



# Semiparametric Methods for Clustered Recurrent Event Data

DOUGLAS E. SCHAUBEL

deschau@umich.edu

*Department of Biostatistics, University of Michigan, 1420 Washington Hts, Ann Arbor, MI, 48109-2029, USA*

JIANWEN CAI

*Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599-7420, USA*

*Received January 10, 2005; Accepted June 2, 2005*

**Abstract.** In biomedical studies, the event of interest is often recurrent and within-subject events cannot usually be assumed independent. In addition, individuals within a cluster might not be independent; for example, in multi-center or familial studies, subjects from the same center or family might be correlated. We propose methods of estimating parameters in two semi-parametric proportional rates/means models for clustered recurrent event data. The first model contains a baseline rate function which is common across clusters, while the second model features cluster-specific baseline rates. Dependence structures for patients-within-cluster and events-within-patient are both unspecified. Estimating equations are derived for the regression parameters. For the common baseline model, an estimator of the baseline mean function is proposed. The asymptotic distributions of the model parameters are derived, while finite-sample properties are assessed through a simulation study. Using data from a national organ failure registry, the proposed methods are applied to the analysis of technique failures among Canadian dialysis patients.

**Keywords:** clustered failure time data, marginal model, ordered event times, proportional rates, robust variance

## 1. Introduction

In many clinical and epidemiologic studies, the event of interest may occur more than once per subject; such outcomes have been termed recurrent events. Examples include hospitalizations, infections, acute myocardial infarctions, and tumor metastases. Since such studies are of fixed length, the experience of each subject is subject to right censoring. Events within a subject cannot usually be assumed to be independent. Therefore, recurrent event data represent a form of multivariate failure time data where the event times are ordered. Moreover, in many biomedical studies, the experience of subjects may be correlated, resulting in clustered data. For example, in familial studies, individuals within a family may be correlated due to shared genetic factors; in a childhood school asthma study, children from the same neighborhood may share certain environmental risk factors (e.g., air particulate levels); or, in a multi-center study of technique failures among patients on dialysis, patients from

the same center may be correlated due to center-specific characteristics with respect to practice patterns. Factors through which patients may be clustered are often unmeasured and usually not of direct interest to the investigators.

The investigation which motivated this example is a retrospective cohort study of technique failures (TF) among patients receiving peritoneal dialysis. Patients who reach end-stage renal disease must either receive a kidney transplant or dialysis in order to remain alive. Due to an increasing shortfall of donor organs, many patients receive dialysis for the rest of their lives. Of the two main dialytic therapies, of which the other is hemodialysis, peritoneal dialysis is less costly and associated with better quality of life. Since funding models for renal centers differ by region, it is of great interest to compare outcomes, such as TF, across regions. Naturally, the analysis must account for the correlation among technique failures within the same patient. It is likely that patients treated at the same renal center are correlated, due to similarities in practice patterns among clinicians from the same center. In addition, patients receiving care at the same center are likely to be from the same catchment area, and hence more likely to share unmeasured socio-economic and clinical characteristics. The research question regards only differences in TF rates among regions, and does not involve the dependence structures, either within-patients or between-patient-within-center; although, not accounting for such dependencies could result in biased variance estimation and hence lead to erroneous conclusions.

Many methods exist for the analysis of multivariate failure times. Methods for (i) single events for clustered subjects and (ii) repeated events among individuals have been broadly grouped under the umbrella of multivariate survival analysis. However, the statistical considerations for these two settings are quite different. Within 'multivariate survival', many available methods can be generally classified as 'conditional' or 'marginal'. Conditional methods (e.g., the frailty model) require correctly specifying the dependence structure and are appealing when the dependence structure is of interest. Marginal models do not incorporate the correlation structure for the purposes of parameter estimation, but account for the dependence in the covariance estimation. The marginal approach is appealing for clustered failure time data when the dependence structure represents a nuisance parameter. Many such existing methods are based on modelling the hazard function (e.g., Wei, Lin and Weissfeld, 1989; Lee, Wei and Amato, 1992; Spiekerman and Lin, 1998; Clegg, Cai and Sen, 1999). For recurrent event data, the mean and rate function are better intuited than the hazard (intensity) function. Moreover, as described, for example, by Lin, Wei, Yang and Ying (2000), the intensity function requires accurately modelling the event history, which could be very complicated. Pepe and Cai (1993) developed semiparametric methods for modelling the rate function, with possibly distinct event number-specific baseline rates. Lawless and Nadeau (1995) considered modelling the mean number of events, and developed the theory for the discrete-time case. Subsequently, Lin, Wei, Yang and Ying (2000) developed the asymptotic theory for the continuous time setting. A detailed review of available methods for recurrent event data is provided by Cook and Lawless (2002). Cai, Wei and Wilcox (2000) developed a class of linear

transformation models for clustered failure time data. Glidden and Vittinghoff (2004) examined marginal and conditional methods for correlated failure time data, where the failure time is univariate and patients are clustered within center.

To the best of our knowledge, all existing marginal methods for the analysis of recurrent event data assume independence among individuals, and, therefore, cannot be directly applied to studies with clustered subjects. In this article, we propose two methods for fitting proportional rates models to recurrent event data wherein study subjects are clustered. The proposed models are semi-parametric in that a functional form is assumed for multiplicative covariate effects, although the baseline rates are left unspecified, as are the dependence structures among correlated events. The first model contains a baseline rate that is common across all clusters, while the second model features cluster-specific baseline rates.

The remainder of this article is organized as follows. In Section 2, we introduce the models and methods for estimating their parameters. Asymptotic results are provided in Section 3, with proofs outlined in the Appendix. Since the estimation procedures do not emit Martingale structures, results from the modern theory of empirical processes are employed. A simulation study is described in Section 4, to study finite-sample properties of the proposed regression parameter estimators. In Section 5, we present an analysis of the above-described retrospective renal failure cohort study. We compare results based on the proposed method with those from methods which incorrectly assume various levels of independence. Concluding remarks are given in Section 6.

## 2. Models and Methods

We now set up the requisite notation. The number of independent clusters is denoted by  $n$ , while the number of subjects in the  $j$ th cluster is denoted by  $n_j$ . For the  $i$ th subject from the  $j$ th cluster, the cumulative number of events, as of time  $t$ , is represented by  $N_{ij}^*(t)$ . The pertinent intensity function is given by  $E[dN_{ij}^*(s)|\mathcal{F}_{ij}(t)]$ , where  $\mathcal{F}_{ij}(t)$  is a filtration containing the event history on the interval  $[0, t)$ ,  $\mathcal{F}_{ij}(t) = \{N_{ij}^*(s); s \in [0, t)\}$ . In the context of recurrent event data, methods for analyzing the intensity function have been proposed by several previous authors; e.g., Andersen & Gill (1982) and Lawless (1987). We consider the rate function,  $E[dN_{ij}^*(t)]$ , a marginal quantity, which can be connected to the intensity function through the relation,

$$E[dN_{ij}^*(t)] = E[E[dN_{ij}^*(t)|\mathcal{F}_{ij}(t)]].$$

That is, the rate function can be considered an expectation across all possible event histories. Specifically, we propose the following proportional rates models:

$$E[dN_{ij}^*(s)|\mathbf{Z}_{ij}(s)] = \exp\{\boldsymbol{\beta}_0^T \mathbf{Z}_{ij}(s)\} d\mu_{0j}(s), \quad (1)$$

$$E[dN_{ij}^*(s)|\mathbf{Z}_{ij}(s)] = \exp\{\boldsymbol{\beta}_0^T \mathbf{Z}_{ij}(s)\} d\mu_0(s), \tag{2}$$

where  $d\mu_j(s)$  and  $d\mu_0(s)$  are unspecified baseline rate functions,  $\boldsymbol{\beta}_0$  is an unknown parameter vector and  $\mathbf{Z}_{ij}(s)$  is the corresponding vector covariates. Presumably, chief interest lies in  $\boldsymbol{\beta}_0$ , with the baseline rate(s) being of secondary or no interest. As implied, time-dependent covariates are permitted, although, in keeping with the spirit of marginal modelling, functions of the event history would not be considered. Lin et al. (2000) studied the independent subjects analog of (2), termed the proportional rates model. The quantity  $E[N_{ij}^*(t)|\mathbf{Z}_{ij}(s), s \in [0, t]]$  always represents a cumulative rate function. If  $\mathbf{Z}_{ij}(s) = \mathbf{z}_{ij}$ , then  $E[N_{ij}^*(t)|\mathbf{z}_{ij}]$  is the mean function. It is also the mean function if all time-dependent covariates are ‘external’ in the sense of Kalbfleisch and Prentice (2002; pp. 196–200); i.e., have paths which are known at time 0.

Events are assumed to be subject to independent right censoring and the censoring time is denoted by  $C_{ij}$ . Then, the observed number of events for subject  $i$  from cluster  $j$  is given by  $N_{ij}(t) = N_{ij}^*(t \wedge C_{ij})$ . It is assumed that each subject’s censoring time is conditionally independent of the recurrent event process given the covariate vector, in that:

$$E[dN_{ij}^*(s)|\mathbf{Z}_{ij}(s), C_{ij} > s] = E[dN_{ij}^*(s)|\mathbf{Z}_{ij}(s)]. \tag{3}$$

Although censoring is assumed to be independent of the events, censoring times for individuals within a cluster need not be independent.

We now describe the methods for parameter estimation, beginning with model (1). Under the set-up under consideration, clusters, not subjects, are the independent units. It is not possible to estimate  $\mu_{0j}(t)$  consistently since each cluster consists of one independent unit. As such, for model (1), we focus on estimating the regression parameter. In practice, whether the common or cluster-specific baseline rate model is assumed, the regression parameter will be of primary interest. Given the conditionally independent censoring described by (3), straightforward conditional expectation arguments can be used to show that the following function has mean zero,

$$\sum_{j=1}^n \sum_{i=1}^{n_j} \int_0^\tau \{\mathbf{Z}_{ij}(s) - \bar{\mathbf{Z}}_j(s; \boldsymbol{\beta}_0)\} dM_{ij}^d(s; \boldsymbol{\beta}_0), \tag{4}$$

where  $\bar{\mathbf{Z}}_j(s; \boldsymbol{\beta}) = \mathbf{S}_j^{(1)}(s; \boldsymbol{\beta}) / \mathbf{S}_j^{(0)}(s; \boldsymbol{\beta})$ , with  $\mathbf{S}_j^{(d)}(s; \boldsymbol{\beta}) = n_j^{-1} \sum_{i=1}^{n_j} I(C_{ij} > s) \mathbf{Z}_{ij}(s)^{\otimes d} \exp\{\boldsymbol{\beta}^T \mathbf{Z}_{ij}(s)\}$  for  $d=0,1,2$ ,  $I(A)$  takes the value 1 when  $A$  occurs and 0 otherwise,  $\mathbf{a}^{\otimes 0} = 1$ ,  $\mathbf{a}^{\otimes 1} = \mathbf{a}$ ,  $\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}^T$ , for a vector  $\mathbf{a}$ , and  $M_{ij}^d(t; \boldsymbol{\beta}) = N_{ij}(t) - \int_0^t I(C_{ij} > s) \exp\{\boldsymbol{\beta}^T \mathbf{Z}_{ij}(s)\} d\mu_{0j}(s)$ . We can simplify (4) such that under model (1),  $\boldsymbol{\beta}_0$  can be estimated by  $\hat{\boldsymbol{\beta}}_d$ , the solution to  $\mathbf{U}_d(\boldsymbol{\beta}) = \mathbf{0}$ , where:

$$\mathbf{U}_d(\boldsymbol{\beta}) = \sum_{j=1}^n \sum_{i=1}^{n_j} \int_0^\tau \{\mathbf{Z}_{ij}(s) - \bar{\mathbf{Z}}_j(s; \boldsymbol{\beta})\} dN_{ij}(s), \tag{5}$$

which is free of the infinite-dimensional baseline rates.

At this point, it is worth noting that we are departing from the traditional use of stratification. For example, Glidden and Vittinghoff (2004) give a detailed examination of methods for the analysis of clustered failure time data. Under the stratified-baseline marginal hazard model they consider, subjects within-cluster are assumed independent, given that allowance is made for cluster-(in this case, center-)specific baselines. On the other hand, we use stratification only to allow for differences in the baseline rates; not to explain the dependence among subjects within a cluster. In practice, suppose there are unmeasured covariates, (say,  $\mathbf{Z}_j$ ) which are either cluster-specific or correlated within cluster, but which are independent of the measured covariates,  $\mathbf{Z}_{ij}$ . Since  $\mathbf{Z}_j$  and  $\mathbf{Z}_{ij}$  are independent, the parameter corresponding to  $\mathbf{Z}_{ij}$ , namely  $\beta_0$ , can be estimated consistently even though  $\mathbf{Z}_j$  is not included in the model. However, although the independence of  $\mathbf{Z}_{ij}$  and  $\mathbf{Z}_j$  remedies concerns with respect to bias, subjects are still not independent within a cluster and such dependence must be captured for the covariance estimator to be consistent. This is a fact we demonstrate in our simulation study in Section 4 and also observe in our real-data analysis in Section 5.

Under model (2), using arguments analogous to those used to derive (5),  $\beta_0$  can be estimated by  $\hat{\beta}_c$ , the solution to  $\mathbf{U}_c(\beta) = \mathbf{0}$ , where:

$$\mathbf{U}_c(\beta) = \sum_{j=1}^n \sum_{i=1}^{n_j} \int_0^\tau \{ \mathbf{Z}_{ij}(s) - \bar{\mathbf{Z}}(s; \beta) \} dN_{ij}(s), \tag{6}$$

with  $\bar{\mathbf{Z}}(s; \beta) = \mathbf{S}^{(1)}(s; \beta) / S^{(0)}(s; \beta)$  and  $\mathbf{S}^{(d)}(s; \beta) = n^{-1} \sum_{j=1}^n \sum_{i=1}^{n_j} I(C_{ij} > s) \mathbf{Z}_{ij}(s)^{\otimes d} \exp\{\beta^T \mathbf{Z}_{ij}(s)\}$  for  $d=0,1,2$ . With respect to the baseline mean, using the fact that  $E[dM_{ij}^c(t; \beta_0) | \mathbf{Z}_{ij}(t)] = 0$ , where  $M_{ij}^c(t; \beta) = N_{ij}(t) - \int_0^t I(C_{ij} > s) \exp\{\beta^T \mathbf{Z}_{ij}(s)\} d\mu_0(s)$  a Breslow–Aalen-type estimator (Breslow, 1972; Aalen, 1978) of  $\mu_0(t)$  is given by:  $\hat{\mu}_0(t; \hat{\beta}_c)$ , where

$$\hat{\mu}_0(t; \beta) = n^{-1} \sum_{j=1}^n \sum_{i=1}^{n_j} \int_0^t S^{(0)}(s; \beta)^{-1} dN_{ij}(s). \tag{7}$$

### 3. Asymptotic Properties

We now describe the large-sample properties of the parameter estimators for models (1) and (2), in that order, beginning with the regularity conditions for model (1):

- (a)  $P(C_{ij} \geq \tau) > 0$ , for some constant,  $\tau > 0$
- (b)  $N_{ij}(\tau) < \eta < \infty$  almost surely, for  $j=1, \dots, n$  and  $i=1, \dots, n_j$
- (c)  $|\mathbf{Z}_{ijk}(0)| + \int_0^\tau |d\mathbf{Z}_{ijk}(s)| < c_Z < \infty$  almost surely
- (d) Positive-definiteness of the matrix,  $\mathbf{\Omega}_d(\beta_0)$ , where:

$$\begin{aligned}\mathbf{\Omega}_d(\boldsymbol{\beta}) &= \lim_{n \rightarrow \infty} \widehat{\mathbf{\Omega}}_d(\boldsymbol{\beta}) \\ \widehat{\mathbf{\Omega}}_d(\boldsymbol{\beta}) &= n^{-1} \sum_{j=1}^n \sum_{i=1}^{n_j} \int_0^\tau \mathbf{V}_j(s; \boldsymbol{\beta}_0) \mathbf{S}_j^{(0)}(s; \boldsymbol{\beta}_0)^{-1} dN_{ij}(s),\end{aligned}\tag{8}$$

$$\text{with } \mathbf{V}_j(s; \boldsymbol{\beta}) = \mathbf{S}_j^{(2)}(s; \boldsymbol{\beta}) \mathbf{S}_j^{(0)}(s; \boldsymbol{\beta})^{-1} - \overline{\mathbf{Z}}_j(s; \boldsymbol{\beta})^{\otimes 2}.$$

(e)  $n_j < \kappa < \infty$ .

(f) Lindeberg Condition:

$$n^{-1} \sum_{j=1}^n E \left[ \boldsymbol{\Psi}_j^d(\boldsymbol{\beta}_0)^{\otimes 2} I(\|\boldsymbol{\Psi}_j^d(\boldsymbol{\beta}_0)\| > n^{1/2} \epsilon) \right] \rightarrow 0,$$

for all  $\epsilon > 0$ , where:

$$\boldsymbol{\Psi}_j^d(\boldsymbol{\beta}) = \sum_{i=1}^{n_j} \int_0^\tau \{ \mathbf{Z}_{ij}(s) - \overline{\mathbf{Z}}_j(s; \boldsymbol{\beta}) \} dM_{ij}^d(s; \boldsymbol{\beta}).\tag{9}$$

Assumption (a) is a standard identifiability condition. The boundedness of the event and covariate processes for an individual in conditions (b) and (c), respectively, is necessary to bound several integrals involved in the consistency and asymptotic normality proofs for  $\widehat{\boldsymbol{\beta}}_d$ . The positive-definiteness of the second derivative matrix in condition (d) is used in the consistency proof for  $\widehat{\boldsymbol{\beta}}_d$  to demonstrate concavity of the function whose derivative generates the estimating equation for  $\widehat{\boldsymbol{\beta}}_d$ . Bounded cluster sizes in condition (e) are out of convenience, to simplify the asymptotics. In practice, this assumption is realistic and does not pose a major constraint. For example, in a multi-center study across several medical centers, the number of patients each unit accepts is either implicitly or explicitly bounded due to equipment and resource limitations. Finally, the Lindeberg-type condition, (f), ensures that no single cluster dominates the parameter estimation.

The asymptotic behavior of the regression parameter estimator in model (1) is summarized by the following theorem.

**THEOREM 1** *Under conditions (a) to (f),  $\widehat{\boldsymbol{\beta}}_d$  converges to  $\boldsymbol{\beta}_0$  almost surely, and  $n^{1/2}(\widehat{\boldsymbol{\beta}}_d - \boldsymbol{\beta}_0)$  converges in distribution to a mean-zero normal random vector with covariance  $\mathbf{\Omega}_d(\boldsymbol{\beta}_0)^{-1} \boldsymbol{\Sigma}_d(\boldsymbol{\beta}_0) \mathbf{\Omega}_d(\boldsymbol{\beta}_0)^{-1}$ , with  $\mathbf{\Omega}_d(\boldsymbol{\beta})$  defined in (8) and  $\boldsymbol{\Sigma}_d(\boldsymbol{\beta}) = \lim_{n \rightarrow \infty} n^{-1} \sum_{j=1}^n E[\boldsymbol{\Psi}_j^d(\boldsymbol{\beta})^{\otimes 2}]$ .*

The asymptotic covariance can be estimated empirically by replacing limiting values with their observed empirical counterparts. The proof of almost sure convergence in Theorem 1 involves several applications of the Strong Law of Large Numbers, followed by results from convex function theory. The proof of asymptotic normality combines the Multivariate Central Limit Theorem with various

applications of results from the modern theory of empirical processes. An outline of the proof of Theorem 1 is provided in the Appendix.

We now provide asymptotic results for the parameter estimators in model (2). Regularity conditions (a), (b), (c) and (e) are as previously-listed. We add the following condition:

(g) There exists  $\mathbf{s}^{(d)}(s; \boldsymbol{\beta})$  such that

$$\sup_{s \in [0, \tau]} \|\mathbf{S}^{(d)}(s; \boldsymbol{\beta}) - \mathbf{s}^{(d)}(s; \boldsymbol{\beta})\| \xrightarrow{a.s.} 0,$$

for  $d=0,1,2$  and all  $\boldsymbol{\beta}$  in a compact set.

Regarding condition (d), positive-definiteness is assumed for the following matrix:

$$\boldsymbol{\Omega}_c(\boldsymbol{\beta}_0) = \int_0^\tau \mathbf{v}(s; \boldsymbol{\beta}_0) s^{(0)}(s; \boldsymbol{\beta}_0) d\mu_0(s), \quad (10)$$

where  $\mathbf{v}(s; \boldsymbol{\beta}) = \mathbf{s}^{(2)}(s; \boldsymbol{\beta})/s^{(0)}(s; \boldsymbol{\beta}) - \bar{\mathbf{z}}(s; \boldsymbol{\beta})^{\otimes 2}$  and  $\bar{\mathbf{z}}(s; \boldsymbol{\beta}) = \mathbf{s}^{(1)}(s; \boldsymbol{\beta})/s^{(0)}(s; \boldsymbol{\beta})$ . In condition (f), we replace  $\boldsymbol{\Psi}_j^d(\boldsymbol{\beta}_0)$  with:

$$\boldsymbol{\Psi}_j^c(\boldsymbol{\beta}) = \sum_{i=1}^{n_j} \int_0^\tau \{\mathbf{Z}_{ij}(s) - \bar{\mathbf{z}}(s; \boldsymbol{\beta})\} dM_{ij}^c(s; \boldsymbol{\beta}). \quad (11)$$

We now describe the essential asymptotic results pertaining to model (2), beginning with the regression parameter estimator.

**THEOREM 2** *Under conditions (a) to (g),  $\widehat{\boldsymbol{\beta}}_c$  is strongly consistent for  $\boldsymbol{\beta}_0$  and  $n^{1/2}(\widehat{\boldsymbol{\beta}}_c - \boldsymbol{\beta}_0)$  is asymptotically normally distributed with mean  $\mathbf{0}$  and covariance  $\boldsymbol{\Omega}_c(\boldsymbol{\beta}_0)^{-1} \boldsymbol{\Sigma}_c(\boldsymbol{\beta}_0) \boldsymbol{\Omega}_c(\boldsymbol{\beta}_0)^{-1}$ , where  $\boldsymbol{\Omega}_c(\boldsymbol{\beta})$  is as defined in (10) and  $\boldsymbol{\Sigma}_c(\boldsymbol{\beta}) = \lim_{n \rightarrow \infty} n^{-1} \sum_{j=1}^n E[\boldsymbol{\Psi}_j^c(\boldsymbol{\beta})^{\otimes 2}]$ .*

The proof of Theorem 2 is similar to that of Theorem 1, but simpler. We next consider the baseline mean estimator for model (2), starting with the definition,  $\widehat{\phi}_0(t) = \{\widehat{\mu}_0(t; \widehat{\boldsymbol{\beta}}_n) - \mu_0(t)\}$ . The asymptotic behavior of  $\{\widehat{\phi}_0(t); t \in [0, \tau]\}$  is summarized by the following theorem.

**THEOREM 3** *Under conditions (a) to (g),  $\widehat{\phi}_0(t)$  converges almost surely to 0, uniformly in  $t \in [0, \tau]$ ; in addition,  $n^{1/2}\widehat{\phi}_0(t)$  converges weakly to a Gaussian process with mean zero and covariance function given by  $\sigma_0(s, t)$ , where:*

$$\sigma_0(s, t) = \lim_{n \rightarrow \infty} n^{-1} \sum_{j=1}^n E[\xi_j(s; \boldsymbol{\beta}_0) \xi_j(t; \boldsymbol{\beta}_0)]$$

$$\xi_j(t; \boldsymbol{\beta}) = \mathbf{h}(t; \boldsymbol{\beta})^T \boldsymbol{\Omega}_c(\boldsymbol{\beta})^{-1} \boldsymbol{\Psi}_j^c(\boldsymbol{\beta}) + \sum_{i=1}^{n_j} \int_0^t s^{(0)}(s; \boldsymbol{\beta})^{-1} dM_{ij}^c(s; \boldsymbol{\beta}),$$

$$\text{with } \mathbf{h}(t; \boldsymbol{\beta}) = - \int_0^t \bar{\mathbf{z}}(s; \boldsymbol{\beta}) d\mu_0(s).$$

Theorem 3 can be proved by decomposing  $\widehat{\phi}_0(t)$  into  $\{\widehat{\mu}_0(t; \widehat{\boldsymbol{\beta}}_n) - \widehat{\mu}_0(t; \boldsymbol{\beta}_0)\} + \{\widehat{\mu}_0(t; \boldsymbol{\beta}_0) - \mu_0(t)\}$  and applying the Uniform Strong Law of Large Numbers (Pollard, 1990) and various empirical process results (van der Vaart and Wellner, 1996). A proof of Theorem 3 is outlined in the Appendix.

#### 4. Simulation Study

Finite-sample properties of the proposed regression parameter estimators were assessed through simulation. First, for each cluster, we generated a latent (frailty) variate,  $R_j$ , inducing positive correlation among subjects within a cluster, with  $R_j \sim \text{Gamma}(\sigma_R^{-2}, \sigma_R^{-2})$ . Next, a subject-specific frailty was generated,  $Q_i \sim \text{Gamma}(\sigma_Q^{-2}, \sigma_Q^{-2})$  to induce correlation among the within-subject event times. Note that  $E[Q_i R_j] = 1$ , while  $\sigma_{QR}^2 \equiv V(Q_i R_j) = (\sigma_Q^2 + 1)(\sigma_R^2 + 1) - 1$ . The magnitude of the within-cluster and within-subject correlation increases with increasing  $\sigma_R^2$  and  $\sigma_Q^2$ , respectively. We chose  $\sigma_R^2 = 0.25, 0.5$  and  $\sigma_Q^2 = 0.5, 1.0$ , since, in most practical situations, the within-subject correlation will be greater than that for between-subject. A covariate,  $Z_{ij}$ , was generated from the Bernoulli(0.5) distribution. For each subject, the  $k$ th event time for the  $i$ th subject within the  $j$ th cluster is given by:

$$T_{i,j,k} = T_{i,j,k-1} - \log\{1 - U_{ijk}\} \{Q_i R_j d\mu_0 \exp\{\beta_0 Z_{ij}\}\}^{-1} \tag{12}$$

where  $U_{ijk}$  are independent Uniform(0,1) variates and  $T_{i,j,0} \equiv 0$ . Setting  $\mu_0(t) = 0.25t, 0.5t$  and  $0.75t$  implies the conditional model,

$$E[N_{ij}^*(t) | Z_{ij}, Q_i, R_j] = Q_i R_j \mu_0 t \exp\{\beta_0 Z_{ij}\},$$

as well as the following marginal model:

$$E[N_{ij}^*(t) | Z_{ij}] = \mu_0 t \exp\{\beta_0 Z_{ij}\}.$$

Censoring times were generated as  $C_{ij} \sim \text{Uniform}(0,5)$ , representing a study of length 5 time units with randomly staggered entry times and administrative censoring only. Observed events were those such that  $T_{i,j,k} < C_{ij}$  and the expected number of observed events per subject ranged from 0.625 to 1.875 for  $Z_{ij} = 0$ , to 1.25 to 3.75 when  $Z_{ij} = 1$ . Each data configuration was replicated 500 times.

The accuracy of the large-sample distributional approximation of  $\widehat{\beta}_c$  from model (2) is assessed in Table 1 for  $n = 50$  and  $n = 100$ , with  $\beta_0$  set to  $\log(2)$  and cluster sizes  $n_j = 2, n_j = 5$  and  $n_j = 20$ . It can be seen that  $\widehat{\beta}_c$  is approximately unbiased for



Table 1. Simulation results: common baseline rate model – bias, accuracy of asymptotic distributional approximation.

$n_j$	$\sigma_R^2$	$\sigma_Q^2$	$d\mu_0(t)/dt$	$n = 50$				$n = 100$				
				Bias	ASE	ESD	CP	Bias	ASE	ESD	CP	
2	0.25	0.5	0.25	-0.006	0.297	0.322	0.92	0.000	0.214	0.218	0.93	
			0.5	0.001	0.253	0.260	0.93	0.007	0.183	0.196	0.93	
			0.75	-0.001	0.239	0.259	0.92	0.002	0.172	0.173	0.96	
	0.5	1.0	0.25	-0.007	0.339	0.376	0.92	-0.007	0.244	0.257	0.94	
			0.5	-0.005	0.300	0.322	0.92	0.007	0.217	0.232	0.93	
			0.75	-0.005	0.286	0.316	0.91	-0.004	0.210	0.215	0.94	
		0.5	0.5	0.25	-0.008	0.320	0.368	0.91	0.002	0.233	0.236	0.94
				0.5	0.002	0.282	0.306	0.93	-0.003	0.205	0.213	0.94
				0.75	-0.010	0.266	0.277	0.94	-0.003	0.193	0.203	0.94
	1.0	0.5	0.25	0.013	0.365	0.394	0.94	-0.007	0.265	0.267	0.95	
			0.5	-0.003	0.329	0.371	0.90	-0.004	0.242	0.256	0.93	
			0.75	-0.009	0.316	0.342	0.91	-0.004	0.230	0.247	0.93	
5	0.25	0.5	0.25	0.000	0.189	0.195	0.94	-0.001	0.135	0.140	0.94	
			0.5	0.000	0.162	0.163	0.94	0.001	0.116	0.119	0.94	
			0.75	0.000	0.151	0.152	0.94	-0.002	0.110	0.110	0.95	
	1.0	0.5	0.25	-0.002	0.217	0.226	0.94	0.000	0.157	0.161	0.94	
			0.5	-0.003	0.193	0.204	0.93	0.003	0.141	0.146	0.94	
			0.75	-0.009	0.186	0.197	0.93	0.004	0.134	0.136	0.94	
	0.5	0.5	0.25	0.006	0.205	0.228	0.92	-0.007	0.148	0.151	0.95	
			0.5	0.006	0.181	0.188	0.93	-0.002	0.130	0.133	0.94	
			0.75	0.002	0.170	0.181	0.93	0.000	0.125	0.123	0.95	
		1.0	0.5	0.25	-0.005	0.236	0.250	0.93	-0.012	0.170	0.176	0.93
				0.5	0.007	0.213	0.221	0.93	0.007	0.154	0.159	0.94
				0.75	0.000	0.203	0.220	0.92	-0.007	0.149	0.151	0.95
20	0.25	0.5	0.25	-0.004	0.095	0.097	0.94	0.000	0.068	0.069	0.95	
			0.5	-0.001	0.081	0.081	0.95	0.000	0.058	0.059	0.95	
			0.75	-0.004	0.076	0.079	0.93	0.000	0.055	0.057	0.95	
	1.0	0.5	0.25	0.000	0.110	0.109	0.95	-0.002	0.079	0.079	0.95	
			0.5	-0.002	0.099	0.099	0.94	0.001	0.071	0.067	0.95	
			0.75	-0.002	0.094	0.099	0.93	-0.001	0.067	0.067	0.95	
	0.5	0.5	0.25	0.001	0.103	0.105	0.94	-0.003	0.074	0.075	0.94	
			0.5	-0.002	0.091	0.097	0.93	0.004	0.065	0.067	0.94	
			0.75	-0.005	0.087	0.091	0.93	-0.001	0.061	0.068	0.94	
		1.0	0.5	0.25	-0.007	0.118	0.130	0.92	0.003	0.086	0.087	0.95
				0.5	0.001	0.107	0.117	0.92	-0.010	0.079	0.074	0.94
				0.75	-0.001	0.104	0.112	0.94	-0.008	0.075	0.075	0.94

all data configurations. The average approximated standard error, denoted by ASE, closely approximates the empirical standard deviation, ESD, with empirical coverage probabilities (CP) close to the nominal value, 0.95. For  $n = 50$ , slight under-coverage occurs, which reduced progressively with increasing cluster size.

We next examined the setting where the baseline rates are cluster-specific by replacing  $\mu_0(t)$  in (12) by:

Table 2. Simulation results: cluster-specific baseline rate model – bias, accuracy of asymptotic distributional approximation.

$n_j$	$\sigma_R^2$	$\sigma_Q^2$	$d\bar{\mu}_0(t)$	$n=50$				$n=100$			
				Bias	ASE	ESD	CP	Bias	ASE	ESD	CP
5	0.25	0.5	0.25	0.007	0.208	0.228	0.93	0.010	0.150	0.150	0.94
			0.5	0.014	0.171	0.179	0.92	-0.007	0.163	0.171	0.93
			0.75	0.007	0.160	0.162	0.94	-0.008	0.115	0.114	0.94
		1.0	0.25	0.017	0.242	0.256	0.92	0.007	0.178	0.188	0.93
			0.5	-0.006	0.212	0.220	0.92	-0.008	0.155	0.161	0.95
			0.75	0.008	0.201	0.206	0.92	-0.003	0.145	0.159	0.93
	0.5	0.5	0.25	0.009	0.214	0.238	0.92	0.020	0.155	0.168	0.92
			0.5	0.016	0.177	0.186	0.94	0.002	0.130	0.128	0.95
			0.75	0.016	0.165	0.180	0.92	0.000	0.121	0.115	0.95
		1.0	0.25	-0.016	0.252	0.266	0.94	0.030	0.185	0.182	0.96
			0.5	-0.020	0.225	0.241	0.92	-0.007	0.163	0.171	0.93
			0.75	-0.011	0.212	0.236	0.92	-0.009	0.156	0.161	0.92
10	0.25	0.5	0.25	0.014	0.137	0.139	0.93	-0.003	0.098	0.103	0.93
			0.5	0.001	0.114	0.117	0.93	0.000	0.082	0.083	0.95
			0.75	0.005	0.107	0.116	0.94	0.003	0.076	0.075	0.96
		1.0	0.25	-0.010	0.159	0.164	0.94	0.008	0.116	0.119	0.94
			0.5	-0.001	0.141	0.145	0.94	-0.003	0.102	0.104	0.94
			0.75	-0.005	0.133	0.138	0.93	-0.002	0.096	0.097	0.94
	0.5	0.5	0.25	0.003	0.141	0.148	0.93	-0.008	0.101	0.105	0.95
			0.5	-0.011	0.119	0.126	0.91	-0.002	0.086	0.087	0.94
			0.75	-0.005	0.109	0.115	0.93	-0.006	0.080	0.077	0.95
		1.0	0.25	-0.009	0.168	0.178	0.93	-0.004	0.121	0.127	0.94
			0.5	-0.008	0.147	0.158	0.93	0.003	0.107	0.107	0.96
			0.75	0.000	0.140	0.154	0.90	-0.005	0.104	0.102	0.93
2,5,10,20,	0.25	0.5	0.25	0.005	0.126	0.126	0.94	0.009	0.090	0.092	0.94
			0.5	-0.002	0.103	0.108	0.95	-0.003	0.074	0.075	0.95
			0.75	0.005	0.095	0.098	0.93	0.002	0.069	0.067	0.95
		1.0	0.25	0.006	0.146	0.140	0.93	-0.009	0.105	0.105	0.94
			0.5	-0.007	0.126	0.135	0.92	-0.001	0.092	0.097	0.93
			0.75	-0.008	0.119	0.123	0.93	-0.003	0.087	0.087	0.95
	0.5	0.5	0.25	-0.004	0.127	0.140	0.91	0.002	0.093	0.093	0.94
			0.5	0.011	0.108	0.114	0.93	0.001	0.079	0.078	0.96
			0.75	-0.005	0.099	0.108	0.93	0.007	0.072	0.073	0.95
		1.0	0.25	-0.002	0.151	0.163	0.92	-0.003	0.111	0.117	0.94
			0.5	0.004	0.132	0.140	0.91	0.003	0.097	0.100	0.94
			0.75	-0.006	0.126	0.132	0.93	-0.005	0.091	0.094	0.95

$$\mu_{0j}(t) = \{1 + (2n)^{-1}\}^{-1} \left\{1 + \frac{j - n/2}{n}\right\} \mu_0(t) \quad (13)$$

We set  $\beta_0 = \log(2)$ ,  $\mu_0(t) = 0.25t$ ,  $\mu_0(t) = 0.5t$  and  $\mu_0(t) = 0.75t$ , with  $n = 50$  then  $n = 100$ . Under (13), the cluster-specific baseline means vary from  $0.5 \times \mu_0(t)$  to  $1.5 \times \mu_0(t)$ . Cluster sizes were set to  $n_j = 5$ ,  $n_j = 10$ , and a third set-up where cluster

Table 3. Simulation results: common baseline model – estimated variance and empirical type-I error rates based on different variance estimators ( $n = 100$ ,  $\beta = 0$ ,  $\mu_{0j}(t) = \mu_0(t)$ ).

$n_j$	$\sigma_R^2$	$\sigma_Q^2$	$d\mu_0(t)/dt$	$100 \times \text{AEV}$			EMPV	Empirical Type-I error		
				Eq. (14)	Eq. (15)	Thm 2		Eq. (14)	Eq. (15)	Thm 2
2	0.25	0.5	0.25	3.59	5.40	5.80	6.08	0.130	0.062	0.052
			0.5	2.05	3.82	4.25	4.72	0.188	0.076	0.058
			0.75	1.53	3.27	3.69	3.89	0.226	0.072	0.052
		1.0	0.25	3.88	7.02	7.41	7.66	0.158	0.064	0.060
			0.5	2.28	5.38	5.78	6.08	0.238	0.078	0.058
			0.75	1.74	4.79	5.19	5.52	0.268	0.080	0.064
	0.5	0.5	0.25	3.81	6.39	7.24	8.37	0.178	0.078	0.066
			0.5	2.19	4.62	5.44	5.70	0.224	0.094	0.066
			0.75	1.65	4.21	5.13	5.65	0.304	0.096	0.066
		1.0	0.25	3.95	8.04	8.88	8.93	0.198	0.066	0.056
			0.5	2.40	6.54	7.32	7.58	0.282	0.068	0.056
			0.75	1.84	5.93	6.72	6.61	0.298	0.068	0.056
5	0.25	0.5	0.25	1.44	2.22	2.96	3.02	0.172	0.092	0.052
			0.5	0.82	1.56	2.31	2.42	0.258	0.130	0.058
			0.75	0.61	1.34	2.09	2.04	0.290	0.124	0.046
		1.0	0.25	1.53	2.88	3.57	3.89	0.218	0.092	0.058
			0.5	0.90	2.17	2.88	3.24	0.326	0.116	0.064
			0.75	0.70	1.99	2.71	2.53	0.284	0.102	0.064
	0.5	0.5	0.25	1.50	2.59	4.05	4.11	0.252	0.116	0.058
			0.5	0.87	1.95	3.47	3.63	0.338	0.140	0.064
			0.75	0.66	1.70	3.18	3.13	0.378	0.148	0.058
		1.0	0.25	1.57	3.42	4.88	4.80	0.260	0.108	0.062
			0.5	0.94	2.69	4.12	3.98	0.338	0.112	0.054
			0.75	0.74	2.48	3.94	3.89	0.386	0.114	0.058
10	0.25	0.5	0.25	0.72	1.11	1.96	2.00	0.246	0.148	0.056
			0.5	0.41	0.79	1.69	1.69	0.346	0.180	0.050
			0.75	0.30	0.68	1.56	1.69	0.412	0.226	0.056
		1.0	0.25	0.76	1.43	2.28	2.35	0.258	0.132	0.066
			0.5	0.45	1.12	1.94	2.06	0.410	0.162	0.048
			0.75	0.35	1.00	1.85	1.74	0.369	0.170	0.059
	0.5	0.5	0.25	0.74	1.31	3.02	3.26	0.346	0.222	0.066
			0.5	0.44	0.99	2.73	2.95	0.436	0.260	0.062
			0.75	0.33	0.88	2.65	2.62	0.504	0.228	0.054
		1.0	0.25	0.77	1.68	3.35	3.62	0.350	0.182	0.066
			0.5	0.48	1.36	3.07	3.53	0.514	0.214	0.066
			0.75	0.37	1.25	2.95	3.12	0.520	0.236	0.052

size is a mixture of  $n_j=2$  (10%),  $n_j=5$  (40%),  $n_j=10$  (40%) and  $n_j=20$  (10%). We examine the performance of model (1) in Table 2, where the average baseline mean is denoted by  $\bar{\mu}_0(t) = n^{-1} \sum_{j=1}^n \mu_{0j}(t)$  and is equal to  $\mu_0(t)$  under (13). Results in Table 2 indicate that  $\hat{\beta}_d$  from model (1) is approximately unbiased, with asymptotic standard errors which yield empirical coverage probabilities close to

the nominal 95%; the degree of coverage was closer to the nominal level for  $n=100$  compared to  $n=50$ .

Next, we set  $\beta_0=0$  and evaluated the performance of three Wald statistics,  $W_n = \hat{\beta}_c^2 / \hat{V}(\hat{\beta}_c)$ , in testing  $H_0: \beta_0=0$  against  $H_1: \beta_0 \neq 0$ . Data were generated from the common baseline model through (12), with  $n=100$  and  $n_j=2, 5,$  and  $10$ . The three test statistics differ in that a different variance estimator was used for each. For the first test statistic, independence is assumed for both subjects within-cluster and events within-subject, with the variance estimator given by:

$$\hat{\Omega}_c(\hat{\beta}_c)^{-1} \left( \sum_{j=1}^n \sum_{i=1}^{n_j} N_{ij}(\tau) \right)^{-1} \sum_{j=1}^n \sum_{i=1}^{n_j} \int_0^\tau \left\{ \mathbf{Z}_{ij} - \bar{\mathbf{Z}}(s; \hat{\beta}_c) \right\} d\hat{M}_{ij}^c(s; \hat{\beta}_c) \left\}^{\otimes 2} \hat{\Omega}_c(\hat{\beta}_c)^{-1}, \quad (14)$$

which is the variance which would be derived under the proportional rates model if one assumed that events-within-subject and subjects-within-cluster were uncorrelated. The second test statistic assumes only that subjects within-cluster are independent, and employs the following variance estimator:

$$\hat{\Omega}_c(\hat{\beta}_c)^{-1} \left( \sum_{j=1}^n n_j \right)^{-1} \sum_{j=1}^n \sum_{i=1}^{n_j} \left\{ \int_0^\tau \left\{ \mathbf{Z}_{ij} - \bar{\mathbf{Z}}(s; \hat{\beta}_c) \right\} d\hat{M}_{ij}^c(s; \hat{\beta}_c) \right\}^{\otimes 2} \hat{\Omega}_c(\hat{\beta}_c)^{-1}, \quad (15)$$

This is the variance estimator proposed by Lin et al. (2000) for recurrent event data, but with independent subjects. The third test statistic is based on the proposed variance estimator corresponding to Theorem 2, specifically,

$$\hat{\Omega}_c(\hat{\beta}_c)^{-1} n^{-1} \sum_{j=1}^n \left\{ \sum_{i=1}^{n_j} \int_0^\tau \left\{ \mathbf{Z}_{ij} - \bar{\mathbf{Z}}(s; \hat{\beta}_c) \right\} d\hat{M}_{ij}^c(s; \hat{\beta}_c) \right\}^{\otimes 2} \hat{\Omega}_c(\hat{\beta}_c)^{-1}. \quad (16)$$

Average estimated variances (AEV) based on (14) and (15), as well as the proposed method are presented in Table 3, along with corresponding empirical Type-I error rates. Empirical variances are also listed and are denoted EMPV. The proposed variance estimator is always closest to the empirical variance and the proposed test statistic rejects  $H_0: \beta_0=0$  with frequency approximately equal to the nominal value, 0.05, for almost all data configurations. Type I error rates based on test statistics assuming within-cluster/within-subject independence, or just within-cluster independence, were often grossly inflated. Although both were often unacceptably greater than the nominal level, the type I error rate for estimator (15) was closer to the nominal value than (14), which is consistent with the fact that its underlying assumptions, while false, are closer to the truth. With cluster size of  $n_j=2$ , the variance estimator which ignores within-cluster correlation actually does not perform very poorly. Although (15) incorrectly assumes that

$$E \left[ \sum_{i=1}^{n_j-1} \int_0^\tau \{ \mathbf{Z}_{ij} - \bar{\mathbf{Z}}(s; \boldsymbol{\beta}_0) \} dM_{ij}^c(s; \boldsymbol{\beta}_0) \sum_{\ell=i+1}^{n_j} \int_0^\tau \{ \mathbf{Z}_{\ell j} - \bar{\mathbf{Z}}(s; \boldsymbol{\beta}_0) \} dM_{\ell j}^c(s; \boldsymbol{\beta}_0) \right]$$

equals zero, few cross-product terms are ignored since cluster sizes are so small. Naturally, for larger cluster sizes, such as  $n_j=5$  and  $n_j=10$ , method (15) does not fare as well. The corresponding results for the cluster-specific baseline rate model (Table 4) were similar, although much less extreme. The stratified model allows the regression parameter to be consistently estimated when the baseline rate is center-specific, as opposed to being equal across all clusters. However, the equality of the baseline rates across clusters and independence of subjects within a cluster are two separate issues and the allowance of cluster-specific baselines need not accommodate the latter. As an example, consider an unmeasured cluster-specific covariate, such as  $R_j$ , for which data are unavailable, which affects the event rate but is independent of the covariate of interest (e.g.,  $\mathbf{Z}_{ij}$ ). Since  $R_j$  and  $\mathbf{Z}_{ij}$  are independent, the regression parameter corresponding to  $\mathbf{Z}_{ij}$  can be estimated consistently even though  $R_j$  is not included in the model. However, the variance will be mis-specified, as demonstrated by the results in Table 4.

Finally, we examined the case where the baseline mean is cluster-specific, but the common baseline mean model is fitted. We chose  $n=50$ ,  $n_j=10$ ,  $\beta_0=\log(2)$ , and  $\mu_{0j}(t)$  as specified in (13), again with  $\mu_0(t)=0.25t$ ,  $0.5t$  and  $0.75t$ . Within this framework, we considered two set-ups. In the first set-up,  $Z_{ij}$  was distributed as Bernoulli(0.5); i.e., constant across all clusters. In the second set-up,  $Z_{ij} \sim \text{Bernoulli}(p_j)$ , where,

Table 4. Simulation results: cluster-specific baseline model estimated variance and empirical type-I error rates based on different variance estimators ( $n=100$ ,  $\beta=0$ ).

$n_j$	$\sigma_R^2$	$\sigma_Q^2$	$d\bar{\mu}_0(t)/dt$	100 × AEV			EMPV	Empirical Type-I error		
				(14)	(15)	(16)		(14)	(15)	(16)
10	0.25	0.5	0.25	0.79	0.96	1.14	1.19	0.090	0.072	0.058
			0.50	0.40	0.63	0.76	0.78	0.156	0.074	0.062
			0.75	0.26	0.52	0.63	0.54	0.182	0.066	0.046
		1.0	0.25	0.79	1.24	1.50	1.59	0.152	0.082	0.062
			0.50	0.39	0.92	1.12	1.20	0.250	0.088	0.064
			0.75	0.26	0.81	0.98	1.12	0.352	0.098	0.064
	0.5	0.5	0.25	0.79	1.01	1.18	1.16	0.094	0.066	0.048
			0.50	0.39	0.68	0.81	0.72	0.148	0.064	0.040
			0.75	0.26	0.57	0.68	0.70	0.238	0.088	0.058
		1.0	0.25	0.40	1.02	1.24	1.28	0.266	0.086	0.062
			0.50	0.79	1.36	1.63	1.41	0.158	0.056	0.040
			0.75	0.26	0.91	1.11	1.25	0.380	0.080	0.064

Table 5. Simulation results: common baseline rate model-bias, accuracy of asymptotic distributional approximation (when baseline rate is actually cluster-specific).

$n_j$	$\sigma_R^2$	$\sigma_Q^2$	$d\bar{\mu}_0(t)/dt$	$\mu_{0j}(t) \neq \mu_0(t), Z_{ij} \sim \text{Ber}(0.5)$				$\mu_{0j}(t) \neq \mu_0(t), Z_{ij} \sim \text{Ber}(p_j)$				
				Bias	ASE	ESD	CP	Bias	ASE	ESD	CP	
10	0.25	0.5	0.25	-0.008	0.138	0.146	0.94	0.169	0.147	0.145	0.77	
			0.50	0.001	0.121	0.129	0.92	0.167	0.134	0.126	0.72	
			0.75	-0.001	0.113	0.121	0.94	0.164	0.134	0.119	0.70	
		1.0	0.25	0.007	0.160	0.158	0.94	0.169	0.166	0.166	0.82	
			0.50	-0.002	0.145	0.153	0.92	0.161	0.152	0.150	0.80	
			0.75	-0.005	0.139	0.147	0.94	0.165	0.158	0.142	0.74	
		0.5	0.5	0.25	0.005	0.149	0.151	0.95	0.149	0.168	0.158	0.84
				0.50	0.012	0.134	0.133	0.96	0.150	0.150	0.145	0.81
				0.75	0.000	0.126	0.132	0.93	0.165	0.144	0.138	0.76
		1.0	0.25	-0.004	0.174	0.186	0.94	0.180	0.197	0.181	0.81	
			0.50	0.008	0.160	0.171	0.92	0.157	0.171	0.165	0.85	
			0.75	0.003	0.150	0.162	0.93	0.153	0.161	0.159	0.84	

$$p_j = 0.5 \times \{1 + (2n)^{-1}\}^{-1} \left\{1 + \frac{j - n/2}{n}\right\}. \quad (17)$$

Note that, under (17),  $n^{-1} \sum_{j=1}^n p_j = 0.5$ . Results are presented in Table 5. Even when the common-baseline model is fitted to a cluster-specific baseline data structure,  $\hat{\beta}_c$  remained unbiased in the setting where the covariate distribution did not differ by cluster. However, in the setting where the covariate's distribution was cluster-specific,  $\hat{\beta}_c$  was greatly biased.

## 5. Analysis of CORR Dialysis Data

We applied the methods to the analysis of technique failures (TF) among Canadian peritoneal dialysis (PD) patients. Data were obtained from the Canadian Organ Replacement Register. Basic demographic and clinical information is collected at the time each patient begins therapy for end-stage renal disease. Follow-up information on treatment assignments and switches, kidney transplants, as well as mortality data are reported by each renal center. The 9861 patients who received PD between January 1, 1990 and December 31, 1997 were included in the analysis, with the observation period concluding on December 31, 1997.

As mentioned previously, the primary goal was to compare covariate-adjusted TF rates by region. Adjustment covariates included age, gender, race, calendar period, underlying disease leading to renal failure, comorbid conditions and type of PD. Patients began follow-up when they initiated PD. Technique failure was considered to be a switch from PD to hemodialysis. Patients were clustered by center. There were  $n=85$  centers; cluster size averaged 155 patients (median of

Table 6. Analysis of CORR dialysis data-estimated regression parameters for region and estimated standard errors by assumed level of independence.

Variance estimator	Independence assumptions	Region	$\widehat{\beta}_c$	$\widehat{SE}(\widehat{\beta}_c)$	$\chi_n^2$	$p$
(14)	Within-Patient	I (ref)	0	–	–	–
		II	–0.012	0.064	0.035	0.85
	Within-Center	III	–0.076	0.061	1.535	0.22
		IV	–0.245	0.067	13.396	< 0.01
	Independence	V	0.015	0.095	0.026	0.87
		VI	–0.186	0.051	13.431	< 0.01
		VII	0.189	0.085	4.919	0.03
(15)	Within-Center	I (ref)	0	–	–	–
		II	–0.012	0.065	0.033	0.86
	Independence	III	–0.076	0.065	1.343	0.25
		IV	–0.245	0.072	11.425	< 0.01
	Independence	V	0.015	0.106	0.021	0.89
		VI	–0.186	0.056	11.089	< 0.01
		VII	0.189	0.093	4.096	0.04
(16)	None	I (ref)	0	–	–	–
		II	–0.012	0.195	0.004	0.95
		III	–0.076	0.185	0.167	0.68
		IV	–0.245	0.159	2.366	0.12
		V	0.015	0.126	0.015	0.90
		VI	–0.186	0.184	1.021	0.31
		VII	0.189	0.236	0.639	0.42

89), and ranged from  $n_j = 5$  to  $n_j = 1332$ . Patients were followed from the time they began PD until the earliest of death, receipt of a kidney transplant, or conclusion of the observation period. In total, there were 3317 technique failures, for an average of 0.33 per patient; the minimum being 0 and maximum being 4. Since our objective was to compare PD technique failure rates between regions, we fitted the following conditional rate model:

$$E[dN_{ij}^*(t) | \mathbf{Z}_{ij}(t), PD_{ij}(t) = 1] = \exp\{\boldsymbol{\beta}_0^T \mathbf{Z}_{ij}(t)\} d\mu_0(t),$$

where  $PD_{ij}(t) = 1$  when the  $j$ th patient from center  $i$  is receiving PD and 0 otherwise. We can interpret  $E[dN_{ij}^*(t) | PD_{ij}(t) = 1]$  as the rate of transfer off PD due to technique failure. Note that the use of such a conditional model avoids issues pertaining to the dependent censoring (e.g., due to death or receipt of a kidney transplant, in which case dialysis ceases, but not due to TF). The proposed model and methods can be applied upon re-defining the risk set process as  $I(C_{ij} \geq t) PD_{ij}(t)$ . The independent censoring assumption pertaining to this analysis can be expressed as:

$$E[dN_{ij}^*(s)|\mathbf{Z}_{ij}(s), PD_{ij}(s) = 1, C_{ij} > s] = E[dN_{ij}^*(s)|\mathbf{Z}_{ij}(s), PD_{ij}(s) = 1].$$

The analysis was based on the common baseline model since the covariate of primary interest, region, is center-specific. If a center-specific baseline model had been used, then differences by region would not be identifiable because there is no variation by region among patients within the same center.

Results of our analysis are displayed in Table 6. The estimated coefficient,  $\widehat{\beta}_c$ , represents the covariate-adjusted log rate ratio for each region compared to Region I (reference). Results are presented for three methods of estimating  $V(\widehat{\beta}_c)$  which correspond to three different levels of assumed independence. The top portion of Table 6 pertains to a model which assumes that all TF's are independent (i.e., within-patient, and between-patient-within-center). It is found that Regions IV and VI have significantly lower TF rates than Region I, while TF rates in Region VII are significantly higher. The middle portion of Table 6 is based on a model which acknowledges lack of within-patient TF independence, but does assume that patients within-center are independent. Estimated standard errors are slightly increased, but are very comparable to those based on the pure independence assumption described in the preceding paragraph; in terms of statistical significance, conclusions remain the same (significantly decreased TF rate for Regions IV and VI; significantly increased TF rate for Region VII). In the bottom section of Table 6,  $V(\widehat{\beta}_c)$  is estimated using Theorem 2, which does not impose any restrictions on the correlation structure either within-patient or within-center. Standard errors estimated by Theorem 2 are notably elevated (differing by multiples of 1.3–3.6) compared to those which assume some degree of independence. No rate ratios approach statistical significance ( $p \geq 0.12$ ).

## 6. Discussion

We propose semiparametric regression methods for clustered recurrent event data. Two models are considered, both of which assume multiplicative covariate effects. The first model allows for distinct cluster-specific baseline event rates, while the second uses a common baseline. For each model, the regression parameter estimators are shown to be consistent and asymptotically normal estimators. For the common baseline model, a baseline mean estimator is proposed, which is shown to be uniformly consistent and, upon normalization, to converge to a zero-mean Gaussian process. Through simulation, both methods are shown to perform well in finite samples. The underestimation of variance is demonstrated in the setting where subjects are clustered, but assumed independent. Finally, the analysis of the dialysis data illustrates the potential to lead to false conclusions if the clustering of the subjects is not taken into account.

The choice between the common and cluster-specific baseline rate models, respectively, would depend on the nature of the data structure and aims of the investiga-



tor. In model (2), comparisons by covariate level are made within a cluster, which would rule out the use of this model if a cluster-level covariate is of interest, as was the case for the dialysis data in Section 5. In situations where cluster-specific baseline rates are distinct, a model which assumes a common baseline rate will yield unbiased estimates of the regression parameter, provided the covariate distributions are equal across clusters. In such cases, the estimated baseline mean can be validly interpreted as having been averaged across clusters. However, in settings where baseline rates and covariate distributions are distinct among clusters, regression coefficients may be greatly biased, if estimated through a common baseline model, as demonstrated in our simulation study.

The proposed methods are flexible in the sense that they make no assumptions about the nature of the dependence structure for either the repeated events within a subject or subjects within the same cluster. However, the lack of involvement of the correlation structure in the parameter estimation can also be seen as a drawback, with respect to lost efficiency. For example, a frailty model which parameterizes both the within-subject and within-cluster correlation may be a valuable method for handling clustered recurrent event data, particularly if the method accommodated hypothesis testing of the frailty parameters. Such a model has not appeared in the published literature and is currently under consideration.

## 7. Appendix

### 7.1. Proof of Theorem 1

We begin by setting:

$$\ell_n(\boldsymbol{\beta}) = \sum_{j=1}^n \sum_{i=1}^{n_j} \int_0^\tau \left\{ \boldsymbol{\beta}^T \mathbf{Z}_{ij}(s) - \log \left( n_j S_j^{(0)}(s; \boldsymbol{\beta}) \right) \right\} dN_{ij}(s),$$

then set  $D_n(\boldsymbol{\beta}) = n^{-1} \{ \ell_n(\boldsymbol{\beta}) - \ell_n(\boldsymbol{\beta}_0) \}$ , such that:

$$D_n(\boldsymbol{\beta}) = n^{-1} \sum_{j=1}^n \sum_{i=1}^{n_j} \int_0^\tau \left\{ (\boldsymbol{\beta} - \boldsymbol{\beta}_0)^T \mathbf{Z}_{ij}(s) + T_j(s; \boldsymbol{\beta}) \right\} dN_{ij}(s)$$

where  $T_j(s; \boldsymbol{\beta}) = \log \left( S_j^{(0)}(s; \boldsymbol{\beta}_0) / S_j^{(0)}(s; \boldsymbol{\beta}) \right)$ . Using conditions (a) to (f), and the Strong Law of Large Numbers (SLLN), it can be shown that  $D_n(\boldsymbol{\beta}) \xrightarrow{a.s.} D(\boldsymbol{\beta})$ , where:

$$D(\boldsymbol{\beta}) = (\boldsymbol{\beta} - \boldsymbol{\beta}_0)^T \bar{\mathbf{P}}(\boldsymbol{\beta}) + \bar{Q}(\boldsymbol{\beta}),$$

with  $\bar{\mathbf{P}}(\boldsymbol{\beta}) = \lim_{n \rightarrow \infty} \bar{\mathbf{P}}_n(\boldsymbol{\beta})$ ,  $\bar{Q}(\boldsymbol{\beta}) = \lim_{n \rightarrow \infty} \bar{Q}_n(\boldsymbol{\beta})$ ,  $\bar{\mathbf{P}}_n(\boldsymbol{\beta}) = n^{-1} \sum_{j=1}^n \mathbf{P}_j(\boldsymbol{\beta})$ ,  $\bar{Q}_n(\boldsymbol{\beta}) = n^{-1} \sum_{j=1}^n Q_j(\boldsymbol{\beta})$  and

$$\begin{aligned}\mathbf{P}_j(\boldsymbol{\beta}) &= \sum_{i=1}^{n_j} \int_0^\tau I(C_{ij} > s) \mathbf{Z}_{ij}(s) \exp\{\boldsymbol{\beta}_0^T \mathbf{Z}_{ij}(s)\} d\mu_{0j}(s) \\ \mathcal{Q}_j(\boldsymbol{\beta}) &= \sum_{i=1}^{n_j} \int_0^\tau I(C_{ij} > s) T_j(s; \boldsymbol{\beta}) \exp\{\boldsymbol{\beta}_0^T \mathbf{Z}_{ij}(s)\} d\mu_{0j}(s)\end{aligned}$$

Let  $\mathcal{B}_r$  be the following compact set:  $\mathcal{B}_r = \{\boldsymbol{\beta} : \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| \leq r\}$ , with boundary  $\partial\mathcal{B}_r = \{\boldsymbol{\beta} : \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| = r\}$ . By Theorem 10.8 of Rockafellar (1970), strong convergence of  $D_n(\boldsymbol{\beta})$  to  $D(\boldsymbol{\beta})$  is uniform on  $\boldsymbol{\beta} \in \mathcal{B}_r$ . Note that  $\partial D(\boldsymbol{\beta})/\partial \boldsymbol{\beta}|_{\boldsymbol{\beta}_0} = \mathbf{0}_{p \times 1}$  and  $\partial^2 D(\boldsymbol{\beta})/\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T|_{\boldsymbol{\beta}_0} = -\boldsymbol{\Omega}_d(\boldsymbol{\beta}_0)$ , which is assumed negative-definite through condition (d). Therefore,  $D(\boldsymbol{\beta})$  has unique maximizer  $\boldsymbol{\beta}_0$ , and in particular  $D(\boldsymbol{\beta}_0) > \sup_{\boldsymbol{\beta} \in \partial\mathcal{B}_r} D(\boldsymbol{\beta})$ . Since  $\sup_{\boldsymbol{\beta} \in \mathcal{B}_r} |D_n(\boldsymbol{\beta}) - D(\boldsymbol{\beta})| \xrightarrow{a.s.} 0$ , for almost all realizations, there exists  $n_0$  such that, for  $n \geq n_0$ ,  $D_n(\boldsymbol{\beta}_0) > D_n(\boldsymbol{\beta})$  for  $\boldsymbol{\beta} \in \partial\mathcal{B}_r$ . Thus, there exists a maximizer of  $D_n(\boldsymbol{\beta})$  in the interior of  $\mathcal{B}_r$ , and arguments in Jacobsen (1989) can be used to demonstrate the uniqueness of this maximizer for large  $n$ . Recall that  $\widehat{\boldsymbol{\beta}}_d$  solves  $\partial D_n(\boldsymbol{\beta})/\partial \boldsymbol{\beta} = \mathbf{0}$ , and hence is the unique maximizer. Since  $r$  is chosen arbitrarily, letting  $r \rightarrow 0$  yields that  $\widehat{\boldsymbol{\beta}}_d \xrightarrow{a.s.} \boldsymbol{\beta}_0$  as  $n \rightarrow \infty$ .

With respect to asymptotic normality, through a first-order Taylor series expansion about  $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ ,

$$\mathbf{U}_d(\widehat{\boldsymbol{\beta}}_d) = \mathbf{U}_d(\boldsymbol{\beta}_0) + \frac{\partial}{\partial \boldsymbol{\beta}^T} \mathbf{U}_d(\boldsymbol{\beta})|_{\boldsymbol{\beta}_*} (\widehat{\boldsymbol{\beta}}_d - \boldsymbol{\beta}_0)$$

where  $\boldsymbol{\beta}_*$  lies on the line segment joining  $\widehat{\boldsymbol{\beta}}_d$  and  $\boldsymbol{\beta}_0$ . Setting  $\widehat{\boldsymbol{\Omega}}_d(\boldsymbol{\beta}_*) = -n^{-1} \partial \mathbf{U}_d(\boldsymbol{\beta})/\partial \boldsymbol{\beta}^T|_{\boldsymbol{\beta}_*}$ , we have  $n^{1/2}(\widehat{\boldsymbol{\beta}}_d - \boldsymbol{\beta}_0) = \widehat{\boldsymbol{\Omega}}_d(\boldsymbol{\beta}_*)^{-1} n^{-1/2} \mathbf{U}_d(\boldsymbol{\beta}_0)$ . Since  $\widehat{\boldsymbol{\beta}}_d \xrightarrow{a.s.} \boldsymbol{\beta}_0$ , with  $\|\boldsymbol{\beta}_* - \boldsymbol{\beta}_0\| \leq \|\widehat{\boldsymbol{\beta}}_d - \boldsymbol{\beta}_0\|$ , as  $n \rightarrow \infty$ ,

$$\widehat{\boldsymbol{\Omega}}_d(\boldsymbol{\beta}_*) \xrightarrow{a.s.} \boldsymbol{\Omega}_d(\boldsymbol{\beta}_0). \quad (18)$$

We now derive the distribution of  $n^{-1/2} \mathbf{U}_d(\boldsymbol{\beta}_0)$ . Considering the integrand of an independent contribution to  $\mathbf{U}_d(\boldsymbol{\beta}_0)$ , provided model (1) and assumption (3) hold,  $E[\mathbf{Z}_{ij}(s) dM_{ij}^d(s; \boldsymbol{\beta}_0) | \mathbf{Z}_{ij}(s)] = \mathbf{0}$ . Additionally,

$$\begin{aligned}E[\overline{\mathbf{Z}}_j(s; \boldsymbol{\beta}_0) dM_{ij}^d(s; \boldsymbol{\beta}_0) | \mathbf{Z}_{ij}(s)] \\ &= E[E[\overline{\mathbf{Z}}_j(s; \boldsymbol{\beta}_0) dM_{ij}^d(s; \boldsymbol{\beta}_0) | \mathbf{Z}_{1j}(s), \dots, \mathbf{Z}_{n_{ij}}(s), C_{1j}, \dots, C_{n_{ij}}]] \\ &= E[E[\overline{\mathbf{Z}}_j(s; \boldsymbol{\beta}_0) | \mathbf{Z}_{1j}(s), \dots, \mathbf{Z}_{n_{ij}}(s), C_{1j}, \dots, C_{n_{ij}}] \\ &\quad \times E[dM_{ij}^d(s; \boldsymbol{\beta}_0) | \mathbf{Z}_{1j}(s), \dots, \mathbf{Z}_{n_{ij}}(s), C_{1j}, \dots, C_{n_{ij}}]] \\ &= E[E[\overline{\mathbf{Z}}_j(s; \boldsymbol{\beta}_0) | \mathbf{Z}_{1j}(s), \dots, \mathbf{Z}_{n_{ij}}(s), C_{1j}, \dots, C_{n_{ij}}] \\ &\quad \times E[dM_{ij}^d(s; \boldsymbol{\beta}_0) | \mathbf{Z}_{ij}(s)]] \\ &= \mathbf{0}.\end{aligned}$$

Hence,  $\mathbf{U}_d(\boldsymbol{\beta}_0)$  is a sum of independent mean-zero random vectors, such that  $n^{-1/2} \mathbf{U}_d(\boldsymbol{\beta}_0)$  converges in distribution to a zero-mean normal with variance  $\boldsymbol{\Sigma}_d(\boldsymbol{\beta}_0)$

by the Lindeberg–Feller Central Limit Theorem (Sen and Singer, 1993). This result, combined with (18) through Slutsky’s Theorem, concludes the proof of asymptotic normality.

**7.2. Proof of Theorem 3**

Set  $\widehat{\phi}_0(t) = \widehat{\phi}_{01}(t) + \widehat{\phi}_{02}(t)$ , where:

$$\begin{aligned} \widehat{\phi}_{01}(t) &= \{\widehat{\mu}_0(t; \widehat{\beta}_c) - \widehat{\mu}_0(t; \beta_0)\} \\ \widehat{\phi}_{02}(t) &= \{\widehat{\mu}_0(t; \beta_0) - \mu_0(t)\}. \end{aligned}$$

By the triangle inequality,

$$\sup_{t \in [0, \tau]} |\widehat{\phi}_0(t)| \leq \sup_{t \in [0, \tau]} |\widehat{\phi}_{01}(t)| + \sup_{t \in [0, \tau]} |\widehat{\phi}_{02}(t)|. \tag{19}$$

Now, one can write:

$$\widehat{\phi}_{01}(t) = n^{-1} \sum_{j=1}^n \sum_{i=1}^{n_j} \int_0^t \{S^{(0)}(s; \widehat{\beta}_c)^{-1} - S^{(0)}(s; \beta_0)^{-1}\} dN_{ij}(s). \tag{20}$$

Based on a linear Taylor series expansion,

$$\{S^{(0)}(s; \widehat{\beta}_c)^{-1} - S^{(0)}(s; \beta_0)^{-1}\} = \frac{-\overline{\mathbf{Z}}(s; \beta_0)^T}{S^{(0)}(s; \beta_0)} (\widehat{\beta}_c - \beta_0) + o(n^{-1/2})$$

almost surely. Substituting this expression back into (20) gives  $\widehat{\phi}_{01}(t) = \widehat{\mathbf{h}}(t; \beta_0)^T (\widehat{\beta}_c - \beta_0) + o(n^{-1/2})$ , where  $\widehat{\mathbf{h}}(t; \beta_0) = -n^{-1} \sum_{j=1}^n \sum_{i=1}^{n_j} \int_0^t \overline{\mathbf{Z}}(s; \beta_0) S^{(0)}(s; \beta_0)^{-1} dN_{ij}(s)$ . Since  $N_{ij}(t)$  is bounded,  $S^{(0)}(s; \beta)$  is bounded away from 0 as  $n \rightarrow \infty$  and  $\mathbf{S}^{(1)}(s; \beta)$  is bounded for all  $\beta$  in a compact set,  $\widehat{\mathbf{h}}(t; \beta_0)$  will be bounded for sufficiently large  $n$ . This, in conjunction with the fact that  $\widehat{\beta}_c \xrightarrow{a.s.} \beta_0$  implies that  $|\widehat{\phi}_{01}(t)| \xrightarrow{a.s.} 0$ . We now consider  $\widehat{\phi}_{02}(t)$ , which can be written:

$$\widehat{\phi}_{02}(t) = n^{-1} \sum_{j=1}^n \sum_{i=1}^{n_j} \int_0^t S^{(0)}(s; \beta_0)^{-1} dM_{ij}^c(s; \beta_0). \tag{21}$$

By the Uniform SLLN (Pollard, 1990, p. 41),  $\left| n^{-1} \sum_{j=1}^n \sum_{i=1}^{n_j} M_{ij}^c(t; \beta_0) \right|$  converges almost surely to 0 for  $t \in [0, \tau]$ . Additionally,  $S^{(0)}(s; \beta)$  is bounded away from 0, while  $S^{(0)}(s; \beta)$  is bounded away from 0 as  $n \rightarrow \infty$ . Combining these facts,

$$\sup_{t \in [0, \tau]} |\widehat{\phi}_{02}(t)| \xrightarrow{a.s.} 0,$$

which, combined with the fact that  $|\widehat{\phi}_{01}(t)| \xrightarrow{a.s.} 0$  for all  $t \in [0, \tau]$  concludes the proof of uniform consistency upon applying (19).

With respect to weak convergence of the process  $\{n^{1/2}\widehat{\phi}_0(t); t \in [0, \tau]\}$ ,

$$\begin{aligned} n^{1/2}\widehat{\phi}_{01}(t) &= \widehat{\mathbf{h}}(t; \boldsymbol{\beta}_0)^T n^{1/2}(\widehat{\boldsymbol{\beta}}_c - \boldsymbol{\beta}_0) + o_p(1) \\ n^{1/2}\widehat{\phi}_{02}(t) &= n^{-1/2} \sum_{j=1}^n \sum_{i=1}^{n_j} \int_0^t S^{(0)}(s; \boldsymbol{\beta}_0)^{-1} dM_{ij}^c(s; \boldsymbol{\beta}_0) + o_p(1). \end{aligned}$$

Using conditions (a)–(h), the SLLN and various results from empirical processes (e.g., Biliias, Gu and Ying, 1997),  $\widehat{\mathbf{h}}(t; \boldsymbol{\beta}_0) \xrightarrow{a.s.} \mathbf{h}(t; \boldsymbol{\beta}_0)$ ; combining this result with the fact that

$$n^{-1/2}\mathbf{U}_c(\boldsymbol{\beta}_0) = n^{-1/2} \sum_{j=1}^n \boldsymbol{\Psi}_j^c(\boldsymbol{\beta}_0) + o_p(1)$$

yields:

$$n^{1/2}\widehat{\phi}_{01}(t) = \mathbf{h}(t; \boldsymbol{\beta}_0)^T \boldsymbol{\Omega}_c(\boldsymbol{\beta}_0)^{-1} n^{-1/2} \sum_{j=1}^n \boldsymbol{\Psi}_j^c(\boldsymbol{\beta}_0) + o_p(1), \quad (22)$$

where  $\boldsymbol{\Psi}_j^c(\boldsymbol{\beta})$  is as defined in (11). With respect to  $n^{1/2}\widehat{\phi}_{02}(t)$ , using results from empirical processes, it can be shown that

$$\sup_{t \in [0, \tau]} \left| n^{-1/2} \sum_{j=1}^n \sum_{i=1}^{n_j} \int_0^t \{S^{(0)}(s; \boldsymbol{\beta}_0)^{-1} - s^{(0)}(s; \boldsymbol{\beta}_0)^{-1}\} dM_{ij}^c(s; \boldsymbol{\beta}_0) \right| \xrightarrow{P} 0.$$

Therefore,

$$n^{1/2}\widehat{\phi}_{02}(t) = n^{-1/2} \sum_{j=1}^n \sum_{i=1}^{n_j} \int_0^t s^{(0)}(s; \boldsymbol{\beta}_0)^{-1} dM_{ij}^c(s; \boldsymbol{\beta}_0) + o_p(1). \quad (23)$$

Combining (22) and (23),  $n^{1/2}\widehat{\phi}(t) = n^{-1/2} \sum_{j=1}^n \xi_j(t; \boldsymbol{\beta}_0) + o_p(1)$ , which behaves asymptotically as a sum of independent mean-zero random variates and, hence, has finite-dimensional distributions which converge to a mean-zero Normal random vector with covariance function  $\sigma_0(s, t)$  by the Lindeberg-Feller CLT. The demonstration that  $\xi_j(t; \boldsymbol{\beta}_0)$  is manageable for  $i=1, \dots, n$ , such that  $n^{1/2}\widehat{\phi}(t)$  is tight, followed by application of the Functional CLT (Pollard, 1990) concludes the proof of weak convergence.

### Acknowledgments

The authors thank the Canadian Organ Replacement Register of the Canadian Institute for Health Information for providing access to the end-stage renal

disease data. They also thank the Associate Editor and Referees for their constructive comments which improved the original version of the manuscript. This research was supported in part by National Institutes of Health grants R01 DK-70869 (DES) and R01 HL-57444 (JC).

## References

- O. O. Aalen, "Nonparametric inference for a family of counting processes," *The Annals of Statistics* vol. 6 pp. 701–726, 1978.
- Y. Biliyas, M. Gu and Z. Ying, "Towards a general asymptotic theory for the Cox model with staggered entry," *The Annals of Statistics* vol. 25 pp. 662–682, 1997.
- N. Breslow, "Contribution to the discussion of the paper by D.R. Cox," *Journal of the Royal Statistical Society, Series B*, vol. 34, pp. 187–220, 1972.
- T. Cai, L. J. Wei and M. Wilcox, "Semiparametric regression analysis for clustered failure time data," *Biometrika* vol. 87 pp. 867–878, 2000.
- L. X. Clegg, J. Cai and P. K. Sen, "A marginal mixed baseline hazards model for multivariate failure time data," *Biometrics* vol. 55 pp. 805–812, 1999.
- R. J. Cook and J. F. Lawless, "Analysis of repeated events," *Statistical Methods in Medical Research* vol. 11 pp. 141–166, 2002.
- D. V. Glidden and E. Vittinghoff, "Modelling clustered survival data from multicentre clinical trials," *Statistics in Medicine* vol. 23 pp. 369–388, 2004.
- M. Jacobsen, "Existence and unicity of MLEs in discrete exponential family distributions," *Scandinavian Journal of Statistics* vol. 16 pp. 335–349, 1989.
- J. D. Kalbfleisch and R. L. Prentice, *The Statistical Analysis of Failure Time Data*, Wiley: New York, NY, 2002.
- J. F. Lawless, "Regression methods for Poisson process data," *Journal of the American Statistical Association* vol. 82 pp. 808–815, 1987.
- J. F. Lawless and C. Nadeau, "Some simple robust methods for the analysis of recurrent events," *Technometrics* vol. 37 pp. 158–168, 1995.
- E. W. Lee, L. J. Wei and D. A. Amato, Cox-type regression analysis for large numbers of small groups of correlated failure time observations in *Survival Analysis: State of the Art*, (J. P. Klein and P. K. Goel, eds.) Kluwer: Dordrecht, pp. 237–247, 1992.
- D. Y. Lin, L. J. Wei, I. Yang and Z. Ying, "Semiparametric regression for the mean and rate functions of recurrent events," *Journal of the Royal Statistical Society: Series B* vol. 62 pp. 711–730, 2000.
- M. S. Pepe and J. Cai, "Some graphical displays and marginal regression analyses for recurrent failure times and time-dependent covariates," *Journal of the American Statistical Association* vol. 88 pp. 811–820, 1993.
- D. Pollard, *Empirical Processes: Theory and Applications*, Institute of Mathematical Statistics: Hayward, CA, 1990.
- R. T. Rockafellar, *Convex Analysis*, Princeton University Press: Princeton, NJ, 1970.
- P. K. Sen and J. M. Singer, *Large Sample Methods in Statistics*, Chapman & Hall: New York, NY, 1993.
- C. F. Spiekerman and D. Y. Lin, "Marginal regression models for multivariate failure time data," *Journal of American Statistical Association* vol. 93 pp. 1164–1175, 1998.
- A. W. van der Vaart and J. A. Wellner, *Weak Convergence and Empirical Processes*, Springer: New York, NY, 1996.
- L. J. Wei, D. Y. Lin and L. Weissfeld, "Regression analysis of multivariate incomplete failure time data by modeling marginal distributions," *Journal of American Statistical Association* vol. 84 pp. 1065–1073, 1989.