Marginal Means/Rates Models for Multiple Type Recurrent Event Data

JIANWEN CAI   cai@bios.unc.edu
Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-7420, USA

DOUGLAS E. SCHAUBEL
Department of Biostatistics, University of Michigan Ann Arbor, MI 48104-2029, USA

Abstract. Recurrent events are frequently observed in biomedical studies, and often more than one type of event is of interest. Follow-up time may be censored due to loss to follow-up or administrative censoring. We propose a class of semi-parametric marginal means/rates models, with a general relative risk form, for assessing the effect of covariates on the censored event processes of interest. We formulate estimating equations for the model parameters, and examine asymptotic properties of the parameter estimators. Finite sample properties of the regression coefficients are examined through simulations. The proposed methods are applied to a retrospective cohort study of risk factors for preschool asthma.

Keywords: marginal means/rates model, multiple type recurrent events, multi-variate survival analysis, semi-parametric model

1. Introduction

In many biomedical studies, subjects may experience the outcome of interest more than once; outcomes of this sort have been termed recurrent events. For example, patients with cerebrovascular disease may experience repeated transient ischemic attacks (Hobson et al., 1993), and HIV patients may experience recurrent opportunistic infections (Li and Lagakos, 1997). Other examples of recurrent events include infections, myocardial infarctions, tumor metastases, and disease relapses/remissions. The structure of recurrent events is that of naturally ordered multivariate failure time data since different events 'within' an individual are correlated.

Often, because subjects' health is assessed in several ways, more than one type of recurrent event may be of interest. For example, in a study of infections following bone marrow transplantation, it is of interest to study different types of recurrent infections simultaneously (e.g., bacterial, fungal and viral infections). Similarly, in a clinical trial regarding the efficacy of nutritional supplement of selenium in relationship to preventing skin cancer (Abu-Libdeh, Turnbull and Clark, 1990), it is of interest to study the recurrence of several types of skin cancer.

As another example, consider a retrospective birth cohort study of childhood asthma outcomes conducted by merging birth information and health administrative

records. The recurrent events of interest are (i) hospitalizations and (ii) physician office visits attributable to asthma. Although both event types relate to the same disease, each has much different implications with respect to disease severity and health care cost. Of interest is the relationship between birth characteristics (e.g., gender, low birth weight, various respiratory disorders) and the mean number of hospitalizations and office visits attributable to preschool asthma.

For single-type recurrent events, there are several estimating procedures proposed in the survival analysis literature. Conditional models (e.g., Prentice, Williams and Peterson, 1981) and marginal models (e.g., Wei, Lin and Weissfeld, 1989) have been proposed to analyze data with a single type of recurrent event. Following the traditional development of survival analysis, these methods are based on modeling the hazard function. Because the mean number of events is a more interpretable quantity than the hazard in the context of recurrent event data, some authors have proposed to model the mean function (e.g., Pepe and Cai, 1993; Lawless and Nadeau, 1995; Lin et al., 2000). Pepe and Cai (1993) proposed semi-parametric procedures for making inferences about the mean function without the Poisson-type assumption. Lawless and Nadeau (1995) have proposed a class of marginal means models for recurrent event data. Lin et al. (2000) studied the theoretical aspects of these robust procedures.

Despite the progress in methodology of analyzing single-type recurrent event data, methods for the analysis of data involving multiple-type recurrent events are limited. Abu-Libdeh, Turnbull and Clark (1990) considered nonhomogeneous Poisson processes with random and fixed effects, with inference based on maximum likelihood. Such parametric approaches require correct specification of the underlying within-subject correlation structure, which could be difficult to achieve with multiple-type recurrent event data. Semi-parametric robust procedures would be desirable if the underlying correlation structure is not of primary interest. The marginal mixed baseline hazard model independently studied by Spiekerman and Lin (1998) and Clegg, Cai and Sen (1999) could potentially be extended to accommodate multiple event types if one is interested in the Cox (1972) type relative risk function. However, the recurrent event mean/rate is often of more interest to investigators compared to the hazard function. In this article, we formulate a class of semi-parametric marginal means/rates regression models for analyzing multiple-type recurrent events data and propose a method of inference for the mean/rate ratio parameters.

The remainder of the article is organized as follows. In Section 2, we introduce notation, formulate the model and propose an estimating procedure for the model parameters. In Section 3, we study the asymptotic properties of the proposed estimators. We conduct simulation studies to evaluate the proposed method in finite samples in Section 4. We illustrate the method in Section 5 by applying it to a preschool asthma study. Finally, we provide some concluding remarks in Section 6.

2. Model and Methods

Suppose that a total of n subjects are observed over time. There are K different types of events of interest, each potentially recurrent and subject to right censoring. We

assume throughout that censoring occurs independently of the event processes under consideration. Let $N_{ik}^*(t) = \int_0^t dN_{ik}^*(s)$ represent the number of events of type k at time t for subject i. Let C_{ik} and $Y_{ik}(s) = I(C_{ik} \geq s)$ denote the event-type-specific censoring time and at-risk function, respectively, where $I(\cdot)$ denotes the indicator function. In practice, censoring times for different event types are usually the same for a subject, i.e., $C_{ik} = C_i$, although this might not always be the case. For example, in record linkage studies, censoring times for different event types may be different since information sources are usually not equally up to date. The observed event processes are denoted by $N_{ik}(t) = \int_0^t Y_{ik}(s)dN_{ik}^*(s)$. Let $\mathbf{Z}_{ik}(t)$ be a $p \times 1$ covariate vector, which may contain external time-dependent covariates; i.e., time-dependent covariates which are not directly involved with the failure mechanism (Kalbfleisch and Prentice, 2002). We model the event-type k mean and rate semi-parametrically by

$$E[N_{ik}^*(t)|\mathbf{Z}_{ik}] = g(\boldsymbol{\beta}_0^T \mathbf{Z}_{ik})\mu_{0k}(t), \tag{1}$$

and

$$E[dN_{ik}^*(t)|\mathbf{Z}_{ik}(t)] = g(\boldsymbol{\beta}_0^T \mathbf{Z}_{ik}(t))d\mu_{0k}(t), \tag{2}$$

respectively, where $\mu_{0k}(t) = \int_0^t d\mu_{0k}(s)$ is an unspecified baseline mean function and $\boldsymbol{\beta}_0$ is a $p \times 1$ vector of parameters of interest. The link function, $g : \mathscr{R} \to \mathscr{R}$, with $g(\cdot) \geq 0$, is pre-specified and assumed to be continuous almost everywhere and twice differentiable. Examples of possible link functions include $g(x) = 1 + x$, $g(x) = e^x$ and $g(x) = \log(1 + e^x)$. Selection of an appropriate link function may be based on prior data or the resulting interpretation of the regression parameters. Note that the incorporation of event-type-specific parameters can be accomplished through appropriate specification of the covariate vectors. The baseline mean functions are allowed to be different for each event type. Since model (1) is a special case of (2), we focus on model (2) hereafter.

Let $\mathbf{S}_k^{(d)}(s; \boldsymbol{\beta}) = n^{-1} \sum_{i=1}^n Y_{ik}(s)\mathbf{Z}_{ik}(s)^{\otimes d}g^{(d)}(\boldsymbol{\beta}^T \mathbf{Z}_{ik}(s))$ for $d = 0, 1, 2$ and $\mathbf{S}_k^{(3)}(s; \boldsymbol{\beta}) = n^{-1} \sum_{i=1}^n Y_{ik}(s)\mathbf{Z}_{ik}(s)^{\otimes 2}g^{(1)}(\boldsymbol{\beta}^T \mathbf{Z}_{ik}(s))^2/g(\boldsymbol{\beta}^T \mathbf{Z}_{ik}(s))$, where $g^{(0)}(x) = g(x)$, $g^{(r)}(x) = d^r g(x)/dx^r$ for $r = 1, 2$, and, for a vector $\mathbf{a}$, $\mathbf{a}^{\otimes 0} = 1$, $\mathbf{a}^{\otimes 1} = \mathbf{a}$, and $\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}^T$. Also, define $\mathbf{E}_k(s; \boldsymbol{\beta}) = \mathbf{S}_k^{(1)}(s; \boldsymbol{\beta})/S_k^{(0)}(s; \boldsymbol{\beta})$ and $\mathbf{V}_k(s; \boldsymbol{\beta}) = \mathbf{S}_k^{(3)}(s; \boldsymbol{\beta})/S_k^{(0)}(s; \boldsymbol{\beta}) - \mathbf{E}_k(s; \boldsymbol{\beta})^{\otimes 2}$. The limiting values of $\mathbf{S}_k^{(d)}(s; \boldsymbol{\beta})$, $\mathbf{E}_k(s; \boldsymbol{\beta})$ and $\mathbf{V}_k(s; \boldsymbol{\beta})$ are given by $\mathbf{s}_k^{(d)}(s; \boldsymbol{\beta})$, $\mathbf{e}_k(s; \boldsymbol{\beta})$ and $\mathbf{v}_k(s; \boldsymbol{\beta})$, respectively.

Analogous to generalized estimating equation methods for other types of response data (Liang and Zeger, 1986), we specify the following estimating equations

$$\sum_{i=1}^n \sum_{k=1}^K \int_0^\tau \mathbf{Z}_{ik}(s)\frac{g^{(1)}(\boldsymbol{\beta}^T \mathbf{Z}_{ik}(s))}{g(\boldsymbol{\beta}^T \mathbf{Z}_{ik}(s))}\left\{dN_{ik}(s) - Y_{ik}(s)g(\boldsymbol{\beta}^T \mathbf{Z}_{ik}(s))d\mu_{0k}(s)\right\} = \mathbf{0}_{p \times 1},$$

$$\tag{3}$$

where $P(Y_{ik}(\tau) = 1) > 0$, for $k = 1, \ldots, K$, and

$$\sum_{i=1}^{n} \int_{0}^{t} \left\{ dN_{ik}(s) - Y_{ik}(s)g(\boldsymbol{\beta}^T \mathbf{Z}_{ik}(s))d\mu_{0k}(s) \right\} = 0. \tag{4}$$

This is in line with the estimating equations of Pepe and Cai (1993) and Lin et al. (2000) for univariate recurrent event processes. Under the model and assuming independent censoring, the left sides of (3) and (4) each have expectation 0. Based on (4), we can express $\mu_{0k}(t)$ in terms of $\boldsymbol{\beta}$ through $d\mu_{0k}(s; \boldsymbol{\beta}) = dN_{\cdot k}(s)/nS_k^{(0)}(s; \boldsymbol{\beta})$, where $dN_{\cdot k}(s) = \sum_{i=1}^{n} dN_{ik}(s)$. Substituting this expression back into (3) yields an estimating equation for $\boldsymbol{\beta}_0$ which is free of $\{\mu_{0k}(\cdot)\}_{k=1}^{K}$:

$$\mathbf{U}_n(\boldsymbol{\beta}) = \sum_{i=1}^{n} \sum_{k=1}^{K} \int_{0}^{\tau} \left\{ \mathbf{Z}_{ik}(s) \frac{g^{(1)}(\boldsymbol{\beta}^T \mathbf{Z}_{ik}(s))}{g(\boldsymbol{\beta}^T \mathbf{Z}_{ik}(s))} - \mathbf{E}_k(s; \boldsymbol{\beta}) \right\} dN_{ik}(s) = \mathbf{0}_{p \times 1}. \tag{5}$$

The Newton–Raphson iterative procedure can be used for solving (5). For a single-type recurrent event process with only unit increments, (5) reduces to the estimating equation examined in Prentice and Self (1983) for the proportional intensity model. Note that the model we propose, like that of Lin et al. (2000), can accommodate increments of any positive size, as would occur in cost data.

Let $\widehat{\boldsymbol{\beta}}_n$ denote the solution to (5). We propose to estimate $\mu_{0k}(t)$ by the following Breslow–Aalen type estimator based on the kth type event:

$$\widehat{\mu}_{0k}(t; \widehat{\boldsymbol{\beta}}_n) = \int_{0}^{t} dN_{\cdot k}(s)/nS_k^{(0)}(s; \widehat{\boldsymbol{\beta}}_n) \tag{6}$$

for $k = 1, \ldots, K$ and $0 \leq t < \tau$. The mean number of type k events by time t for a subject with covariate pattern $\mathbf{Z}_{ik}(\cdot) = \mathbf{z}_{ik}(\cdot)$ is estimated by $\int_{0}^{t} g(\widehat{\boldsymbol{\beta}}_n^T \mathbf{z}_{ik}(s))d\widehat{\mu}_{0k}(s; \widehat{\boldsymbol{\beta}}_n)$.

3.   Asymptotic Properties

We assume throughout that the following set of conditions holds:

(a) $\{N_{ik}(\cdot), Y_{ik}(\cdot), \mathbf{Z}_{ik}(\cdot)\}_{k=1}^{K}$ are independent and identically distributed for $i = 1, \ldots, n$

(b) $P(Y_{ik}(\tau) = 1) > 0$.

(c) $|Z_{ik\ell}(0)| + \int_{0}^{\tau} |dZ_{ik\ell}(s)| < c_Z < \infty$ almost surely, for $\ell = 1, \ldots, p$.

(d) Positive-definiteness of the matrix $\mathbf{A}(\boldsymbol{\beta_0}) = \sum_{k=1}^{K} \int_{0}^{\tau} \mathbf{v}_k(s; \boldsymbol{\beta_0})s_k^{(0)}(s; \boldsymbol{\beta_0})d\mu_{0k}(s)$.

(e) $N_{ik}(\tau)$ is bounded by a constant almost surely for $i = 1, \ldots, n$, $k = 1, \ldots, K$.

(f) For $\boldsymbol{\beta} \in \mathscr{B}_0$, where $\mathscr{B}_0$ is a small neighborhood about $\boldsymbol{\beta}_0$, $g(\boldsymbol{\beta}^T \mathbf{Z}_{ik}(s))$ is locally bounded away from 0.

(g) $\mathbf{s}_k^{(d)}(s; \boldsymbol{\beta})$ and $\mathbf{r}_k^{(r)}(s; \boldsymbol{\beta})$, for $d = 0, 1, 2, 3$ and $r = 0, 1, 2$, are continuous functions of $\boldsymbol{\beta} \in \mathscr{B}_0$ uniformly in $s \in [0, \tau]$ and are bounded on $\mathscr{B}_0 \times [0, \tau]$, with

$$\mathbf{s}_k^{(1)}(s; \boldsymbol{\beta}) = \partial s_k^{(0)}(s; \boldsymbol{\beta})/\partial \boldsymbol{\beta}, \quad \mathbf{s}_k^{(2)}(s; \boldsymbol{\beta}) = \partial^2 s_k^{(0)}(s; \boldsymbol{\beta})/\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T,$$
$$\mathbf{r}_k^{(1)}(s; \boldsymbol{\beta}) = \partial r_k^{(0)}(s; \boldsymbol{\beta})/\partial \boldsymbol{\beta}, \quad \mathbf{r}_k^{(2)}(s; \boldsymbol{\beta}) = \partial^2 r_k^{(0)}(s; \boldsymbol{\beta})/\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T,$$

where $\mathbf{r}_k^{(r)}(s; \boldsymbol{\beta})$ is defined below. Conditions (a) and (c) imply the following as $n \to \infty$ for $\boldsymbol{\beta} \in \mathcal{B}_0$, $d = 0, 1, 2, 3$ and $r = 0, 1, 2$:

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

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

where $\| \mathbf{a} \| = (\mathbf{a}^T \mathbf{a})^{1/2}$ and

$$R_k^{(0)}(s; \boldsymbol{\beta}) = n^{-1} \sum_{i=1}^{n} Y_{ik}(s) \log \left\{ \frac{g(\boldsymbol{\beta}^T \mathbf{Z}_{ik}(s))}{g(\boldsymbol{\beta}_0^T \mathbf{Z}_{ik}(s))} \right\} g(\boldsymbol{\beta}_0^T \mathbf{Z}_{ik}(s)),$$

$$\mathbf{R}_k^{(1)}(s; \boldsymbol{\beta}) = \frac{\partial}{\partial \boldsymbol{\beta}} R_k^{(0)}(s; \boldsymbol{\beta}) = n^{-1} \sum_{i=1}^{n} Y_{ik}(s) \mathbf{Z}_{ik}(s) \frac{g^{(1)}(\boldsymbol{\beta}^T \mathbf{Z}_{ik}(s))}{g(\boldsymbol{\beta}^T \mathbf{Z}_{ik}(s))} g(\boldsymbol{\beta}_0^T \mathbf{Z}_{ik}(s)),$$

$$\mathbf{R}_k^{(2)}(s; \boldsymbol{\beta}) = \frac{\partial^2}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} R_k^{(0)}(s; \boldsymbol{\beta})$$

$$= n^{-1} \sum_{i=1}^{n} Y_{ik}(s) \mathbf{Z}_{ik}(s)^{\otimes 2}$$

$$\times \left\{ \frac{g^{(2)}(\boldsymbol{\beta}^T \mathbf{Z}_{ik}(s))}{g(\boldsymbol{\beta}^T \mathbf{Z}_{ik}(s))} - \left[\frac{g^{(1)}(\boldsymbol{\beta}^T \mathbf{Z}_{ik}(s))}{g(\boldsymbol{\beta}^T \mathbf{Z}_{ik}(s))} \right]^2 \right\} g(\boldsymbol{\beta}_0^T \mathbf{Z}_{ik}(s)).$$

We summarize the asymptotic properties of $\widehat{\boldsymbol{\beta}}_n$ and $\widehat{\mu}_{0k}(t; \widehat{\boldsymbol{\beta}}_n)$ in the following theorems.

THEOREM 1 $\widehat{\boldsymbol{\beta}}_n$ *is a consistent estimator of* $\boldsymbol{\beta}_0$ *(i.e.,* $\widehat{\boldsymbol{\beta}}_n \xrightarrow{a.s.} \boldsymbol{\beta}_0$*), and* $\sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \xrightarrow{D} N_p(\mathbf{0}_{p \times 1}, \sum(\boldsymbol{\beta}_0))$*, where* $\sum(\boldsymbol{\beta}) = \mathbf{A}(\boldsymbol{\beta})^{-1} \mathbf{B}(\boldsymbol{\beta}) \mathbf{A}(\boldsymbol{\beta})^{-1}$*, with* $\mathbf{A}(\boldsymbol{\beta})$ *defined by condition (d) and* $\mathbf{B}(\boldsymbol{\beta}) = E\left[\left(\sum_{k=1}^{K} \mathbf{U}_{1,k}(\boldsymbol{\beta}) \right)^{\otimes 2} \right]$*, where*

$$\mathbf{U}_{i,k}(\boldsymbol{\beta}) = \int_0^\tau \left\{ \mathbf{Z}_{ik}(s) \frac{g^{(1)}(\boldsymbol{\beta}^T \mathbf{Z}_{ik}(s))}{g(\boldsymbol{\beta}^T \mathbf{Z}_{ik}(s))} - \mathbf{e}_k(s; \boldsymbol{\beta}) \right\} dM_{ik}(s; \boldsymbol{\beta}), \tag{7}$$

with $dM_{ik}(s; \boldsymbol{\beta}) = dN_{ik}(s) - Y_{ik}(s) g(\boldsymbol{\beta}^T \mathbf{Z}_{ik}(s)) d\mu_{0k}(s)$.

Consistency follows from the fact that $n^{-1}\{X_n(\boldsymbol{\beta}) - X_n(\boldsymbol{\beta}_0)\}$ converges to a concave function with unique maximizer $\widehat{\boldsymbol{\beta}}_n$, where $\partial X_n(\boldsymbol{\beta})/\partial \boldsymbol{\beta} = \mathbf{U}_n(\boldsymbol{\beta})$. Asymptotically, $n^{1/2}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)$ behaves as a scaled normalized sum of independent and identically distributed random vectors. An outline of a proof of Theorem 1 is provided in the Appendix. Consistent estimators of $\mathbf{A}(\boldsymbol{\beta}_0)$ and $\mathbf{B}(\boldsymbol{\beta}_0)$ are given by

$$\mathbf{A}_n(\widehat{\boldsymbol{\beta}}_n) = n^{-1} \sum_{k=1}^{K} \int_0^\tau \mathbf{V}_k(s; \widehat{\boldsymbol{\beta}}_n) \, dN \cdot_k (s) \tag{8}$$

and

$$\mathbf{B}_n(\widehat{\boldsymbol{\beta}}_n) = \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{k=1}^{K}\int_0^\tau\left\{\mathbf{Z}_{ik}(s)\frac{g^{(1)}(\widehat{\boldsymbol{\beta}}_n^T\mathbf{Z}_{ik}(s))}{g(\widehat{\boldsymbol{\beta}}_n^T\mathbf{Z}_{ik}(s))} - \mathbf{E}_k(s;\widehat{\boldsymbol{\beta}}_n)\right\}d\widehat{M}_{ik}(s;\widehat{\boldsymbol{\beta}}_n)\right)^{\otimes 2},$$

where $d\widehat{M}_{ik}(s;\boldsymbol{\beta}) = dN_{ik}(s) - Y_{ik}(s)g(\boldsymbol{\beta}^T\mathbf{Z}_{ik}(s))d\widehat{\mu}_{0k}(s;\boldsymbol{\beta})$ and $\widehat{\mu}_{0k}(t;\boldsymbol{\beta})$ is given as in (6). The consistency of $\mathbf{A}_n(\widehat{\boldsymbol{\beta}}_n)$ and $\mathbf{B}_n(\widehat{\boldsymbol{\beta}}_n)$ can be shown using repeated applications of the Strong Law of Large Numbers and Lemma 1 from Lin et al. (2000).

We now consider the limiting properties of $\widehat{\mu}_{0k}(t;\widehat{\boldsymbol{\beta}}_n)$ and $\mathbf{W}_n(t) = [W_{1:n}(t), \ldots, W_{K:n}(t)]^T$, where $W_{k:n}(t) = \sqrt{n}\left(\widehat{\mu}_{0k}(t;\widehat{\boldsymbol{\beta}}_n) - \mu_{0k}(t)\right)$ for $k = 1, \ldots, K$. Set $\mathbf{W}(t) = [W_1(t), \ldots, W_K(t)]^T$. To begin, we define an appropriate metric on the pertinent function space. Let $f_k(t) \in D[0,\tau]$, where $D[0,\tau]$ is the space of cadlag functions such that $f_k(\cdot) : [0,\tau] \to \mathscr{R}$. Set $\mathbf{f}(t) = [f_1(t), \ldots, f_K(t)]^T$ such that $\mathbf{f}(\cdot) : [0,\tau] \to \mathscr{R}^K$. Define $D[0,\tau]^K$ to be the space consisting of such functions and, following the approach of Spiekerman and Lin (1998), equip $D[0,\tau]^K$ with the metric, $\rho(p,q) = \max\left\{\sup_{t\in[0,\tau]}|p_k(t) - q_k(t)|\right\}_{k=1}^{K}$ for $p(\cdot), q(\cdot) \in D[0,\tau]^K$. Having defined an appropriate metric space with respect to which convergence can be discussed, we state the following results regarding the baseline mean estimators.

THEOREM 2 $\widehat{\mu}_{0k}(t;\widehat{\boldsymbol{\beta}}_n)$ converges almost surely to $\mu_{0k}(t)$ uniformly in $t \in [0,\tau]$, and $\mathbf{W}_n(t)$ converges weakly to a zero-mean Gaussian field in $D[0,\tau]^K$, with covariance function $\mathrm{Cov}(W_k(s), W_\ell(t))$ given by $c_{k\ell}(s,t;\boldsymbol{\beta}_0) = E[C_{1k}(s;\boldsymbol{\beta}_0)C_{1\ell}(t;\boldsymbol{\beta}_0)]$, with

$$C_{ik}(t;\boldsymbol{\beta}) = \int_0^t \frac{dM_{ik}(s;\boldsymbol{\beta})}{s_k^{(0)}(s;\boldsymbol{\beta})} + \mathbf{h}_k^T(t;\boldsymbol{\beta})\mathbf{A}(\boldsymbol{\beta})^{-1}\sum_{\ell=1}^{K}\mathbf{U}_{i,\ell}(\boldsymbol{\beta}), \tag{9}$$

and $\mathbf{h}_k(t;\boldsymbol{\beta}) = -\int_0^t \mathbf{e}_k(s;\boldsymbol{\beta})d\mu_{0k}(s)$.

The covariance function can be consistently estimated by replacing limiting quantities with their respective sample counterparts; i.e., $c_{k\ell}(s,t;\boldsymbol{\beta}_0)$ can be consistently estimated by its empirical counterpart $\widehat{c}_{k\ell}(s,t;\widehat{\boldsymbol{\beta}}_n) = n^{-1}\sum_{i=1}^{n}\widehat{C}_{ik}(s;\widehat{\boldsymbol{\beta}}_n)\widehat{C}_{i\ell}(t;\widehat{\boldsymbol{\beta}}_n)$, where for $i = 1, \ldots, n$ and $k = 1, \ldots, K$,

$$\widehat{C}_{ik}(t;\widehat{\boldsymbol{\beta}}_n) = \int_0^t \frac{d\widehat{M}_{ik}(s;\widehat{\boldsymbol{\beta}}_n)}{S_k^{(0)}(s;\widehat{\boldsymbol{\beta}}_n)} + \mathbf{H}_k^T(t;\widehat{\boldsymbol{\beta}}_n)\mathbf{A}(\widehat{\boldsymbol{\beta}}_n)^{-1}\sum_{\ell=1}^{K}\widehat{\mathbf{U}}_{i,\ell}(\widehat{\boldsymbol{\beta}}_n), \tag{10}$$

where $\mathbf{H}_k(t;\widehat{\boldsymbol{\beta}}_n) = -\int_0^t \mathbf{E}_k(s;\widehat{\boldsymbol{\beta}}_n)d\widehat{\mu}_{0k}(s,\widehat{\boldsymbol{\beta}}_n)$,

$$\widehat{\mathbf{U}}_{i,k}(\widehat{\boldsymbol{\beta}}_n) = \int_0^\tau\left\{\mathbf{Z}_{ik}(s)\frac{g^{(1)}(\widehat{\boldsymbol{\beta}}_n^T\mathbf{Z}_{ik}(s))}{g(\widehat{\boldsymbol{\beta}}_n^T\mathbf{Z}_{ik}(s))} - \mathbf{E}_k(s;\widehat{\boldsymbol{\beta}}_n)\right\}d\widehat{M}_{ik}(s;\widehat{\boldsymbol{\beta}}_n),$$

$d\widehat{M}_{ik}(s;\widehat{\boldsymbol{\beta}}_n) = dN_{ik}(s) - Y_{ik}(s)g(\widehat{\boldsymbol{\beta}}_n^T\mathbf{Z}_{ik}(s))d\widehat{\mu}_{0k}(s;\widehat{\boldsymbol{\beta}}_n)$, and $\widehat{\mu}_{0k}(t;\widehat{\boldsymbol{\beta}}_n)$ is given as in (6).

The proof of Theorem 2 proceeds by representing $W_{k:n}(t)$ as a normalized average of independent and identically distributed random variates, applying the Cramer-Wold Theorem (Sen and Singer, 1993), and then applying results from empirical

processes (van der Vaart and Wellner, 1996). An outline of the proof is provided in the Appendix.

4. Simulation Studies

Simulation studies were conducted to examine the finite sample properties of the proposed estimators. Specifically, we simulated multiple event times for two event types with $n = 30, 50, 100$ or 200 subjects. Event times were generated from the following mixed effects marginal rates model: $E[dN_{ik}^*(s)|\mathbf{Z}_{ik}(s), Q_{ik}] = Q_{ik}g(\boldsymbol{\beta}_0^T\mathbf{Z}_{ik}(s))d\mu_{0k}(s)$ for $i = 1, \ldots, n$ and $k = 1, 2$. We considered the cases: $d\mu_{0k}(t) = d\mu_{0k}$, with $\mu_{01}(t) = 0.25t$ and $\mu_{02}(t) = 0.50t$, $Q_{ik} \equiv Q_i \sim \Gamma(\sigma_Q^{-2}, \sigma_Q^{-2})$, $g(\boldsymbol{\beta}_0^T\mathbf{Z}_{ik}(s)) = e^{\beta_0 Z_i}$, with $Z_i = 0$ and $Z_i = 1$ for $n/2$ subjects each. We considered $\beta_0 = 0$ and $\beta_0 = \log(2)$. With respect to the kth event type, the above-listed proportional means model corresponds to the mixed Poisson model: $\left(\int_0^t dN_{ik}^*(s)|Q_i\right) \sim$ Poisson($Q_i t e^{\beta_0 Z_i}d\mu_{0k}$); i.e., a homogeneous Poisson process (Chiang, 1980) for each subject i, given Q_i. Employing the probability integral transform (Casella and Berger, 1990; p. 52–54), the $(\ell + 1)$th type k event time for subject i is given by $T_{i,k,\ell+1} = T_{i,k,\ell} - \log(U_{i,k,\ell+1})\{Q_i e^{\beta_0 Z_i}d\mu_{0k}\}^{-1}$, where $U_{i,k,\ell} \sim$ Uniform$(0, 1)$ with $T_{i,k,0} = 0$. The study duration $\tau = 5$ was employed, with censoring times for each subject generated as $C_{ik} = C_i \sim$ Uniform$(0, \tau)$, and generated independently of both event processes. The expected number of events per subject ranged from 1.25 ($Z_i = 0$) to 2.5 ($Z_i = 1$) for $k = 1$, and 2.5 ($Z_i = 0$) to 5.0 ($Z_i = 1$) for $k = 2$. The random effect, Q_i, for which $E[Q_i] = 1$ and $V[Q_i] = \sigma_Q^2$, induces positive correlation among within-subject event times, with larger σ_Q^2 corresponding to higher correlation. Values of σ_Q^2 employed were 0, 0.5, 1.0, and 2.0. Note that when $\sigma_Q^2 = 0$, i.e., $Q_i = 1$ for all i, intra-subject event times are independent.

As suggested by a referee, we also empirically assessed the efficiency of our proposed semi-parametric method relative to a fully parametric approach. We estimated the parameters $\theta = \sigma^{-2}$, $\lambda_1 = d\mu_{01}$, $\lambda_2 = d\mu_{02}$, and β through maximum likelihood. Using the multiple event-type analog of the approach described in Cook and Lawless (2002), the likelihood contribution of subject i is given by:

$$L_i = \left\{\prod_{k=1}^2 \lambda_k^{N_{ik}(C_i)}\right\}e^{\beta Z_i N_i(C_i)} \int_0^\infty q^{N_i(C_i)}e^{-qC_i \exp\{\beta Z_i\}(\lambda_1+\lambda_2)}\frac{q^{\theta-1}e^{-q\theta}\theta^\theta}{\Gamma(\theta)}dq,$$

where $N_i(t) = N_{i1}(t) + N_{i2}(t)$ is the total number of events for subject i up to time t. The log-likelihood contribution for the ith individual is given by:

$$\ell_i = \log\Gamma(N_i(C_i) + \theta) - \log\Gamma(\theta) + \sum_{k=1}^2 N_{ik}(C_i)\log\lambda_k$$

$$+ \beta Z_i N_i(C_i) - N_i(C_i)\log\theta$$

$$- \{N_i(C_i) + \theta\}\log\{\theta^{-1}C_i e^{\beta Z_i}(\lambda_1 + \lambda_2) + 1\}.$$

The score equations were solved using the Newton–Raphson method. The efficiency of the proposed estimator was estimated as the ratio of the mean square error for $\widehat{\beta}_{\mathrm{MLE}}$ to that of $\widehat{\beta}_n$.

For each parameter combination, we ran $R = 2500$ simulations. The degree of bias was assessed by comparing the average value of $\widehat{\beta}_n$ across all simulated replicates, $\overline{\beta}_n = R^{-1} \sum_{r=1}^{R} \widehat{\beta}_n^{[r]}$, with the true value, β_0, where $\widehat{\beta}_n^{[r]}$ is the estimate of β_0 based on the rth simulation ($r = 1, \ldots, R$). The appropriateness of the asymptotic approximation of the distribution of $\widehat{\beta}_n$ was assessed by comparing the average large-sample robust standard error estimate, $\mathrm{SE}(\widehat{\beta}_n) = R^{-1} \sum_{r=1}^{R} \mathrm{SE}(\widehat{\beta})_n^{[r]}$, to the sample standard deviation of the R estimates, $\sigma(\widehat{\beta}_n) = \left[(R-1)^{-1} \sum_{r=1}^{R} (\widehat{\beta}_n^{[r]} - \overline{\beta}_n)^2 \right]^{1/2}$. The accuracy of the normal approximation was also assessed by comparing empirical 95%

Table 1. Summary statistics for simulation studies.

β_0	n	σ_Q^2	$\overline{\beta}_n$	$\sigma(\widehat{\beta}_n)$	$\mathrm{SE}(\widehat{\beta}_n)$	$\mathrm{CP}_n(\beta_0)$	$\mathrm{EFF}(\widehat{\beta}_n : \widehat{\beta}_{\mathrm{MLE}})$
0	30	0	0.005	0.292	0.258	0.919	0.98
		0.5	−0.006	0.423	0.372	0.910	0.88
		1	0.003	0.525	0.448	0.902	0.86
		2	0.005	0.712	0.557	0.877	0.85
	50	0	0.000	0.219	0.202	0.928	0.97
		0.5	0.002	0.323	0.294	0.924	0.92
		1	−0.005	0.391	0.359	0.928	0.89
		2	0.009	0.538	0.455	0.896	0.81
	100	0	0.002	0.144	0.145	0.954	1.00
		0.5	−0.004	0.219	0.213	0.943	0.94
		1	−0.007	0.279	0.264	0.931	0.89
		2	0.004	0.368	0.337	0.923	0.86
	200	0	−0.003	0.103	0.103	0.948	1.00
		0.5	0.008	0.155	0.153	0.941	0.97
		1	0.004	0.197	0.190	0.942	0.91
		2	0.006	0.254	0.246	0.941	0.84
0.693	30	0	0.704	0.243	0.224	0.921	0.98
		0.5	0.700	0.399	0.346	0.907	0.83
		1	0.699	0.505	0.427	0.891	0.86
		2	0.692	0.677	0.537	0.872	0.86
	50	0	0.692	0.182	0.176	0.934	0.99
		0.5	0.697	0.309	0.277	0.914	0.90
		1	0.700	0.379	0.347	0.918	0.86
		2	0.693	0.515	0.441	0.897	0.86
	100	0	0.698	0.127	0.125	0.943	1.00
		0.5	0.684	0.210	0.201	0.933	0.95
		1	0.692	0.274	0.253	0.924	0.89
		2	0.699	0.349	0.330	0.930	0.83
	200	0	0.693	0.090	0.089	0.951	1.00
		0.5	0.696	0.146	0.144	0.944	0.91
		1	0.695	0.186	0.183	0.944	0.88
		2	0.698	0.250	0.240	0.939	0.85

confidence interval coverage rate, $CP_n(\beta_0) = R^{-1} \sum_{r=1}^{R} I(\widehat{\beta}_n^{[r]} - 1.96 \, \mathrm{SE}(\widehat{\beta})_n^{[r]} \leq \beta_0 \leq \widehat{\beta}_n^{[r]} + 1.96 \, \mathrm{SE}(\widehat{\beta})_n^{[r]})$, with the nominal value of 0.95.

Results of the simulations are presented in Table 1. For all trials, $\widehat{\beta}_n$ was approximately unbiased for β_0, even for $n = 30$. When the sample size is small, for example, $n = 30$ or 50, the accuracy with which $\mathrm{SE}(\widehat{\beta}_n)$ approximated $\sigma(\widehat{\beta}_n)$ decreased as the intra-subject event time correlation increased, and consequently the 95% confidence interval coverage rate also decreased as event time correlation increased. The 95% confidence interval coverage rate increased notably when going from $n = 50$ to $n = 100$. For a given n and σ^2, coverage probability was similar for $\beta = 0$ and $\beta = \log(2)$. The efficiency of the proposed regression parameter estimator relative to the MLE ranged from 0.81 to 1.00. Generally, for fixed n, efficiency decreased as within-subject correlation increased.

5. Application: Risk Factors for Preschool Asthma

Data used in this analysis were obtained from Manitoba Health, a provincial health administration organization in Canada. The purpose of the investigation was to quantify birth conditions as risk factors associated with asthma during the 0–4 age interval. Details regarding data collection are available in the original article (Schaubel et al., 1996). Briefly, a birth cohort of children born in the 1984/85 fiscal year (April 1, 1984 to March 31, 1985) was assembled by combining three administrative databases: a birth information file, hospital separation records, and physician claims records. The birth file contained information on several factors previously suspected of being associated with childhood asthma, including low birth weight, prematurity, and various neonatal respiratory disorders. Dates of hospitalizations and physician office visits for asthma were extracted, and linked with the birth file based on a unique personal health identification number assigned to each child at birth. Children were followed retrospectively from birth, and each was censored some time between the attainment of ages 4 and 5.

In this study, there are two event processes of interest: (i) physician office visits and (ii) hospitalizations attributable to asthma. Although both types of events are related to the same underlying disease, the two processes have much different implications in terms of implied disease severity and associated heath care costs. Risk factors of interest include low (<2500 g) birth weight (LBW), respiratory distress syndrome (RDS), transient tachypnea of the newborn (TTN), mild-to-moderate birth asphyxia (ASPH-M), severe birth asphyxia (ASPH-S) and male gender (SEX = M).

A total of 5907 physician office visits and 376 hospitalizations for asthma were observed from the birth cohort ($n = 16,207$). There were 1472 children with at least 1 office visit and 207 children with at least 1 hospitalization. The number of office visits (hospitalizations) among the children with at least 1 office visit (hospitalization) ranged from 1(1) to 51(13). Among the adverse birth conditions studied, the most common was ASPH-M (5.7% of children), followed by low birth weight (4.3%).

Table 2. Risk factors for preschool asthma: physician office visits and hospitalizations.

Factors	Coefficient	SE_N	SE_R	Mean ratio (95% CI)
Birth weight (low(<2.5kg) = 1, normal = 0)				
Office visits	0.447	0.055	0.151	1.56 (1.16, 2.10)
Hospitalizations	0.764	0.191	0.265	2.15 (1.28, 3.61)
Respiratory distress syndrome (yes = 1, no = 0)				
Office visits	0.982	0.068	0.204	2.67 (1.79, 3.98)
Hospitalizations	1.320	0.225	0.316	3.74 (2.02, 6.95)
Transient tachypnea (yes = 1, no = 0)	0.266	0.074	0.248	1.31 (0.80, 2.12)
Mild-to-moderate birth asphyxia (yes = 1, no = 0)	0.244	0.049	0.178	1.28 (0.90, 1.81)
Severe birth asphyxia (yes = 1, no=0)	0.664	0.064	0.243	1.94 (1.21, 3.13)
Gender (male = 1, female = 0)	0.682	0.027	0.090	1.98 (1.66, 2.36)

Transient tachypnea was experienced by 2.0% of the cohort, while severe birth asphyxia was experienced by 1.9%. The least frequent condition was respiratory distress syndrome, observed in 1.5% of the cohort.

A marginal means model was fitted to the above-described data using a log-linear link function, i.e., $E[N_{ik}^*(t)|\mathbf{Z}_{ik}] = \mu_{0k}(t)e^{\boldsymbol{\beta}^T\mathbf{Z}_{ik}}$. Covariates included in the model were indicators for LBW, RDS, TTN, ASPH-S, ASPH-M and male gender. A model was first fitted with distinct coefficients for each event type. The final model contained event-specific parameters for a covariate only when the coefficient estimates were significantly different for physician office visits and hospitalizations based on the results of the first model. The final model contained separate coefficients for LBW and RDS.

Mean ratios (MR $= e^\beta$) based on the model are listed in Table 2. Both "naive" and robust standard errors (denoted by SE_n and SE_r, respectively) are presented, where the former do not account for the correlation among events for the same subject. ASPH-S (MR $= 1.94$), male gender (MR $= 1.98$), LBW and RDS were each associated with significantly increased mean number of asthma-attributable physician office visits and hospitalizations. For both LBW and RDS, the effect on mean number of hospitalizations was notably greater (LBW: MR $= 2.15$; RDS: MR $= 3.74$) than that for office visits (LBW: MR $= 1.56$; RDS: MR $= 2.67$). The within-subject event time correlations were extremely high, as reflected by the ratio of robust to naive standard errors, which ranged from 1.39 for LBW (hospitalization) to 3.77 for ASPH-S.

A smoothed estimate (loess) of the baseline rate function based on $d\widehat{\mu}_{0k}(t;\widehat{\boldsymbol{\beta}}_n)$ for each event type is displayed in Figure 1. The frequency of the physician visits increases steadily from birth, reaches a maximum at approximately age 3, and decreases thereafter. The frequency for hospitalizations increases sharply during the $0-2$ age interval, reaches a maximum during age 2, and decreases steadily thereafter. The estimated baseline mean number of asthma-related physician visits and hospitalizations per 1000 children and the corresponding 95% confidence intervals are plotted in Figure 2.
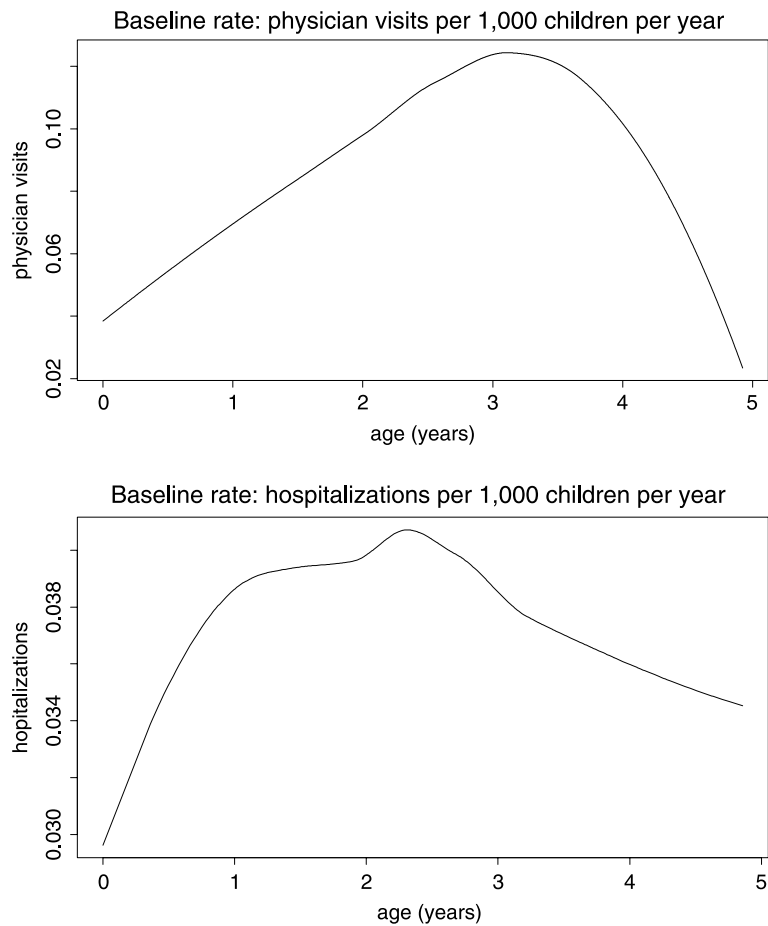
Figure 1. Smoothed estimate of baseline rates of physician office visits and hospitalizations per 1000 children per year.

6. Discussion

We have proposed a semi-parametric marginal means/rates model for the analysis of recurrent event data when events of multiple types are of interest. Estimating equations for the model parameters have been proposed. Parameter estimates were shown to be consistent and asymptotically normally distributed, with covariance matrices which can be estimated consistently. Simulation results indicate that large sample approximations to the distribution of the regression parameter estimator are reasonable for moderate and, in some cases, very small
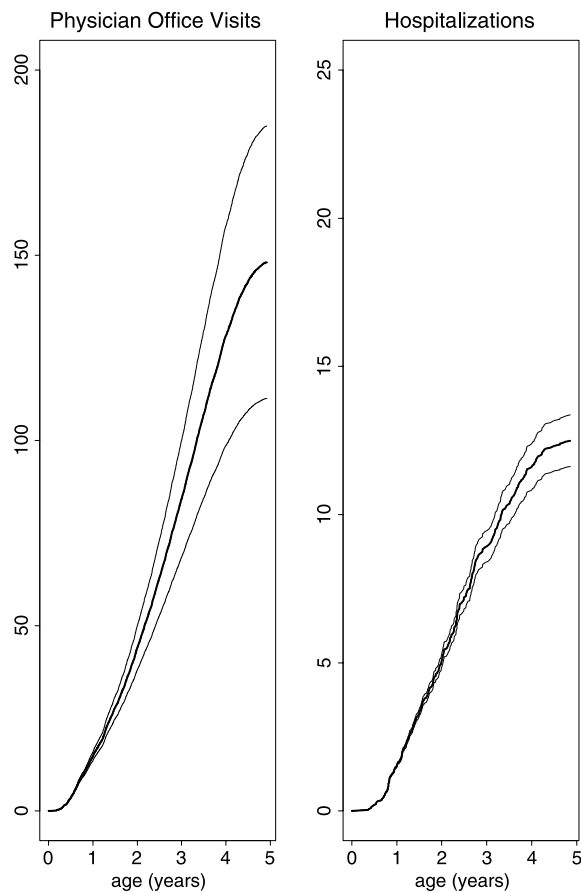
Figure 2. Estimated baseline mean number of physician office visits and hospitalizations per 1000 children with 95% confidence limits.

sample sizes. We applied a special case of the proposed model to a retrospective cohort study, and found that several neonatal characteristics were significantly associated with increased mean number of physician visits and hospitalizations for preschool asthma.

The proposed model is semi-parametric in the sense that the form of the baseline mean function is unspecified while the functional form which relates the covariate vector to the mean is specified. An alternative would be a fully parametric model, as discussed in Lawless and Nadeau (1995) and Cook and Lawless (2002). A semi-parametric model would be preferable if the investigators are uncertain about the form of the mean function for the baseline group. Our simulation results indicate that, when the parametric model is correctly specified,

efficiency loss under a semi-parametric model is modest. Often in practice, it is unlikely that the analyst would have sufficient knowledge of the event process under study to correctly specify a parametric model, provided such a model even exists.

The link function in the proposed model is not restricted to be in the exponential form. Although it is widely applied, mostly due to familiarity and the availability of statistical software, the exponential form may not accurately describe the relationship between covariates and the mean. Although formal methods of comparing link functions are not yet available, it would be possible to compare the appropriateness of various link functions through metrics based on the error terms, $\widehat{M}_{ik}(\tau; \widehat{\boldsymbol{\beta}}_n)$, examples of which include $\sum_{i=1}^{n} \sum_{k=1}^{K} \widehat{M}_{ik}(\tau; \widehat{\boldsymbol{\beta}}_n)^2$ and $\sup_{0 \leq t \leq \tau} \sum_{i=1}^{n} \sum_{k=1}^{K} \left| \widehat{M}_{ik}(t; \widehat{\boldsymbol{\beta}}_n) \right|$.

Our simulation results indicate that the normal approximation to the sampling distribution of the regression parameter estimator is appropriate for moderate sample sizes. In very small samples, the robust variance estimator for the regression coefficients in the proposed model is much smaller than the sampling variability and correspondingly the 95% confidence interval coverage rate is much smaller than the nominal value. The degree to which asymptotic approximations are appropriate in finite samples depends on the sample size and the strength of dependence among events within subjects. When the sample size is too small and the asymptotic approximation does not work well, a bootstrap method (Efron, 1981) could be an alternative.

Our proposed method assumes that the censoring and event processes are independent. For the retrospective cohort study described in Section 4, this assumption is reasonable, since, despite their debilitating effect and associated health care costs, asthma episodes among children rarely result in death. However, in many other studies, this independent censoring assumption might not hold, especially when one of the events under consideration is terminal in nature, i.e., its occurrence precludes subsequent events. For example, if the events of interest are hospitalizations for various diseases and if death probability increases with the number of hospitalizations, then the estimates of regression coefficients based on the assumption of independent censoring could be biased. As discussed by Lawless (1995), one way to correct the problem is to explicitly model the censoring mechanism. For single event-type, various approaches have been suggested by Cook and Lawless (1997). Ghosh and Lin (2000) have considered the non-parametric case, while Wang, Qin and Chiang (2001) have developed methods for informative censoring.

Acknowledgments

Appendix

Proof of Theorem 1:

Let $X_n(\boldsymbol{\beta}) = \sum_{i=1}^n \sum_{k=1}^K \int_0^\tau \left\{ \log g(\boldsymbol{\beta}^T \mathbf{Z}_{ik}(s)) - \log\left(nS_k^{(0)}(s;\boldsymbol{\beta})\right) \right\} dN_{ik}(s)$ and $\Delta_n(\boldsymbol{\beta}) = \frac{1}{n}\{X_n(\boldsymbol{\beta}) - X_n(\boldsymbol{\beta}_0)\}$;
note that $\mathbf{U}_n(\boldsymbol{\beta}) = \partial X_n(\boldsymbol{\beta})/\partial\boldsymbol{\beta}$. We can write $\Delta_n(\boldsymbol{\beta}) = \Delta_{1:n}(\boldsymbol{\beta}) + \Delta_{2:n}(\boldsymbol{\beta})$, where

$$\Delta_{1:n}(\boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^n \sum_{k=1}^K \int_0^\tau \left\{ \log\left(\frac{g(\boldsymbol{\beta}^T \mathbf{Z}_{ik}(s))}{g(\boldsymbol{\beta}_0^T \mathbf{Z}_{ik}(s))}\right) - \log\left(\frac{S_k^{(0)}(s;\boldsymbol{\beta})}{S_k^{(0)}(s;\boldsymbol{\beta}_0)}\right) \right\} dM_{ik}(s;\boldsymbol{\beta}_0),$$

$$\Delta_{2:n}(\boldsymbol{\beta}) = \sum_{k=1}^K \int_0^\tau \left\{ R_k^{(0)}(s;\boldsymbol{\beta}) - \log\left(\frac{S_k^{(0)}(s;\boldsymbol{\beta})}{S_k^{(0)}(s;\boldsymbol{\beta}_0)}\right) \right\} S_k^{(0)}(s;\boldsymbol{\beta}_0)\, d\mu_{0k}(s).$$

Given conditions (a), (c) and (e), by the Strong Law of Large Numbers, $\Delta_{1:n}(\boldsymbol{\beta}) \xrightarrow{a.s.} 0$, while

$$\Delta_{2:n}(\boldsymbol{\beta}) \xrightarrow{a.s.} \sum_{k=1}^K \int_0^\tau \left\{ r_k^{(0)}(s;\boldsymbol{\beta}) - \log\left(\frac{s_k^{(0)}(s;\boldsymbol{\beta})}{s_k^{(0)}(s;\boldsymbol{\beta}_0)}\right) \right\} s_k^{(0)}(s;\boldsymbol{\beta}_0)\, d\mu_{0k}(s) \equiv \Delta(\boldsymbol{\beta}).$$

Therefore, as $n \to \infty$, $\Delta_n(\boldsymbol{\beta}) \xrightarrow{a.s.} \Delta(\boldsymbol{\beta})$, which has first and second derivatives:

$$\frac{\partial}{\partial\boldsymbol{\beta}}\Delta(\boldsymbol{\beta}) = \sum_{k=1}^K \int_0^\tau \left\{ \mathbf{r}_k^{(1)}(s;\boldsymbol{\beta}) - \mathbf{e}_k(s;\boldsymbol{\beta}) s_k^{(0)}(s;\boldsymbol{\beta}_0) \right\} d\mu_{0k}(s),$$

$$\frac{\partial^2}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}\Delta(\boldsymbol{\beta}) = \sum_{k=1}^K \int_0^\tau \left\{ \mathbf{r}_k^{(2)}(s;\boldsymbol{\beta}) - \left(\frac{\mathbf{s}_k^{(2)}(s;\boldsymbol{\beta})}{s_k^{(0)}(s;\boldsymbol{\beta})} - \mathbf{e}_k(s;\boldsymbol{\beta})^{\otimes 2}\right) s_k^{(0)}(s;\boldsymbol{\beta}_0) \right\} d\mu_{0k}(s).$$

Now, evaluated at $\boldsymbol{\beta} = \boldsymbol{\beta}_0$, $\partial\Delta(\boldsymbol{\beta})/\partial\boldsymbol{\beta} = \mathbf{0}_{p\times 1}$, while $-\partial^2\Delta(\boldsymbol{\beta})/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T = \mathbf{A}(\boldsymbol{\beta}_0)$, which is positive-definite by condition (d). Therefore, $\Delta(\boldsymbol{\beta})$ has a local maximum at $\boldsymbol{\beta} = \boldsymbol{\beta}_0$. Set $\mathscr{B}_\delta = \{\boldsymbol{\beta} : \| \boldsymbol{\beta} - \boldsymbol{\beta}_0 \| \leq \delta\}$ for arbitrary $\delta > 0$. Thus, $\Delta(\boldsymbol{\beta}_0) \geq \Delta(\boldsymbol{\beta})$ for $\boldsymbol{\beta} \in \mathscr{B}_\delta$, $\Delta(\boldsymbol{\beta}_0) > \Delta(\boldsymbol{\beta})$ for $\boldsymbol{\beta} \in \partial\mathscr{B}_\delta$, where $\partial\mathscr{B}_\delta = \{\boldsymbol{\beta} : \| \boldsymbol{\beta} - \boldsymbol{\beta}_0 \| = \delta\}$. Using the SLLN and continuity arguments,

$$\| \Delta_n(\boldsymbol{\beta}) - \Delta_n(\boldsymbol{\beta}_0) \| \xrightarrow{a.s.} \| \Delta(\boldsymbol{\beta}) - \Delta(\boldsymbol{\beta}_0) \|.$$

Therefore, $\Delta_n(\boldsymbol{\beta}_0) \geq \Delta_n(\boldsymbol{\beta})$ for all $\boldsymbol{\beta} \in \mathscr{B}_\delta$, with $\Delta_n(\boldsymbol{\beta}_0) > \Delta_n(\boldsymbol{\beta})$ when $\boldsymbol{\beta} \in \partial\mathscr{B}_\delta$. Thus, $\Delta_n(\boldsymbol{\beta})$ has a maximum which is not on the boundary, implying that there is an interior point of $\mathscr{B}_\delta$ which corresponds to a local maximum of $\Delta_n(\boldsymbol{\beta})$. But, $\partial\Delta_n(\boldsymbol{\beta})/\partial\boldsymbol{\beta} = \mathbf{0}_{p\times 1}$ at $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}_n$, meaning that $\widehat{\boldsymbol{\beta}}_n$ is the local maximum. Since δ was arbitrary, letting $\delta \to 0$ demonstrates that $\widehat{\boldsymbol{\beta}}_n \xrightarrow{a.s.} \boldsymbol{\beta}_0$.

By the consistency of $\widehat{\boldsymbol{\beta}}_n$, Taylor expansion of $\mathbf{U}_n(\widehat{\boldsymbol{\beta}}_n)$ at $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ gives

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

where $\boldsymbol{\beta}_*$ lies between $\widehat{\boldsymbol{\beta}}_n$ and $\boldsymbol{\beta}_0$ in $\mathscr{R}^p$. Since $\mathbf{U}_n(\widehat{\boldsymbol{\beta}}_n) = 0$, we have

$$(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) = -\left\{ \frac{\partial}{\partial \boldsymbol{\beta}^T} \mathbf{U}_n(\boldsymbol{\beta})|_{\boldsymbol{\beta}_*} \right\}^{-1} \mathbf{U}_n(\boldsymbol{\beta}_0).$$

Setting $\mathbf{I}_n(\boldsymbol{\beta}_*) = -\partial \mathbf{U}_n(\boldsymbol{\beta})/\partial \boldsymbol{\beta}^T|_{\boldsymbol{\beta}_*}$, and $\mathbf{A}_n(\boldsymbol{\beta}) = n^{-1}\mathbf{I}_n(\boldsymbol{\beta})$, we have

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) = \{\mathbf{A}_n(\boldsymbol{\beta}_*)\}^{-1} n^{-1/2} \mathbf{U}_n(\boldsymbol{\beta}_0). \tag{A.1}$$

Using arguments illustrated in the Appendix of Lin et al. (2000), it can be shown that

$$\sup_{t\in[0,\tau]} \left\| n^{-1/2} \sum_{i=1}^n \int_0^t \{\mathbf{E}_k(s; \boldsymbol{\beta}_0) - \mathbf{e}_k(s; \boldsymbol{\beta}_0)\} dM_{ik}(s; \boldsymbol{\beta}_0) \right\| \xrightarrow{P} 0.$$

Hence, $\mathbf{U}_n(\boldsymbol{\beta}_0)$ is asymptotically equivalent to $\sum_{k=1}^K \mathbf{U}_{k:n}(\boldsymbol{\beta}_0)$, where

$$\mathbf{U}_{k:n}(\boldsymbol{\beta}_0) = \sum_{i=1}^n \int_0^\tau \left\{ \mathbf{Z}_{ik}(s) \frac{g^{(1)}(\boldsymbol{\beta}_0^T \mathbf{Z}_{ik}(s))}{g(\boldsymbol{\beta}_0^T \mathbf{Z}_{ik}(s))} - \mathbf{e}_k(s; \boldsymbol{\beta}_0) \right\} dM_{ik}(s; \boldsymbol{\beta}_0).$$

Since $\mathbf{Z}_{ik}(s)$ and $g^{(r)}(\boldsymbol{\beta}^T \mathbf{Z}_{ik}(s))$ are locally bounded for $r = 0, 1, 2$ and $\mathbf{U}_{k:n}(\boldsymbol{\beta}_0)$ is a sum of independent and identically distributed random vectors with $E[\mathbf{U}_{k:n}(\boldsymbol{\beta}_0)] = \mathbf{0}_{p\times 1}$, by the Multivariate Central Limit Theorem,

$$n^{-1/2}\mathbf{U}_n(\boldsymbol{\beta}_0) \xrightarrow{D} N_p(\mathbf{0}_{p\times 1}, \mathbf{B}(\boldsymbol{\beta}_0)), \tag{A.2}$$

where

$$\mathbf{B}(\boldsymbol{\beta}) = E\left[\left(\sum_{k=1}^K \int_0^\tau \left\{ \mathbf{Z}_{1k}(s) \frac{g^{(1)}(\boldsymbol{\beta}^T \mathbf{Z}_{1k}(s))}{g(\boldsymbol{\beta}^T \mathbf{Z}_{1k}(s))} - \mathbf{e}_k(s; \boldsymbol{\beta}) \right\} dM_{1k}(s; \boldsymbol{\beta}) \right)^{\otimes 2} \right].$$

Set $\mathbf{A}_n(\boldsymbol{\beta}) = \sum_{k=1}^K \mathbf{A}_{k:n}(\boldsymbol{\beta})$, where $\mathbf{A}_{k:n}(\boldsymbol{\beta}) = \sum_{\ell=1}^4 \mathbf{A}_{k\ell:n}(\boldsymbol{\beta})$, with

$$\mathbf{A}_{k1:n}(\boldsymbol{\beta}) = -n^{-1} \sum_{i=1}^n \int_0^\tau \mathbf{Z}_{ik}(s)^{\otimes 2}$$

$$\times \left\{ \frac{g^{(2)}(\boldsymbol{\beta}^T \mathbf{Z}_{ik}(s))}{g(\boldsymbol{\beta}^T \mathbf{Z}_{ik}(s))} - \left(\frac{g^{(1)}(\boldsymbol{\beta}^T \mathbf{Z}_{ik}(s))}{g(\boldsymbol{\beta}^T \mathbf{Z}_{ik}(s))} \right)^2 \right\} dM_{ik}(s; \boldsymbol{\beta}_0),$$

$$\mathbf{A}_{k2:n}(\boldsymbol{\beta}) = -n^{-1} \sum_{i=1}^n \int_0^\tau \mathbf{Z}_{ik}(s)^{\otimes 2} \left\{ \frac{g^{(2)}(\boldsymbol{\beta}^T \mathbf{Z}_{ik}(s))}{g(\boldsymbol{\beta}^T \mathbf{Z}_{ik}(s))} - \left(\frac{g^{(1)}(\boldsymbol{\beta}^T \mathbf{Z}_{ik}(s))}{g(\boldsymbol{\beta}^T \mathbf{Z}_{ik}(s))} \right)^2 \right\}$$

$$\times Y_{ik}(s) g(\boldsymbol{\beta}_0^T \mathbf{Z}_{ik}(s)) d\mu_{0k}(s),$$

$$\mathbf{A}_{k3:n}(\boldsymbol{\beta}) = n^{-1} \sum_{i=1}^n \int_0^\tau \left\{ \frac{\mathbf{S}_k^{(2)}(s; \boldsymbol{\beta})}{S_k^{(0)}(s; \boldsymbol{\beta})} - \mathbf{E}_k(s; \boldsymbol{\beta})^{\otimes 2} \right\} dM_{ik}(s; \boldsymbol{\beta}_0),$$

$$\mathbf{A}_{k4:n}(\boldsymbol{\beta}) = n^{-1} \sum_{i=1}^n \int_0^\tau \left\{ \frac{\mathbf{S}_k^{(2)}(s; \boldsymbol{\beta})}{S_k^{(0)}(s; \boldsymbol{\beta})} - \mathbf{E}_k(s; \boldsymbol{\beta})^{\otimes 2} \right\} Y_{ik}(s) g(\boldsymbol{\beta}_0^T \mathbf{Z}_{ik}(s)) d\mu_{0k}(s).$$

By repeated applications of the Strong Law of Large Numbers and Lemma 1 of Lin et al. (2000), $\mathbf{A}_{k1:n}(\boldsymbol{\beta})$ and $\mathbf{A}_{k3:n}(\boldsymbol{\beta})$ converge in probability to $\mathbf{0}_{p \times p}$, while

$$\mathbf{A}_{k2:n}(\boldsymbol{\beta}) = \int_0^\tau \mathbf{R}_k^{(2)}(s; \boldsymbol{\beta}) d\mu_{0k}(s) \xrightarrow{P} \int_0^\tau \mathbf{r}_k^{(2)}(s; \boldsymbol{\beta}) d\mu_{0k}(s)$$

and

$$\mathbf{A}_{k4:n}(\boldsymbol{\beta}) \xrightarrow{P} \int_0^\tau \left\{ \frac{\mathbf{s}_k^{(2)}(s; \boldsymbol{\beta})}{s_k^{(0)}(s; \boldsymbol{\beta})} - \mathbf{e}_k(s; \boldsymbol{\beta})^{\otimes 2} \right\} s_k^{(0)}(s; \boldsymbol{\beta}) d\mu_{0k}(s).$$

Now, since $\widehat{\boldsymbol{\beta}}_n \xrightarrow{a.s.} \boldsymbol{\beta}_0$, and since $||\boldsymbol{\beta}_* - \boldsymbol{\beta}_0|| \leq ||\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0||$,

$$\begin{aligned}
\mathbf{A}_{k:n}(\boldsymbol{\beta}_*) \xrightarrow{P} & \int_0^\tau \left\{ \frac{\mathbf{s}_k^{(2)}(s; \boldsymbol{\beta}_0)}{s_k^{(0)}(s; \boldsymbol{\beta}_0)} - \left(\frac{\mathbf{s}_k^{(1)}(s; \boldsymbol{\beta}_0)}{s_k^{(0)}(s; \boldsymbol{\beta}_0)} \right)^{\otimes 2} \right\} s_k^{(0)}(s; \boldsymbol{\beta}_0) d\mu_{0k}(s) \\
& - \int_0^\tau \mathbf{r}_k^{(2)}(s; \boldsymbol{\beta}_0) d\mu_{0k}(s) \\
= & \int_0^\tau \left\{ \frac{\mathbf{s}_k^{(3)}(s; \boldsymbol{\beta}_0)}{\mathbf{s}_k^{(0)}(s; \boldsymbol{\beta}_0)} - \left(\frac{\mathbf{s}_k^{(1)}(s; \boldsymbol{\beta}_0)}{s_k^{(0)}(s; \boldsymbol{\beta}_0)} \right)^{\otimes 2} \right\} s_k^{(0)}(s; \boldsymbol{\beta}_0) d\mu_{0k}(s) \\
= & \int_0^\tau \mathbf{v}_k(s; \boldsymbol{\beta}_0) s_k^{(0)}(s; \boldsymbol{\beta}_0) d\mu_{0k}(s) \equiv \mathbf{A}_k(\boldsymbol{\beta}_0).
\end{aligned}$$

Thus,

$$\mathbf{A}_n(\boldsymbol{\beta}_*) \xrightarrow{P} \sum_{k=1}^K \mathbf{A}_k(\boldsymbol{\beta}_0) \equiv \mathbf{A}(\boldsymbol{\beta}_0). \tag{A.3}$$

By the results in (A.1), (A.2) and (A.3), applying Slutsky's Theorem (Sen and Singer, 1993), $n^{1/2}\left(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\right) \xrightarrow{D} N_p(\mathbf{0}_{p \times 1}, \Sigma(\boldsymbol{\beta}_0))$ where $\Sigma(\boldsymbol{\beta}) = \mathbf{A}(\boldsymbol{\beta})^{-1}\mathbf{B}(\boldsymbol{\beta})\mathbf{A}(\boldsymbol{\beta})^{-1}$.

Proof of Theorem 2: Applying a linear Taylor series expansion, followed by the triangle inequality:

$$\begin{aligned}
\left| \widehat{\mu}_{0k}(t; \widehat{\boldsymbol{\beta}}_n) - \mu_{0k}(t) \right| \leq & \left| n^{-1} \sum_{i=1}^n \int_0^t \frac{dM_{ik}(s; \boldsymbol{\beta}_0)}{S_k^{(0)}(s; \boldsymbol{\beta}_0)} \right| \\
& + \left| n^{-1} \sum_{i=1}^n \int_0^t \frac{\mathbf{S}_k^{(1)}(s; \boldsymbol{\beta}_*)^T}{S_k^{(0)}(s; \boldsymbol{\beta}_*)^2} dM_{ik}(s; \boldsymbol{\beta}_0)(\widehat{\boldsymbol{\beta}}_n + \boldsymbol{\beta}_0) \right| \\
& + \left| \int_0^t \frac{S_k^{(0)}(s; \boldsymbol{\beta}_0)\mathbf{S}_k^{(1)}(s; \boldsymbol{\beta}_*)^T}{S_k^{(0)}(s; \boldsymbol{\beta}_*)^2} d\mu_{0k}(s)(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) \right|.
\end{aligned}$$

By the Uniform Strong Law of Large Numbers (USLLN; Pollard, 1990) and Lemma 1 of Lin et al. (2000), the first term converges almost surely to 0 uniformly

in t. Regarding the second and third terms, the integrals are bounded asymptotically, as the integrands are bounded almost surely under conditions (c), (e), (f) and (g). This, combined with the strong consistency result from Theorem 1, implies that the second and third terms converge to 0 almost surely as $n \to \infty$. Therefore, $\widehat{\mu}_{0k}(t; \widehat{\boldsymbol{\beta}}_n) \xrightarrow{a.s.} \mu_{0k}(t)$ uniformly in t. Additionally,

$$
\begin{aligned}
W_{k:n}(t) =& \sqrt{n}\Big(\widehat{\mu}_{0k}(t; \widehat{\boldsymbol{\beta}}_n) - \mu_{0k}(t)\Big) \\
=& n^{-1/2} \sum_{i=1}^{n} \int_0^t \frac{dM_{ik}(s; \boldsymbol{\beta}_0)}{S_k^{(0)}(s; \boldsymbol{\beta}_0)} + n^{1/2}\mathbf{H}_k(t; \boldsymbol{\beta}_0)^T\Big(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\Big) + o_p(1) \\
=& n^{-1/2} \sum_{i=1}^{n} \int_0^t \frac{dM_{ik}(s; \boldsymbol{\beta}_0)}{S_k^{(0)}(s; \boldsymbol{\beta}_0)} + \mathbf{H}_k(t; \boldsymbol{\beta}_0)^T\mathbf{A}(\boldsymbol{\beta}_0)^{-1}n^{-1/2}\mathbf{U}_n(\boldsymbol{\beta}_0) + o_p(1) \\
=& n^{-1/2} \sum_{i=1}^{n} C_{ik}(t; \boldsymbol{\beta}_0) + o_p(1),
\end{aligned}
$$

where $\mathbf{H}_k(t; \boldsymbol{\beta}) = - \int_0^t \mathbf{E}_k(s; \boldsymbol{\beta})d\widehat{\mu}_{0k}(s; \boldsymbol{\beta})$ and $C_{ik}(t; \boldsymbol{\beta}_0)$ is as defined in (9). The last equality was derived through repeated applications of the USLLN and Lemma 1 of Lin et al. (2000). Since $W_{k:n}(t)$ behaves asymptotically as a scaled average of independent and identically distributed random variables, convergence in finite-dimensional distributions follows from the Multivariate Central Limit Theorem, while convergence in finite dimensional distributions of $\mathbf{W}_n(\cdot)$ follows from the Cramer-Wold Theorem (Sen and Singer, 1993). Tightness of $\mathbf{W}_n(\cdot)$ follows the theorem used in Example 2.11.16 of van der Vaart and Wellner (1996) due to the boundedness of $N_{ik}(t)$ for $t \in [0, \tau]$; hence $\mathbf{W}_n(\cdot)$ converges weakly to a zero-mean Gaussian process with covariance function $c_{k\ell}(s, t; \boldsymbol{\beta}_0)$. $\qquad\square$

References

H. Abu-Libdeh, B. W. Turnbull, and L. C. Clark, "Analysis of multi-type recurrent events in longitudinal studies: Application to a skin cancer prevention trial," *Biometrics* vol. 46 pp. 1017–1034, 1990.

G. Casella and R. L. Berger, *Statistical Inference*, Duxbury Press: Toronto, ON, 1990.

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.

C. L. Chiang, *An Introduction to Stochastic Processes and their Applications*, Kreiger: New York, NY, 1980.

R. J. Cook and J. F. Lawless, "Marginal analysis of recurrent events and a terminating event," *Stat. Med.* vol. 16 pp. 911–924, 1997.

R. J. Cook and J. F. Lawless, "Analysis of repeated events," *Stat. Meth. Med. Res.* vol. 11 pp. 141–166, 2002.

D. R. Cox, "Regression models and life-tables (with discussion)," *J. Roy. Stat. Soc.: Series B* vol. 34 pp. 187–220, 1972.

B. Efron, "Censored data and the bootstrap," *J. Am. Stat. Assoc.* vol. 76 pp. 312–319, 1981.

D. Ghosh and D. Y. Lin, "Nonparametric analysis of recurrent events and death," *Biometrics* vol. 56 pp. 554–562, 2000.

R. W. Hobson, D. G. Weiss, W. S. Fields, J. Goldstone, W. S. Moore, J. B. Towne, C. B. Wright and the Veterans Affairs Cooperative Study Group, ''Effect of carotid endarterectomy for asymptomatic carotid stenosis,'' *New Engl. J. Med.* vol. 328 pp. 221–227, 1993.

J. D. Kalbfleisch and R. L. Prentice, *The Statistical Analysis of Failure Time Data*, Wiley: New York, NY, 2002.

J. F. Lawless, ''The analysis of recurrent events for multiple subjects,'' *Appl. Stat.* vol. 44 pp. 487–498, 1995.

J. F. Lawless and C. Nadeau, ''Some simple robust methods for the analysis of recurrent events,'' *Technometrics* vol. 37 pp. 158–168, 1995.

Q. H. Li and S. W. Lagakos, ''Use of the Wei-Lin-Weissfeld method for the analysis of a recurring and a terminating event,'' *Stat. Med.* vol. 16 pp. 925–940, 1997.

K. Y. Liang and S. L. Zeger, ''Longitudinal data analysis using generalized linear models,'' *Biometrika* vol. 73 pp. 13–22, 1986.

D. Y. Lin, L. J. Wei, I. Yang, and Z. Ying, ''Semiparametric regression for the mean and rate functions of recurrent events,'' *J. Roy. Stat. Soc.: Ser. 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,'' *J. Am. Stat. Assoc.* vol. 88 pp. 811–820, 1993.

D. Pollard, *Empirical Processes: Theory and Applications*, Institute of Mathematical Statistics: Hayward, CA, 1990.

R. L. Prentice, B. J. Williams, and A. V. Peterson, ''On the regression analysis of multivariate failure time data,'' *Biometrika* vol. 68 pp. 373–389, 1981.

R. L. Prentice and S. G. Self, ''Asymptotic distribution theory for Cox-type regression models with general relative risk form,'' *Ann. of Stat.* vol. 11 pp. 804–812, 1983.

D. Schaubel, H. Johansen, M. Dutta, M. Desmeules, A. Becker and Y. Mao. ''Neonatal characteristics as risk factors for preschool asthma,'' *J. Asthma* vol. 33 pp. 255–264, 1996.

P. K. Sen and J. M. Singer, *Large Sample Methods in Statistics*, Chapman and Hall: New York, NY, 1993.

C. F. Spiekerman and D. Y. Lin, ''Marginal regression models for multivariate failure time data,'' *J. Am. Stat. Assoc.* 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.

M.-C.Wang, J. Qin, and C.-T. Chiang, ''Analyzing recurrent event data with informative censoring,'' *J. Am. Stat. Assoc.* vol. 96 pp. 1057–1065, 2001.

L. J. Wei, D. Y. Lin, and L. Weissfeld, ''Regression analysis of multivariate incomplete failure time data by modeling marginal distributions,'' *J. Am. Stat. Assoc.* vol. 84 pp. 1065–1073, 1989.