Biometrika (2004), **91**, 4, *pp.* 801–818

Printed in Great Britain

Additive hazards model with multivariate failure time data

By GUOSHENG YIN

Department of Biostatistics, The University of Texas M. D. Anderson Cancer Center, Houston, Texas 77030, U.S.A.

gyin@odin.mdacc.tmc.edu

AND JIANWEN CAI

Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-7420, U.S.A.

cai@bios.unc.edu

Summary

Marginal additive hazards models are considered for multivariate survival data in which individuals may experience events of several types and there may also be correlation between individuals. Estimators are proposed for the parameters of such models and for the baseline hazard functions. The estimators of the regression coefficients are shown asymptotically to follow a multivariate normal distribution with a sandwich-type covariance matrix that can be consistently estimated. The estimated baseline and subject-specific cumulative hazard processes are shown to converge weakly to a zero-mean Gaussian random field. The weak convergence properties for the corresponding survival processes are established. A resampling technique is proposed for constructing simultaneous confidence bands for the survival curve of a specific subject. The methodology is extended to a multivariate version of a class of partly parametric additive hazards model. Simulation studies are conducted to assess finite sample properties, and the method is illustrated with an application to development of coronary heart diseases and cardiovascular accidents in the Framingham Heart Study.

Some key words: Censoring; Confidence band; Correlated survival data; Counting process; Estimating equation; Semiparametric; Survival function.

1. Introduction

The additive hazards model, in which covariate effects are expressed through hazard differences rather than hazard ratios as in Cox's (1972, 1975) proportional hazards model, has often been suggested; see for example Breslow & Day (1980, pp. 53–9; 1987, pp. 122–31) and Cox & Oakes (1984, pp. 73–4). O'Neill (1986) has shown that use of the proportional hazards model can result in serious bias when the additive hazards model is correct.

For independent survival data subject to right-censoring, semiparametric estimation of the additive hazards model when the baseline hazard function is unspecified has been studied by many authors. Lin & Ying (1994) derived large-sample theory paralleling the martingale approach developed by Andersen & Gill (1982) for Cox's model. The additive hazards model has been applied to interval censored data by Lin et al. (1998) and Martinussen & Scheike (2002), to measurement error problems by Kulich & Lin (2000), to frailty models by Lin & Ying (1997) and to cumulative incidence rates by Shen & Chen (1999). Extensions of the additive hazards model which allow time-varying coefficients have been proposed by several authors. Huffer & McKeague (1991) studied weighted least squares estimation for a nonparametric additive risk model first proposed by Aalen (1980, 1989). McKeague & Sasieni (1994) developed a partly parametric additive hazards model that includes both time-dependent and constant regression coefficients. Klein & Moeschberger (1997, Ch. 10) summarised this work; see also Scheike (2002).

All this work has assumed mutual independence of the survival times. Correlated or clustered survival data are often analysed by frailty models, described for example in Hougaard (2000), or by marginal models, reviewed by Lin (1994). Marginal models for events of different types occurring to the same subject will usually involve different baseline hazards for each type of event; see Wei et al. (1989). In contrast, Lee et al. (1992) discussed marginal models for highly stratified, i.e. clustered, data with events of the same type, so that a single baseline hazard function is appropriate. In this paper, we formulate and analyse a marginal additive hazards model for survival data which include both clustering of individuals and events of several types. Our work complements that of Spiekerman & Lin (1998) and Clegg et al. (1999), who developed marginal proportional hazards models for data with the same structure. We also discuss an extension of McKeague & Sasieni's (1994) partly parametric model to this problem. We apply our methods to data from the Framingham Heart Study. Our analysis concerns two types of event, coronary heart disease and cerebrovascular accident, and allows for clustering of events among siblings.

2. THE ADDITIVE HAZARDS MODEL AND INFERENCE PROCEDURES

Let T_{ikl} $(l=1,\ldots,L;\ k=1,\ldots,K;\ i=1,\ldots,n)$ be the failure time for failure type k of subject l in cluster i, and let $Z_{ikl}(t)$ be the $p\times 1$ bounded and possibly external time-dependent covariate vector. Correspondingly, let C_{ikl} be the censoring time, and let $X_{ikl}=\min(T_{ikl}, C_{ikl})$ be the observed time. The censoring indicator is $\Delta_{ikl}=I(T_{ikl}\leqslant C_{ikl})$, where $I(.)$ is the indicator function. For technical reasons, we let each cluster potentially have the same number of subjects, that is L and K are fixed, while we allow the cluster sizes to change by setting $C_{ikl}=0$ whenever T_{ikl} is missing. For some constant τ, $\{T_i, C_i, Z_i(t); t\in[0,\tau]\}$ are assumed to be independent and identically distributed for $i=1,\ldots,n$, where $T_i=\{(T_{i11},\ldots,T_{i1L}),\ldots,(T_{iK1},\ldots,T_{iKL})\}'$, and C_i and $Z_i(t)$ are defined similarly. Assume that T_i and C_i are conditionally independent given $Z_i(t)$.

Let $\lambda_{ikl}(t; Z_{ikl})$ denote the marginal hazard for the failure time for failure type k of subject l in cluster i. We propose the following additive hazards model,

$$\lambda_{ikl}(t; Z_{ikl})=\lambda_{0k}(t)+\beta_{0k}'Z_{ikl}(t), \tag{2·1}$$

where the prime denotes transpose, $\lambda_{0k}(t)$ is the unknown and unspecified baseline hazard function for failure type k, and β_{0k} is the $p\times 1$ regression coefficient vector. The baseline cumulative hazard function for failure type k is $\Lambda_{0k}(t)=\int_0^t \lambda_{0k}(u)du$. When the cluster size is one, that is $L=1$, model (2·1) reduces to a distinct baseline hazards model,

$$\lambda_{ik}(t; Z_{ik})=\lambda_{0k}(t)+\beta_{0k}'Z_{ik}(t),$$

and when there is only one event type, that is $K=1$, to a common baseline hazards model,

$$\lambda_{il}(t; Z_{il})=\lambda_{0}(t)+\beta_{0}'Z_{il}(t).$$

The counting process is denoted by $N_{ikl}(t)=I(X_{ikl}\leqslant t, \Delta_{ikl}=1)$, the at-risk process by $Y_{ikl}(t)=I(X_{ikl}\geqslant t)$ and the marginal filtration by

$$\mathscr{F}_{ikl}(t)=\sigma\{(N_{ikl}(u), Y_{ikl}(u+), Z_{ikl}(u+)), 0\leqslant u\leqslant t\}.$$

By the Doob–Meyer decomposition (Fleming & Harrington, 1991, pp. 31–42),

$$N_{ikl}(t)=M_{ikl}(t)+\int_0^t Y_{ikl}(u)\lambda_{ikl}(u; Z_{ikl})du, \tag{2·2}$$

where $M_{ikl}(t)$ is a local square-integrable martingale with respect to $\mathscr{F}_{ikl}(t)$. As a result of the underlying correlation, $M_{ikl}(t)$ is not a martingale with respect to the joint filtration generated by all the failure, censoring and covariate information up to time t.

It follows from (2·2) that $dM_{ikl}(t)=dN_{ikl}(t)-Y_{ikl}(t)d\Lambda_{0k}(t)-Y_{ikl}(t)\beta_{0k}'Z_{ikl}(t)dt$. Let $\hat{\beta}_k$ denote the estimator of the true regression parameter β_{0k}. Under the working independence assumption, the baseline cumulative hazard function for the kth failure type can be estimated by

$$\hat{\Lambda}_{0k}(t;\hat{\beta}_k)=\int_0^t \frac{\sum_{i=1}^n\sum_{l=1}^L\{dN_{ikl}(u)-Y_{ikl}(u)\hat{\beta}_k'Z_{ikl}(u)du\}}{\sum_{i=1}^n\sum_{l=1}^L Y_{ikl}(u)}. \tag{2·3}$$

Observe that $\sum_{i=1}^n\sum_{l=1}^L\int_0^\tau Z_{ikl}(t)dM_{ikl}(t)$ is the sum of martingale integrals and therefore has mean zero. A natural estimating function for β_{0k} is

$$U_k(\beta)=\sum_{i=1}^n\sum_{l=1}^L\int_0^\tau Z_{ikl}(t)\{dN_{ikl}(t)-Y_{ikl}(t)\beta'Z_{ikl}(t)dt-Y_{ikl}(t)d\hat{\Lambda}_{0k}(t;\beta)\}.$$

If we substitute (2·3) into the above estimating function, then, after some algebra, we have

$$U_k(\beta)=\sum_{i=1}^n\sum_{l=1}^L\int_0^\tau\{Z_{ikl}(t)-\bar{Z}_k(t)\}\{dN_{ikl}(t)-Y_{ikl}(t)\beta'Z_{ikl}(t)dt\}, \tag{2·4}$$

where

$$\bar{Z}_k(t)=\frac{\sum_{i=1}^n\sum_{l=1}^L Y_{ikl}(t)Z_{ikl}(t)}{\sum_{i=1}^n\sum_{l=1}^L Y_{ikl}(t)}.$$

Setting $U_k(\beta)=0$ and solving for β, we obtain

$$\hat{\beta}_k=\left[\sum_{i=1}^n\sum_{l=1}^L\int_0^\tau Y_{ikl}(t)\{Z_{ikl}(t)-\bar{Z}_k(t)\}^{\otimes 2}dt\right]^{-1}\left[\sum_{i=1}^n\sum_{l=1}^L\int_0^\tau\{Z_{ikl}(t)-\bar{Z}_k(t)\}dN_{ikl}(t)\right],$$

where $a^{\otimes 2}=aa'$. Define $A_k=n^{-1}\sum_{i=1}^n\sum_{l=1}^L\int_0^\tau Y_{ikl}(t)\{Z_{ikl}(t)-\bar{Z}_k(t)\}^{\otimes 2}dt$. Then the matrix A_k converges in probability to a nonsingular deterministic matrix denoted by $\mathscr{A}_k$. Assume that $n^{-1}\sum_{i=1}^n\sum_{l=1}^L Y_{ikl}(t)$ uniformly converges to $\pi_k(t)$, and $n^{-1}\sum_{i=1}^n\sum_{l=1}^L Y_{ikl}(t)Z_{ikl}(t)$ uniformly converges to $\tilde{z}_k(t)$ for $t\in[0,\tau]$. Simple algebraic manipulation yields

$$U_k(\beta_{0k})=\sum_{i=1}^n\sum_{l=1}^L\int_0^\tau\{Z_{ikl}(t)-\bar{Z}_k(t)\}dM_{ikl}(t).$$

It is shown in Appendix 1 that $n^{-\frac{1}{2}}U_k(\beta_{0k})$ is asymptotically equivalent to

$$n^{-\frac{1}{2}}\sum_{i=1}^n\mathscr{U}_{ik}(\beta_{0k}),$$

where

$$\mathscr{U}_{ik}(\beta_{0k}) = \sum_{l=1}^{L} \int_0^\tau \left\{ Z_{ikl}(t) - \frac{\tilde{z}_k(t)}{\pi_k(t)} \right\} dM_{ikl}(t). \qquad (2\cdot5)$$

Therefore, $n^{-\frac{1}{2}} U_k(\beta_{0k})$ is essentially a sum of independent and identically distributed random variables, which entails that $n^{-\frac{1}{2}} \{U_1(\beta_{01}), \ldots, U_K(\beta_{0K})\}'$ asymptotically follows a $(p \times K)$-variate normal distribution, by the multivariate central limit theorem. The asymptotic properties of the regression coefficient estimates are given in the following theorem.

THEOREM 1. *Under the regularity conditions given in Appendix* 1, *as* $n \to \infty$,

$$n^{\frac{1}{2}} \{(\hat{\beta}_1' - \beta_{01}'), \ldots, (\hat{\beta}_K' - \beta_{0K}')\}'$$

converges in distribution to a zero-mean $(p \times K)$*-dimensional normal random vector. For* $j, k = 1, \ldots, K$, *the variance-covariance matrix between* $n^{\frac{1}{2}}(\hat{\beta}_j - \beta_{0j})$ *and* $n^{\frac{1}{2}}(\hat{\beta}_k - \beta_{0k})$ *is* $D_{jk}(\beta_{0j}, \beta_{0k}) = \mathscr{A}_j^{-1} E\{\mathscr{U}_{1j}(\beta_{0j}) \mathscr{U}_{1k}'(\beta_{0k})\} \mathscr{A}_k^{-1}$.

A consistent estimator of the covariance matrix is given by

$$\hat{D}_{jk}(\hat{\beta}_j, \hat{\beta}_k) = A_j^{-1} \left\{ n^{-1} \sum_{i=1}^{n} \hat{\mathscr{U}}_{ij}(\hat{\beta}_j) \hat{\mathscr{U}}_{ik}'(\hat{\beta}_k) \right\} A_k^{-1},$$

where $\hat{\mathscr{U}}_{ik}(\hat{\beta}_k) = \sum_{l=1}^{L} \int_0^\tau \{Z_{ikl}(t) - \bar{Z}_k(t)\} d\hat{M}_{ikl}(t)$ and

$$\hat{M}_{ikl}(t) = N_{ikl}(t) - \int_0^t Y_{ikl}(u) \hat{\beta}_k' Z_{ikl}(u) du - \int_0^t Y_{ikl}(u) d\hat{\Lambda}_{0k}(u; \hat{\beta}_k).$$

Based on model (2·1), for a specific subject with the covariate vector $z_0(t)$, the cumulative hazard function can be estimated by $\hat{\Lambda}_k(t; \hat{\beta}_k, z_0) = \hat{\Lambda}_{0k}(t; \hat{\beta}_k) + \int_0^t \hat{\beta}_k' z_0(u) du$, and the survival function by $\hat{S}_k(t; z_0) = \exp\{-\hat{\Lambda}_k(t; \hat{\beta}_k, z_0)\}$. For ease of exposition, we suppress the argument $\hat{\beta}_k$ in the sequel unless it is necessary to state it explicitly. The baseline cumulative hazard estimator in (2·3) might not be nondecreasing in t. To ensure monotonicity, we make a minor modification, which still preserves the asymptotic properties, that is $\hat{\Lambda}_{0k}^*(t) = \max_{s \leqslant t} \hat{\Lambda}_{0k}(s)$ and $\hat{S}_k^*(t; z_0) = \min_{s \leqslant t} \hat{S}_k(s; z_0)$, for $k = 1, \ldots, K$. Following similar arguments to those in Lin & Ying (1994), we can show that $\hat{\Lambda}_{0k}^*(t)$ and $\hat{\Lambda}_{0k}(t)$ are asymptotically equivalent in the sense that $\hat{\Lambda}_{0k}^*(t) - \hat{\Lambda}_{0k}(t) = o_p(n^{-\frac{1}{2}})$. A similar modification is applied to the upper and lower simultaneous confidence bands constructed in the next section.

3. CONFIDENCE BANDS FOR SURVIVAL CURVES

We now consider prediction of the survival curve for a specific pattern of covariates. We define the baseline stochastic processes jointly across all K failure types as

$$W_n(t) = n^{\frac{1}{2}} [\{\hat{\Lambda}_{01}(t) - \Lambda_{01}(t)\}, \ldots, \{\hat{\Lambda}_{0K}(t) - \Lambda_{0K}(t)\}]'.$$

Let $\mathscr{D}[0, \tau]^K$ be a metric space consisting of right-continuous functions $\{f_1(t), \ldots, f_K(t)\}'$ with left-side limits, where $f_k(t): [0, \tau] \to \mathscr{R}$ for $k = 1, \ldots, K$. The metric is defined as $d(f, g) = \max_{t \in [0,\tau]} \{|f_k(t) - g_k(t)|, 1 \leqslant k \leqslant K\}$ for $f, g \in \mathscr{D}[0, \tau]^K$.

Theorem 2. *As $n \to \infty$, $W_n(t)$ converges weakly to a zero-mean Gaussian random field $\mathscr{W}(t)$ in $\mathscr{D}[0, \tau]^K$, where $\mathscr{W}(t) = \{\mathscr{W}_1(t), \ldots, \mathscr{W}_k(t)\}'$. The covariance function between $\mathscr{W}_j(s)$ and $\mathscr{W}_k(t)$ is $\xi_{jk}(s, t) = E\{\Psi_{1j}(s)\Psi_{1k}(t)\}$, for $j, k = 1, \ldots, K$, where*

$$\Psi_{ik}(t) = \int_0^t \frac{dM_{ik.}(u)}{\pi_k(u)} - \mathscr{C}_k'(t)\mathscr{A}_k^{-1} \sum_{l=1}^{L} \int_0^\tau \left\{ Z_{ikl}(t) - \frac{\tilde{z}_k(t)}{\pi_k(t)} \right\} dM_{ikl}(t) \tag{3·1}$$

and $\mathscr{C}_k(t) = \int_0^t \{\tilde{z}_k(u)/\pi_k(u)\}du$.

The dot notation means summation over the corresponding subindex; for example, $M_{ik.}(t) = \sum_{l=1}^{L} M_{ikl}(t)$ and $Y_{.k.}(t) = \sum_{i=1}^{n}\sum_{l=1}^{L} Y_{ikl}(t)$. A consistent estimator of the covariance matrix is $\hat{\xi}_{jk}(s, t) = n^{-1}\sum_{i=1}^{n}\{\hat{\Psi}_{ij}(s)\hat{\Psi}_{ik}(t)\}$, where $\hat{\Psi}_{ik}(t)$ is defined by replacing $\pi_k(t)$, $\tilde{z}_k(t)/\pi_k(t)$, $\mathscr{C}_k(t)$, $\mathscr{A}_k$ and $M_{ikl}(t)$ with their corresponding empirical estimators $Y_{.k.}(t)/n$, $\bar{Z}_k(t)$, $C_k(t) = \int_0^t \bar{Z}_k(u)du$, A_k and $\hat{M}_{ikl}(t)$ in (3·1).

The proof of Theorem 2 is based on verifying the finite-dimensional distribution convergence and the tightness condition (Billingsley, 1999) as outlined in Appendix 2. For a given covariate vector $z_0(t)$, the subject-specific stochastic processes are defined as $W_n(t; z_0) = n^{\frac{1}{2}}[\{\hat{\Lambda}_1(t; z_0) - \Lambda_1(t; z_0)\}, \ldots, \{\hat{\Lambda}_K(t; z_0) - \Lambda_K(t; z_0)\}]'$.

Theorem 3. *As $n \to \infty$, $W_n(t; z_0)$ converges weakly to a zero-mean Gaussian random field $\mathscr{W}(t; z_0)$ in $\mathscr{D}[0, \tau]^K$, where $\mathscr{W}(t; z_0) = \{\mathscr{W}_1(t; z_0), \ldots, \mathscr{W}_K(t; z_0)\}'$. The covariance function between $\mathscr{W}_j(s; z_0)$ and $\mathscr{W}_k(t; z_0)$ is $\xi_{jk}(s, t; z_0) = E\{\Psi_{1j}(s; z_0)\Psi_{1k}(t; z_0)\}$, for $j, k = 1, \ldots, K$, where*

$$\Psi_{ik}(t; z_0) = \int_0^t \frac{dM_{ik.}(u)}{\pi_k(u)} + \mathscr{G}_k'(t)\mathscr{A}_k^{-1} \sum_{l=1}^{L} \int_0^\tau \left\{ Z_{ikl}(t) - \frac{\tilde{z}_k(t)}{\pi_k(t)} \right\} dM_{ikl}(t), \tag{3·2}$$

and $\mathscr{G}_k(t; z_0) = \int_0^t \{z_0(u) - \tilde{z}_k(u)/\pi_k(u)\}du$.

The proof is omitted as it is similar to that of Theorem 2. The covariance function can be consistently estimated by its empirical counterpart

$$\hat{\xi}_{jk}(s, t; z_0) = n^{-1} \sum_{i=1}^{n} \{\hat{\Psi}_{ij}(s; z_0)\hat{\Psi}_{ik}(t; z_0)\},$$

where $\hat{\Psi}_{ik}(t; z_0)$ is defined by replacing $\pi_k(t)$, $\tilde{z}_k(t)/\pi_k(t)$, $\mathscr{G}_k(t; z_0)$, $\mathscr{A}_k$ and $M_{ikl}(t)$ with $Y_{.k.}(t)/n$, $\bar{Z}_k(t)$, $G_k(t; z_0) = \int_0^t \{z_0(u) - \bar{Z}_k(u)\}du$, A_k and $\hat{M}_{ikl}(t)$ in (3·2). By the functional delta method, the joint survival processes for all K failure types,

$$n^{\frac{1}{2}}[\{\hat{S}_1(t; z_0) - S_1(t; z_0)\}, \ldots, \{\hat{S}_K(t; z_0) - S_K(t; z_0)\}]',$$

converge weakly to a zero-mean Gaussian random field in $\mathscr{D}[0, \tau]^K$. A consistent estimator of the covariance function is given by $\hat{S}_j(s; z_0)\hat{S}_k(t; z_0)\hat{\xi}_{jk}(s, t; z_0)$.

The pointwise confidence interval for the survival function at each fixed time point t can be constructed using the asymptotic properties of $\hat{S}_k(t; z_0)$. Construction of simultaneous confidence bands for all t involves the distribution of functionals of the limiting distribution of $W_n(t; z_0)$, which does not have an independent increment structure. We use a simulation technique of Spiekerman & Lin (1998) to approximate the distribution of $W_n(t; z_0)$. Define $\hat{W}_n(t; z_0) = \{\hat{W}_1(t; z_0), \ldots, \hat{W}_K(t; z_0)\}'$, where $\hat{W}_k(t; z_0) = n^{-\frac{1}{2}}\sum_{i=1}^{n}\hat{\Psi}_{ik}(t; z_0)Q_i$, and the Q_i's are generated independently from the standard normal distribution $N(0, 1)$. Conditional on the observed data, the only random variable in $\widehat{W}_k(t; z_0)$ is Q_i such that $\widehat{W}_k(t; z_0)$ is the sum of the independent normal random variables. The conditional covariance function of $\widehat{W}_j(s; z_0)$ and $\widehat{W}_k(t; z_0)$ is $n^{-1}\sum_{i=1}^{n}\{\widehat{\Psi}_{ij}(s; z_0)\widehat{\Psi}_{ik}(t; z_0)\}$. Therefore, given the data, $\widehat{W}_n(t; z_0)$ has the same limiting distribution as $W_n(t; z_0)$. We generate a large number of random samples of $(Q_1, \ldots, Q_n)$ from $N(0, 1)$, while fixing the data at their observed values. The next theorem justifies the resampling method.

THEOREM 4. *Conditional on the observed data*

$$\{(N_{ikl}(t), Y_{ikl}(t), Z_{ikl}(t)); t \in [0, \tau]; i = 1, \ldots, n; k = 1, \ldots, K; l = 1, \ldots, L\},$$

$\widehat{W}_n(t; z_0)$ *converges weakly to the same zero-mean Gaussian random field* $\mathscr{W}(t; z_0)$ *as* $W_n(t; z_0)$ *in* $\mathscr{D}[0, \tau]^K$.

The proof is outlined in Appendix 3. To construct the confidence band for $S_k(t; z_0)$, we start by defining a class of transformed processes for failure type k, namely

$$B(t; z_0) = n^{\frac{1}{2}} g(t; z_0)[\phi\{\widehat{\Lambda}_k(t; z_0)\} - \phi\{\Lambda_k(t; z_0)\}],$$

where $g(.)$ is a known weight function which converges uniformly to a nonnegative bounded function on $[t_1, t_2]$ $(0 \leqslant t_1 \leqslant t_2 \leqslant \tau)$, and $\phi(.)$ is a known transformation function with a nonzero and continuous derivative $\phi'(.)$, such as $\phi(.) = \log(.)$ or $\phi(.) = \log\{-\log(.)\}$. The logarithmic transformation can restrict the confidence band to the range $[0, 1]$, and can improve the coverage probabilities in small samples (Kalbfleisch & Prentice, 2002). The functional delta method yields that $B(t; z_0)$ is asymptotically equivalent to $g(t; z_0)\phi'\{\widehat{\Lambda}_k(t; z_0)\}W_k(t; z_0)$.

The $(1-\alpha)$ confidence band for $\phi\{\Lambda_k(t; z_0)\}$ is then given by

$$\phi\{\widehat{\Lambda}_k(t; z_0)\} \mp n^{-\frac{1}{2}} q_\alpha / g(t; z_0), \tag{3·3}$$

where q_α is defined by $\mathrm{pr}\{\sup_{i,l(t_1 \leqslant X_{ikl} \leqslant t_2)}|B(X_{ikl}; z_0)| > q_\alpha\} = \alpha$. The critical constant q_α can be obtained by choosing the $(1-\alpha)$th quantile from the large number of copies of $\{\sup_{i,l(t_1 \leqslant X_{ikl} \leqslant t_2)}|B(X_{ikl}; z_0)|\}$. Appropriately chosen weight functions may narrow the width of the confidence band at the time range of interest. Let $\hat{\sigma}_k^2(t; z_0)$ be the variance function estimator at time t of the process $W_k(t; z_0)$, that is $\hat{\sigma}_k^2(t; z_0) = \hat{\xi}_{kk}(t, t; z_0)$ for $k = 1, \ldots, K$. The equal-precision band (Nair, 1984) defines the weight function to be $g(t; z_0) = \widehat{\Lambda}_k(t; z_0)/\hat{\sigma}_k(t; z_0)$, while the Hall–Wellner band (Hall & Wellner, 1980) requires that $g(t; z_0) = \widehat{\Lambda}_k(t; z_0)/\{1 + \hat{\sigma}_k^2(t; z_0)\}$. The valid range of confidence bands is usually restricted to the first and last uncensored observations. The range might be further restricted to $[t_1, t_2]$ (Nair, 1984; Chen & Ying, 1996), because of the unstable estimation in the tails. For $b = 1$ and $b = 2$, t_b can be obtained by solving $c_b = \hat{\sigma}^2(t_b; z_0)/\{1 + \hat{\sigma}^2(t_b; z_0)\}$, where (c_1, c_2) may be prespecified to be (0·1, 0·9).

4. PARTLY PARAMETRIC ADDITIVE HAZARDS MODELS

We extend the methods to the partly parametric additive hazards model, which McKeague & Sasieni (1994) studied for independent failure time data. When the effect of some covariates, for example R_{ikl}, may vary over time, a $q \times 1$ vector of time-varying coefficients $\alpha_{0k}(t)$ is introduced in the model, so that

$$\lambda_{ikl}(t; R_{ikl}, Z_{ikl}) = \alpha_{0k}(t)' R_{ikl} + \beta_{0k}' Z_{ikl}. \tag{4·1}$$

For the kth failure type, the loglikelihood function based on (4·1) is given by

$$l_k(\beta, \alpha) = \sum_{i=1}^{n} \sum_{l=1}^{L} \left[\Delta_{ikl} \log \{\alpha(X_{ikl})' R_{ikl} + \beta' Z_{ikl}\} - \int_0^\tau Y_{ikl}(t)\{\alpha(t)' R_{ikl} + \beta' Z_{ikl}\} dt \right].$$

The derivative with respect to β is

$$\frac{\partial l_k(\beta, \alpha)}{\partial \beta} = \sum_{i=1}^{n} \sum_{l=1}^{L} \left\{ \frac{\Delta_{ikl} Z_{ikl}}{\lambda_{ikl}(X_{ikl})} - \int_0^\tau \frac{Y_{ikl}(t) Z_{ikl} Z'_{ikl} \beta}{\lambda_{ikl}(t)} dt - \int_0^\tau \frac{Y_{ikl}(t) Z_{ikl} R'_{ikl} \alpha(t)}{\lambda_{ikl}(t)} dt \right\}.$$

Let $\mathbb{Z}_k = \{Z_{1k1} Y_{1k1}(t), \ldots, Z_{nkL} Y_{nkL}(t)\}'$, let $\mathbb{R}_k$ be defined similarly, let

$$\lambda_k^{-1}(t) = \text{diag}\,\{1/\lambda_{ikl}(t), i = 1, \ldots, n; l = 1, \ldots, L\}$$

be an $(nL \times nL)$ diagonal matrix, and let $N_k(t) = \{N_{1k1}(t), \ldots, N_{nkL}(t)\}'$ be an (nL)-dimensional counting process vector. If we define $\Omega_k(t) = \int_0^t \alpha_k(u) du$ and set $\partial l_k(\beta,\alpha)/\partial\beta = 0$, we have

$$\beta_k = \left\{ \int_0^\tau \mathbb{Z}'_k \lambda_k^{-1}(t) \mathbb{Z}_k dt \right\}^{-1} \left\{ \int_0^\tau \mathbb{Z}'_k \lambda_k^{-1}(t) dN_k(t) - \int_0^\tau \mathbb{Z}'_k \lambda_k^{-1}(t) \mathbb{R}_k d\Omega_k(t) \right\}.$$

For $\alpha(t)$, consider the submodel $\alpha(t) = b_0(t) + \eta b(t)$, where η is a scalar and $b(t)$ is a given $(q \times 1)$ vector of functions. The derivative with respect to η is

$$\frac{\partial l_k(\beta, \eta)}{\partial \eta} = \sum_{i=1}^{n} \sum_{l=1}^{L} \left\{ \frac{\Delta_{ikl} b(X_{ikl})' R_{ikl}}{\lambda_{ikl}(X_{ikl})} - \int_0^\tau \frac{Y_{ikl}(t) b(t)' R_{ikl} R'_{ikl} \alpha(t)}{\lambda_{ikl}(t)} dt - \int_0^\tau \frac{Y_{ikl}(t) b(t)' R_{ikl} Z'_{ikl} \beta}{\lambda_{ikl}(t)} dt \right\}.$$

Since this submodel is a special case of model (4·1), an estimator for (4·1) should work on all submodels. If we solve $\partial l_k(\beta, \eta)/\partial \eta = 0$ for all vector-valued functions $b(t)$, we obtain

$$\Omega_k(t) = \int_0^t \{\mathbb{R}'_k \lambda_k^{-1}(u) \mathbb{R}_k\}^{-1} \{\mathbb{R}'_k \lambda_k^{-1}(u) dN_k(u) - \mathbb{R}'_k \lambda_k^{-1}(u) \mathbb{Z}_k \beta_k du\}.$$

By plugging $\Omega_k(t)$ into the expression for β_k, we finally have

$$\beta_k = \left\{ \int_0^\tau \mathbb{Z}'_k H_k(t) \mathbb{Z}_k dt \right\}^{-1} \int_0^\tau \mathbb{Z}'_k H_k(t) dN_k(t), \tag{4·2}$$

where $H_k(t) = \lambda_k^{-1}(t) - \lambda_k^{-1}(t)\mathbb{R}_k\{\mathbb{R}'_k \lambda_k^{-1}(t) \mathbb{R}_k\}^{-1} \mathbb{R}'_k \lambda_k^{-1}(t)$. As shown in (4·2), β_k resembles a weighted least squares estimator, but β_k still depends on the unknown $\lambda_k^{-1}(t)$. When we replace $\lambda_k^{-1}(t)$ by the identity matrix I, so that $H_k(t) = I - \mathbb{R}_k(\mathbb{R}'_k\mathbb{R}_k)^{-1}\mathbb{R}'_k$, β_k reduces to the ordinary least squares estimator. The estimation algorithm proceeds as follows: first use I instead of $\lambda_k^{-1}(t)$ to obtain the initial values for β_k and $\Omega_k(t)$, denoted by $\beta_k^{(0)}$ and $\Omega_k^{(0)}(t)$; next, based on $\Omega_k^{(0)}(t)$, estimate $\alpha_k(t)$ nonparametrically using the kernel smoothing method (Bowman & Azzalini, 1997); then obtain the estimator $\hat{\lambda}_k^{-1}(t)$ based on (4·1); and finally substitute $\lambda_k^{-1}(t)$ by $\hat{\lambda}_k^{-1}(t)$ to obtain the final estimators $\hat{\beta}_k$ and $\hat{\Omega}_k(t)$.

Noting that $\hat{H}_k(t)$ is orthogonal to $\mathbb{R}_k$, and using (4·1) and the decomposition

$$M_{ikl}(t) = N_{ikl}(t) - \int_0^t Y_{ikl}(u)\lambda_{ikl}(u; R_{ikl}, Z_{ikl})du,$$

we have

$$\int_0^\tau \mathbb{Z}_k' \hat{H}_k(t) dM_k(t) = \int_0^\tau \mathbb{Z}_k' \hat{H}_k(t) dN_k(t) - \left\{ \int_0^\tau \mathbb{Z}_k' \hat{H}_k(t) \mathbb{Z}_k dt \right\} \beta_{0k},$$

and thus

$$n^{\frac{1}{2}}(\hat{\beta}_k - \beta_{0k}) = \left\{ n^{-1} \int_0^\tau \mathbb{Z}_k' \hat{H}_k(t) \mathbb{Z}_k dt \right\}^{-1} n^{-\frac{1}{2}} \int_0^\tau \mathbb{Z}_k' \hat{H}_k(t) dM_k(t),$$

where $M_k(t) = \{M_{1k1}(t), \ldots, M_{nkL}(t)\}'$. By arguments similar to those in Appendix 1, the asymptotic normality of $n^{\frac{1}{2}}(\hat{\beta}_1', \ldots, \hat{\beta}_K')'$ can be proved. Furthermore, by noting that

$$n^{\frac{1}{2}}\{\hat{\Omega}_k(t) - \Omega_{0k}(t)\} = n^{\frac{1}{2}} \int_0^t \{\mathbb{R}_k' \hat{\lambda}_k(u)^{-1} \mathbb{R}_k\}^{-1} \mathbb{R}_k \hat{\lambda}_k(u)^{-1} dM_k(u)$$
$$- \left[\int_0^t \{\mathbb{R}_k' \hat{\lambda}_k(u)^{-1} \mathbb{R}_k\}^{-1} \mathbb{R}_k \hat{\lambda}_k(u)^{-1} \mathbb{Z}_k' du \right] n^{\frac{1}{2}}(\hat{\beta}_k - \beta_{0k}),$$

we can show that $n^{\frac{1}{2}}\{\hat{\Omega}_k(t) - \Omega_{0k}(t)\}$ converges weakly to a zero-mean Gaussian random field based on arguments similar to those in Appendix 2.

The estimation procedures for (2·1) and (4·1) are very different. The former mimics the partial likelihood score function, and the latter involves the nonparametric hazard estimation. A larger sample size would be required to estimate the hazard function reliably in a nonparametric manner.

5. Simulation studies

To investigate the properties of our proposed method with practical sample sizes, we conducted extensive simulation studies. For the first part of the simulations, we chose marginal exponential distributions for the two distinct failure types, $K = 2$, and a constant cluster size of two, $L = 2$, within each failure type. We generated the failure times for the ith cluster, T_{i11}, T_{i12}, T_{i21} and T_{i22} $(i = 1, \ldots, n)$, from the multivariate Clayton–Oakes model (Clayton & Cuzick, 1985; Oakes, 1989). The joint survival function was given by

$$\text{pr}(T_{i11} > t_{i11}, T_{i12} > t_{i12}, T_{i21} > t_{i21}, T_{i22} > t_{i22} | Z_{i11}, Z_{i12}, Z_{i21}, Z_{i22})$$
$$= \left[\sum_{k=1}^{2} \sum_{l=1}^{2} \exp\left\{ \frac{(\lambda_{0k} + \beta' Z_{ikl}) t_{ikl}}{\theta} \right\} - 3 \right]^{-\theta},$$

where $\theta > 0$, and smaller θ induced larger correlation. The parameter θ was preset to be 3·9, 1·2 or 0·31, which corresponded to the within cluster correlation of the failure times of $\rho = 0{\cdot}2$, 0·5 or 0·8 for the first failure type, and $\rho = 0{\cdot}23$, 0·56 or 0·87 for the second failure type, respectively. Different baseline hazards were assumed for the two failure types, namely $\lambda_{01} = 2$ and $\lambda_{02} = 4$. Two covariates were included in the model: one was a binary variable, Z_1, taking the value of 0 or 1 with probability 0·5, and the other was a continuous variable, Z_2, generated independently from Un(0, 5). The true regression coefficients were preset at $\beta_1 = 1$, categorical, and $\beta_2 = 0{\cdot}5$, continuous. The censoring times were generated independently from Un(0, a), with $a = 1{\cdot}1$, 0·43, 0·16 to achieve approximately 25%, 50% and 75% censoring rates for $k = 1$, and 16%, 37% and 65% for $k = 2$. For each simulation configuration, 500 replicated samples were generated, and the number of clusters, n, was 100 and 150. We evaluated the small sample size properties of $\hat{\beta}$ and the coverage properties of the robust equal-precision, EP_R, and Hall–Wellner, HW_R, 95% confidence bands. The robust bands were constructed with adjustment for the intraclass correlation, while the naive method did not take the correlation into account. To illustrate the empirical coverage rates of the simultaneous confidence bands, we set $z_0 = (1, 0)'$ to estimate the survival curve and its bands. We independently generated 1000 simple random samples $(Q_1, \ldots, Q_n)$ from $N(0, 1)$ for obtaining the critical constant q_α in (3·3).

For each of the data realisations, we obtained the pointwise estimates of the regression coefficients and the sandwich-type variance estimators. We calculated the sample standard deviations, SD, the average of the estimated standard errors, SE, and the 95% nominal level coverage rates, CR. As shown in Table 1, both of the point estimators $\hat{\beta}_1$ and $\hat{\beta}_2$ are approximately unbiased and approach the true values as the sample size increases. Comparing the columns 'SD' and 'SE' suggests that the variance estimators provide good estimation of the variability of the regression coefficient estimators. The 95% confidence interval coverage rates are close to the nominal level, which ensures the adequacy of the asymptotic approximations for practical use. The variation of $\hat{\beta}$ becomes smaller with an increased sample size and becomes larger with an increased censoring rate. We also examined the scenarios with $n = 25$ and 50, where similar conclusions were drawn. The equal-precision and Hall–Wellner 95% confidence bands were constructed, and the corresponding simultaneous coverage rates, EP_R and HW_R, were calculated. The empirical coverage rates of the confidence bands were satisfactory and close to 95%. The results provide empirical evidence that the approximation method by resampling many Q_i's from $N(0, 1)$ works well.

We carried out another set of simulations with a single failure type and the cluster size of four, that is $K = 1$ and $L = 4$. The baseline hazard function was chosen to be from a Weibull distribution, for example $\lambda_0(t) = \gamma t$. The failure times of the ith cluster $(T_{i1}, T_{i2}, T_{i3}, T_{i4})$ were generated from the Clayton–Oakes model with

$$\text{pr}(T_{i1} > t_{i1}, T_{i2} > t_{i2}, T_{i3} > t_{i3}, T_{i4} > t_{i4} \mid Z_{i1}, Z_{i2}, Z_{i3}, Z_{i4}) = \left\{ \sum_{l=1}^{4} \exp\left(\frac{\gamma t_{il}^2}{2\theta} + \frac{\beta' Z_{il} t_{il}}{\theta} \right) - 3 \right\}^{-\theta},$$

where we chose $\gamma = 1$. The correlation parameter θ was preset at 3·6, 1 or 0·24 for the within-cluster correlation of $\rho = 0{\cdot}2$, 0·5 or 0·8, respectively. Two covariates were generated independently, of which Z_1 is a categorical variate taking the value of 0 or 1 with probability 0·5, and $Z_2 \sim \text{Un}(0, 2)$. The true regression coefficients were prespecified as $\beta_1 = 0{\cdot}2$ and $\beta_2 = 0{\cdot}7$. The censoring times were generated independently from Un(0, a) with a prespecified to be 3, 1·4 or 0·62 to achieve approximately 25%, 50% or 75%

Table 1: *First simulation study. Estimates and confidence bands for cluster size of* 2 *with* $\beta_1 = 1$ *and* $\beta_2 = 0{\cdot}5$

n	ρ	$c\%$	$\hat\beta_1$	SD	SE	CR	$\hat\beta_2$	SD	SE	CR	EP_R	HW_R
						Failure Type 1						
100	0·2	25	1·024	0·613	0·606	94·6	0·502	0·212	0·208	95·4	96·0	94·6
		50	0·986	0·721	0·744	96·2	0·496	0·260	0·258	95·8	92·2	90·0
		75	1·031	1·090	1·066	95·4	0·480	0·373	0·370	95·6	94·2	93·0
	0·5	25	1·031	0·621	0·607	93·6	0·501	0·215	0·208	94·2	95·4	94·0
		50	0·991	0·730	0·746	94·8	0·497	0·260	0·258	94·6	94·8	92·6
		75	1·019	1·085	1·065	95·8	0·474	0·383	0·369	94·2	92·8	91·6
	0·8	25	1·028	0·611	0·606	94·2	0·506	0·220	0·208	93·6	97·0	94·2
		50	0·977	0·737	0·747	96·4	0·502	0·263	0·260	96·4	94·4	92·8
		75	0·990	1·049	1·061	94·6	0·482	0·382	0·370	93·6	93·8	92·4
150	0·2	25	1·041	0·529	0·496	92·8	0·508	0·189	0·172	92·4	97·0	95·2
		50	1·064	0·643	0·614	93·6	0·511	0·227	0·213	94·0	93·6	91·4
		75	1·037	0·885	0·874	95·0	0·518	0·298	0·304	95·2	96·6	94·0
	0·5	25	1·050	0·514	0·499	94·0	0·514	0·193	0·174	92·6	96·0	95·0
		50	1·074	0·625	0·616	95·2	0·514	0·229	0·214	93·4	94·8	93·8
		75	1·049	0·884	0·878	95·2	0·521	0·295	0·304	95·6	95·2	93·0
	0·8	25	1·057	0·509	0·502	95·4	0·521	0·191	0·176	92·0	96·0	95·2
		50	1·078	0·618	0·619	95·4	0·519	0·233	0·216	94·2	94·2	94·4
		75	1·075	0·866	0·879	94·8	0·540	0·296	0·306	95·8	96·0	94·6
						Failure Type 2						
100	0·23	16	1·011	0·911	0·888	94·4	0·512	0·333	0·308	91·8	95·4	94·2
		37	1·002	1·086	1·028	94·4	0·505	0·377	0·357	93·4	94·2	94·0
		65	1·029	1·417	1·382	96·0	0·541	0·500	0·482	93·4	94·0	92·4
	0·56	16	0·979	0·909	0·883	93·8	0·505	0·336	0·305	92·2	94·6	93·0
		37	1·008	1·052	1·027	95·0	0·500	0·376	0·356	93·4	94·0	92·8
		65	1·014	1·453	1·381	95·6	0·519	0·503	0·481	95·4	95·2	95·0
	0·87	16	0·982	0·871	0·873	95·4	0·496	0·324	0·303	93·0	97·0	94·6
		37	0·986	1·044	1·021	93·8	0·486	0·356	0·354	95·2	96·2	94·6
		65	1·002	1·503	1·381	94·2	0·482	0·499	0·483	95·4	93·2	93·0
150	0·23	16	1·008	0·760	0·730	93·2	0·508	0·267	0·253	95·0	93·8	91·6
		37	0·962	0·878	0·846	94·4	0·508	0·312	0·293	92·4	94·6	92·0
		65	0·916	1·178	1·134	94·4	0·498	0·408	0·393	93·8	94·4	91·0
	0·56	16	1·005	0·739	0·731	93·8	0·505	0·272	0·253	91·6	95·2	92·8
		37	0·977	0·866	0·847	93·8	0·503	0·308	0·294	92·4	94·2	91·6
		65	0·925	1·192	1·139	93·4	0·504	0·414	0·396	95·0	93·8	93·0
	0·87	16	1·015	0·723	0·727	96·2	0·501	0·266	0·252	92·0	97·0	93·4
		37	0·967	0·839	0·847	95·2	0·496	0·296	0·294	93·6	96·6	94·0
		65	0·946	1·137	1·143	95·2	0·500	0·434	0·398	92·6	94·4	92·6

SD, standard deviation; SE, average of estimated standard errors; CR, 95% coverage rate; EP_R and HW_R, simultaneous coverage rates of equal-precision and Hall–Wellner bands, respectively.

censoring rates. The number of clusters was $n = 50$ and 100, and 500 simulations were performed for each scenario. We estimated the survival curve for $z_0 = (1, 0{\cdot}5)'$ and constructed the corresponding 95% simultaneous confidence bands. We also constructed the naive bands, EP_N and HW_N, by ignoring the underlying failure time correlation. Observing $E\{M_{ikl}(t)\} = 0$ and var $\{M_{ikl}(t)\} = E\{N_{ikl}(t)\}$, in the naive method we replace $M_{ikl}(t)$ by

$N_{ikl}(t)Q_{ikl}$ in (3·2), where the Q_{ikl}'s were generated independently from $N(0, 1)$ (Lin et al., 1993; 1994). The results are summarised in Table 2, which shows the appropriateness of the asymptotic approximation for the regression parameters and the simulation methods for the construction of the confidence bands with finite sample sizes. A comparison between the robust and naive bands indicates that the naive method performs poorly, especially when the underlying failure time correlation is high.

Table 2: *Second simulation study. Estimates and confidence bands for cluster size of* 4 *with* $\beta_1 = 0{\cdot}2$ *and* $\beta_2 = 0{\cdot}7$

ρ	$c\%$	$\hat\beta_1$	SD	SE	CR	$\hat\beta_2$	SD	SE	CR	EP_R	HW_R	EP_N	HW_N
						$n = 50$							
0·2	0	0·202	0·189	0·178	93·0	0·711	0·180	0·167	92·6	94·8	93·0	94·4	93·8
	25	0·207	0·209	0·197	92·2	0·724	0·192	0·181	93·0	94·8	93·8	94·6	93·6
	50	0·205	0·220	0·219	95·6	0·698	0·198	0·200	94·8	91·2	91·0	93·0	92·6
	75	0·225	0·292	0·271	93·6	0·728	0·249	0·244	93·8	95·4	92·0	92·2	92·0
0·5	0	0·202	0·184	0·174	93·6	0·724	0·193	0·172	91·4	95·8	93·8	93·4	91·6
	25	0·212	0·212	0·198	93·4	0·730	0·199	0·187	92·8	94·4	92·8	93·0	92·2
	50	0·219	0·236	0·219	92·8	0·724	0·224	0·206	93·4	92·4	90·0	91·0	90·0
	75	0·221	0·295	0·272	92·6	0·723	0·261	0·250	93·0	95·6	90·4	92·4	90·6
0·8	0	0·209	0·182	0·171	92·6	0·747	0·204	0·178	91·6	94·6	91·4	85·8	84·0
	25	0·214	0·206	0·196	93·0	0·729	0·206	0·193	93·2	93·2	92·4	88·0	86·6
	50	0·213	0·234	0·219	93·4	0·718	0·220	0·215	95·2	92·2	90·8	87·4	86·2
	75	0·216	0·282	0·271	93·2	0·722	0·269	0·262	94·6	95·8	93·0	89·2	87·2
						$n = 100$							
0·2	0	0·205	0·129	0·129	94·8	0·700	0·124	0·121	94·4	95·8	93·4	95·2	93·2
	25	0·197	0·138	0·141	95·2	0·703	0·139	0·130	93·2	96·6	94·4	96·4	93·2
	50	0·203	0·159	0·156	95·4	0·710	0·149	0·142	93·6	92·6	90·6	93·2	93·0
	75	0·206	0·200	0·194	93·4	0·707	0·185	0·172	93·6	96·0	91·8	94·2	91·8
0·5	0	0·208	0·130	0·128	95·0	0·710	0·132	0·126	91·6	95·2	92·2	92·4	91·2
	25	0·200	0·140	0·141	95·4	0·710	0·146	0·135	93·2	97·6	94·6	93·6	89·4
	50	0·201	0·159	0·156	94·6	0·712	0·153	0·146	94·6	95·6	92·2	92·6	91·8
	75	0·206	0·201	0·194	94·0	0·702	0·183	0·177	94·0	96·0	92·8	92·4	92·0
0·8	0	0·216	0·131	0·128	93·8	0·722	0·138	0·130	93·0	96·2	93·4	86·6	81·8
	25	0·205	0·139	0·140	94·6	0·715	0·146	0·139	94·0	96·2	94·0	88·0	83·6
	50	0·200	0·153	0·156	96·0	0·707	0·155	0·152	94·2	96·6	94·4	89·0	86·6
	75	0·208	0·194	0·194	95·4	0·707	0·186	0·185	94·4	97·0	94·6	90·8	88·4

SD, standard deviation; SE, average of estimated standard errors; CR, 95% coverage rate; EP_R and HW_R are the simultaneous coverage rates of equal-precision and Hall–Wellner bands, and EP_N and HW_N are the naive ones, respectively.

6. Framingham Heart Study

We applied our inference procedures to data from the Framingham Heart Study. The objective of the study was to identify the risk factors or characteristics that contribute to cardiovascular disease by following a large number of disease-free participants, those with no overt symptom and who had not suffered a heart attack or stroke, over a long period of time. The study was initiated in 1948 and the subjects were examined every two years. Multiple failure outcomes were recorded from the same subject, e.g. coronary heart disease and cerebrovascular accident.

We considered the first manifestations of coronary heart disease and cerebrovascular accident as two different events, that is $K=2$. The times to event were recorded in years. In this analysis, we had 1571 individuals, of whom 233 experienced coronary heart disease but not cerebrovascular accident, 34 experienced cerebrovascular accident but not coronary heart disease, and 17 experienced both. There were 113 sibling clusters of size 2, 24 of size 3 and 3 of size 4. For comparison, we also fitted the Cox-type marginal regression model. The analyses based on the additive and multiplicative hazards models are summarised in Table 3. The absolute values of the parameter estimates from the additive hazards model are much smaller than those from the Cox model, which is often the case. The estimates from the additive hazards and Cox models have the same signs, indicating the same directions of the covariate effects, while the *p*-values for the two models differ.

Table 3. *Analysis of data from the Framingham Heart Study under the additive and multiplicative hazards models*

	Additive hazards model					
	Coronary heart disease			Cerebrovascular accident		
Covariate	$\hat{\beta}\times 10^{-3}$	SE $\times 10^{-3}$	*p*-value	$\hat{\beta}\times 10^{-3}$	SE $\times 10^{-3}$	*p*-value
Smoke (yes = 1, no = 0)	2·4562	1·4038	0·080	0·9713	0·6215	0·118
Sex (female = 1, male = 0)	−5·1987	1·4343	<0·001	−0·3027	0·6206	0·626
Body mass index	0·2499	0·0776	0·001	0·0871	0·0428	0·042
Cholesterol	0·0363	0·0060	<0·001	0·0037	0·0030	0·218
Systolic blood pressure	0·1023	0·0321	0·001	0·0204	0·0119	0·086
Diastolic blood pressure	0·0800	0·0455	0·079	0·0378	0·0174	0·029
Waiting time (years)	0·0127	0·0614	0·836	0·0337	0·0199	0·091
	Multiplicative hazards model					
	Coronary heart disease			Cerebrovascular accident		
Covariate	$\hat{\beta}$	SE	*p*-value	$\hat{\beta}$	SE	*p*-value
Smoke (yes = 1, no = 0)	0·3157	0·1414	0·026	0·6412	0·3204	0·045
Sex (female = 1, male = 0)	−0·6203	0·1335	<0·001	−0·1979	0·2822	0·480
Body mass index	0·0308	0·0179	0·085	0·0486	0·0359	0·180
Cholesterol	0·0042	0·0016	0·007	0·0021	0·0033	0·530
Systolic blood pressure	0·0120	0·0072	0·096	0·0114	0·0154	0·460
Diastolic blood pressure	0·0095	0·0122	0·440	0·0246	0·0249	0·320
Waiting time (years)	0·0010	0·0200	0·960	0·0248	0·0458	0·590

SE, standard error.

To illustrate the prediction of the survival probability for a given subject, Fig. 1 shows the estimated survival curves for a male smoker with body mass index of 35 kg/m^2, cholesterol level of 360 mg/dl, systolic blood pressure of 160 mm Hg, diastolic blood pressure of 90 mm Hg and waiting time of 10 years. The pointwise confidence intervals from the additive hazards model are much narrower than those from the Cox model. The equal-precision and Hall–Wellner 95% confidence bands are constructed for the same subject with 10 000 copies of Q_i random samples based on the simulation method. Figure 2 presents the 95% simultaneous confidence bands with the estimated survival curves of coronary heart disease and cerebrovascular accident. The equal-precision band is narrower in the two tails and the Hall–Wellner band is narrower in the middle of the range.

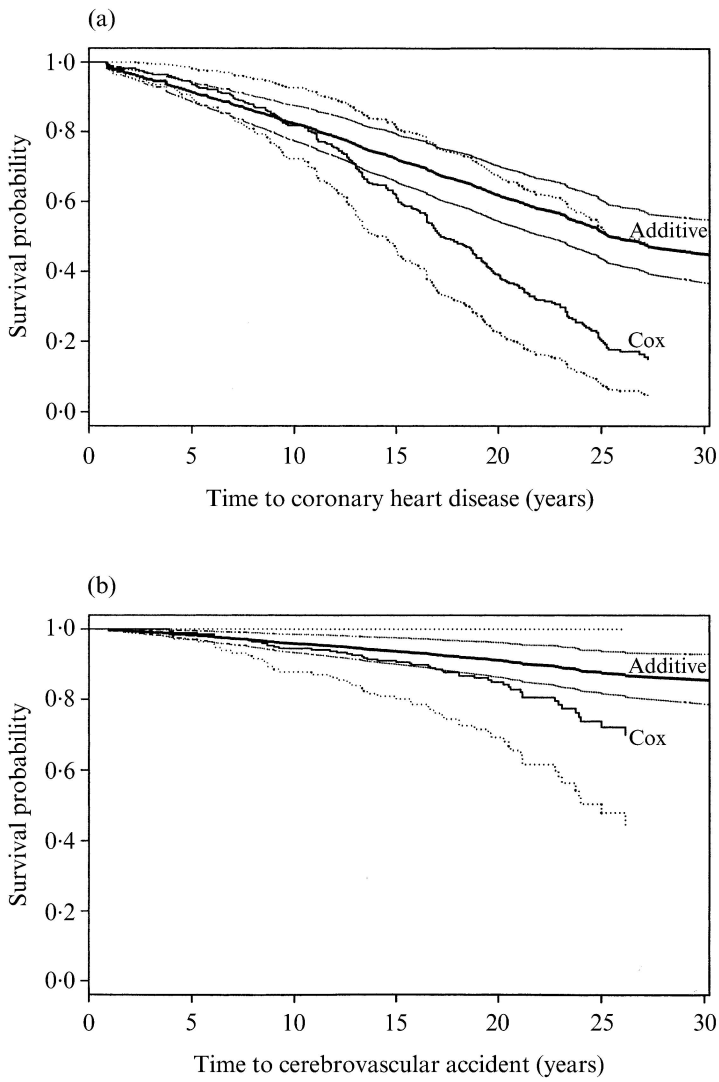

Fig. 1: Framingham Heart Study. Survival curves under the additive hazards (bold solid curve) and Cox (thin solid curve) models, for (a) coronary heart disease and (b) cerebrovascular accident of a subject with covariates smoke = 1, sex = male, body mass index = 35, cholesterol = 360, systolic blood pressure = 160, diastolic blood pressure = 90, waiting time = 10, and the corresponding 95% pointwise confidence intervals (dotted curves).

7. Remarks

Unlike maximisation of the partial likelihood function in the Cox model, solution of the estimating equation (2·4) does not require any iterative numerical procedure: the parameter estimates have a closed form. In principal, efficiency could be gained by incorporating an appropriate weight into (2·4). However, as pointed out by Lin & Ying (1994) for the case of independent data, the efficiency loss is usually very small.

As demonstrated in the simulation study, ignoring the correlation would result in under-coverage of the true survival curve, which becomes severe when there is high correlation.

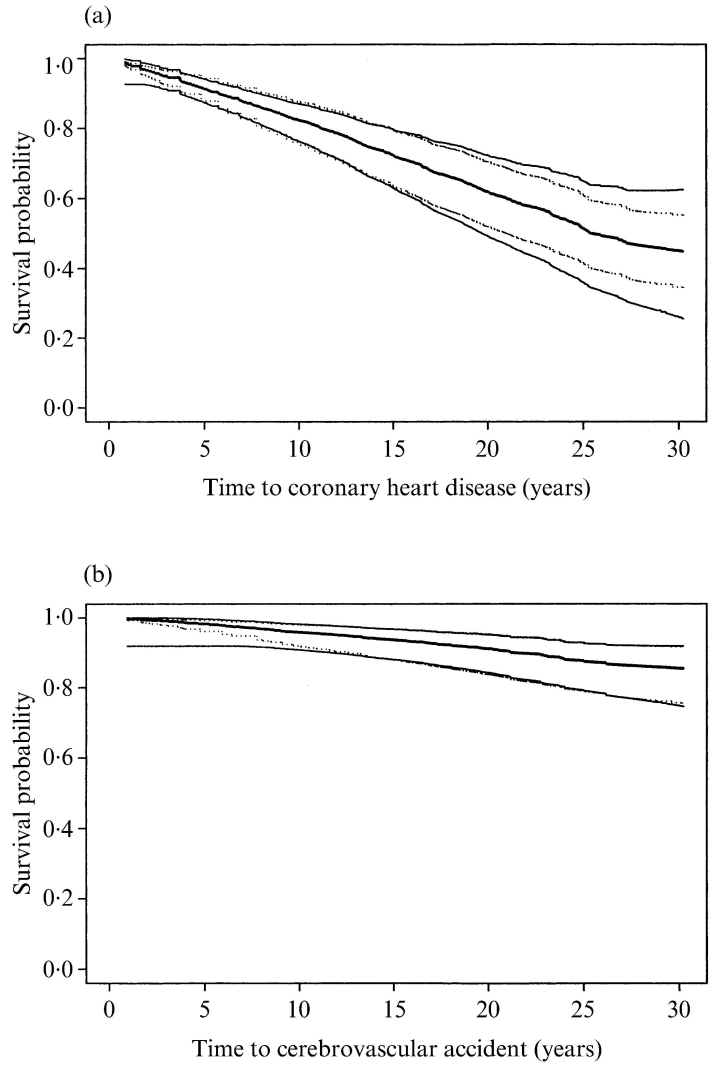

Fig. 2: Framingham Heart Study. Equal-precision (dotted curves) and Hall–Wellner (solid curves) 95% confidence bands for the survival curves for (a) coronary heart disease and (b) cerebrovascular accident, under the additive hazards model. The middle, bold solid curve is the point estimate of the survival curve for the subject with $z_0 = (1, 0, 35, 360, 160, 90, 10)'$.

An alternative to our proposed resampling method could be the bootstrap method using the clusters as the sampling units to preserve the intracluster correlation.

Acknowledgement

We would like to thank the editor, an associate editor and a referee for their critical reading of the manuscript and constructive comments which led to great improvement of the earlier versions. This research was partially supported by a grant from the National Institutes of Health.

Appendix 1

Large-sample distribution of $(\hat\beta_1', \ldots, \hat\beta_K')'$

For $l=1,\ldots,L$, $k=1,\ldots,K$, $i=1,\ldots,n$ and some constant $\tau>0$, we assume the following set of regularity conditions throughout this paper: $\text{pr}\{Y_{ikl}(t)=1, t\in[0,\tau]\}>0$; $\int_0^\tau \lambda_{0k}(t)dt<\infty$ for each k; the covariate vector $Z_{ikl}(t)$ is bounded for $t\in[0,\tau]$; and $\mathscr{A}_k$ is positive definite.

By the functional central limit theorem (Pollard, 1990, Theorem 10.6), it can be shown that $n^{-1/2}\sum_{i=1}^n\sum_{l=1}^L M_{ikl}(t)$ converges in distribution to a zero-mean Gaussian process with continuous sample paths. The strong representation theorem (Pollard, 1990, Theorem 9.4) and Lemma A.3 of Bilias et al. (1997) entail that

$$n^{-1/2}\sum_{i=1}^n\sum_{l=1}^L\int_0^\tau\left\{\bar Z_k(t)-\frac{\tilde z_k(t)}{\pi_k(t)}\right\}dM_{ikl}(t)\to 0, \qquad (\text{A1·1})$$

in probability, as $n\to\infty$. For failure type k, we have

$$\begin{aligned}n^{1/2}(\hat\beta_k-\beta_{0k}) &= n^{-1/2}A_k^{-1}\sum_{i=1}^n\sum_{l=1}^L\left[\int_0^\tau\{Z_{ikl}(t)-\bar Z_k(t)\}dN_{ikl}(t)\right.\\ &\qquad\left.-\int_0^\tau Y_{ikl}(t)\{Z_{ikl}(t)-\bar Z_k(t)\}^{\otimes 2}\beta_{0k}dt\right]\\ &= n^{-1/2}A_k^{-1}\sum_{i=1}^n\sum_{l=1}^L\int_0^\tau\{Z_{ikl}(t)-\bar Z_k(t)\}dM_{ikl}(t).\end{aligned}$$

Under the regularity conditions, if we apply (A1·1), the above quantity can be shown to be asymptotically equivalent to $n^{-1/2}A_k^{-1}\sum_{i=1}^n\mathscr{U}_{ik}(\beta_{0k})$, where $\mathscr{U}_{ik}(\beta_{0k})$ is defined as in (2·5). Note that A_k converges in probability to $\mathscr{A}_k$, and $\mathscr{U}_{ik}(\beta_{0k})$ $(i=1,\ldots,n)$ are independent and identically distributed random vectors. By the multivariate central limit theorem and Slutsky's theorem, Theorem 1 follows.

Appendix 2

Weak convergence properties of $W_n(t)$

With similar arguments to those in Appendix 1, we can show that, as $n\to\infty$,

$$n^{-1/2}\sum_{i=1}^n\int_0^t\left\{\frac{1}{n^{-1}Y_{.k.}(u)}-\frac{1}{\pi_k(u)}\right\}dM_{ik.}(u)\to 0, \qquad (\text{A2·1})$$

in probability. For failure type k, if we plug in (2·3), the baseline cumulative hazard process can be written as

$$\begin{aligned}W_{nk}(t) &= n^{1/2}\{\hat\Lambda_{0k}(t;\hat\beta_k)-\hat\Lambda_{0k}(t;\beta_{0k})\}+n^{1/2}\{\hat\Lambda_{0k}(t;\beta_{0k})-\Lambda_{0k}(t)\}\\ &= n^{1/2}\int_0^t\frac{\sum_{i=1}^n\sum_{l=1}^L Y_{ikl}(u)(\beta_{0k}-\hat\beta_k)'Z_{ikl}(u)du}{\sum_{i=1}^n\sum_{l=1}^L Y_{ikl}(u)}\\ &\quad+n^{1/2}\int_0^t\frac{\sum_{i=1}^n\sum_{l=1}^L\{dN_{ikl}(u)-Y_{ikl}(u)\beta_{0k}'Z_{ikl}(u)du-Y_{ikl}(u)\lambda_{0k}(u)du\}}{\sum_{i=1}^n\sum_{l=1}^L Y_{ikl}(u)}\\ &= n^{-1/2}\sum_{i=1}^n\left[\int_0^t\frac{dM_{ik.}(u)}{n^{-1}Y_{.k.}(u)}-C_k'(t)A_k^{-1}\sum_{l=1}^L\int_0^\tau\{Z_{ikl}(t)-\bar Z_k(t)\}dM_{ikl}(t)\right].\end{aligned}$$

Coupling with (A1·1), it can be shown that $W_{nk}(t)$ is asymptotically equivalent to $n^{-1/2}\sum_{i=1}^n\Psi_{ik}(t)$, where $\{\Psi_{1k}(t),\ldots,\Psi_{nk}(t)\}$ given in (3·1), are independent and identically distributed random variables, for each fixed t. Hence, for any finite number of time points $(t_1,\ldots,t_m)$, it can be shown that the joint distribution of $\{W_{nk}(t_1), \ldots, W_{nk}(t_m)\}'$ is asymptotically normal with mean zero; that is the finite-dimensional distributional convergence follows from the multivariate central limit theorem. Next, we prove the tightness condition of $W_{nk}(t)$ in order to ensure the weak convergence property of the processes. Since the space of $\mathscr{D}[0, \tau]^K$ is equipped with the uniform metric, the tightness of $W_{nk}(t)$ follows from the tightness of $W_{nk}^{(1)}(t) = n^{1/2}\int_0^t dM_{.k.}(u)/Y_{.k.}(u)$ and $W_{nk}^{(2)}(t) = n^{1/2}(\hat{\beta}_k - \beta_{0k})'C_k(t)$. It follows from the weak convergence of $n^{-1/2}M_{.k.}(t)$ and (A2·1) that $W_{nk}^{(1)}(t)$ converges weakly to a zero-mean Gaussian process. Thus, $W_{nk}^{(1)}(t)$ is tight by Theorem 10.2 of Pollard (1990), and the tightness of $W_{nk}^{(2)}(t)$ follows from Theorem 1. Hence, $W_n(t)$ converges to a zero-mean Gaussian random field in $\mathscr{D}[0, \tau]^K$.

APPENDIX 3

Weak convergence properties of $\hat{W}_n(t; z_0)$

Without loss of generality, we now prove that, conditional on the data, $\hat{W}_n(t)$ converges weakly to the same zero-mean Gaussian random field $\mathscr{W}(t)$ as $W_n(t)$ in $\mathscr{D}[0, \tau]^K$. The proof for the weak convergence of $\hat{W}_n(t; z_0)$, follows similar arguments. Note that $\hat{W}_n(t) = \{\hat{W}_{n1}(t), \ldots, \hat{W}_{nK}(t)\}'$, where $\hat{W}_{nk}(t) = n^{-1/2}\sum_{i=1}^n \hat{\Psi}_{ik}(t)Q_i$, for $k = 1, \ldots, K$, and the Q_i's are generated independently from $N(0, 1)$. Define $\tilde{W}_n(t) = \{\tilde{W}_{n1}(t), \ldots, \tilde{W}_{nK}(t)\}'$, where $\tilde{W}_{nk}(t) = n^{-1/2}\sum_{i=1}^n \Psi_{ik}(t)Q_i$. By the proof of Theorem 2, $n^{-1/2}\{\Psi_{.1}(t), \ldots, \Psi_{.K}(t)\}'$ converges weakly to $\mathscr{W}(t)$ unconditionally. Based on the conditional multiplier central limit theorem in van der Vaart & Wellner (1996, Theorem 2.9.6), $\tilde{W}_n(t)$ converges weakly in probability to $\mathscr{W}(t)$ conditional on the data. It suffices to prove that $\|\tilde{W}_{nk}(t) - \hat{W}_{nk}(t)\| \to 0$ in probability, where $\|f(t)\| = \sup_{t \in [0,\tau]}|f(t)|$ for a function $f: [0, \tau] \to \mathscr{R}$.

Now, we consider $\|\tilde{W}_{nk}(t) - \hat{W}_{nk}(t)\| \leqslant \|W_{nk}^{(d,1)}(t)\| + \|W_{nk}^{(d,2)}(t)\|$, where

$$W_{nk}^{(d,1)}(t) = n^{-1/2}\sum_{i=1}^n \sum_{l=1}^L \int_0^t \left\{\frac{Q_i dM_{ikl}(u)}{\pi_k(u)} - \frac{Q_i d\hat{M}_{ikl}(u)}{n^{-1}Y_{.k.}(u)}\right\},$$

$$W_{nk}^{(d,2)}(t) = n^{-1/2}\sum_{i=1}^n Q_i C_k'(t)A_k^{-1}\sum_{l=1}^L \int_0^\tau \{Z_{ikl}(t) - \bar{Z}_k(t)\}d\hat{M}_{ikl}(t)$$

$$- n^{-1/2}\sum_{i=1}^n Q_i \mathscr{C}_k'(t)\mathscr{A}_k^{-1}\sum_{l=1}^L \int_0^\tau \left\{Z_{ikl}(t) - \frac{\tilde{z}_k(t)}{\pi_k(t)}\right\}dM_{ikl}(t).$$

Some algebraic manipulation yields that,

$$\|W_{nk}^{(d,1)}(t)\| \leqslant \left\| n^{-1/2}\sum_{i=1}^n \sum_{l=1}^L Q_i\left\{\frac{1}{\pi_k(X_{ikl})} - \frac{1}{n^{-1}Y_{.k.}(X_{ikl})}\right\}\Delta_{ikl}I(X_{ikl} \leqslant t)\right\|$$
$$+ \left\| n^{-1/2}\sum_{i=1}^n \int_0^t Q_i Y_{ik.}(u)\left\{\frac{d\hat{\Lambda}_{0k}(u)}{n^{-1}Y_{.k.}(u)} - \frac{d\Lambda_{0k}(u)}{\pi_k(u)}\right\}\right\|$$
$$+ \left\| n^{-1/2}\sum_{i=1}^n \sum_{l=1}^L \int_0^t Q_i Y_{ikl}(u)\left\{\frac{\hat{\beta}_k'}{n^{-1}Y_{.k.}(u)} - \frac{\beta_{0k}'}{\pi_k(u)}\right\}Z_{ikl}(u)du\right\|. \quad (A3·1)$$

The first term on the right-hand side of (A3·1) asymptotically converges to zero in probability by Lemma A.3 in Spiekerman & Lin (1998) and the fact that, uniformly,

$$\|n^{-1}Y_{.k.}(t) - \pi_k(t)\| \to 0, \quad (A3·2)$$

in probability. The second term of (A3·1) converges to zero by Lemma A.3 and Theorem 2 in Spiekerman & Lin (1998) and (A3·2). The third term of (A3·1) converges to zero by Lemma A.3 and Theorem 2 in Spiekerman & Lin (1998), Theorem 1 and (A3·2).

Note that $\| W_{nk}^{(d,2)}(t) \|$ is bounded above by

$$\left\| \{C_k'(t)A_k^{-1} - \mathscr{C}_k'(t)\mathscr{A}_k^{-1}\} n^{-1/2} \sum_{i=1}^{n} Q_i \hat{s}_{ik.} \right\| + \left\| \mathscr{C}_k'(t)\mathscr{A}_k^{-1} n^{-1/2} \sum_{i=1}^{n} Q_i(\hat{s}_{ik.} - s_{ik.}) \right\|, \quad \text{(A3·3)}$$

where

$$\hat{s}_{ik.} = \sum_{l=1}^{L} \int_0^\tau \{Z_{ikl}(t) - \bar{Z}_k(t)\} d\hat{M}_{ikl}(t), \quad s_{ik.} = \sum_{l=1}^{L} \int_0^\tau \left\{ Z_{ikl}(t) - \frac{\tilde{z}_k(t)}{\pi_k(t)} \right\} dM_{ikl}(t).$$

By the uniform convergence of $C_k(t)$ to $\mathscr{C}_k(t)$, and A_k to $\mathscr{A}_k$, and the fact that $n^{-1/2}\sum_{i=1}^{n} Q_i \hat{s}_{ik.}$ converges to a normal distribution, the first term in (A3·3) converges to zero. The second term in (A3·3) goes to zero because $n^{-1/2}\sum_{i=1}^{n} Q_i(\hat{s}_{ik.} - s_{ik.}) \to 0$ in probability, and thus the proof is complete.

References

Aalen, O. O. (1980). A model for nonparametric regression analysis of counting processes. In *Lecture Notes in Statistics*, **2**, Ed. W. Klonecki, A. Kosek and J. Rosinski, pp. 1–25. New York: Springer.

Aalen, O. O. (1989). A linear regression model for the analysis of life times. *Statist. Med.* **8**, 907–25.

Andersen, P. K. & Gill, R. D. (1982). Cox's regression model for counting processes: A large-sample study. *Ann. Statist.* **10**, 1100–20.

Bilias, Y., Gu, M. & Ying, Z. (1997). Towards a general asymptotic theory for Cox model with staggered entry. *Ann. Statist.* **25**, 662–82.

Billingsley, P. (1999). *Convergence of Probability Measures*, 2nd ed. New York: Wiley.

Bowman, A. W. & Azzalini, A. (1997). *Applied Smoothing Techniques for Data Analysis, The Kernel Approach with S-Plus Illustrations*. Oxford: Clarendon.

Breslow, N. E. & Day, N. E. (1980). *Statistical Methods in Cancer Research*, **1**, *The Design and Analysis of Case-Control Studies*. Lyon: IARC.

Breslow, N. E. & Day, N. E. (1987). *Statistical Methods in Cancer Reearch*, **2**, *The Design and Analysis of Case-Control Studies*. Lyon: IARC.

Chen, K. & Ying, Z. (1996). A counterexample to a conjecture concerning the Hall–Wellner band. *Ann. Statist.* **24**, 641–6.

Clayton, D. G. & Cuzick, J. (1985). Multivariate generalizations of the proportional hazards model (with Discussion). *J. R. Statist. Soc.* A **148**, 82–117.

Clegg, L. X., Cai, J. & Sen, P. K. (1999). A marginal mixed baseline hazards model for multivariate failure time data. *Biometrics* **55**, 805–12.

Cox, D. R. (1972). Regression models and life tables (with Discussion). *J. R. Statist. Soc.* B **34**, 187–220.

Cox, D. R. (1975). Partial likelihood. *Biometrika* **62**, 269–76.

Cox, D. R. & Oakes, D. (1984). *Analysis of Survival Data*. New York: Chapman.

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

Hall, W. J. & Wellner, J. A. (1980). Confidence bands for a survival curve from censored data. *Biometrika* **67**, 133–43.

Hougaard, P. (2000). *Analysis of Multivariate Survival Data*. New York: Springer.

Huffer, F. W. & McKeague, I. W. (1991). Weighted least squares estimation for Aalen's additive risk model. *J. Am. Statist. Assoc.* **86**, 38–53.

Kalbfleisch, J. D. & Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data*, 2nd ed. New York: Wiley.

Klein, J. P. & Moeschberger, M. (1997). *Survival Analysis: Techniques for Censored and Truncated Data*. New York: Springer.

Kulich, M. & Lin, D. Y. (2000). Additive hazard regression with covariate measurement error. *J. Am. Statist. Assoc.* **95**, 238–48.

Lee, E. W., Wei, L. J. & 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*, Ed. J. P. Klein and P. K. Goel, pp. 237–47. Dordrecht: Kluwer.

Lin, D. Y. (1994). Cox regression analysis of multivariate failure time data: the marginal approach. *Statist. Med.* **13**, 2233–47.

Lin, D. Y. & Ying, Z. (1994). Semiparametric analysis of the additive risk model. *Biometrika* **81**, 61–71.

Lin, D. Y. & Ying, Z. (1997). Additive hazards regression models for survival data. In *Proceedings of the First Seattle Symposium in Biostatistics: Survival Analysis*, Ed. D. Y. Lin and T. R. Fleming, pp. 185–98. New York: Springer.

LIN, D. Y., FLEMING, T. R. & WEI, L. J. (1994). Confidence bands for survival curves under the proportional hazards model. *Biometrika* **81**, 73–81.

LIN, D. Y., OAKES, D. & YING, Z. (1998). Additive hazards regression for current status data. *Biometrika* **85**, 289–98.

LIN, D. Y., WEI, L. J. & YING, Z. (1993). Checking the Cox model with cumulative sum of martingale-based residuals. *Biometrika* **80**, 557–72.

MARTINUSSEN, T. & SCHEIKE, T. H. (2002). Efficient estimation in additive hazards regression with current status data. *Biometrika* **89**, 649–58.

MCKEAGUE, I. W. & SASIENI, P. D. (1994). A partly parametric additive risk model. *Biometrika* **81**, 501–14.

NAIR, V. N. (1984). Confidence bands for survival functions with censored data: a comparative study. *Technometrics* **26**, 265–75.

OAKES, D. (1989). Bivariate survival models induced by frailties. *J. Am. Statist. Assoc.* **84**, 487–93.

O'NEILL, T. J. (1986). Inconsistency of the misspecified proportional hazards model. *Statist. Prob. Lett.* **4**, 219–22.

POLLARD, D. (1990). *Empirical Processes: Theories and Applications.* Hayward, CA: Institute of Mathematical Statistics.

SCHEIKE, T. H. (2002). The additive nonparametric and semiparametric Aalen model as the rate function for a counting process. *Lifetime Data Anal.* **8**, 247–62.

SHEN, Y. & CHENG, S. C. (1999). Confidence bands for cumulative incidence curves under the additive risk model. *Biometrics* **55**, 1093–100.

SPIEKERMAN, C. F. & LIN, D. Y. (1998). Marginal regression models for multivariate failure time data. *J. Am. Statist. Assoc.* **93**, 1164–75.

WEI, L. J., LIN, D. Y. & WEISSFELD, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *J. Am. Statist. Assoc.* **84**, 1065–73.

VAN DER VAART, A. W. & WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics.* New York: Springer-Verlag.

[*Received October* 2002. *Revised February* 2004]