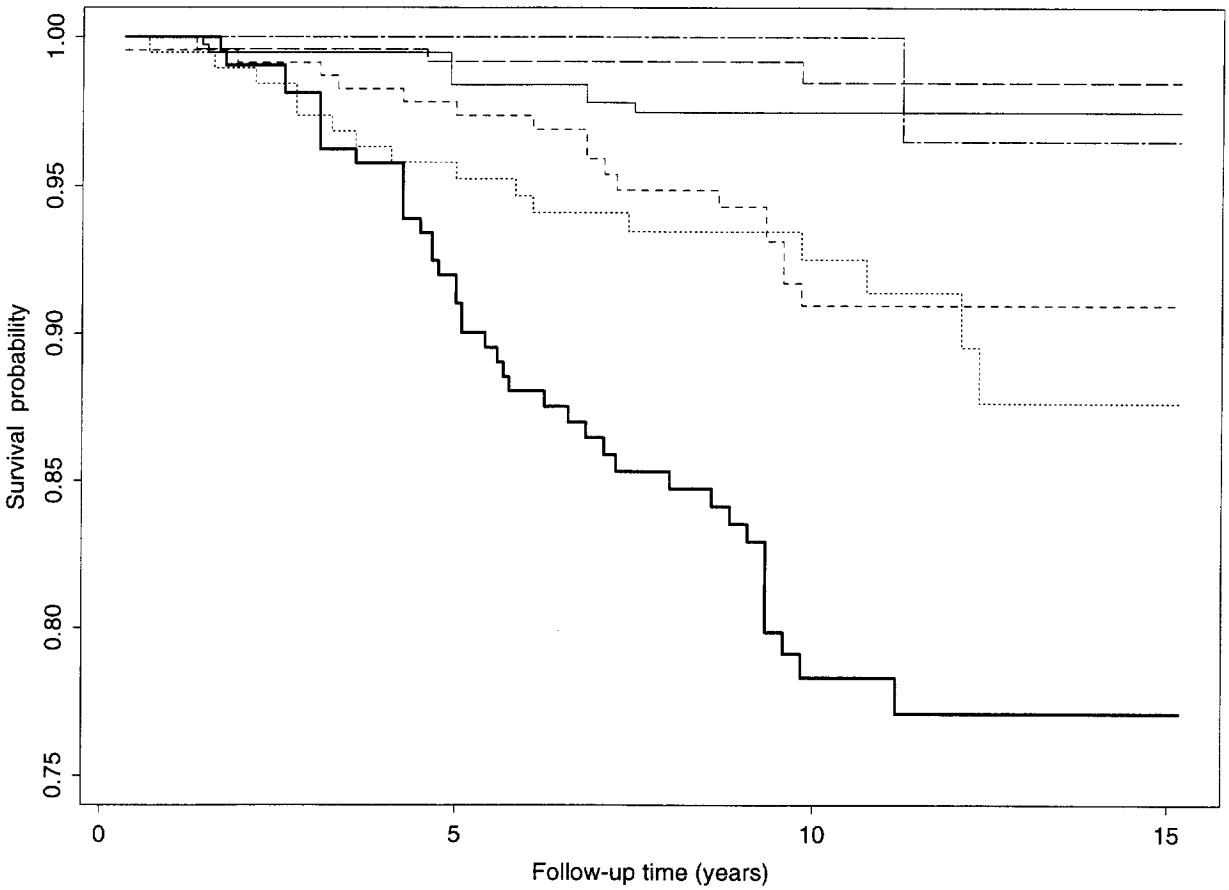

Figure 1. Baseline Survival Function Estimates for the Six Types of Teeth. ——, Upper molars; · · ·, lower molars; - - -, upper premolars; – – –, lower premolars; ——, upper anteriors; – · – ·, lower anteriors. The baseline corresponds to a 45-year-old smoker with satisfactory oral hygiene.

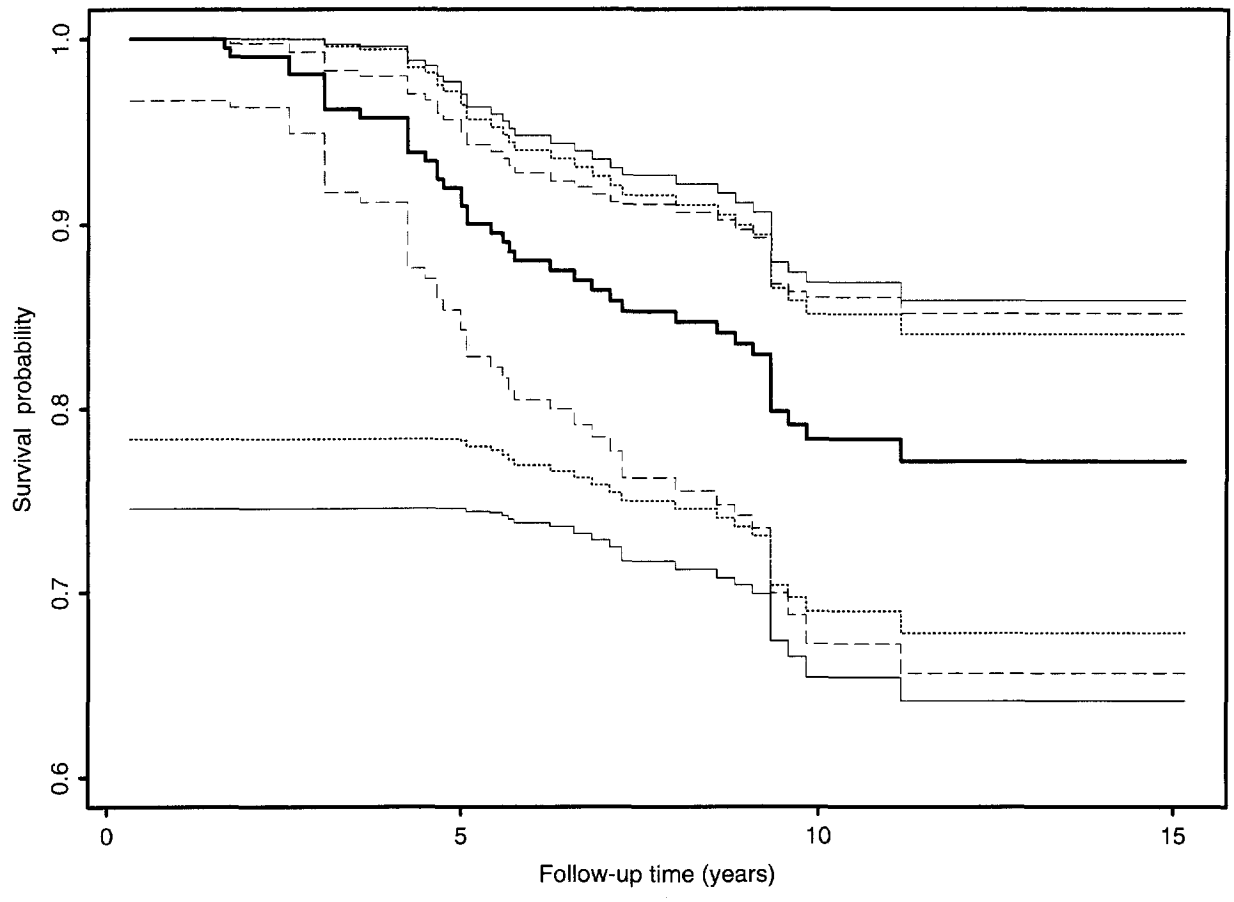

Figure 2. Estimation of the Survival Function of the Upper-Molar Teeth for a 45-Year-Old Smoker With Satisfactory Oral Hygiene. The point estimate is shown by the middle bold solid curve, the robust 95% confidence band by the outside solid curves, the naive 95% confidence band by the dotted curves, and the robust pointwise 95% confidence limits by the dashed curves.

$\sigma\{N_{ikl}(s), Y_{ikl}(s), \mathbf{Z}_{ikl}(s)\colon\ 0 \le s \le t\}$; however, due to the intraclass dependence, $M_{ikl}(t)$ $(i = 1, \ldots, n; k = 1, \ldots, K; l = 1, \ldots, L)$ are not martingales with respect to $\mathcal{F}_t = \vee_{i=1}^{n} \vee_{k=1}^{K} \vee_{l=1}^{L} \mathcal{F}_{t,ikl}$, the joint filtration generated by all of the failure, censoring, and covariate histories up to time t.

It is convenient to introduce the following notation: for $k = 1, \ldots, K$,

$$\mathbf{S}_k^{(r)}(\boldsymbol{\beta}, t) = n^{-1} \sum_{i=1}^{n} \sum_{l=1}^{L} Y_{ikl}(t) e^{\boldsymbol{\beta}^T \mathbf{Z}_{ikl}(t)} \mathbf{Z}_{ikl}(t)^{\otimes r},$$

$$\mathbf{s}_k^{(r)}(\boldsymbol{\beta}, t) = \mathcal{E}\{\mathbf{S}_k^{(r)}(\boldsymbol{\beta}, t)\},$$

$$\mathbf{E}_k(\boldsymbol{\beta}, t) = \frac{\mathbf{S}_k^{(1)}(\boldsymbol{\beta}, t)}{S_k^{(0)}(\boldsymbol{\beta}, t)}, \qquad \mathbf{e}_k(\boldsymbol{\beta}, t) = \frac{\mathbf{s}_k^{(1)}(\boldsymbol{\beta}, t)}{s_k^{(0)}(\boldsymbol{\beta}, t)},$$

$$\mathbf{V}_k(\boldsymbol{\beta}, t) = \frac{\mathbf{S}_k^{(2)}(\boldsymbol{\beta}, t)}{S_k^{(0)}(\boldsymbol{\beta}, t)} - \mathbf{E}_k(\boldsymbol{\beta}, t)^{\otimes 2},$$

and

$$\mathbf{v}_k(\boldsymbol{\beta}, t) = \frac{\mathbf{s}_k^{(2)}(\boldsymbol{\beta}, t)}{s_k^{(0)}(\boldsymbol{\beta}, t)} - \mathbf{e}_k(\boldsymbol{\beta}, t)^{\otimes 2},$$

where $\mathbf{a}^{\otimes 0} = 1$, $\mathbf{a}^{\otimes 1} = \mathbf{a}$, and $\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}^T$ and $\mathcal{E}$ denotes expectation. We denote the summation over a subscript by replacing that subscript with "$\cdot$".

The following set of conditions are assumed throughout the article. For some constant $\tau > 0$:

a. $\Pr\{Y_{ikl}(t) = 1, \text{ for all } t \in [0, \tau]\} > 0$ for all i, k and l.

b. $|Z_{jikl}(0)| + \int_0^\tau |\, dZ_{jikl}(u)| < B_Z$ a.s. for all j, i, k, l, and some constant $B_Z < \infty$.

c. $\mathbf{A} = \sum_{k=1}^{K} \int_0^\tau \mathbf{v}_k(\boldsymbol{\beta}_0, u) s_k^{(0)}(\boldsymbol{\beta}_0, u) \lambda_{0k}(u)\, du$ is positive definite.

Conditions a and b, together with the iid assumption, entail the following:

d. $\int_0^\tau \lambda_{0k}(u)\, du < \infty$ for each k.

e. There exists a neighborhood $\mathcal{B}$ of $\boldsymbol{\beta}_0$ such that for $r = 0, 1, 2$ and $k = 1, \ldots, K$,

$$\sup_{t \in [0,\tau], \boldsymbol{\beta} \in \mathcal{B}} \|\mathbf{S}_k^{(r)}(\boldsymbol{\beta}, t) - \mathbf{s}_k^{(r)}(\boldsymbol{\beta}, t)\|_E \xrightarrow{p} \mathbf{0},$$

where $\|\mathbf{a}\|_E = (\mathbf{a}^T \mathbf{a})^{1/2}$ for a column vector $\mathbf{a}$.

f. $\mathbf{s}_k^{(r)}(\boldsymbol{\beta}, t)$ $(k = 1, \ldots, K; r = 0, 1, 2)$ are continuous functions of $\boldsymbol{\beta} \in \mathcal{B}$ uniformly in $t \in [0, \tau]$ and are bounded on $\mathcal{B} \times [0, \tau]$, $s_k^{(0)}(\boldsymbol{\beta}, t)$ $(k = 1, \ldots, K)$ are

bounded away from 0 on $\mathcal{B} \times [0, \tau]$, and

$$\mathbf{s}_k^{(1)}(\boldsymbol{\beta}, t) = \frac{\partial}{\partial \boldsymbol{\beta}} s_k^{(0)}(\boldsymbol{\beta}, t)$$

and

$$\mathbf{s}_k^{(2)}(\boldsymbol{\beta}, t) = \frac{\partial^2}{\partial \boldsymbol{\beta}^2} s_k^{(0)}(\boldsymbol{\beta}, t)$$

for $k = 1, \ldots, K, \boldsymbol{\beta} \in \mathcal{B}$, and $t \in [0, \tau]$.

2.2 Construction of Estimators for $\boldsymbol{\beta}_0$ and $\Lambda_{0k}(\cdot)$ ($k = 1, \ldots, K$)

Under the independence working assumption, the "quasi-partial likelihood" for $\boldsymbol{\beta}_0$ is

$$\mathcal{L}(\boldsymbol{\beta}) = \prod_{i=1}^{n} \prod_{k=1}^{K} \prod_{l=1}^{L} \left\{ \frac{e^{\boldsymbol{\beta}^T \mathbf{Z}_{ikl}(X_{ikl})}}{n S_k^{(0)}(\boldsymbol{\beta}, X_{ikl})} \right\}^{\Delta_{ikl}},$$

which is the partial likelihood function for a stratified Cox model with K independent strata and nL independent observations in each stratum (Andersen, Borgan, Gill, and Keiding 1993, p. 482). The logarithm of $\mathcal{L}(\boldsymbol{\beta})$ is

$$l(\boldsymbol{\beta}) = \sum_{i=1}^{n} \sum_{k=1}^{K} \sum_{l=1}^{L} \int_0^\tau [\boldsymbol{\beta}^T \mathbf{Z}_{ikl}(u) - \log\{n S_k^{(0)}(\boldsymbol{\beta}, u)\}] \times dN_{ikl}(u).$$

The first and minus second derivatives of $l(\boldsymbol{\beta})$ are

$$\mathbf{U}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \sum_{k=1}^{K} \sum_{l=1}^{L} \int_0^\tau \{\mathbf{Z}_{ikl}(u) - \mathbf{E}_k(\boldsymbol{\beta}, u)\} \, dN_{ikl}(u)$$

and

$$\mathcal{I}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \sum_{k=1}^{K} \sum_{l=1}^{L} \int_0^\tau \mathbf{V}_k(\boldsymbol{\beta}, u) \, dN_{ikl}(u).$$

It is easy to see that $\mathcal{I}(\boldsymbol{\beta})$ is positive semi-definite. Let $\hat{\boldsymbol{\beta}}$ be the root of $\mathbf{U}(\boldsymbol{\beta})$, which is unique if $\mathcal{I}(\boldsymbol{\beta})$ is nonsingular. The corresponding Aalen–Breslow type estimators for $\Lambda_{0k}(t)$ $(k = 1, \ldots, K)$ are

$$\hat{\Lambda}_{0k}(t; \hat{\boldsymbol{\beta}}) = \int_0^t \frac{dN_{\cdot k \cdot}(u)}{n S_k^{(0)}(\hat{\boldsymbol{\beta}}, u)}.$$

It follows from (4) that

$$\mathbf{U}(\boldsymbol{\beta}_0) = \sum_{i=1}^{n} \sum_{k=1}^{K} \sum_{l=1}^{L} \int_0^\tau \{\mathbf{Z}_{ikl}(u) - \mathbf{E}_k(\boldsymbol{\beta}_0, u)\} \, dM_{ikl}(u), \tag{5}$$

and, given $Y_{\cdot k \cdot}(t) > 0$,

$$\hat{\Lambda}_{0k}(t; \boldsymbol{\beta}_0) - \Lambda_{0k}(t) = \int_0^t \frac{dM_{\cdot k \cdot}(u)}{n S_k^{(0)}(\boldsymbol{\beta}_0, u)}. \tag{6}$$

As mentioned in Section 2.1, $M_{ikl}(t)$ $(i = 1, \ldots, n; k = 1, \ldots, K; l = 1, \ldots, L)$ are not martingales with respect to the joint filtration $\mathcal{F}_t$; therefore, the familiar martingale convergence theorems cannot be directly applied to (5) or (6). Thus we use other tools, including those from the theory of empirical processes, to study the asymptotic properties of the proposed estimators.

2.3 Asymptotic Properties of $\hat{\boldsymbol{\beta}}$ and $\mathbf{U}(\boldsymbol{\beta}_0)$

In this section we state and prove the main theorems for $\hat{\boldsymbol{\beta}}$ and $\mathbf{U}(\boldsymbol{\beta}_0)$. Because these results are natural extensions of those of Wei et al. (1989) and Lee et al. (1992), we keep our discussion fairly brief. It should be noted that more rigorous and complete proofs are provided here than in those papers. In fact, one of the main motivations behind this work was to fill some important gaps in the existing proofs. To avoid unnecessary technical distractions, we relegate to Appendix A several lemmas, which are useful in this section as well as in Sections 2.5 and 2.6 and Appendix B. The following lemma is proven in Appendix B.

Lemma 1. The estimator $\hat{\boldsymbol{\beta}}$ converges in probability to $\boldsymbol{\beta}_0$.

Theorem 1. The random vector $n^{-1/2}\mathbf{U}(\boldsymbol{\beta}_0)$ converges weakly to a p-variate normal vector with mean $\mathbf{0}$ and covariance matrix $\mathbf{B} = \mathcal{E}(\mathbf{w}_{1\cdot\cdot}^{\otimes 2})$, where $\mathbf{w}_{ikl} = \int_0^\tau \{\mathbf{Z}_{ikl}(u) - \mathbf{e}_k(\boldsymbol{\beta}_0, u)\} \, dM_{ikl}(u)$.

Proof. By Lemma A.1,

$$n^{-1/2} \int_0^\tau \{\mathbf{E}_k(\boldsymbol{\beta}_0, u) - \mathbf{e}_k(\boldsymbol{\beta}_0, u)\} \, dM_{\cdot k \cdot}(u) \xrightarrow{p} \mathbf{0}, \tag{7}$$

which implies that

$$\mathbf{U}(\boldsymbol{\beta}_0) = \sum_{i=1}^{n} \sum_{k=1}^{K} \sum_{l=1}^{L} \int_0^\tau \{\mathbf{Z}_{ikl}(u) - \mathbf{e}_k(\boldsymbol{\beta}_0, u)\} \times dM_{ikl}(u) + o_p(n^{1/2}). \tag{8}$$

Because (8) is essentially a sum of n iid random vectors with zero mean and finite variance, the desired asymptotic normality follows from the multivariate central limit theorem.

Remark 1. In the case of $L \geq 2$, (7) is not a trivial result. Lee et al. (1992) also used this result to study the asymptotic properties of $\mathbf{U}(\boldsymbol{\beta}_0)$ under model (1) (i.e., $K = 1$ and $L \geq 2$), but did not provide a justification for it. The proof of Lemma A.1 given in Appendix A fills this critical gap.

Corollary 1. The random vector $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ converges weakly to a p-variate normal vector with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Omega} = \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}$.

Proof. By the Taylor series expansion,

$$n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = \{\mathcal{I}(\boldsymbol{\beta}^*)/n\}^{-1} n^{-1/2} \mathbf{U}(\boldsymbol{\beta}_0), \tag{9}$$

where $\boldsymbol{\beta}^*$ is on the line segment between $\hat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}_0$. The desired convergence then follows from Theorem 1 provided

that

$$n^{-1}\mathcal{I}(\boldsymbol{\beta}^*) \xrightarrow{p} \mathbf{A}. \tag{10}$$

The proof of (10) is deferred until the next section.

2.4 Asymptotic Properties of $\hat{\Lambda}_{0k}(\cdot;\hat{\boldsymbol{\beta}})$ ($k = 1, \ldots, K$)

In this and the following sections, we study the estimation of the cumulative baseline hazard functions $\Lambda_{0k}(\cdot)$ $(k = 1, \ldots, K)$ and related quantities, which is the main methodological contribution of our work. The proofs of the theorems are quite technical and tedious and thus are relegated to Appendix B. In the sequel, we use the norm $\|f(t)\| = \sup_{t\in[0,\tau]} |f(t)|$ for a function $f\colon [0,\tau] \to \mathcal{R}$. We first give a theorem that we use to establish the uniform consistency of $\hat{\Lambda}_{0k}(\cdot;\hat{\boldsymbol{\beta}})$ $(k = 1, \ldots, K)$ as well as other results.

Theorem 2. Let $f_n(\cdot)$ $(n = 1, 2, \ldots)$ be a sequence of possibly random, left-continuous functions with right-side limits such that $\int_0^\tau |df_n(u)| = O_p(n^{1/2})$ and $\|f_n(t)\| = O_p(1)$. Also, suppose that $\boldsymbol{\beta}^*$ converges in probability to $\boldsymbol{\beta}_0$. Then for $k = 1, \ldots, K$,

$$\left\| \int_0^t f_n(u)\, d\hat{\Lambda}_{0k}(u;\boldsymbol{\beta}^*) - \int_0^t f_n(u)\lambda_{0k}(u)\, du \right\| \xrightarrow{p} 0. \tag{11}$$

Furthermore, if $\|f_n(t) - f(t)\| \to^p 0$, then

$$\left\| \int_0^t f_n(u)\, d\hat{\Lambda}_{0k}(u;\boldsymbol{\beta}^*) - \int_0^t f_n(u)\lambda_{0k}(u)\, du \right\| \xrightarrow{p} 0. \tag{12}$$

Theorem 2, together with Lemma 1, implies (10) as well as the following result.

Corollary 2. For each $k = 1, \ldots, K$, the estimator $\hat{\Lambda}_{0k}(t;\hat{\boldsymbol{\beta}})$ converges in probability to $\Lambda_{0k}(t)$ uniformly in $t \in [0,\tau]$.

We now study the weak convergence of $\hat{\Lambda}_{0k}(\cdot;\hat{\boldsymbol{\beta}})$ $(k = 1, \ldots, K)$. Let $\mathbf{W}(t) = n^{1/2}[\{\hat{\Lambda}_{01}(t;\hat{\boldsymbol{\beta}}) - \Lambda_{01}(t)\}, \ldots, \{\hat{\Lambda}_{0K}(t;\hat{\boldsymbol{\beta}}) - \Lambda_{0K}(t)\}]^T$. Also, let $\boldsymbol{\mathcal{W}}(t) = \{\mathcal{W}_1(t), \ldots, \mathcal{W}_K(t)\}^T$ be a zero-mean Gaussian random field, the covariance function between $\mathcal{W}_j(s)$ and $\mathcal{W}_k(t)$ $(1 \le j,k \le K; 0 \le s,t \le \tau)$ being $\xi_{jk}(s,t) = \mathcal{E}\{\Psi_{1j}(s)\Psi_{1k}(t)\}$, where for $i = 1, \ldots, n$ and $k = 1, \ldots, K$,

$$\Psi_{ik}(t) = \int_0^t \frac{dM_{ik\cdot}(u)}{s_k^{(0)}(\boldsymbol{\beta}_0, u)} + \mathbf{h}_k(t)^T \mathbf{A}^{-1}\mathbf{w}_{i\cdot},$$

and $\mathbf{h}_k(t) = -\int_0^t \mathbf{e}_k(\boldsymbol{\beta}_0, u)\lambda_{0k}(u)\, du$. In addition, let $\mathcal{D}[0,\tau]^K$ be a space consisting of functions $f\colon [0,\tau] \to \mathcal{R}^K$ such that $\mathbf{f}(t) = \{f_1(t), \ldots, f_K(t)\}^T$, where for each $k = 1, \ldots, K$, $f_k\colon [0,\tau] \to \mathcal{R}$ is right-continuous with left-side limits. Make $\mathcal{D}[0,\tau]^K$ a metric space by equipping it with the metric $\rho_K(\mathbf{f},\mathbf{g})$, where $\rho_K(\mathbf{f},\mathbf{g}) = \max\{\|f_k(t) - g_k(t)\|\colon 1 \le k \le K\}$ for $\mathbf{f},\mathbf{g} \in \mathcal{D}[0,\tau]^K$.

Theorem 3. The random field $\mathbf{W}(t)$ converges weakly to $\boldsymbol{\mathcal{W}}(t)$ in $\mathcal{D}[0,\tau]^K$.

2.5 Variance and Covariance Estimation

It is natural to estimate $\mathbf{A}$ and $\mathbf{B}$ by $\hat{\mathbf{A}}(\hat{\boldsymbol{\beta}}) = \mathcal{I}(\hat{\boldsymbol{\beta}})/n$ and $\hat{\mathbf{B}}(\hat{\boldsymbol{\beta}}) = n^{-1}\sum_{i=1}^n \hat{\mathbf{w}}_{i\cdot}^{\otimes 2}$, where $\hat{\mathbf{w}}_{ikl} = \int_0^\tau \{\mathbf{Z}_{ikl}(u) - \mathbf{E}_k(\hat{\boldsymbol{\beta}}, u)\} d\hat{M}_{ikl}(u)$ and $\hat{M}_{ikl}(t) = N_{ikl}(t) - \int_0^t Y_{ikl}(u)\, e^{\hat{\boldsymbol{\beta}}^T\mathbf{Z}_{ikl}(u)}\, d\hat{\Lambda}_{0k}(u;\hat{\boldsymbol{\beta}})$. The consistency of $\hat{\mathbf{A}}(\hat{\boldsymbol{\beta}})$ follows from (10) and Lemma 1, whereas that of $\hat{\mathbf{B}}(\hat{\boldsymbol{\beta}})$ follows from Lemma A.2. Thus we have the following result.

Corollary 3. The covariance matrix estimator $\hat{\mathbf{A}}(\hat{\boldsymbol{\beta}})^{-1}\hat{\mathbf{B}}(\hat{\boldsymbol{\beta}})\hat{\mathbf{A}}(\hat{\boldsymbol{\beta}})^{-1}$ converges in probability to $\boldsymbol{\Omega}$.

Remark 2. The consistency of the covariance matrix estimators of $\hat{\boldsymbol{\beta}}$ proposed by Lee et al. (1992) and Wei et al. (1989) for models (1) and (2) has not been established before, but now follows immediately from Corollary 3.

Similarly, we estimate the limiting covariance function $\xi_{jk}(s,t)(1 \le j,k \le K; s,t \in [0,\tau])$ by its empirical counterpart $\hat{\xi}_{jk}(s,t) = n^{-1}\sum_{i=1}^n \hat{\Psi}_{ij}(s)\hat{\Psi}_{ik}(t)$, where for $i = 1, \ldots, n$ and $k = 1, \ldots, K$,

$$\hat{\Psi}_{ik}(t) = \int_0^t \frac{d\hat{M}_{ik\cdot}(u)}{S_k^{(0)}(\hat{\boldsymbol{\beta}}, u)} + \mathbf{H}_k(\hat{\boldsymbol{\beta}}, t)^T \hat{\mathbf{A}}(\hat{\boldsymbol{\beta}})^{-1}\hat{\mathbf{w}}_{i\cdot},$$

and $\mathbf{H}_k(\boldsymbol{\beta}, t) = -\int_0^t \mathbf{E}_k(\boldsymbol{\beta}, u)\, d\hat{\Lambda}_{0k}(u;\boldsymbol{\beta})$. The following result follows from Lemma A.2, Corollary 3, and the proof of Theorem 3.

Corollary 4. For any $1 \le j,k \le K$, the covariance function estimator $\hat{\xi}_{jk}(s,t)$ converges in probability to $\xi_{jk}(s,t)$ uniformly in $s,t \in [0,\tau]$.

2.6 Simultaneous Inference on $\Lambda_{0k}(\cdot)$ ($k = 1, \ldots, K$)

Theorem 3 and Corollary 4 enable one to make inference about the $\Lambda_{0k}(\cdot)$ at fixed time points. To draw more general simultaneous inference, such as constructing confidence bands for $\Lambda_{0k}(\cdot)$ or testing the equality of $\Lambda_{0j}(\cdot)$ and $\Lambda_{0k}(\cdot)$ $(j \ne k)$ over the entire time span of interest, we need to evaluate the probability distribution of $\boldsymbol{\mathcal{W}}(\cdot)$. This cannot be done analytically even if $K = 1$, due to the complicated nature of the covariance function. We instead develop a resampling method to approximate the distribution of $\boldsymbol{\mathcal{W}}(\cdot)$.

Define $\hat{\mathbf{W}}(t) = \{\hat{W}_1(t), \ldots, \hat{W}_K(t)\}^T$, where $\hat{W}_k(t) = n^{-1/2}\sum_{i=1}^n \hat{\Psi}_{ik}(t)G_i$ $(k = 1, \ldots, K)$ and $(G_1, \ldots, G_n)$ are independent standard normal variables that are independent of the data $\{N_{ikl}(t), Y_{ikl}(t), \mathbf{Z}_{ikl}(t); t \in [0,\tau]\}$ $(i = 1, \ldots, n; k = 1, \ldots, K; l = 1, \ldots, L)$. The following theorem states that $\hat{\mathbf{W}}$ and $\mathbf{W}$ have the same limiting distribution. It is proved in Appendix B.

Theorem 4. Conditional on the data $\{N_{ikl}(t), Y_{ikl}(t), \mathbf{Z}_{ikl}(t); t \in [0,\tau]\}$ $(i = 1, \ldots, n; k = 1, \ldots, K; l = 1, \ldots, L)$, the random field $\hat{\mathbf{W}}$ converges weakly to $\boldsymbol{\mathcal{W}}$ in $\mathcal{D}[0,\tau]^K$ in probability.

Remark 3. Theorem 4 involves the concept of conditional weak convergence in probability (van der Vaart and Wellner 1996, sec. 2.9), and says that, for any continuous bounded function $f\colon \mathcal{D}[0,\tau]^K \to \mathcal{R}$, the condi-

tional expectation $\mathcal{E}\{f(\hat{\mathbf{W}})|\mathcal{X}\}$ converges in probability to $\mathcal{E}\{f(\boldsymbol{W})\}$, where $\mathcal{X}$ denotes the σ field generated by the data $\{N_{ikl}(t), Y_{ikl}(t), \mathbf{Z}_{ikl}(t); t \in [0,\tau]\}$ $(i = 1,\ldots,n; k = 1,\ldots,K; l = 1,\ldots,L)$.

Theorem 4 provides the theoretical basis for our resampling method. To approximate the distribution of $\mathbf{W}(\cdot)$, we obtain a large number of realizations from $\hat{\mathbf{W}}(\cdot)$ by repeatedly generating normal random samples $(G_1,\ldots,G_n)$ while fixing the data $\{N_{ikl}(t), Y_{ikl}(t), \mathbf{Z}_{ikl}(t); t \in [0,\tau]\}$ $(i = 1,\ldots,n; k = 1,\ldots,K; l = 1,\ldots,L)$ at their observed values.

This resampling scheme shares the spirit of that used by Lin et al. (1994) for constructing the confidence bands with univariate failure time data. Due to the intracluster dependence of multivariate failure times, however, the approximation developed here differs substantially from that of Lin et al. (1994), and the proofs given in Appendix B are more difficult. In fact, Lin et al. did not rigorously justify their resampling scheme. For the special case of $K = L = 1$, the $\mathbf{W}(\cdot)$ process considered here reduces to that of Lin et al., and our $\hat{\mathbf{W}}(\cdot)$ is asymptotically equivalent to, though still numerically different from, Lin et al.'s approximating process.

To construct confidence bands for $\Lambda_{0k}(\cdot)$, it is useful to consider the transformed process $W_k^*(t) = n^{1/2}q(t)[\phi\{\hat{\Lambda}_{0k}(t;\hat{\boldsymbol{\beta}})\} - \phi\{\Lambda_{0k}(t)\}]$, where ϕ is a known function with nonzero continuous derivative ϕ' on $[\Lambda_{0k}(t_1), \Lambda_{0k}(t_2)]$ $(0 \le t_1 \le t_2 \le \tau)$, and the weight function $q(\cdot)$ converges in probability to a nonnegative function uniformly on $[t_1, t_2]$. By the functional δ method (Andersen et al. 1993, sec. II.8), $W_k^*(t) = q(t)\phi'\{\hat{\Lambda}_{0k}(t;\hat{\boldsymbol{\beta}})\}W_k(t) + o_p(1)$, whose distribution can be approximated by that of $\hat{W}_k^*(t) = q(t)\phi'\{\hat{\Lambda}_{0k}(t;\hat{\boldsymbol{\beta}})\}\hat{W}_k(t)$. Let c_α be the boundary value satisfying $\Pr\{\sup_{t\in[t_1,t_2]}|\hat{W}_k^*(t)| > c_\alpha\} = \alpha$, the probability being estimated through simulation. Then an approximate $(1-\alpha)100\%$ confidence band for $\phi\{\Lambda_{0k}(t)\}$ on $[t_1, t_2]$ is $\phi\{\hat{\Lambda}_{0k}(t;\hat{\boldsymbol{\beta}})\} \mp n^{-1/2}c_\alpha/q(t)$. The choices of the transformation ϕ and weight function q were discussed by Lin et al., and the same principles apply here.

A subtle correction is recommended in evaluating the distribution of the supremum of the $|W_k^*(\cdot)|$ process. If $q(\cdot)$ is a step function that changes values only at observation times, then one can write $\sup_t |W_k^*(t)|$ as

$$\max_{ijl}(|W_k^*(X_{ijl})| \vee |n^{1/2}q(X_{ijl}-)[\phi\{\hat{\Lambda}_{0k}(X_{ijl}-;\hat{\boldsymbol{\beta}})\} - \phi\{\hat{\Lambda}_{0k}(X_{ijl};\hat{\boldsymbol{\beta}})\}] + q(X_{ijl}-)W_k^*(X_{ijl})/q(X_{ijl})|),$$

where $a \vee b = \max(a,b)$. Thus we estimate c_α to be the $(1-\alpha)100$th percentile of 1,000 simulated realizations of

$$\max_{ijl}(|\hat{W}_k^*(X_{ijl})| \vee |n^{1/2}q(X_{ijl}-)[\phi\{\hat{\Lambda}_{0k}(X_{ijl}-;\hat{\boldsymbol{\beta}})\} - \phi\{\hat{\Lambda}_{0k}(X_{ijl};\hat{\boldsymbol{\beta}})\}] + q(X_{ijl}-)\hat{W}_k^*(X_{ijl})/q(X_{ijl})|).$$

Our experience indicates that this approach results in better small-sample performance than the more obvious choice of $\max_{ijl}\{|\hat{W}_k^*(X_{ijl})| \vee |\hat{W}_k^*(X_{ijl}-)|\}$, although the two approaches are asymptotically equivalent.

3. NUMERICAL RESULTS

3.1 Simulation Studies

Extensive simulation studies were conducted to assess the finite-sample behavior of the inference procedures proposed in Section 2. The results of our studies on the estimation of $\boldsymbol{\beta}_0$ under model (3) are similar to those of Lin (1994) for the special cases of models (1) and (2). They showed that the asymptotic approximations are adequate for practical use and that ignoring the intracluster dependence could yield misleading variance–covariance estimators. The details are omitted here but have been provided by Spiekerman (1995).

To evaluate the performance of the proposed confidence bands, we generated multivariate failure times from model (1) under two families of multivariate distributions: the Clayton (1978) family with joint survival function

$$\Pr(T_1 > t_1,\ldots,T_L > t_L|Z_1,\ldots,Z_L) = \left[\sum_{l=1}^{L}\exp\{-(1-\theta)e^{\beta_0 Z_l}t_l\} - L + 1\right]^{1/(1-\theta)},$$

and the Hougaard (1986) family with joint survival function

$$\Pr(T_1 > t_1,\ldots,T_L > t_L|Z_1,\ldots,Z_L) = \exp\left[-\left\{\sum_{l=1}^{L}(t_l e^{\beta_0 Z_l})^{1/\gamma}\right\}^{\gamma}\right].$$

We set θ to 1.0, 1.67, and 3.0, corresponding to Kendall's τ of 0 (independence), .25, and .5 under $Z_l = 0$ $(l = 1,\ldots,L)$, and set γ to .7, .5, and .4, corresponding to Kendall's τ of .3, .5, and .6 under $Z_l = 0$ $(l = 1,\ldots,L)$. A single dichotomous covariate was included in the model, and $\boldsymbol{\beta}_0$ was set to log 2. The covariate values were generated by two designs. Design 1 represents a matched study in which half the members of each cluster have $Z = 0$ and half have $Z = 1$. Under design 2, Z represents a cluster-level covariate whose values are the same for all members of the same cluster, and we set $\Pr(Z = 0) = \Pr(Z = 1) = 1/2$ for each cluster. The failure times within each cluster were potentially censored by a common uniform$[0, c]$ random variable independent of the failure times. The censoring parameter c was chosen to achieve 25%, 50%, or 75% censorship.

We studied primarily the Hall–Wellner type band based on the log transformation of the cumulative hazard function. Specifically, we chose $\phi(\cdot) = \log(\cdot)$ and $q(t) = \hat{\Lambda}_0(t;\hat{\boldsymbol{\beta}})/\{1+\hat{\xi}_{11}(t,t)\}$. The resulting 95% confidence band for the baseline survival function $S_0(t) = \Pr(T > t|Z = 0)$ takes the form $\hat{S}_0(t)^{\exp\{\pm n^{-1/2}c_{.05}/q(t)\}}$, where $\hat{S}_0(t) = \exp\{-\hat{\Lambda}_0(t;\hat{\boldsymbol{\beta}})\}$. To reduce tail instabilities, we set $t_1 = 0$ and $t_2 = .9c$ and restricted each band to be between the first and last uncensored failure times. For comparison, we also evaluated the corresponding Hall–Wellner type band of Lin et al. (1994, eq. 2.4). We refer to the latter as the naive confidence band, because no adjustment is made to account for the intracluster dependence.

Tables 1 and 2 display the empirical performance of the 95% confidence bands for the true baseline survival function under Clayton's and Hougaard's families of distributions. Each entry was based on 1,000 simulated datasets.

Table 1. Empirical Coverage Percentages of the 95% Confidence Bands for the Survival Function Under the Clayton Family of Distributions

			Design 1				Design 2			
			$L = 2$		$L = 4$		$L = 2$		$L = 4$	
θ	n	Censoring %	Robust	Naive	Robust	Naive	Robust	Naive	Robust	Naive
1.00	50	25	93.9	95.6	93.5	95.3	93.4	96.5	93.5	97.1
		50	95.9	97.3	95.7	97.4	94.2	96.3	92.5	96.4
		75	95.4	96.1	95.3	96.8	94.8	96.5	95.8	97.6
	100	25	95.1	95.8	94.3	95.1	92.8	94.0	92.9	94.7
		50	95.7	96.5	96.1	96.7	95.6	97.2	94.9	96.1
		75	96.4	96.8	95.8	96.5	95.8	96.7	95.7	96.6
	200	25	95.4	95.3	95.3	96.0	95.4	95.6	94.4	94.6
		50	95.9	96.0	95.6	95.9	94.8	95.4	94.8	94.8
		75	95.7	95.7	94.3	94.7	95.4	96.2	95.5	96.2
1.67	50	25	94.6	96.6	93.0	92.8	94.3	94.0	92.9	87.1
		50	94.9	95.7	94.8	94.6	93.9	94.7	92.7	89.5
		75	96.4	97.0	94.6	94.4	95.4	96.2	94.1	92.5
	100	25	95.4	96.3	93.0	90.6	94.1	91.7	94.1	83.5
		50	96.2	96.4	95.9	94.9	95.2	94.1	94.5	88.5
		75	95.5	96.6	95.5	95.0	95.2	94.6	94.7	91.5
	200	25	94.3	94.3	94.5	89.2	94.1	91.5	94.0	83.1
		50	95.9	95.9	95.0	91.8	95.7	93.5	95.1	87.9
		75	96.3	96.6	94.4	92.7	94.9	94.2	93.3	89.4
3.00	50	25	94.6	95.9	91.8	89.2	94.5	90.2	89.4	75.6
		50	94.0	96.1	95.4	93.3	93.5	93.5	92.7	83.8
		75	96.1	96.6	95.7	92.8	94.7	94.1	93.7	88.2
	100	25	94.0	94.2	93.3	86.2	93.8	86.3	95.0	72.5
		50	95.4	95.7	95.9	91.9	95.2	91.5	95.5	81.2
		75	96.0	96.2	95.5	92.4	95.0	93.3	94.9	87.5
	200	25	94.6	94.1	94.7	86.0	94.2	85.1	94.8	71.2
		50	96.3	96.6	95.2	91.0	94.7	90.3	95.4	79.7
		75	96.0	96.4	95.1	91.9	96.2	93.5	95.9	86.5

For each dataset, 1,000 simulations were used to estimate the boundary values. The "continuity correction" described at the end of Section 2.6 was used for both the robust and naive bands. The proposed (robust) confidence bands maintain coverage probabilities near the nominal level in virtually all cases. The naive bands perform well under independence, but in general do not have proper coverage probabilities when the failure times are correlated. The undercoverage of the naive bands is the most severe under design 2 with $L = 4$.

3.2 A Real Example

We now apply the proposed methods to a dental study conducted by Dr. Michael K. McGuire of the University of Texas to assess the role of commonly measured clinical

Table 2. Empirical Coverage Percentages of the 95% Confidence Bands for the Survival Function Under the Hougaard Family of Distributions

			Design 1				Design 2			
			$L = 2$		$L = 4$		$L = 2$		$L = 4$	
γ	n	Censoring %	Robust	Naive	Robust	Naive	Robust	Naive	Robust	Naive
.7	50	25	93.7	95.2	93.7	87.7	94.4	93.0	92.7	86.5
		50	94.4	95.8	95.8	92.3	94.4	94.2	95.0	88.6
		75	95.5	96.8	95.1	90.8	95.7	93.8	93.3	87.0
	100	25	93.2	94.2	94.3	81.5	94.7	91.2	92.7	84.4
		50	94.0	94.9	94.1	85.0	95.5	92.7	94.7	86.2
		75	94.9	96.5	94.0	87.5	95.7	92.7	94.1	85.3
.5	50	25	94.4	96.0	92.1	90.2	93.2	90.6	92.2	78.5
		50	94.0	95.9	93.9	93.0	94.9	91.5	94.5	81.6
		75	95.5	95.7	95.3	93.4	95.6	92.5	93.7	81.8
	100	25	93.0	93.6	94.6	89.2	94.3	89.1	95.0	77.9
		50	95.3	96.0	95.7	92.3	95.0	91.1	94.3	78.4
		75	95.1	96.2	95.1	92.3	96.0	91.3	94.1	80.6
.4	50	25	93.0	94.8	91.5	89.7	94.1	89.8	93.7	79.1
		50	94.9	96.2	95.4	93.8	94.3	90.9	96.1	81.9
		75	96.4	97.1	95.5	93.4	94.4	91.4	95.1	82.6
	100	25	93.3	93.2	94.2	88.2	94.5	87.3	94.5	76.1
		50	94.8	95.6	95.8	92.5	94.9	88.9	95.3	78.8
		75	96.0	96.5	95.8	92.1	94.7	90.3	95.5	80.2

parameters in predicting tooth survival (McGuire and Nunn 1996). The database consists of 100 consecutive patients from Dr. McGuire's appointment book who had at least 5 years of maintenance care. All of these patients had been initially diagnosed with moderate to severe chronic adult periodontitis.

For this illustration, we confine our attention to the effects of behavioral factors, such as cigarette smoking and oral hygiene, on tooth survival. For each tooth, the failure time is defined as the time to tooth loss measured from the initiation of active periodontal therapy under Dr. McGuire. The failure times for the teeth of the same patient are anticipated to be strongly correlated.

As is commonly done in the analysis of dental data, third molars (i.e., wisdom teeth) are excluded from our analysis. One patient had all 20 surviving teeth extracted at a single visit. We exclude this individual from our analysis because of the extreme leverage on the regression parameter estimates. The remaining data contain follow-up on 2,413 teeth, with failures observed on 93 teeth. The follow-up time ranged from .33 to 15.17 years.

After the exclusion of wisdom teeth, each patient has 28 possible teeth. The 28 tooth positions fall into six groups or types: upper and lower molars, upper and lower premolars, and upper and lower anteriors (Harty and Ogston 1992). The survival distributions are expected to differ appreciably between these groups but to be the same or similar within the same group (Hujoel et al. 1998). Thus we consider model (3) with six separate baseline hazard functions for the six types of teeth. We include three covariates in the model: *smoking*, coded as 0 if the patient was a smoker and 1 otherwise; *hygiene*, coded as 1 if the patient exhibited poor oral hygiene and 0 otherwise; and log(*age*), centered at the sample mean, log 45, to make the baseline survival functions more meaningful. Due to the small numbers of observed failures, especially in the anterior teeth, we impose a common regression parameter vector for the six types of teeth.

The regression results are summarized in Table 3. The relative risk of tooth loss for cigarette smoking is 2.45, with a robust 95% confidence interval of (1.21, 4.95). Due to the strong dependence among failure times of the same patient, the robust standard error estimates for the three regression parameter estimates are substantially higher than their naive counterparts. In fact, the effects of oral hygiene and age are significant at the 5% level according to the naive Wald tests, but not according to robust tests. Incidentally, none of the second-order terms was found to be significant. Furthermore, the p values for testing the significance of dummy time-dependent covariates of the form $Z * \log t$ are .93, .71, and .55 with respect to *smoking*, *hygiene*, and log(*age*), providing no evidence against the proportional hazards assumption.

Table 3. Estimates of the Regression Parameters for the Dental Study

		Robust		Naive	
Covariate	*Estimate*	*SE*	*P value*	*SE*	*P value*
Smoking	−.898	.359	.012	.214	<.001
Hygiene	.699	.531	.188	.274	.011
Log(*age*/45)	1.205	.681	.077	.499	.016

NOTE: "Estimate" denotes the point estimate of the regression parameter and "SE" denotes the estimated standard error. The "*P* value" is two-sided.

The six baseline survival function estimates are exhibited in Figure 1. The codings of the covariates imply that the baseline pertains to a subject who is a 45-year-old smoker with satisfactory oral hygiene. There appear to be considerable differences between the tooth groups. Furthermore, these differences do not satisfy proportional hazards, as is evident from the figure and confirmed by numerical tests; the p value for testing the significance of $Z * \log t$ is .039 with respect to the indicator covariate of upper molar versus lower anterior. Thus it would be inappropriate to use model (1) with indicator covariates to represent the differences between the tooth groups.

Figure 2 displays the point estimate along with the 95% pointwise confidence limits and simultaneous confidence bands for the baseline survival function of the upper molars. It is not surprising that the robust pointwise confidence limits are much narrower than the robust confidence band. The naive confidence band lies within the robust band, and even within the robust pointwise confidence limits after year 9. The naive pointwise confidence limits, which are not shown in the figure to avoid overcrowding, are much narrower than their robust counterparts. This figure demonstrates the importance of adjusting for intraclass dependence as well as multiple comparisons.

Instead of model (3) with $K = 6$, one could use model (2) with 28 different baseline hazard functions, but this would result in less parsimonious summarization of the data and also incur loss of statistical efficiency. The small numbers of observed failures, especially for the individual tooth positions in the upper and lower anteriors, further favor using model (3) over (2).

4. DISCUSSION

The marginal modelling methodology as described by Lee et al. (1992) and Wei et al. (1989) has been implemented in the recent releases of SAS, S-PLUS, STATA and other statistical software packages, which is likely to increase its popularity. This article complements the work of Lee et al. and Wei et al. in several key aspects. First, it offers additional modeling capabilities by allowing separate baseline hazard functions among different strata and imposing the same baseline hazard function within each stratum. Second, it provides a rigorous asymptotic theory for the estimation of the regression parameters, filling several important gaps in the existing proofs for the special cases of models (1) and (2). Third, it establishes the asymptotic properties of the Aalen–Breslow type estimators for the cumulative baseline hazard functions and develops the corresponding inference procedures. We hope that these new results will facilitate further research and applications of statistical models and methods for analyzing multivariate failure time data.

In Section 2 we focused on the estimation of the baseline survival function. Often, one is interested in estimating the survival function associated with a given set of covariate values, say $\mathbf{z}_0$. This is particularly useful in predicting

the survival experience for future patients. Note that the survival function associated with $\mathbf{z}_0$ is equal to the baseline survival function if the covariates are centered at $\mathbf{z}_0$. Thus it does not lose any generality by concentrating on the estimation of the baseline survival function, though simple modifications can be explicitly made to the formulas in Section 2 to estimate covariate-specific survival functions.

The resampling method developed in Section 2.6 may also be used to compare $\Lambda_{0j}(t)$ and $\Lambda_{0k}(t)$ $(j \neq k)$ or the corresponding survival functions over the time interval $[t_1, t_2]$ $(0 \leq t_1 < t_2 \leq \tau)$. Specifically, one can construct confidence bands for the difference $\phi\{\Lambda_{0j}(\cdot)\} - \phi\{\Lambda_{0k}(\cdot)\}$ by considering the process $W_j^*(\cdot) - W_k^*(\cdot)$, whose distribution can again be approximated by simulation. As a by-product, one can generate Kolmogorov–Smirnov type statistics to test the equality of $\Lambda_{0j}(\cdot)$ and $\Lambda_{0k}(\cdot)$. These tests are consistent against the general alternative $\Lambda_{0j}(t) \neq \Lambda_{0k}(t)$ for any $t \in [t_1, t_2]$.

The desired asymptotic results were derived under a reasonable set of conditions. These results also hold under other sets of conditions. In particular, one may relax the iid assumption on the $(T_i, C_i, \mathbf{Z}_i)$'s merely by assuming independence of the $(T_i, C_i, \mathbf{Z}_i)$'s. Without the iid assumption, however, additional regularity conditions and more complicated notation would be required. Specifically, we would need to redefine $\mathbf{s}_k^{(r)}(\boldsymbol{\beta}, t) = \lim_{n\to\infty} n^{-1} \sum_{i=1}^n \sum_{l=1}^L \mathcal{E}\{Y_{ikl}(t) e^{\boldsymbol{\beta}^T \mathbf{Z}_{ikl}(t)} \mathbf{Z}_{ikl}(t)^{\otimes r}\}$ $(k = 1, \ldots, K; r = 0, 1, 2)$, $\mathbf{B} = \lim_{n\to\infty} n^{-1} \sum_{i=1}^n \mathcal{E}(\mathbf{w}_{i\cdot\cdot}^{\otimes 2})$, and $\xi_{jk}(t, s) = \lim_{n\to\infty} n^{-1} \sum_{i=1}^n \mathcal{E}\{\Psi_{ij}(s)\Psi_{ik}(t)\}$ $(1 \leq j, k \leq K; s, t \in [0, \tau])$, and assume that these limits exist.

Cai and Prentice (1995) explored alternative methods for estimating $\boldsymbol{\beta}_0$ under model (2), and Liang et al. (1993) did so for model (1). The latter authors also suggested an estimator for Λ_0, but did not investigate its properties. In the one-sample case with dependent failure time observations, Ying and Wei (1994) showed that the Kaplan–Meier estimator remains consistent and asymptotically normal.

In some applications, it is of scientific interest to assess the strength of dependency among related failure times with adjustment for the effects of covariates. One promising approach, as alluded by Bandeen-Roche and Liang (1996), is to characterize the dependency with the association parameter θ of the Clayton (1978) model while formulating the marginal distributions with model (1), (2), or (3). The estimation theory for $\boldsymbol{\beta}_0$ and Λ_{0k} $(k = 1, \ldots, K)$ developed in this article will be useful in studying inference procedures for θ under this formulation.

APPENDIX A: SOME USEFUL LEMMAS

Lemma A.1. If f_n $(n = 1, 2, \ldots)$ is a sequence of random functions on $[0, \tau]$ that satisfies

$$\int_0^\tau |df_n(u)| = O_p(1)$$

and

$$\|f_n(t)\| = o_p(1), \tag{A.1}$$

then for $k = 1, \ldots, K$ and $l = 1, \ldots, L$, $\|n^{-1/2} \int_0^t f_n(u)\, dM_{\cdot kl}(u)\| \to^p 0$.

Proof. Fix k and l, and let $\varepsilon > 0$ and $\delta > 0$. Because $n^{-1/2} M_{\cdot kl}(t)$ is a martingale with respect to the filtration $\vee_{i=1}^n \mathcal{F}_{t,ikl}$, standard martingale arguments yield

$$\|M_{\cdot kl}(t)\| = O_p(n^{1/2}). \tag{A.2}$$

It then follows from integration by parts that

$$\left\| n^{-1/2} \int_0^t f_n(u)\, dM_{\cdot kl}(u) \right\| \leq \left\| n^{-1/2} \int_0^t M_{\cdot kl}(u)\, df_n(u) \right\| + o_p(1). \tag{A.3}$$

By (A.1) and (A.2), there exists a constant, say η, such that

$$\limsup_n \Pr\left\{ \int_0^\tau |df_n(u)| \vee \|n^{-1/2} M_{\cdot kl}(t)\| > \eta \right\} < \delta. \tag{A.4}$$

The fact that $n^{-1/2} M_{\cdot kl}(t)$ is a martingale with respect to the marginal filtration implies that there exists a finite set of points $0 = t_0 < \cdots < t_m = \tau$ such that

$$\limsup_n \Pr\left\{ \|n^{-1/2} M_{\cdot kl}(t) - M_n(t)\| > \frac{\varepsilon}{2\eta} \right\} < \delta, \tag{A.5}$$

where $M_n(t) = n^{-1/2} \sum_{j=1}^m M_{\cdot kl}(t_{j-1}) 1(t_{j-1} \leq t < t_j)$ (Pollard 1984, p. 180). The first term on the right side of (A.3) is

$$\begin{aligned} &\left\| \int_0^t \{n^{-1/2} M_{\cdot kl}(u) - M_n(u)\}\, df_n(u) + \int_0^t M_n(u)\, df_n(u) \right\| \\ &\quad \leq \|n^{-1/2} M_{\cdot kl}(t) - M_n(t)\| \int_0^\tau |df_n(u)| \\ &\quad\quad + 2m \|n^{-1/2} M_{\cdot kl}(t)\| \|f_n(t)\|. \end{aligned} \tag{A.6}$$

By (A.1),

$$\limsup_n \Pr\left\{ \|f_n(t)\| > \frac{\varepsilon}{4m\eta} \right\} = 0. \tag{A.7}$$

Thus, by (A.4), (A.5), (A.7), and the fact that δ is arbitrary, the right side of (A.6) converges in probability to 0. This completes the proof.

Lemma A.2. Let f_n and f be as described in Theorem 2. Then for $1 \leq j, k \leq K$ and $1 \leq l, m \leq L$,

$$\left\| n^{-1} \sum_{i=1}^n \left\{ \int_0^t f_n(u)\, d\hat{M}_{ikl}(u) \int_0^s f_n(v)\, d\hat{M}_{ijm}(v) - \int_0^t f(u)\, dM_{ikl}(u) \int_0^s f(v)\, dM_{ijm}(v) \right\} \right\|_2 \xrightarrow{p} 0, \tag{A.8}$$

where $\|f(s,t)\|_2 = \sup_{0 \leq s,t \leq \tau} |f(s,t)|$.

Proof. Fix k, j, l, and m. Then

$$\left\| n^{-1} \sum_{i=1}^n \int_0^t f_n(u)\, d\hat{M}_{ikl}(u) \times \left\{ \int_0^s f_n(v)\, d\hat{M}_{ijm}(v) - \int_0^s f(v)\, dM_{ijm}(v) \right\} \right\|_2$$

$$\leq \left\| n^{-1} \sum_{i=1}^{n} J_{ikl}(t) \int_0^s \{f_n(v) - f(v)\}\, dN_{ijm}(v) \right\|_2$$

$$+ \left\| \int_0^s \bar{J}(v,t;\hat{\boldsymbol{\beta}}) f_n(v)\, d\hat{\Lambda}_{0j}(v;\hat{\boldsymbol{\beta}}) \right.$$

$$\left. - \int_0^s \bar{J}(v,t;\boldsymbol{\beta}_0) f(v) \lambda_{0j}(v)\, dv \right\|_2, \qquad \text{(A.9)}$$

where $J_{ikl}(t) = \int_0^t f_n(u)\, d\hat{M}_{ikl}(u)$ and $\bar{J}(v,t;\boldsymbol{\beta}) = n^{-1} \sum_{i=1}^n J_{ikl}(t) Y_{ijm}(v) e^{\boldsymbol{\beta}^T \mathbf{Z}_{ijm}(v)}$. By conditions b, e, and f,

$$\max_i \|J_{ikl}(t)\| = O_p(1). \qquad \text{(A.10)}$$

Thus, by the consistency of $\hat{\boldsymbol{\beta}}$ and condition b, $\|\bar{J}(s,t;\hat{\boldsymbol{\beta}}) - \bar{J}(s,t;\boldsymbol{\beta}_0)\|_2 = o_p(1)$. A minor extension of Theorem 2 can then be used to show that the second term on the right side of (A.9) converges to 0. The condition $\|f_n(t) - f(t)\| \xrightarrow{p} 0$, together with (A.10), implies that the first term on the right side of (A.9) is also asymptotically negligible. Thus (A.9) $\xrightarrow{p}$ 0. Likewise, $\|n^{-1} \sum_{i=1}^n \int_0^s f(v)\, d\hat{M}_{ijm}(v) \{\int_0^t f_n(u)\, d\hat{M}_{ikl}(u) - \int_0^t f(u)\, dM_{ikl}(u)\}\|_2 = o_p(1)$. Hence (A.8) holds.

Lemma A.3. Let $X_1, \ldots, X_n$ be a sequence of real-valued random variables, let $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$ be a sequence of random p vectors, and let h_n: $R^{p+1} \to R$ be a sequence of random functions such that $\{X_i, \mathbf{Y}_i, h_n(X_i, \mathbf{Y}_i)\}$ $(i = 1, \ldots, n)$ are measurable with respect to the increasing σ field $\mathcal{X}_n$. Also assume that $H_n \equiv \{n^{-1} \sum_{i=1}^n h_n(X_i, \mathbf{Y}_i)^2\}^{1/2} = o_p(1)$. Let $G_1, \ldots, G_n$ be independent standard normal random variables that are independent of $\mathcal{X}_\infty$. Then

$$\left\| n^{-1/2} \sum_{i=1}^{n} h_n(X_i, \mathbf{Y}_i) 1(X_i \leq t) G_i \right\| = o_p(1). \qquad \text{(A.11)}$$

If $H_n = O_p(1)$, then so is (A.11).

Proof. Let $X_{[1]} \leq \cdots \leq X_{[n]}$ be the order statistics of the X_i, and let $\mathbf{Y}_{[i]} = \mathbf{Y}_j$ if $X_{[i]} = X_j$. Because the G_i are independent of $\mathcal{X}_\infty$, we have

$$\Pr\left\{ \left\| n^{-1/2} \sum_{i=1}^{n} h_n(X_i, \mathbf{Y}_i) 1(X_i \leq t) G_i \right\| > \varepsilon \right\}$$

$$\leq \Pr\left\{ \max_{1 \leq k \leq n} \left| n^{-1/2} \sum_{i=1}^{k} h_n(X_{[i]}, \mathbf{Y}_{[i]}) G_i \right| > \varepsilon \right\}$$

$$\leq 4 \Pr\left\{ n^{-1/2} \left| \sum_{i=1}^{n} h_n(X_{[i]}, \mathbf{Y}_{[i]}) G_i \right| > \varepsilon \right\},$$

where the last inequality follows from the reflection principle (Shiryayev 1984, lem. 1, p. 372) after conditioning on $\mathcal{X}_n$. A simple characteristic function argument can then be used to show that $n^{-1/2} \sum_{i=1}^n h_n(X_{[i]}, \mathbf{Y}_{[i]}) G_i \xrightarrow{p} 0$ if $H_n \xrightarrow{p} 0$. For the case with $H_n = O_p(1)$, note that $\Pr\{n^{-1/2} | \sum_{i=1}^n h_n(X_{[i]}, \mathbf{Y}_{[i]}) G_i| > \eta^2\} = \mathcal{E}\{\Pr(H_n |G_1| > \eta^2 | \mathcal{X}_n)\} \leq \Pr(H_n > \eta) + \Pr(|G_1| > \eta)$, which can be made arbitrarily small for large η.

APPENDIX B: PROOFS OF LEMMA 1 AND THEOREMS 2, 3, AND 4

Proof of Lemma 1

Note that $n^{-1}\{l(\boldsymbol{\beta}) - l(\boldsymbol{\beta}_0)\}$ equals

$$\sum_{k=1}^{K} \int_0^\tau \left[(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^T \mathbf{S}_k^{(1)}(\boldsymbol{\beta}_0, u) \right.$$

$$\left. - \log\left\{ \frac{S_k^{(0)}(\boldsymbol{\beta}, u)}{S_k^{(0)}(\boldsymbol{\beta}_0, u)} \right\} S_k^{(0)}(\boldsymbol{\beta}_0, u) \right] \lambda_{0k}(u)\, du$$

$$+ n^{-1} \sum_{i=1}^{n} \sum_{k=1}^{K} \sum_{l=1}^{L} \int_0^\tau \left[(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^T \mathbf{Z}_{ikl}(u) \right.$$

$$\left. - \log\left\{ \frac{S_k^{(0)}(\boldsymbol{\beta}, u)}{S_k^{(0)}(\boldsymbol{\beta}_0, u)} \right\} \right] dM_{ikl}(u).$$

By the weak law of large numbers and Lemma A.1, the second term in the display converges to 0 in probability. In view of conditions e and f, the first term converges in probability to

$$\sum_{k=1}^{K} \int_0^\tau \left[(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^T \mathbf{s}_k^{(1)}(\boldsymbol{\beta}_0, u) - \log\left\{ \frac{s_k^{(0)}(\boldsymbol{\beta}, u)}{s_k^{(0)}(\boldsymbol{\beta}_0, u)} \right\} s_k^{(0)}(\boldsymbol{\beta}_0, u) \right]$$

$$\times \lambda_{0k}(u)\, du.$$

It then follows from the proof of lemma 3.1 of Andersen and Gill (1982) that $\hat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}_0$.

Proof of Theorem 2

Fix k. The hypotheses on f_n and conditions b, e, and f imply that

$$\int_0^\tau |d\{f_n(u)/S_k^{(0)}(\boldsymbol{\beta}_0, u)\}| = O_p(n^{1/2}),$$

$$\|f_n(t)/S_k^{(0)}(\boldsymbol{\beta}_0, t)\| = O_p(1). \qquad \text{(B.1)}$$

Due to condition (i),

$$\int_0^t f_n(u)\, d\hat{\Lambda}_{0k}(u;\boldsymbol{\beta}_0) - \int_0^t f_n(u) \lambda_{0k}(u)\, du$$

$$= n^{-1} \int_0^t \{f_n(u)/S_k^{(0)}(\boldsymbol{\beta}_0, u)\}\, dM_{\cdot k \cdot}(u) + o_p(1), \qquad \text{(B.2)}$$

which, by (B.1) and Lemma A.1, converges in probability to 0 uniformly over $[0, \tau]$. Conditions e and f entail that $\|S_k^{(0)}(\boldsymbol{\beta}^*, t)^{-1} - S_k^{(0)}(\boldsymbol{\beta}_0, t)^{-1}\| = o_p(1)$. Thus

$$\left\| \int_0^t f_n(u)\, d\hat{\Lambda}_{0k}(u;\boldsymbol{\beta}^*) - \int_0^t f_n(u)\, d\hat{\Lambda}_{0k}(u;\boldsymbol{\beta}_0) \right\| \xrightarrow{p} 0,$$

which, combined with the convergence to 0 of (B.2), completes the proof of (11). The triangle inequality, (11), and condition d yield (12).

Proof of Theorem 3

Clearly,

$$n^{1/2}\{\hat{\Lambda}_{0k}(t;\hat{\boldsymbol{\beta}}) - \Lambda_{0k}(t)\}$$

$$= n^{1/2}\{\hat{\Lambda}_{0k}(t;\boldsymbol{\beta}_0) - \Lambda_{0k}(t)\}$$

$$+ n^{1/2}\{\hat{\Lambda}_{0k}(t;\hat{\boldsymbol{\beta}}) - \hat{\Lambda}_{0k}(t;\boldsymbol{\beta}_0)\}. \qquad \text{(B.3)}$$

By Lemma A.1 and the arguments given in the proof of Theorem 2, the first term on the right side of (B.3) can be written as

$$n^{1/2}\{\hat{\Lambda}_{0k}(t;\boldsymbol{\beta}_0) - \Lambda_{0k}(t)\} = n^{-1/2}\int_0^t \frac{dM_{\cdot k\cdot}(u)}{s_k^{(0)}(\boldsymbol{\beta}_0,u)} + o_p(1), \quad \text{(B.4)}$$

where, throughout this proof, $o_p(1)$ is uniform in $t \le \tau$. By the Taylor series expansion, the second term on the right side of (B.3) becomes

$$n^{1/2}\{\hat{\Lambda}_{0k}(t;\hat{\boldsymbol{\beta}}) - \hat{\Lambda}_{0k}(t;\boldsymbol{\beta}_0)\} = \mathbf{H}_k(\boldsymbol{\beta}^*,t)^T n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0), \quad \text{(B.5)}$$

where $\boldsymbol{\beta}^*$ is on the line segment between $\hat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}_0$. By Lemma 1, Theorem 2, and condition e, $\mathbf{H}_k(\boldsymbol{\beta}^*,t) = \mathbf{h}_k(t) + o_p(1)$. It then follows from (B.3)–(B.5) and (8)–(10) that

$$n^{1/2}\{\hat{\Lambda}_{0k}(t;\hat{\boldsymbol{\beta}}) - \Lambda_{0k}(t)\} = n^{-1/2}\sum_{i=1}^n \Psi_{ik}(t) + o_p(1),$$

which is essentially a sum of n iid random variables. Thus, by the multivariate central limit theorem, the finite-dimensional distributions of $\mathbf{W}(t)$ are asymptotically the same as those of $\boldsymbol{\mathcal{W}}(t)$.

We now show the tightness of $\mathbf{W}$. For each k and l, define $Q_{1,kl}(t) = \int_0^t s_k^{(0)}(\boldsymbol{\beta}_0,u)^{-1} d\{n^{-1/2}M_{\cdot kl}(u)\}$, and $Q_{2,k}(t) = \mathbf{h}_k(t)^T n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$. Obviously, $n^{1/2}\{\hat{\Lambda}_{0k}(t;\hat{\boldsymbol{\beta}}) - \Lambda_{0k}(t)\} = Q_{1,k\cdot}(t) + Q_{2,k}(t) + o_p(1)$. Because $\mathcal{D}[0,\tau]^K$ has been defined using the uniform metric, the tightness of $\mathbf{W}$ will follow from the tightness of $Q_{1,kl}$ and $Q_{2,k}$ $(k = 1,\ldots,K; l = 1,\ldots,L)$. By conditions d and f, $Q_{1,kl}$ is a square-integrable martingale with respect to the filtration $\vee_{i=1}^n \mathcal{F}_{t,ikl}$. The tightness of $Q_{1,kl}$ thus follows from standard martingale proofs (Pollard 1984, thm. VIII.13). The tightness of $Q_{2,k}(\cdot)$ follows easily from condition f and Corollary 1.

Proof of Theorem 4

Define $\tilde{\mathbf{W}}(t) = \{\tilde{W}_1(t),\ldots,\tilde{W}_K(t)\}^T$, where $\tilde{W}_k(t) = n^{-1/2}\sum_{i=1}^n \Psi_{ik}(t)G_i$ $(k = 1,\ldots,K)$. By the proof of Theorem 3, $n^{-1/2}\{\Psi_{\cdot 1}(t),\ldots,\Psi_{\cdot K}(t)\}^T$ converges weakly to $\boldsymbol{\mathcal{W}}(t)$ unconditionally. Thus, by the conditional multiplier central limit theorem (van der Vaart and Wellner 1996, thm. 2.9.6), $\tilde{\mathbf{W}}$ converges weakly in probability to $\boldsymbol{\mathcal{W}}$ conditional on the data. To complete the proof, it suffices to show that $\|\tilde{W}_k(t) - \hat{W}_k(t)\| \xrightarrow{P} 0$, where the convergence is not conditional on the data.

Note that $\|\tilde{W}_k(t) - \hat{W}_k(t)\|$ is bounded above by

$$\begin{aligned}
&\sum_{l=1}^L \left\| n^{-1/2}\sum_{i=1}^n \{S_k^{(0)}(\hat{\boldsymbol{\beta}},X_{ikl})^{-1} - s_k^{(0)}(\boldsymbol{\beta}_0,X_{ikl})^{-1}\} \right.\\
&\qquad \left. \times\ \Delta_{ikl}1(X_{ikl}\le t)G_i \right\|\\
&+ \sum_{l=1}^L \left\| \int_0^t \left\{ n^{-1/2}\sum_{i=1}^n e^{\boldsymbol{\beta}_0^T\mathbf{Z}_{ikl}(u)}Y_{ikl}(u)G_i \right\} \right.\\
&\qquad \left. \times \left\{ \frac{d\hat{\Lambda}_{0k}(u;\hat{\boldsymbol{\beta}})}{S_k^{(0)}(\hat{\boldsymbol{\beta}},u)} - \frac{d\Lambda_{0k}(u)}{s_k^{(0)}(\boldsymbol{\beta}_0,u)} \right\} \right\|\\
&+ \sum_{l=1}^L \left\| n^{-1/2}\sum_{i=1}^n \{e^{\hat{\boldsymbol{\beta}}^T\mathbf{Z}_{ikl}(t)} - e^{\boldsymbol{\beta}_0^T\mathbf{Z}_{ikl}(t)}\}Y_{ikl}(t)G_i \right\|\\
&\qquad \times \int_0^\tau S_k^{(0)}(\hat{\boldsymbol{\beta}},u)^{-1}\, d\hat{\Lambda}_{0k}(u;\hat{\boldsymbol{\beta}})\\
&+ \left\| \{\mathbf{H}_k(\hat{\boldsymbol{\beta}},t)^T\hat{\mathbf{A}}(\hat{\boldsymbol{\beta}})^{-1} - \mathbf{h}_k(t)^T\mathbf{A}^{-1}\} n^{-1/2}\sum_{i=1}^n \hat{\mathbf{w}}_{i\cdot\cdot}G_i \right\|\\
&+ \left\| \mathbf{h}_k(t)^T\mathbf{A}^{-1}n^{-1/2}\sum_{i=1}^n (\hat{\mathbf{w}}_{i\cdot\cdot} - \mathbf{w}_{i\cdot\cdot})G_i \right\|. \quad \text{(B.6)}
\end{aligned}$$

The first term of (B.6) converges to 0 in probability by Lemma A.3 and the fact that

$$\|S_k^{(0)}(\hat{\boldsymbol{\beta}},t)^{-1} - s_k^{(0)}(\boldsymbol{\beta}_0,t)^{-1}\| \xrightarrow{P} 0. \quad \text{(B.7)}$$

The second term of (B.6) converges to 0 by Theorem 2, Lemma A.3, and (B.7). The normed factor in the third term converges to 0 by Lemmas A.3 and 1. This fact, together with conditions e and f and Theorem 2, implies the asymptotic negligibility of the third term. The asymptotic negligibility of the fourth term follows from condition c, the uniform convergence of $\mathbf{H}_k(\hat{\boldsymbol{\beta}},t)$ and $\hat{\mathbf{A}}(\hat{\boldsymbol{\beta}})$, and the fact that $n^{-1/2}\sum_{i=1}^n \hat{\mathbf{w}}_{i\cdot\cdot}G_i$ converges in distribution. The last term goes to 0 because $n^{-1/2}\sum_{i=1}^n(\hat{\mathbf{w}}_{i\cdot\cdot} - \mathbf{w}_{i\cdot\cdot})G_i \xrightarrow{P} 0$. This completes our proof.

[Received April 1997. Revised March 1998.]

REFERENCES

Andersen, P. K., Borgan, Ø., Gill, R. D., and Keiding, N. (1993), *Statistical Models Based on Counting Processes*, New York: Springer-Verlag.

Andersen, P. K., and Gill, R. D. (1982), "Cox's Regression Model for Counting Processes: A Large-Sample Study," *The Annals of Statistics*, 10, 1100–1120.

Bandeen-Roche, K. J., and Liang, K.-Y. (1996), "Modelling 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.

Clayton, D. G. (1978), "A Model for Association in Bivariate Life Tables and Its Applications in Epidemiological Studies of Familial Tendency in Chronic Disease Incidence," *Biometrika*, 65, 141–151.

Harty, F. J., and Ogston, R. (1992), *Concise Illustrated Dental Dictionary*, Oxford, U.K.: Wright.

Hougaard, P. (1986), "A Class of Multivariate Failure Time Distributions," *Biometrika*, 73, 671–678.

Hujoel, P. P., Loe, H., Anerud, A., Boysen, H., and Leroux, B. G. (1998), "Forty-Five Year Tooth Survival Probabilities Among Men in Oslo, Norway," *Journal of Dental Research*, in press.

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

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 *Survival Analysis: State of the Art*, eds. J. P. Klein and P. K. Goel, Dordrecht: Kluwer Academic, pp. 237–247.

Liang, K.-Y., Self, S. G., and Chang, Y.-C. (1993), "Modelling Marginal Hazards in Multivariate Failure-Time Data." *Journal of the Royal Statistical Society*, 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., Fleming, T. R., and Wei, L. J. (1994), "Confidence Bands for Survival Curves Under the Proportional Hazards Model," *Biometrika*, 81, 73–81.

McGuire, M. K., and Nunn, M. E. (1996), "Prognosis Versus Actual Outcome III: The Effectiveness of Clinical Parameters in Accurately Predicting Tooth Survival," *Journal of Periodontology*, 67, 666–674.

Pollard, D. (1984), *Convergence of Stochastic Processes*, New York: Springer-Verlag.

Shiryayev, A. N. (1984), *Probability*, New York: Springer-Verlag.
Spiekerman, C. F. (1995), "Regression Methods for Correlated Survival Data," unpublished Ph.D. dissertation, University of Washington.
van der Vaart, A., and Wellner, J. A. (1996), *Weak Convergence and Empirical Processes*, New York: Springer-Verlag.
Wei, L. J., Lin, D. Y., and Weissfeld, L. (1989), "Regression Analysis of Multivariate Incomplete Failure Time Data by Modeling Marginal Distributions," *Journal of the American Statistical Association*, 84, 1065–1073.
Ying, Z., and Wei, L. J. (1994), "The Kaplan–Meier Estimate for Dependent Failure Time Observations," *Journal of Multivariate Analysis*, 50, 17–29.