Biometrics 63, 663–672
September 2007
DOI: 10.1111/j.1541-0420.2006.00730.x

Marginal Analysis of Correlated Failure Time Data with Informative Cluster Sizes

Xiuyu J. Cong,[1,*] Guosheng Yin,[2,] and Yu Shen[2,***]**

[1]Department of Biometrics and Data Management, Boehringer Ingelheim Pharmaceuticals, P.O. Box 368, Ridgefield, Connecticut 06877, U.S.A.
[2]Department of Biostatistics, University of Texas M.D. Anderson Cancer Center, Houston, Texas 77030, U.S.A.
[*]*email:* xcong@rdg.boehringer-ingelheim.com
[**]*email:* gsyin@mdanderson.org
[***]*email:* yshen@mdanderson.org

Summary. We consider modeling correlated survival data when cluster sizes may be informative to the outcome of interest based on a within-cluster resampling (WCR) approach and a weighted score function (WSF) method. We derive the large sample properties for the WCR estimators under the Cox proportional hazards model. We establish consistency and asymptotic normality of the regression coefficient estimators, and the weak convergence property of the estimated baseline cumulative hazard function. The WSF method is to incorporate the inverse of cluster sizes as weights in the score function. We conduct simulation studies to assess and compare the finite-sample behaviors of the estimators and apply the proposed methods to a dental study as an illustration.

Key words: Correlated failure time; Multivariate survival analysis; Weighted score function; Within-cluster resampling.

1. Introduction

Correlated survival data often arise in biomedical research settings. For example, in randomized multicenter clinical trials, patients are recruited and grouped by study centers. In dental or family disease studies, all teeth of the same person or all members of the same family are naturally clustered together. Observations within the same cluster are likely to be correlated, where the correlation needs to be accounted for in statistical estimation and inference.

An interesting problem in clustered/correlated survival data, which is often ignored, is the possible informativeness of cluster sizes. Cluster size is informative when the outcome of interest among individuals in a cluster is associated with the size of that cluster. For example, in a toxicology study assessing the effect of mother mice being exposed to certain toxicant, mothers that are particularly susceptible to the toxicant may produce more offspring with birth defects that have lower survival probabilities and meanwhile may experience more fetal resorptions, hence reducing the litter size. In this case, pups of a smaller litter tend to have shorter survival, thus the cluster size is informative to the effect of toxicant exposure on offspring survival. Another example is found in a dental study, where the effects of behavioral factors such as cigarette smoking and hygiene status may predict tooth survival for patients with chronic periodontitis (McGuire and Nunn, 1996). The outcome of interest was the time to tooth loss from initiation of therapy. Shown in Figure 1 are the Kaplan–Meier curves for molar teeth stratified by the cluster sizes (the number of molars of a patient ≤ 6, $=7$, or $=8$). We can see that patients with more teeth have higher molar survival probability. As a result, cluster size is informative to tooth survival.

Marginal models (MMs) have been proposed and widely used for analyzing clustered survival data. In analyzing multivariate failure times where individuals may experience different types of failures, Wei, Lin, and Weissfeld (1989) proposed the marginal proportional hazards model (Cox, 1975) separately for each failure type, while the covariance matrix was estimated jointly across all failure types to adjust for the correlation. For clustered or highly stratified survival data, Lee Wei, and Amato (1992) proposed a multiplicative intensity model, for which they estimated the regression coefficients assuming independence among observations and provided a "sandwich" form of covariance matrix estimator. Spiekerman and Lin (1998) and Clegg, Cai, and Sen (1999) presented a more general marginal regression model for multivariate failure time data. Alternative MMs including the accelerated failure time model, linear transformation model, and additive hazards model have been studied for analyzing correlated survival data, see for example, Lin and Wei (1992), Lee, Wei, and Ying (1993), Chen and Wei (1997), Cai, Wei, and Wilcox (2000), Yin and Cai (2004), and Lu (2005).

However, all the aforementioned methods do not take into account the possible informativeness of cluster sizes. In fact,

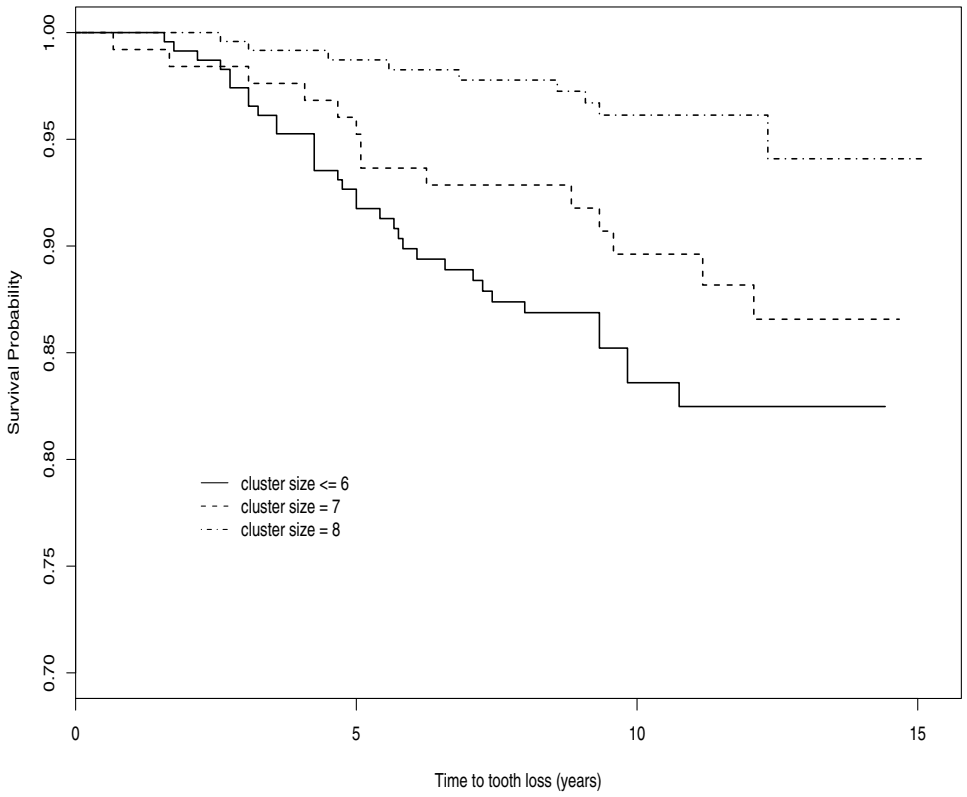

Figure 1. Kaplan–Meier survival curves for molars stratified by cluster sizes in the dental study.

when cluster size is informative, the standard MMs tend to overweigh the large clusters because each individual observation contributes equally in the likelihood function, and therefore produce biased estimates. Hoffman, Sen, and Weinberg (2001) proposed a within-cluster resampling (WCR) procedure, where one observation is randomly sampled from each cluster. The observations in the resampled dataset are thus independent and standard methods can be readily applied. By resampling the observed data with replacement many times, the final estimator can be obtained through averaging over the estimates from the resampled data. Williamson, Datta, and Satten (2003) proposed a modified generalized estimating equation (GEE) method (Liang and Zeger, 1986), where the estimating equation is inversely weighted by cluster sizes. Both methods can adequately account for informative cluster sizes by assigning equal weights to clusters and produce valid inference for the parameters of interest. Follmann, Proschan, and Leifer (2003) established the asymptotic theories and broad applications of the WCR approach, referred as multiple outputation. More recently, Benhin, Rao, and Scott (2005) gave a thorough and in-depth discussion on a mean estimating equation approach that is in essence analogous to the work of Williamson et al. (2003).

The theories of the WCR and weighted estimating equation approaches are well developed in the GEE context. For correlated survival data with informative cluster sizes, the problem is much more complicated and challenging due to the existence of random censoring and the infinite dimensionality of the unknown hazard function. In this article, we investigate the WCR method for correlated survival data with informative cluster sizes, such that the resampled independent data can be analyzed using the conventional Cox model. We also generalize the standard MM of Lee et al. (1992) by incorporating the inverse of cluster sizes as weights into the score function to account for informative cluster sizes.

The rest of the article is organized as follows. In Section 2, we introduce the WCR method under the Cox proportional hazards model and derive the large sample properties for the estimators of the regression coefficients and the baseline cumulative hazard function. We also study the weighted score function (WSF) for correlated survival data to adjust for informative cluster sizes. In Section 3, we conduct simulation studies to assess the finite sample properties of the proposed methods. In Section 4, we apply the two proposed methods to a dental study and compare the results with those from the standard MMs. We provide some discussion in Section 5, and outline the technical proofs in the Web Appendices.

2. Proposed Methods

2.1 *Notation*

Let $i = 1, \ldots, m$ index the clusters which are assumed to be independent of each other, and $j = 1, \ldots, n_i$ denote the individuals within the ith cluster. Let T_{ij} and C_{ij} be the failure and censoring times for the jth individual in the ith cluster,

respectively. Let $\mathbf{Z}_{ij}(t)$ denote a p-vector of possibly time-dependent covariates, where $t \in [0, \tau]$ for some finite constant $\tau > 0$. We assume that T_{ij} is conditionally independent of C_{ij} given $\mathbf{Z}_{ij}(t)$. The observed times are $X_{ij} = \min(T_{ij}, C_{ij})$, with the failure indicator $\Delta_{ij} = I(X_{ij} = T_{ij})$, where $I(\cdot)$ is the indicator function. Within each cluster, we assume exchangeability among individuals.

We assume that the cluster size is informative to survival probability. In other words, the survival probabilities of individuals in a cluster depend on the size of that cluster. However, the causes for cluster sizes being informative can be complicated and usually unknown, because some latent variables may implicitly affect the baseline hazard for each cluster and/or the covariates. For example, the marginal hazard function may be associated with the cluster size through the following frailty model:

$$\lambda(t \mid \mathbf{Z}_{ij}, w_i) = \lambda_0(t) w_i \exp\{\boldsymbol{\gamma}_0' \mathbf{Z}_{ij}(t)\},$$

where the latent variable w_i depends on cluster size, and $\lambda_0(t)$ is an unspecified baseline hazard function.

If cluster sizes are ignorable (noninformative to survival), the usual marginal proportional hazards model (Lee et al., (1992) is applicable, given by

$$\lambda(t \mid \mathbf{Z}_{ij}) = \lambda_0(t) \exp\{\boldsymbol{\beta}_0' \mathbf{Z}_{ij}(t)\}, \tag{1}$$

where $\boldsymbol{\beta}_0$ is the regression coefficient vector. However, when cluster sizes are informative, the estimates and inference based on equation (1) may be incorrect. To account for informative cluster sizes, we propose two different approaches in this regard.

We define the counting process $N_{ij}(t) = I(X_{ij} \le t, \Delta_{ij} = 1)$, the at-risk process $Y_{ij}(t) = I(X_{ij} \ge t)$, and

$$M_{ij}(t) = N_{ij}(t) - \int_0^t Y_{ij}(u) \exp\{\boldsymbol{\beta}_0' \mathbf{Z}_{ij}(u)\} \lambda_0(u)\, du.$$

Note that $M_{ij}(t)$ is a local square-integrable martingale with respect to the marginal filtration $\mathcal{F}_{ij}(t) = \sigma\{N_{ij}(u), Y_{ij}(u), \mathbf{Z}_{ij}(u) : 0 \le u \le t\}$. However, due to the within-cluster dependence, $M_{ij}(t)$ is not a martingale with respect to the joint filtration generated by the history of all the failure, censoring and covariate information up to time t.

2.2 *Within-Cluster Resampling*

We randomly sample, with replacement, one individual from each of the m clusters. The bth resampled dataset denoted by $\{X_i^b, \Delta_i^b, \mathbf{Z}_i^b(t); i = 1, \ldots, m, t \in [0, \tau]\}$, consists of m independent observations, which can be analyzed using the Cox proportional hazards model for independent failure time data. For $b = 1, \ldots, B$, where B is a large fixed number, we introduce the following necessary notation:

$$\mathbf{S}_b^{(k)}(\boldsymbol{\beta}, t) = m^{-1} \sum_{i=1}^{m} Y_i^b(t) \big\{\mathbf{Z}_i^b(t)\big\}^{\otimes k} \exp\big\{\boldsymbol{\beta}' \mathbf{Z}_i^b(t)\big\},$$

$$\mathbf{s}^{(k)}(\boldsymbol{\beta}, t) = E\big\{\mathbf{S}_b^{(k)}(\boldsymbol{\beta}, t)\big\}, \quad \mathbf{e}(\boldsymbol{\beta}, t) = \mathbf{s}^{(1)}(\boldsymbol{\beta}, t)/s^{(0)}(\boldsymbol{\beta}, t),$$

$$\mathbf{V}_b(\boldsymbol{\beta}, t) = \mathbf{S}_b^{(2)}(\boldsymbol{\beta}, t)\big/S_b^{(0)}(\boldsymbol{\beta}, t) - \big\{\mathbf{S}_b^{(1)}(\boldsymbol{\beta}, t)\big/S_b^{(0)}(\boldsymbol{\beta}, t)\big\}^{\otimes 2},$$

where $\mathbf{a}^{\otimes k} = 1$, $\mathbf{a}$, $\mathbf{a}\mathbf{a}'$, for $k = 0, 1, 2$.

For the bth resampled data, the partial likelihood function is given by

$$L_b(\boldsymbol{\beta}) = \prod_{i=1}^{m} \left[\frac{\exp\big\{\boldsymbol{\beta}' \mathbf{Z}_i^b\big(X_i^b\big)\big\}}{m S_b^{(0)}\big(\boldsymbol{\beta}, X_i^b\big)} \right]^{\Delta_i^b}, \tag{2}$$

and accordingly, the score function is

$$\mathbf{U}_b(\boldsymbol{\beta}) = \sum_{i=1}^{m} \int_0^{\tau} \left\{ \mathbf{Z}_i^b(t) - \frac{\mathbf{S}_b^{(1)}(\boldsymbol{\beta}, t)}{S_b^{(0)}(\boldsymbol{\beta}, t)} \right\} dN_i^b(t). \tag{3}$$

Solving $\mathbf{U}_b(\boldsymbol{\beta}) = 0$, we obtain a consistent estimator for $\boldsymbol{\beta}$, denoted as $\hat{\boldsymbol{\beta}}_b$. The baseline cumulative hazard $\Lambda_0(t) = \int_0^t \lambda_0(u)\, du$ can be estimated by the Breslow–Aalen estimator, which for the bth resampled dataset is given by

$$\hat{\Lambda}_0^b(t, \hat{\boldsymbol{\beta}}_b) = \sum_{i=1}^{m} \int_0^t \frac{dN_i^b(u)}{\sum_{j=1}^{m} Y_j^b(u) \exp\{\hat{\boldsymbol{\beta}}_b' \mathbf{Z}_j^b(u)\}}.$$

After repeating this procedure B times, the WCR estimator for $\boldsymbol{\beta}$ is constructed as the average of the B resample-based estimates,

$$\bar{\boldsymbol{\beta}}_{\text{wcr}} = \frac{1}{B} \sum_{b=1}^{B} \hat{\boldsymbol{\beta}}_b, \tag{4}$$

and similarly, the WCR estimator for $\Lambda_0(t)$ is

$$\bar{\Lambda}_0(t, \hat{\boldsymbol{\beta}}) = \frac{1}{B} \sum_{b=1}^{B} \hat{\Lambda}_0^b(t, \hat{\boldsymbol{\beta}}_b), \tag{5}$$

where $\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{\beta}}_1, \ldots, \hat{\boldsymbol{\beta}}_B)$.

An attractive feature of the WCR method is that the estimates can be obtained by maximizing the standard partial likelihood function for independent data without specifying any correlation matrix. Simply by averaging over the estimates from the resampled data, we can obtain a consistent estimator for $\boldsymbol{\beta}$, and the variance–covariance matrix of $\bar{\boldsymbol{\beta}}_{\text{wcr}}$ in a relatively straightforward fashion, as shown in the next section.

2.3 *Inference Procedures*

Under certain regularity conditions (Anderson and Gill, 1982; Fleming and Harrington, 1990, p. 289–290), for each resampled dataset, $\hat{\boldsymbol{\beta}}_b$ is consistent and asymptotically normal. To prove the asymptotic normality of $\bar{\boldsymbol{\beta}}_{\text{wcr}}$, the central limit theorem (CLT) cannot be directly applied because $\bar{\boldsymbol{\beta}}_{\text{wcr}}$ is the average of B identically distributed but dependent maximum partial likelihood estimators. Following similar arguments as in Hoffman et al. (2001), we can rewrite $\bar{\boldsymbol{\beta}}_{\text{wcr}}$ as the sum of m independent cluster-specific terms so that the multivariate CLT can be applied. The asymptotic normality is stated in the following theorem, for which the proof is outlined in Web Appendix A.

THEOREM 1: *Under regularity conditions, as* $m \to \infty$, $\sqrt{m}(\bar{\boldsymbol{\beta}}_{\text{wcr}} - \boldsymbol{\beta}_0) \to N_p(\mathbf{0}, \boldsymbol{\Sigma})$ *in distribution, where* $\boldsymbol{\Sigma}$ *is a finite and positive definite matrix.*

As B increases, the covariance matrix of $\bar{\boldsymbol{\beta}}_{\text{wcr}}$ converges to $\boldsymbol{\Sigma}$. A consistent estimator for $\boldsymbol{\Sigma}$ is given as

$$\hat{\boldsymbol{\Sigma}} = \frac{m}{B}\left\{\sum_{b=1}^{B}\hat{\boldsymbol{\Sigma}}_b - (B-1)\hat{\boldsymbol{\Omega}}\right\}, \tag{6}$$

where $\hat{\boldsymbol{\Sigma}}_b$ is the estimated variance–covariance matrix for $\hat{\boldsymbol{\beta}}_b$ given by

$$\hat{\boldsymbol{\Sigma}}_b = \left\{\sum_{i=1}^{m}\int_0^{\tau}\mathbf{V}_b(\hat{\boldsymbol{\beta}}_b, t)\, dN_i^b(t)\right\}^{-1}, \tag{7}$$

and $\hat{\boldsymbol{\Omega}}$ is the estimated covariance matrix among the B resample-based estimates $\hat{\boldsymbol{\beta}}_b$,

$$\hat{\boldsymbol{\Omega}} = (B-1)^{-1}\sum_{b=1}^{B}(\hat{\boldsymbol{\beta}}_b - \bar{\boldsymbol{\beta}}_{\text{wcr}})(\hat{\boldsymbol{\beta}}_b - \bar{\boldsymbol{\beta}}_{\text{wcr}})'.$$

The consistency of $\hat{\boldsymbol{\Sigma}}$ is given in Theorem 2, with a sketched proof in Web Appendix B.

THEOREM 2: *Under regularity conditions,* $\hat{\boldsymbol{\Sigma}}$ *is consistent.*

Let $W(t) = \sqrt{m}\{\bar{\Lambda}_0(t, \hat{\boldsymbol{\beta}}) - \Lambda_0(t)\}, t \in [0, \tau]$, and let $\mathcal{W}(t)$ be a zero-mean Gaussian process with a finite covariance function.

THEOREM 3: *The random process* $W(t)$ *converges weakly to* $\mathcal{W}(t)$ *for* $t \in [0, \tau]$.

Proof of Theorem 3 is provided in Web Appendix C. Similar to equation (6), the covariance function of $\bar{\Lambda}_0(t, \hat{\boldsymbol{\beta}})$, between time t_1 and t_2, can be estimated by

$$\frac{m}{B}\left[\sum_{b=1}^{B}\text{Cov}\left\{\hat{\Lambda}_0^b(t_1, \hat{\boldsymbol{\beta}}_b), \hat{\Lambda}_0^b(t_2, \hat{\boldsymbol{\beta}}_b)\right\} - (B-1)\sum_{b=1}^{B}\left\{\hat{\Lambda}_0^b(t_1, \hat{\boldsymbol{\beta}}_b) - \bar{\Lambda}_0(t_1, \hat{\boldsymbol{\beta}})\right\}\left\{\hat{\Lambda}_0^b(t_2, \hat{\boldsymbol{\beta}}_b) - \bar{\Lambda}_0(t_2, \hat{\boldsymbol{\beta}})\right\}\right], \tag{8}$$

where $\text{Cov}\{\hat{\Lambda}_0^b(t_1, \hat{\boldsymbol{\beta}}_b), \hat{\Lambda}_0^b(t_2, \hat{\boldsymbol{\beta}}_b)\}$ is given by

$$\int_0^{\min(t_1,t_2)}\frac{d\hat{\Lambda}_0^b(u)}{S_b^{(0)}(\hat{\boldsymbol{\beta}}_b, u)} + \mathbf{H}'(\hat{\boldsymbol{\beta}}_b, t_1)\hat{\boldsymbol{\Sigma}}_b\mathbf{H}(\hat{\boldsymbol{\beta}}_b, t_2),$$

and

$$\mathbf{H}(\hat{\boldsymbol{\beta}}_b, t) = -\sum_{i=1}^{m}\int_0^t\frac{\mathbf{S}_b^{(1)}(\hat{\boldsymbol{\beta}}_b, u)}{m\left\{S_b^{(0)}(\hat{\boldsymbol{\beta}}_b, u)\right\}^2}\, dN_i^b(u).$$

2.4 *Weighted Score Function*

In this section, we consider extending the work of Williamson et al. (2003) to the clustered survival data with informative cluster sizes. We propose the following WSF:

$$\mathbf{U}(\boldsymbol{\beta}) = \sum_{i=1}^{m}\frac{1}{n_i}\sum_{j=1}^{n_i}\int_0^{\tau}\left[\mathbf{Z}_{ij}(t) - \frac{\sum_{k=1}^{m}\frac{1}{n_k}\sum_{l=1}^{n_k}Y_{kl}(t)\mathbf{Z}_{kl}(t)\exp\{\boldsymbol{\beta}'\mathbf{Z}_{kl}(t)\}}{\sum_{k=1}^{m}\frac{1}{n_k}\sum_{l=1}^{n_k}Y_{kl}(t)\exp\{\boldsymbol{\beta}'\mathbf{Z}_{kl}(t)\}}\right] \times dN_{ij}(t). \tag{9}$$

By setting $\mathbf{U}(\boldsymbol{\beta}) = \mathbf{0}$, we can obtain the estimators from the WSF, denoted by $\hat{\boldsymbol{\beta}}_{\text{wsf}}$. To account for the informative cluster sizes, the contribution of each individual to the overall estimating equation is inversely weighted by the corresponding cluster size. Similar to the WCR approach, this avoids overweighing larger clusters as opposed to the standard unweighted MMs. Asymptotic properties of this estimator can be developed similarly as in Lee et al. (1992) and following arguments similar to those in Cai and Prentice (1997) by letting the weight equal the inverse of each cluster size.

The covariance matrix of the limiting distribution of $\sqrt{m}(\hat{\boldsymbol{\beta}}_{\text{wsf}} - \boldsymbol{\beta}_0)$ can be consistently estimated by

$$\mathbf{J}^{-1}(\hat{\boldsymbol{\beta}}_{\text{wsf}})\hat{\mathbf{V}}\mathbf{J}^{-1}(\hat{\boldsymbol{\beta}}_{\text{wsf}}), \tag{10}$$

where $\mathbf{J}(\cdot)$ is the information matrix, and

$$\hat{\mathbf{V}} = \frac{1}{m}\sum_{i=1}^{m}\frac{1}{n_i^2}\sum_{j=1}^{n_i}\sum_{k=1}^{n_i}\hat{\mathbf{U}}_{ij}(\tau)\hat{\mathbf{U}}_{ik}'(\tau),$$

$$\hat{\mathbf{U}}_{ij}(\tau) = \int_0^{\tau}\left[\mathbf{Z}_{ij}(t) - \frac{\sum_{k=1}^{m}\frac{1}{n_k}\sum_{l=1}^{n_k}Y_{kl}(t)\mathbf{Z}_{kl}(t)\exp\{\hat{\boldsymbol{\beta}}_{\text{wsf}}'\mathbf{Z}_{kl}(t)\}}{\sum_{k=1}^{m}\frac{1}{n_k}\sum_{l=1}^{n_k}Y_{kl}(t)\exp\{\hat{\boldsymbol{\beta}}_{\text{wsf}}'\mathbf{Z}_{kl}(t)\}}\right]d\hat{M}_{ij}(t).$$

The baseline cumulative hazard function at time t can be estimated using the weighted Breslow–Aalen estimator,

$$\hat{\Lambda}_0(t, \hat{\boldsymbol{\beta}}_{\text{wsf}}) = \sum_{i=1}^{m}\frac{1}{n_i}\sum_{j=1}^{n_i}\int_0^t\frac{dN_{ij}(u)}{\sum_{k=1}^{m}\frac{1}{n_k}\sum_{l=1}^{n_k}Y_{kl}(u)\exp\{\hat{\boldsymbol{\beta}}_{\text{wsf}}'\mathbf{Z}_{kl}(u)\}}.$$

The covariance function can be estimated as

$$\text{Cov}\left\{\hat{\Lambda}_0(t_1, \hat{\boldsymbol{\beta}}_{\text{wsf}}), \hat{\Lambda}_0(t_2, \hat{\boldsymbol{\beta}}_{\text{wsf}})\right\} = \sum_{i=1}^{m}\psi_i(t_1)\psi_i(t_2), \tag{11}$$

where

$$\psi_i(t) = \frac{1}{n_i}\sum_{j=1}^{n_i}\int_0^t\frac{d\hat{M}_{ij}(u)}{\sum_{k=1}^{m}\frac{1}{n_k}\sum_{l=1}^{n_k}Y_{kl}(u)\exp\{\hat{\boldsymbol{\beta}}_{\text{wsf}}'\mathbf{Z}_{kl}(u)\}} - \mathbf{H}'(\hat{\boldsymbol{\beta}}_{\text{wsf}}, t)\mathbf{J}^{-1}(\hat{\boldsymbol{\beta}}_{\text{wsf}})\hat{\mathbf{U}}_i(t),$$

$\hat{\mathbf{U}}_i(t) = \frac{1}{n_i}\sum_{j=1}^{n_i}\hat{\mathbf{U}}_{ij}(t)$ and

$$\mathbf{H}(\hat{\boldsymbol{\beta}}_{\text{wsf}}, t) = \sum_{i=1}^{m}\frac{1}{n_i}\sum_{j=1}^{n_i}\int_0^t \frac{\sum_{k=1}^{m}\frac{1}{n_k}\sum_{l=1}^{n_k} Y_{kl}(u)\mathbf{Z}_{kl}(u)\exp\{\hat{\boldsymbol{\beta}}'_{\text{wsf}}\mathbf{Z}_{kl}(u)\}}{\left[\sum_{k=1}^{m}\frac{1}{n_k}\sum_{l=1}^{n_k} Y_{kl}(u)\exp\{\hat{\boldsymbol{\beta}}'_{\text{wsf}}\mathbf{Z}_{kl}(u)\}\right]^2} dN_{ij}(u).$$

The WCR estimator and the WSF estimator are asymptotically equivalent in the sense that as $m \to \infty$, $\|\hat{\boldsymbol{\beta}}_{\text{wsf}} - \tilde{\boldsymbol{\beta}}_{\text{wcr}}\| \to 0$, where $\tilde{\boldsymbol{\beta}}_{\text{wcr}}$ is defined as $\lim_{B\to\infty}\bar{\boldsymbol{\beta}}_{\text{wcr}}$. In Web Appendix D, we outline the proof that the limiting distributions are the same for $\hat{\boldsymbol{\beta}}_{\text{wsf}}$ and $\tilde{\boldsymbol{\beta}}_{\text{wcr}}$, as $m \to \infty$.

3. Simulation Studies

We conducted simulation studies to assess the performance of our proposed methods. We considered different scenarios by varying the within-cluster correlation, censoring rate, and more importantly, informative and noninformative cluster sizes on survival.

Correlated failure times were simulated using the Cox proportional hazards model with positive stable frailty (Hougaard, 1996),

$$\lambda(t \mid \mathbf{Z}_{ij}, w_i) = \lambda_0(t) w_i \exp(\boldsymbol{\gamma}_0'\mathbf{Z}_{ij}),$$

where w_i follows a positive stable distribution with parameter α, $\alpha \in (0, 1)$. For cluster i, we first generated the frailty w_i, given which the survival times of the individuals within this cluster can be independently simulated. A positive stable variable can be generated using the following representation (Chambers, Mallows, and Stuck, 1976; Nolan, 2006):

$$W = (a(\theta)/\xi)^{(1-\alpha)/\alpha},$$

where θ and ξ are independent, θ is uniform on $(0, \pi)$, ξ is exponential with mean one, and

$$a(\theta) = \frac{(\sin(1-\alpha)\theta)(\sin\alpha\theta)^{\alpha/(1-\alpha)}}{(\sin\theta)^{1/(1-\alpha)}}.$$

The parameter α represents the degree of correlation between cluster members, with $\alpha \to 0$ giving the maximal positive dependence, and $\alpha \to 1$ corresponding to the independent case. We let α equal 0.5 and 0.75, corresponding to the within-cluster Pearson's correlation coefficient $\rho = 0.3$ and 0.15, respectively. For ease of exposition, we considered a constant baseline hazard, $\lambda_0(t) = 1$. After integrating out the frailty, the true marginal regression parameters are $\boldsymbol{\beta}_0 = \alpha\boldsymbol{\gamma}_0$ and the true baseline cumulative hazard at time t is $\Lambda_0(t) = t^\alpha$.

To induce informative cluster sizes, we let the size of each cluster depend on the value of the generated frailty such that

$$n_i = (k/10) + 2, \quad \text{if } q_k \le w_i < q_{k+10}, \quad \text{for } k = 0, 10, \ldots, 90,$$

where q_k is the kth percentile of the frailty distribution. As a result, the cluster sizes vary from 2 to 11, depending on which percentile range the frailty values fall in. For the noninformative cases, the cluster sizes are randomly taken from $\{2, \ldots, 11\}$ regardless of the frailty values.

The censoring times were generated independently from a uniform distribution, Unif(0, c), where the value of c can be selected to achieve desired censoring rates. Two covariates were included in the simulation: one was a binary variable, Z_1, taking a value of 0 or 1 with probability 0.5, which may represent the treatment or control group; the other was a continuous variable, Z_2, independently generated from Unif(0, 1). We took the number of clusters $m = 200$ or 300. For each setup, we simulated 1000 datasets and analyzed each dataset using the WCR method with $B = 2000$ resamples, the WSF and the standard MM by Lee et al. (1992) under the working independence assumption.

For each data realization, we obtained the point estimates of the regression coefficients using all three methods, $\bar{\boldsymbol{\beta}}_{\text{wcr}}$, $\hat{\boldsymbol{\beta}}_{\text{wsf}}$, and $\hat{\boldsymbol{\beta}}_{\text{mm}}$, where $\hat{\boldsymbol{\beta}}_{\text{mm}}$ denotes the estimator under the standard MM. We also calculated the sample standard deviation (SD) over the 1000 simulations, the mean standard error (SE) and the 95% confidence interval coverage rate for each estimated coefficient. As shown in Tables 1 and 2, when the cluster size is informative, the point estimates of the coefficients using the WCR and WSF methods are approximately unbiased and the 95% confidence interval coverage rates are close to the nominal value, whereas the MM estimates are substantially biased. On the other hand, when the cluster size is not informative, all point estimates are approximately unbiased and the coverage rates of all three methods are reasonably close to the nominal level. In both tables, the variation of the parameter estimates decreases when the number of clusters increases, and increases when the censoring rate increases. The sample SDs are close to the mean SEs for the WCR and WSF methods, which suggests that our variance estimators (6) and (10) provide good estimates for the variability of $\bar{\boldsymbol{\beta}}_{\text{wcr}}$ and $\hat{\boldsymbol{\beta}}_{\text{wsf}}$. The WSF method performs relatively better than WCR when m is small, but when m is large, the estimators from these two methods are very close to each other.

In Figure 2, we show the quantile–quantile (Q–Q) plots for the estimated regression coefficient $\hat{\beta}_1$ after being standardized versus a standard normal distribution. All six Q–Q plots appear to lie closely on a straight line, which indicates that the parameter estimators approximately follow normal distributions. The deviation from the diagonal line (the solid line) in the plot in the lower left corner shows the bias of the MM estimator when the cluster size is informative.

For the baseline cumulative hazard, we computed the pointwise estimates and SEs for some selected time points, which were chosen to be the 20th, 40th, 60th, and 80th percentiles of the underlying true failure time distribution. In Table 3, we show the results with $m = 200$ and 300, $\alpha = 0.5$, 50% censoring rate, and informative and noninformative cluster sizes. The WCR estimator, $\bar{\Lambda}_0(t, \hat{\boldsymbol{\beta}})$, and the WSF estimator $\hat{\Lambda}_0(t, \hat{\boldsymbol{\beta}}_{\text{wsf}})$ provide pointwise estimates that are very close to the true cumulative baseline hazard regardless of whether the cluster size is informative or not. The MM estimator is substantially biased with poor 95% confidence interval coverage rates when the cluster size is informative, but performs comparably with the other two when the cluster size is not informative. The WCR and WSF methods yield similar pointwise estimates for the baseline cumulative hazard, however,

Table 1

Simulation results for β_1*, coefficient for the binary covariate, with true values* 0.5, 0.75 *for* $\alpha = 0.5$, 0.75: *estimate* (*Est*)*, SD, SE, and* 95% *confidence interval coverage rate* (*CR*) *in percentage*

		WCR				WSF				MM			
True	cen%	Est	SD	SE	CR	Est	SD	SE	CR	Est	SD	SE	CR
$m = 200$, Informative cluster size													
0.5	25	0.500	0.090	0.089	94.0	0.499	0.089	0.090	94.7	0.466	0.067	0.068	92.1
	50	0.498	0.105	0.099	92.1	0.495	0.104	0.101	94.0	0.470	0.085	0.083	92.4
0.75	25	0.757	0.089	0.084	93.2	0.753	0.089	0.086	94.1	0.710	0.071	0.068	89.1
	50	0.756	0.099	0.095	94.7	0.751	0.098	0.097	95.4	0.708	0.079	0.080	91.4
$m = 300$, Informative cluster size													
0.5	25	0.501	0.074	0.072	93.6	0.500	0.073	0.072	93.9	0.464	0.053	0.053	88.7
	50	0.501	0.076	0.073	93.8	0.500	0.076	0.074	94.4	0.467	0.056	0.055	89.5
0.75	25	0.751	0.069	0.068	94.8	0.748	0.068	0.068	95.8	0.705	0.051	0.054	87.8
	50	0.751	0.072	0.072	94.7	0.748	0.072	0.073	95.2	0.706	0.056	0.058	89.3
$m = 200$, Noninformative cluster size													
0.5	25	0.502	0.076	0.075	95.2	0.499	0.076	0.076	95.6	0.500	0.068	0.068	94.9
	50	0.506	0.087	0.082	92.9	0.503	0.086	0.083	93.9	0.503	0.076	0.074	93.8
0.75	25	0.754	0.075	0.074	94.0	0.748	0.074	0.076	95.5	0.750	0.068	0.068	95.0
	50	0.759	0.088	0.082	92.5	0.754	0.087	0.084	93.6	0.754	0.075	0.075	95.3
$m = 300$, Noninformative cluster size													
0.5	25	0.505	0.062	0.062	94.9	0.503	0.062	0.063	95.2	0.504	0.055	0.056	95.5
	50	0.506	0.068	0.067	93.6	0.505	0.067	0.068	94.7	0.505	0.060	0.061	95.7
0.75	25	0.759	0.063	0.061	94.2	0.754	0.062	0.062	95.0	0.754	0.056	0.056	95.3
	50	0.758	0.069	0.067	93.4	0.755	0.069	0.068	95.4	0.754	0.061	0.061	95.1

Table 2

Simulation results for β_2*, coefficient for the continuous covariate, with true values* 0.5, 0.75 *for* $\alpha =$ 0.5, 0.75: *estimate* (*Est*)*, SD, SE, and* 95% *confidence interval coverage rate* (*CR*) *in percentage*

		WCR				WSF				MM			
True	cen%	Est	SD	SE	CR	Est	SD	SE	CR	Est	SD	SE	CR
$m = 200$, Informative cluster size													
0.5	25	0.501	0.154	0.150	93.5	0.499	0.153	0.152	95.0	0.466	0.111	0.113	94.4
	50	0.508	0.175	0.165	92.3	0.506	0.173	0.170	94.7	0.479	0.140	0.137	93.7
0.75	25	0.765	0.147	0.142	93.5	0.759	0.145	0.144	94.2	0.715	0.110	0.112	94.9
	50	0.754	0.166	0.157	92.0	0.748	0.165	0.163	93.8	0.708	0.131	0.131	93.3
$m = 300$, Informative cluster size													
0.5	25	0.504	0.120	0.121	95.4	0.502	0.120	0.121	95.8	0.466	0.086	0.087	92.9
	50	0.505	0.124	0.123	94.7	0.504	0.124	0.124	96.2	0.469	0.092	0.092	94.0
0.75	25	0.753	0.121	0.114	92.9	0.749	0.121	0.116	94.2	0.704	0.090	0.089	91.1
	50	0.753	0.125	0.120	93.8	0.749	0.124	0.122	95.2	0.705	0.096	0.096	93.1
$m = 200$, Noninformative cluster size													
0.5	25	0.503	0.133	0.125	92.2	0.500	0.131	0.128	94.7	0.500	0.114	0.113	94.8
	50	0.511	0.142	0.137	93.7	0.508	0.140	0.139	95.6	0.509	0.121	0.123	95.5
0.75	25	0.755	0.127	0.123	92.6	0.749	0.124	0.126	95.3	0.749	0.110	0.112	94.4
	50	0.763	0.141	0.137	93.0	0.757	0.139	0.139	95.1	0.758	0.120	0.123	96.1
$m = 300$, Noninformative cluster size													
0.5	25	0.501	0.107	0.104	93.9	0.499	0.106	0.105	95.0	0.503	0.091	0.093	95.7
	50	0.499	0.115	0.112	94.3	0.496	0.114	0.114	94.9	0.500	0.099	0.100	94.9
0.75	25	0.752	0.107	0.102	91.9	0.748	0.106	0.103	93.3	0.752	0.092	0.091	94.7
	50	0.750	0.118	0.112	93.1	0.746	0.117	0.114	94.0	0.751	0.102	0.100	94.6

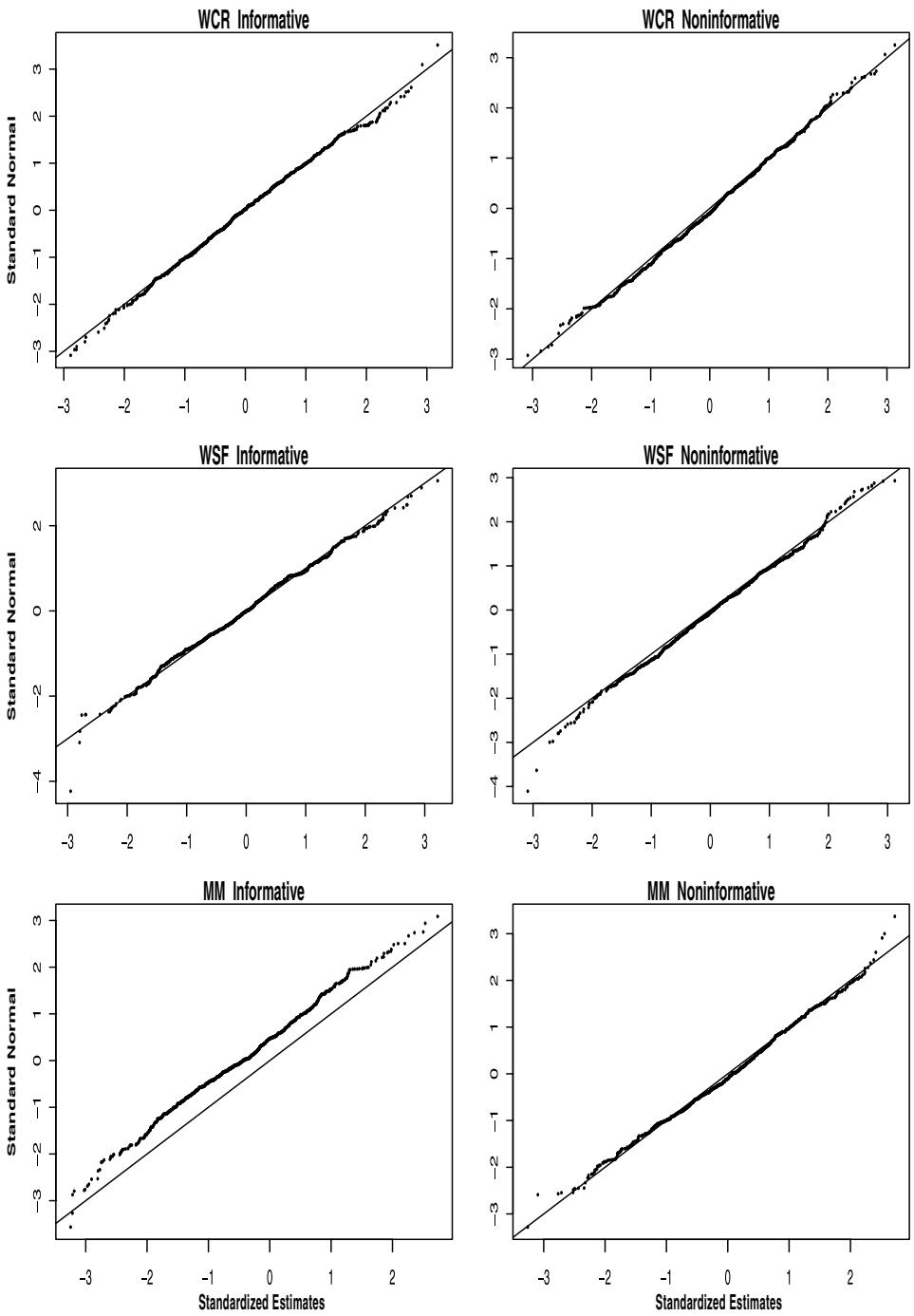

Figure 2. Q–Q plots for the WCR, WSF, and MM estimators, $\bar{\beta}_{\text{wcr}}, \hat{\beta}_{\text{wsf}}$, and $\hat{\beta}_{\text{mm}}$ for Z_1, based on 1000 simulations with $m = 200$, $\alpha = 0.5$, and 50% censoring rate.

the WCR estimator has better 95% confidence interval coverage rates. The similarity between the columns SD and SE suggests that equations (8) and (11) serve as good estimators for the variance of $\bar{\Lambda}_0(t, \hat{\boldsymbol{\beta}})$ and $\hat{\Lambda}_0(t, \hat{\boldsymbol{\beta}}_{\text{wsf}})$, respectively.

4. Data Application

We applied the WCR, WSF, and MM methods to the dental study described in Spiekerman and Lin (1998). The original study was conducted by McGuire and Nunn (1996) to assess the effect of some commonly measured risk factors in predicting tooth survival. The dataset consisted of 100 consecutive patients from Dr McGuire's appointment book. All of these patients had been diagnosed with moderate to severe chronic adult periodontitis and had received at least 5 years of maintenance care. For this analysis, we considered two risk factors, age and smoking status (0 = smoker, and 1 otherwise), to

Table 3
Simulation results for the cumulative baseline hazard $\Lambda_0(t)$: estimate (Est), SD, SE, 95% confidence interval coverage rate (CR) in percentage with $\alpha = 0.5$ and 50% censoring rate

		WCR				WSF				MM			
t	$\Lambda_0(t)$	Est	SD	SE	CR	Est	SD	SE	CR	Est	SD	SE	CR
$m = 200$, Informative cluster size													
0.04	0.200	0.203	0.031	0.031	95.7	0.202	0.031	0.030	94.5	0.332	0.046	0.043	10.5
0.15	0.387	0.392	0.052	0.052	95.1	0.390	0.051	0.048	93.5	0.618	0.069	0.063	2.6
0.25	0.500	0.507	0.063	0.064	95.7	0.503	0.062	0.059	93.4	0.782	0.080	0.074	1.7
0.50	0.707	0.719	0.085	0.087	95.8	0.711	0.084	0.079	94.5	1.077	0.101	0.092	0.8
$m = 300$, Informative cluster size													
0.04	0.200	0.201	0.024	0.025	96.2	0.200	0.024	0.024	95.7	0.330	0.035	0.035	1.6
0.15	0.387	0.389	0.040	0.042	95.7	0.387	0.040	0.039	94.7	0.614	0.053	0.051	0.2
0.25	0.500	0.503	0.050	0.052	95.6	0.500	0.050	0.048	93.7	0.780	0.063	0.060	0.1
0.50	0.707	0.713	0.068	0.070	96.1	0.708	0.068	0.064	94.4	1.074	0.081	0.075	0.1
$m = 200$, Noninformative cluster size													
0.04	0.200	0.203	0.031	0.031	94.8	0.202	0.031	0.028	92.6	0.202	0.031	0.029	92.9
0.15	0.387	0.393	0.052	0.051	93.8	0.390	0.051	0.046	90.7	0.390	0.050	0.045	92.6
0.25	0.500	0.508	0.064	0.063	94.0	0.503	0.063	0.056	91.1	0.503	0.059	0.054	93.0
0.50	0.707	0.718	0.084	0.084	94.2	0.710	0.083	0.074	90.6	0.710	0.078	0.069	92.5
$m = 300$, Noninformative cluster size													
0.04	0.200	0.201	0.025	0.025	95.2	0.201	0.025	0.023	94.3	0.200	0.024	0.023	93.6
0.15	0.387	0.390	0.042	0.041	95.0	0.389	0.041	0.038	92.6	0.388	0.040	0.037	92.2
0.25	0.500	0.505	0.051	0.051	94.7	0.502	0.050	0.046	92.3	0.500	0.048	0.044	92.1
0.50	0.707	0.716	0.067	0.068	94.6	0.710	0.067	0.060	92.6	0.708	0.063	0.057	91.3

predict the tooth survival. The failure time for each tooth was defined as the time to tooth loss measured from the initiation of active periodontal therapy.

In this analysis, we focused on the upper and lower molars (a normal person should have a maximum of eight molars). We had 96 patients with both upper and lower molars, that is, the number of clusters was 96. The cluster sizes ranged from 1 to 8. The total number of teeth was 598 with 58 observed failures. As illustrated in the introduction section, the cluster size might be informative to tooth survival. Patients who had more teeth tend to have higher tooth survival probability; more precisely, larger cluster sizes indicated better survival.

We performed $B = 10{,}000$ resamplings for the WCR method. The analysis results are summarized in Table 4, where we compare the estimates of the regression coefficients using the WCR and WSF methods with those using the standard MM. The point estimates for the effect of smoking status are similar using WCR and WSF but different from that obtained by the MM method. This might be due to the informativeness of cluster sizes. The hazard ratio of tooth loss for cigarette smoking is 1.647 (WCR and WSF) and 1.747 (MM) with overlapping 95% confidence intervals. Smoking is a very important factor that hastens tooth loss. The estimates of the age effect from the three methods are consistent, indicating that older patients would lose their teeth significantly sooner than younger patients.

As suggested by the associate editor, we also fit the data with cluster size included as a covariate in the standard MM. The estimated regression coefficients (SE, p-value) for smoking status, age, and cluster size are −0.440 (0.352, 0.106), 1.478 (0.764, 0.026), and −0.166 (0.085, 0.026), respectively. The cluster size effect is statistically significant and the hazard of tooth loss decreases with increasing cluster size. This is consistent with our observation from the Kaplan–Meier curves in Figure 1. However, the interpretations of these estimates are different. For the WCR and WSF methods, the inference focuses on the effect of age and smoking status on a randomly selected molar tooth from a randomly selected person, while the informative cluster size is a nuisance variable but

Table 4
Estimates of the regression coefficients for the dental study

	WCR			WSF			MM		
Covariate	Estimate	SE	P-value	Estimate	SE	P-value	Estimate	SE	P-value
Smoking status	−0.501	0.304	0.049	−0.500	0.342	0.071	−0.558	0.340	0.050
Age	1.877	0.984	0.028	1.721	0.743	0.010	1.759	0.724	0.008

properly adjusted. When taking the cluster size (the number of molars) as a covariate in the standard MM, the covariate effects of age and smoking status are based on a randomly selected molar given a certain number of molars that a person has. One feature of the WCR and WSF models is that the effect of cluster size is considered as a nuisance. We do not need to explicitly specify the correlation structure between failure times and cluster size, which is often unknown and difficult to model. Moreover, the WCR and WSF methods provide valid estimates regardless of whether the cluster size is informative or not.

5. Discussion

When cluster sizes are informative to the correlated survival outcomes, the estimated regression coefficients could be substantially biased when using the standard MM approaches. In contrast, both the WCR and WSF methods provide valid estimates in the presence of informative cluster size and do not require specification of the dependence structure of cluster size and the outcome of interest. Simulation studies have shown that the estimates under the WCR and WSF methods are approximately unbiased with reasonable 95% confidence interval coverages.

One advantage of the WCR method over WSF is that by sampling one observation from each cluster, the estimation problem reduces to the independent case; therefore, standard software can be easily applied regardless of the underlying correlation structure between the cluster size and failure time. However, WCR is computationally intensive due to the resampling scheme. Follmann et al. (2003) recommended that it would be enough resamples when the WCR inference is similar to that based on enumeration. Based on our experience in the simulations, 2000 resamples are usually sufficient for practical use.

The variance estimators defined in equations (6) and (8) involve the subtraction of two terms, thus it is possible to obtain a negative estimator for the variance. In our simulation, however, this is very rare. In particular, the occurrence frequency is less than one out of a thousand when the number of resamplings is large. Another potential problem in applying the WCR method is that the estimator from a single resampled dataset might be unstable under heavy censoring, because each resampled dataset only consists of partial information from the original data.

6. Supplementary Materials

Web Appendices referenced in Sections 2–4 are available under the paper information link at the Biometrics web site `http://www.tibs.org/biometrics`.

Acknowledgements

The authors would like to thank Drs C. Spiekerman and D. Y. Lin for the dental dataset, an associate editor, and a referee for their constructive comments and suggestions that have greatly improved the manuscript. This research is partly supported by GY's grant from M. D. Anderson Cancer Center and YS's NIH grant R01 CA79466. This article was done while XJC. was a Ph.D. student at Rice University.

References

Anderson, P. and Gill, R. (1982). Cox's regression model for counting processes: A large sample study. *The Annals of Statistics* **10,** 1100–1120.

Benhin, E., Rao, N. K., and Scott, A. J. (2005). Mean estimating equation approach to analysing cluster-correlated data with nonignorable cluster sizes. *Biometrika* **92,** 435–450.

Cai, J. and Prentice, R. (1997). Regression estimation using multivariate failure time data and a common baseline hazard function model. *Lifetime Data Analysis* **3,** 197–213.

Cai, T., Wei, L. J., and Wilcox, M. (2000). Semiparametric regression analysis for clustered failure time data. *Biometrika* **87,** 867–878.

Chambers, J. M., Mallows, C. L., and Stuck, B. W. (1976). A method for simulating stable random variables. *Journal of the American Statistical Association* **71,** 340–344.

Chen, L. and Wei, L. J. (1997). Analysis of multivariate survival times with non-proportional hazards models. In *Proceedings of the First Seattle Symposium in Biostatistics: Survival Analysis*, D. Y. Lin and T. R. Fleming (eds). New York: Springer-Verlag.

Clegg, L. X., Cai, J., and Sen, P. (1999). A marginal mixed baseline hazards model for multivariate failure time data. *Biometrics* **55,** 805–812.

Cox, D. R. (1975). Partial likelihood. *Biometrika* **62,** 269–276.

Fleming, T. and Harrington, D. (1990). *Counting Processes and Survival Analysis*. New York: Wiley.

Follmann, D., Proschan, M., and Leifer, E. (2003). Multiple outputation: Inference for complex clustered data by averaging analyses from independent data. *Biometrics* **59,** 420–429.

Hoffman, E., Sen, P., and Weinberg, C. (2001). Within-cluster resampling. *Biometrika* **88,** 1121–1134.

Hougaard, P. (1986). A class of multivariate failure time distribution. *Biometrika* **73,** 671–678.

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 Arts*, J. P. Klein and P. K. Goel (eds). Dordrecht: Kluwer Academic Publishers.

Lee, E. W., Wei, L. J., and Ying, Z. (1993). Linear regression analysis for highly stratified failure time data. *Journal of the American Statistical Association* **88,** 557–565.

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

Lin, J. S. and Wei, L. J. (1992). Linear regression analysis for multivariate failure time observations. *Journal of American Statistical Association* **87,** 1091–1097.

Lu, W. (2005). Marginal regression of multivariate event times based on linear transformation models. *Lifetime Data Analysis* **11,** 389–404.

McGuire, M. K. and Nunn, M. E. (1996). Prognosis versus actual outcome III: The effects of clinical parameters

in accurately predicting tooth survival. *Journal of Periodontology* **67,** 666–674.

Nolan, J. P. (2006). *Stable Distributions—Models for Heavy Tailed Data*. Boston: Birkhäuser. In progress, Chapter 1 online at `cademic2.american.edu/~jpnolan`.

Spiekerman, C. and Lin, D. (1998). Marginal regression models for multivariate failure time data. *Journal of the American Statistical Association* **93,** 1164–1175.

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.

Williamson, J., Datta, S., and Satten, G. (2003). Marginal analyses of clustered data when cluster size is informative. *Biometrics* **59,** 36–42.

Yin, G. and Cai, J. (2004). Additive hazards model with multivariate failure time data. *Biometrika* **91,** 801–818.

Received November 2005. *Revised September* 2006.
Accepted September 2006.