MARGINAL REGRESSION MODELS FOR RECURRENT AND TERMINAL EVENTS

Debashis Ghosh and D. Y. Lin

University of Michigan and University of North Carolina

Abstract: A major complication in the analysis of recurrent event data from medical studies is the presence of death. We consider the marginal mean function for the cumulative number of recurrent events over time, acknowledging the fact that death precludes further recurrences. We specify that covariates have multiplicative effects on an arbitrary baseline mean function while leaving the stochastic structure of the recurrent event process completely unspecified. We then propose estimators for the regression parameters and the baseline mean function under this semiparametric model. The asymptotic properties of these estimators are established. Joint inferences about recurrent events and death are also discussed. The finite-sample behavior of the proposed inference procedures is assessed through simulation studies. An application to a well-known bladder tumor study is provided.

Key words and phrases: Censoring, competing risks, counting process, empirical process, multiple events, survival analysis.

1. Introduction

In many longitudinal studies, the event of interest may recur on the same subject. Medical examples of recurrent events include repeated opportunistic infections in HIV-infected subjects (Li and Lagakos (1997)), recurrent seizures in epileptic patients (Albert (1991)) and tumor recurrences in cancer patients (Byar (1980)). The recurrence of serious events often elevates the risk of death so that the subject may experience death, which precludes further recurrent events.

The data on recurrent events provide richer information about disease progression than those of a single event. Statistical analysis of recurrent event data has received tremendous attention. There exist regression methods for studying the gap times between events (Prentice, Williams and Peterson (1981)), the marginal hazards for individual recurrences (Wei, Lin and Weissfeld (1989)) and the intensity/rate functions of the recurrent event process (Andersen and Gill (1982), Pepe and Cai (1993), Lawless, Nadeau and Cook (1997), Lin, Wei, Yang and Ying (2000)). All these methods, however, deal primarily with recurrent events that are not terminated by death.

Some efforts have been put forth recently on the regression analysis of recurrent events in the presence of death. Li and Lagakos (1997) adapted the

method of Wei et al. (1989) by treating death as a censoring variable for recurrent events, or by defining the failure time for each recurrence as the minimum of the recurrent event time and survival time. Cook and Lawless (1997) studied the mean and rate functions of recurrent events among survivors at certain time points. Neither of these two approaches yields results that pertain to the subject's ultimate recurrence experience.

In this article, we focus on the marginal mean of the cumulative number of recurrent events over time. This mean function incorporates the fact that a subject who dies cannot experience further recurrent events and thus characterizes the subject's ultimate recurrence experience in the presence of death. Nonparametric inferences for this mean function in the one- and two-sample settings have recently been studied by Cook and Lawless (1997), Ghosh and Lin (2000) and Strawderman (2000). In this article, we propose semiparametric regression models which specify multiplicative covariate effects on the marginal mean function. We develop two procedures for estimating the regression parameters and the mean function: one is based on the familiar inverse probability of censoring weighting, and one is a novel approach based on modelling survival time. The asymptotic and finite-sample properties of the resultant estimators are studied. We also provide model checking techniques as well as methods for joint inferences on the covariate effects for recurrent events and death.

In the next section, we present the semiparametric regression models for the mean function along with the corresponding inference procedures. We report in Section 3 on the results of some simulation studies. In Section 4, we apply the proposed methods to data from a cancer clinical trial. Some concluding remarks are made in Section 5.

2. Regression Methods

2.1. Data structures and regression models

Let $N^*(t)$ be the number of recurrent events over the time interval $[0, t]$, let D be the survival time, and let $\mathbf{Z}$ be a $p \times 1$ vector of covariates. Naturally, a subject who dies cannot experience further recurrent events so that $N^*(t)$ does not jump after D. We wish to formulate the effects of $\mathbf{Z}$ on the marginal distribution of $N^*(\cdot)$ without specifying the nature of dependence among recurrent events, or that between recurrent events and death. Define $\mu_{\mathbf{Z}}(t) = E\{N^*(t)|\mathbf{Z}\}$, which is the marginal expected number of recurrent events up to t associated with $\mathbf{Z}$, and which acknowledges the fact that there is no further recurrence after death. We formulate $\mu_{\mathbf{Z}}(t)$ through the semiparametric model

$$\mu_{\mathbf{Z}}(t) = e^{\boldsymbol{\beta}_0^T \mathbf{z}} \mu_0(t), \tag{1}$$

where $\mu_0(\cdot)$ is an unspecified continuous function, and $\boldsymbol{\beta}_0$ is a $p \times 1$ vector of unknown regression parameters.

It is implicitly assumed in the above description that covariates are time-invariant. To accommodate time-varying covariates, we consider the rate function $d\mu_{\mathbf{Z}}(t) \equiv E\{dN^*(t)|\mathbf{Z}(s) : s \geq 0\}$, where $\mathbf{Z}(\cdot)$ is a p-dimensional external covariate process (Kalbfleisch and Prentice (1980, Section 5.2.1)). We then generalize (1) as follows

$$d\mu_{\mathbf{Z}}(t) = e^{\boldsymbol{\beta}_0^T \mathbf{Z}(t)} d\mu_0(t). \qquad (2)$$

Under (2), $\mu_{\mathbf{Z}}(t) = \int_0^t e^{\boldsymbol{\beta}_0^T \mathbf{Z}(s)} d\mu_0(s)$, which reduces to (1) if covariates are all time-invariant.

Models (1) and (2) specify that covariates have multiplicative effects on the mean and rate functions of recurrent events, respectively, and are thus referred to as the proportional means and rates models. In the absence of death, these models have been studied by Pepe and Cai (1993), Lawless, Nadean and Cook (1997) and Lin et al. (2000). The presence of death poses serious new challenges.

In most applications, the follow-up is limited so that $N^*(\cdot)$ may be censored. In fact, it is the combination of censoring and death that creates the biggest challenge in the estimation of models (1) and (2). Let C denote the follow-up or censoring time. It is assumed that $N^*(\cdot)$ is independent of C conditional on $\mathbf{Z}(\cdot)$. Note that $N^*(\cdot)$ can only be observed up to C and that in general only the minimum of D and C is known. Write $X = D \wedge C$, $\delta = I(D \leq C)$ and $N(t) = N^*(t \wedge C)$, where $a \wedge b = \min(a, b)$ and $I(\cdot)$ is the indicator function. For a random sample of n subjects, the data consist of $\{N_i(\cdot), X_i, \delta_i, \mathbf{Z}_i(\cdot)\}$, $i = 1, \ldots, n$. Our task is to derive estimation procedures for $\boldsymbol{\beta}_0$ and $\mu_0(\cdot)$ of models (1) and (2) based on $\{N_i(\cdot), X_i, \delta_i, \mathbf{Z}_i(\cdot)\}$, $i = 1, \ldots, n$.

2.2. Estimation method when censoring times are known

We first consider the simplified setting in which censoring times are known for all subjects, including those who die during the study. This will be the case if, for instance, censoring is caused solely by the termination of the study so that C_i is the difference between the date of study termination and that of study entry for the ith subject. One can then use the following estimating function to estimate $\boldsymbol{\beta}_0$:

$$\mathbf{U}(\boldsymbol{\beta}) = \sum_{i=1}^n \int_0^{\tau^*} \left\{ \mathbf{Z}_i(t) - \frac{\sum_{j=1}^n I(C_j \geq t)\mathbf{Z}_j(t)e^{\boldsymbol{\beta}^T \mathbf{Z}_j(t)}}{\sum_{j=1}^n I(C_j \geq t)e^{\boldsymbol{\beta}^T \mathbf{Z}_j(t)}} \right\} I(C_i \geq t)dN_i(t), \quad (3)$$

where τ^* is a constant such that $\Pr(C_i \geq \tau^*) > 0$, $i = 1, \ldots, n$. Simple algebraic manipulation yields

$$\mathbf{U}(\boldsymbol{\beta}_0) = \sum_{i=1}^{n} \int_0^{\tau^*} \left\{ \mathbf{Z}_i(t) - \frac{\sum_{j=1}^{n} I(C_j \geq t) \mathbf{Z}_j(t) e^{\boldsymbol{\beta}_0^T \mathbf{z}_j(t)}}{\sum_{j=1}^{n} I(C_j \geq t) e^{\boldsymbol{\beta}_0^T \mathbf{z}_j(t)}} \right\} I(C_i \geq t) \{ dN_i^*(t) - e^{\boldsymbol{\beta}_0^T \mathbf{z}_i(t)} d\mu_0(t) \},$$

which is a sum of integrals with respect to zero-mean processes. Using empirical-process arguments such as those of Lin et al. (2000), one can show that $n^{-1/2} \mathbf{U}(\boldsymbol{\beta}_0)$ is asymptotically zero-mean normal, so that the solution to $\mathbf{U}(\boldsymbol{\beta}_0) = \mathbf{0}$ is consistent and asymptotically normal. In fact, the asymptotic results described by Pepe and Cai (1993), Lawless et al. (1997) and Lin et al. (2000) are applicable here because those results require only that the censoring times be known and allow $N^*(\cdot)$ to be an arbitrary process satisfying model (2). Furthermore, the Breslow-type estimator, as given in (2.3) of Lin et al. (2000), continues to be consistent and asymptotically Gaussian.

In virtually all practical situations, there is potential loss to follow-up. Thus in general, C_i is unknown if the ith subject dies before he/she is censored, and (3) cannot be evaluated. We consider two modifications of (3) which replace $I(C_i \geq t)$, $i = 1, \ldots, n$, by observable quantities with the same expectations: the first modification is related to the familiar inverse probability of censoring weighting (IPCW) technique (Robins and Rotnitzky (1992)), and the second involves modeling the survival distribution and is referred to as inverse probability of survival weighting (IPSW).

2.3. IPCW method

Suppose that C_i, $i = 1, \ldots, n$, have a common distribution with survival function $G(t) \equiv \Pr(C \geq t)$, and that C and D are independent. Consider the quantity $w_i(t) = I(C_i \geq D_i \wedge t) G(t) / G(X_i \wedge t)$, which reduces to $I(C_i \geq t)$ in the absence of death, i.e., $D_i = \infty$. By the law of conditional expectations, $E\{w_i(t)\} = E[E\{I(C_i \geq D_i \wedge t) G(t) / G(D_i \wedge t) | D_i\}] = G(t) E\{G(D_i \wedge t) / G(D_i \wedge t)\} = G(t)$, which is the expectation of $I(C_i \geq t)$. Since G is unknown, but can be estimated by the Kaplan-Meier estimator $\widehat{G}$ say, we approximate $w_i(t)$ by $\widehat{w}_i(t) \equiv I(C_i \geq D_i \wedge t) \widehat{G}(t) / \widehat{G}(X_i \wedge t)$.

It is possible to allow C to depend on $\mathbf{Z}(\cdot)$ but require that C and D be conditionally independent given $\mathbf{Z}(\cdot)$. Define $w_i^C(t) = I(C_i \geq D_i \wedge t) G(t|\mathbf{Z}_i) / G(X_i \wedge t|\mathbf{Z}_i)$, where $G(t|\mathbf{Z})$ is the survival function of C conditional on $\mathbf{Z}(\cdot)$. Again, by the law of conditional expectations, $E\{w_i^C(t)|\mathbf{Z}_i\} = G(t|\mathbf{Z}_i)$. It is convenient to formulate $G(t|\mathbf{Z})$ through the proportional hazards model (Cox (1972)):

$$\lambda^C(t|\mathbf{Z}) = \lambda_0^C(t) e^{\boldsymbol{\gamma}_C^T \mathbf{Z}(t)}, \tag{4}$$

where $\lambda^C(t|\mathbf{Z})$ is the hazard function corresponding to $G(t|\mathbf{Z})$, $\lambda_0^C(\cdot)$ is an unspecified baseline hazard function, and $\boldsymbol{\gamma}_C$ is a $p \times 1$ vector of unknown regression parameters. Let $\widehat{G}(t|\mathbf{Z}) = \exp\{-\int_0^t e^{\widehat{\boldsymbol{\gamma}}_C^T \mathbf{Z}(u)} d\widehat{\Lambda}_0^C(u)\}$, where $\widehat{\boldsymbol{\gamma}}_C$ and $\widehat{\Lambda}_0^C(\cdot)$ are the maximum partial likelihood (Cox (1975)) and Breslow (1974) estimators of $\boldsymbol{\gamma}_C$ and $\Lambda_0^C(t) \equiv \int_0^t \lambda_0^C(u) du$. We then approximate $w_i^C(t)$ by $\widehat{w}_i^C(t) \equiv I(C_i \geq D_i \wedge t)\widehat{G}(t|\mathbf{Z}_i)/\widehat{G}(X_i \wedge t|\mathbf{Z}_i)$, $i = 1, \ldots, n$.

By replacing $I(C_i \geq t)$ in (3) with $\widehat{w}_i^C(t)$, $i = 1, \ldots, n$, we obtain the following estimating function for $\boldsymbol{\beta}_0$:

$$\mathbf{U}^C(\boldsymbol{\beta}) = \sum_{i=1}^n \int_0^\tau \{\mathbf{Z}_i(t) - \overline{\mathbf{Z}}^C(\boldsymbol{\beta}, t)\}\widehat{w}_i^C(t) dN_i(t), \tag{5}$$

where $\overline{\mathbf{Z}}^C(\boldsymbol{\beta}, t) = \tilde{\mathbf{S}}^{(1)}(\boldsymbol{\beta}, t)/\tilde{S}^{(0)}(\boldsymbol{\beta}, t)$, and $\tilde{\mathbf{S}}^{(k)}(\boldsymbol{\beta}, t) = n^{-1}\sum_{j=1}^n \widehat{w}_j^C(t)\mathbf{Z}_j(t)^{\otimes k} e^{\boldsymbol{\beta}^T \mathbf{Z}_j(t)}$, $k = 0, 1, 2$, with $\mathbf{a}^{\otimes 0} = 1$, $\mathbf{a}^{\otimes 1} = \mathbf{a}$ and $\mathbf{a}^{\otimes 2} = \mathbf{a}\mathbf{a}^T$. For technical reasons, the constant $\tau > 0$ is chosen such that $\Pr(X_i \geq \tau|\mathbf{Z}_i) > 0$, $i = 1, \ldots, n$. Let $\widehat{\boldsymbol{\beta}}_C$ be the solution to $\mathbf{U}^C(\boldsymbol{\beta}) = \mathbf{0}$. The corresponding estimator of the baseline mean function $\mu_0(\cdot)$ is given by the Breslow-type estimator

$$\widehat{\mu}_0^C(t) \equiv \sum_{i=1}^n \int_0^t \frac{\widehat{w}_i^C(u) dN_i(u)}{n\tilde{S}^{(0)}(\widehat{\boldsymbol{\beta}}_C, u)}, \quad 0 \leq t \leq \tau, \tag{6}$$

which, in the absence of death, reduces to (2.3) of Lin et al. (2000).

Remark 1. The replacement of $I(C_i \geq t)$ with $\widehat{w}_i^C(t)$ is reminiscent of the inverse probability of censoring weighting (IPCW) technique (Robins and Rotnitzky (1992)), which has been used by various authors (e.g., Lin and Ying (1993), Cheng, Wei and Ying (1995), Fine and Gray (1999)) in different contexts.

2.4. IPSW method

The IPCW method requires modeling the censoring distribution, which is a nuisance. In this section we develop an alternative method that involves modeling the survival distribution, which, unlike censoring, is of clinical interest. This method also lends itself to the joint inferences of recurrent events and death to be discussed in Section 2.8.

As in the previous section, we would like to replace $I(C_i \geq t)$ by an observable quantity with the same expectation. Since X_i is always observed, we substitute $I(X_i \geq t)$ for $I(C_i \geq t)$ in (3), and divide it by $S(t|\mathbf{Z}_i) \equiv \Pr(D_i \geq t|\mathbf{Z}_i)$. Write $w_i^D(t) = I(X_i \geq t)/S(t|\mathbf{Z}_i)$. Assume that D and C are independent conditional on $\mathbf{Z}(\cdot)$. It then follows that $E\{w_i^D(t)|\mathbf{Z}_i\} = E\{I(X_i \geq t)|\mathbf{Z}_i\}/S(t|\mathbf{Z}_i) = S(t|\mathbf{Z}_i)G(t|\mathbf{Z}_i)/S(t|\mathbf{Z}_i) = G(t|\mathbf{Z}_i)$.

Remark 2. While it may be reasonable to assume that censoring is independent of covariates, the same cannot be said of survival. Thus, we do not consider the case in which survival does not depend on covariates.

We specify the proportional hazards model for the survival distribution:

$$\lambda^D(t|\mathbf{Z}) = \lambda_0^D(t)e^{\boldsymbol{\gamma}_D^T \mathbf{Z}(t)}, \tag{7}$$

where $\lambda^D(t|\mathbf{Z})$ is the hazard function corresponding to $S(t|\mathbf{Z})$, $\lambda_0^D(t)$ is an unspecified baseline hazard function, and $\boldsymbol{\gamma}_D$ is a $p \times 1$ vector of regression coefficients. Let $\widehat{S}(t|\mathbf{Z}) = \exp\{-\int_0^t e^{\widehat{\boldsymbol{\gamma}}_D^T \mathbf{Z}(u)} d\widehat{\Lambda}_0^D(u)\}$, where $\widehat{\boldsymbol{\gamma}}_D$ and $\widehat{\Lambda}_0^D(t)$ are the maximum partial likelihood and Breslow estimators of $\boldsymbol{\gamma}_D$ and $\Lambda_0^D(t) \equiv \int_0^t \lambda_0^D(u)du$. We then approximate $w_i^D(t)$ by $\widehat{w}_i^D(t) \equiv I(X_i \geq t)/\widehat{S}(t|\mathbf{Z}_i)$ and modify (3) as

$$\mathbf{U}^D(\boldsymbol{\beta}) = \sum_{i=1}^n \int_0^\tau \{\mathbf{Z}_i(t) - \overline{\mathbf{Z}}^D(\boldsymbol{\beta}, t)\}\widehat{w}_i^D(t)dN_i(t), \tag{8}$$

where $\overline{\mathbf{Z}}^D(\boldsymbol{\beta}, t) = \widehat{\mathbf{S}}^{(1)}(\boldsymbol{\beta}, t)/\widehat{S}^{(0)}(\boldsymbol{\beta}, t)$, and $\widehat{\mathbf{S}}^{(k)}(\boldsymbol{\beta}, t) = n^{-1}\sum_{j=1}^n \widehat{w}_j^D(t)\mathbf{Z}_j(t)^{\otimes k}e^{\boldsymbol{\beta}^T \mathbf{Z}_j(t)}$, $k = 0, 1, 2$. Let $\widehat{\boldsymbol{\beta}}_D$ be the solution to $\mathbf{U}^D(\boldsymbol{\beta}) = \mathbf{0}$. The corresponding estimator of $\mu_0(\cdot)$ is given by

$$\widehat{\mu}_0^D(t) \equiv \sum_{i=1}^n \int_0^t \frac{\widehat{w}_i^D(u)dN_i(u)}{n\widehat{S}^{(0)}(\widehat{\boldsymbol{\beta}}_D, u)}, \quad 0 \leq t \leq \tau. \tag{9}$$

Remark 3. We refer to the technique used in (8) and (9) as the inverse probability of survival weighting (IPSW), which shares the spirit of the IPCW technique.

2.5. Asymptotic results for the IPCW method

We impose regularity conditions, similar to those of Andersen and Gill (1982, Thm 4.1).

A. $\{N_i(\cdot), X_i, \delta_i, \mathbf{Z}_i(\cdot)\}$ $(i = 1, \ldots, n)$ are independent and identically distributed (i.i.d.).

B. There exists a $\tau > 0$ such that $P(X_i \geq \tau|\mathbf{Z}_i) > 0$ $(i = 1, \ldots, n)$.

C. $N_i(\tau)$, $i = 1, \ldots, n$, are bounded.

D. $\mathbf{Z}_i(\cdot)$, $i = 1, \ldots, n$, have bounded total variations, i.e., $|Z_{ji}(0)| + \int_0^\tau |dZ_{ji}(t)| \leq K$ for all $j = 1, \ldots, p$ and $i = 1, \ldots, n$, where Z_{ji} is the jth component of $\mathbf{Z}_i$ and $0 < K < \infty$.

E. $\mathbf{A} \equiv \mathrm{E}[\int_0^\tau \{\mathbf{Z}(t) - \overline{\mathbf{z}}(\boldsymbol{\beta}_0, t)\}^{\otimes 2}G(t|\mathbf{Z})e^{\boldsymbol{\beta}_0^T \mathbf{Z}(t)}d\mu_0(t)]$ is positive definite, where $\overline{\mathbf{z}}(\boldsymbol{\beta}, t)$ is the limit of $\overline{\mathbf{Z}}^C(\boldsymbol{\beta}, t)$.

It is useful to introduce notation: for $i = 1, \ldots, n$, let

$$M_i(t) = \int_0^t w_i^C(u)\{dN_i(u) - e^{\boldsymbol{\beta}_0^T \mathbf{Z}_i(u)}d\mu_0(u)\}, \tag{10}$$

$N_i^C(t) = I(X_i \leq t, \delta_i = 0)$, and $M_i^C(t) = N_i^C(t) - \int_0^t Y_i(u)e^{\boldsymbol{\gamma}_C^T \mathbf{Z}_i(u)}\lambda_0^C(u)du$, where $Y_i(t) = I(X_i \geq t)$. Also, let $\widehat{M}_i(t) = \int_0^t \widehat{w}_i^C(u)\{dN_i(u) - e^{\widehat{\boldsymbol{\beta}}_C^T \mathbf{Z}_i(u)}d\widehat{\mu}_0^C(u)\}$, and $\widehat{M}_i^C(t) = N_i^C(t) - \int_0^t Y_i(u)e^{\widehat{\boldsymbol{\gamma}}_C^T \mathbf{Z}_i(u)}d\widehat{\Lambda}_0^C(u)$. We first state the asymptotic properties of $\widehat{\boldsymbol{\beta}}_C$.

Theorem 1. *The estimator $\widehat{\boldsymbol{\beta}}_C$ is strongly consistent, i.e., $\widehat{\boldsymbol{\beta}}_C \rightarrow_{a.s.} \boldsymbol{\beta}_0$. Furthermore, $n^{1/2}(\widehat{\boldsymbol{\beta}}_C - \boldsymbol{\beta}_0)$ converges in distribution to a zero-mean normal random vector with a covariance matrix that can be consistently estimated by $\widehat{\mathbf{A}}_C^{-1}\widehat{\boldsymbol{\Sigma}}_C\widehat{\mathbf{A}}_C^{-1}$, where $\widehat{\mathbf{A}}_C = -n^{-1}\partial\mathbf{U}^C(\widehat{\boldsymbol{\beta}}_C)/\partial\boldsymbol{\beta}$, $\widehat{\boldsymbol{\Sigma}}_C = n^{-1}\sum_{i=1}^n(\widehat{\boldsymbol{\eta}}_i^C + \widehat{\boldsymbol{\psi}}_i^C)^{\otimes 2}$, $\widehat{\boldsymbol{\eta}}_i^C = \int_0^\tau\{\mathbf{Z}_i(t) - \overline{\mathbf{Z}}^C(\widehat{\boldsymbol{\beta}}_C, t)\}d\widehat{M}_i(t)$,*

$$\widehat{\boldsymbol{\psi}}_i^C = \int_0^\tau \widehat{\mathbf{B}}_C\left\{\mathbf{Z}_i(t) - \frac{\widehat{\mathbf{R}}^{(1)}(\widehat{\boldsymbol{\gamma}}_C, t)}{\widehat{R}^{(0)}(\widehat{\boldsymbol{\gamma}}_C, t)}\right\}d\widehat{M}_i^C(t) + \int_0^\tau \frac{\widehat{\mathbf{q}}^C(t)}{\widehat{R}^{(0)}(\widehat{\boldsymbol{\gamma}}_C, t)}d\widehat{M}_i^C(t),$$

$$\widehat{\mathbf{B}}_C = -n^{-1}\sum_{i=1}^n\int_0^\tau\{\mathbf{Z}_i(t) - \overline{\mathbf{Z}}^C(\widehat{\boldsymbol{\beta}}_C, t)\}\widehat{\mathbf{g}}^C(X_i, t; \mathbf{Z}_i)^T\widehat{\boldsymbol{\Omega}}_C^{-1}I(t > X_i)d\widehat{M}_i(t),$$

$$\widehat{\mathbf{g}}^C(X_i, t; \mathbf{Z}_i) = \int_{X_i}^t e^{\widehat{\boldsymbol{\gamma}}_C^T \mathbf{Z}_i(u)}\left\{\mathbf{Z}_i(u) - \frac{\widehat{\mathbf{R}}^{(1)}(\widehat{\boldsymbol{\gamma}}_C, u)}{\widehat{R}^{(0)}(\widehat{\boldsymbol{\gamma}}_C, u)}\right\}d\widehat{\Lambda}_0^C(u),$$

$$\widehat{\mathbf{q}}^C(t) = -n^{-1}\sum_{i=1}^n\int_0^\tau\{\mathbf{Z}_i(u) - \overline{\mathbf{Z}}^C(\widehat{\boldsymbol{\beta}}_C, u)\}e^{\widehat{\boldsymbol{\gamma}}_C^T \mathbf{Z}_i(t)}I(u \geq t > X_i)d\widehat{M}_i(u),$$

$$\widehat{\boldsymbol{\Omega}}_C = n^{-1}\sum_{i=1}^n\int_0^\tau\left[\frac{\mathbf{R}^{(2)}(\widehat{\boldsymbol{\gamma}}_C, t)}{R^{(0)}(\widehat{\boldsymbol{\gamma}}_C, t)} - \left\{\frac{\mathbf{R}^{(1)}(\widehat{\boldsymbol{\gamma}}_C, t)}{R^{(0)}(\widehat{\boldsymbol{\gamma}}_C, t)}\right\}^{\otimes 2}\right]dN_i^C(t),$$

and $\widehat{\mathbf{R}}^{(k)}(\boldsymbol{\gamma}, t) = n^{-1}\sum_{j=1}^n Y_j(t)\mathbf{Z}_j(t)^{\otimes k}e^{\boldsymbol{\gamma}^T \mathbf{Z}_j(t)}$, $k = 0, 1, 2$.

The proofs of theorems are relegated to the Appendix.

Remark 4. If $\widehat{w}_i^C(t)$, $i = 1, \ldots, n$, in (5) are replaced by $\widehat{w}_i(t)$, then the conclusion of Theorem 1 continues to hold, but with

$$\widehat{\boldsymbol{\eta}}_i^C = \int_0^\tau\left\{\mathbf{Z}_i(t) - \frac{\sum_{j=1}^n\widehat{w}_j(t)\mathbf{Z}_j(t)e^{\widehat{\boldsymbol{\beta}}_C^T \mathbf{Z}_j(t)}}{\sum_{j=1}^n\widehat{w}_j(t)e^{\widehat{\boldsymbol{\beta}}_C^T \mathbf{Z}_j(t)}}\right\}d\tilde{M}_i(t), \quad \widehat{\boldsymbol{\psi}}_i^C = \int_0^\tau\frac{\widehat{\mathbf{q}}(t)}{\sum_{j=1}^n Y_j(t)}d\tilde{M}_i^C(t),$$

where

$$\widehat{\mathbf{q}}(t) = -n^{-1}\sum_{i=1}^n\int_0^\tau\left\{\mathbf{Z}_i(u) - \frac{\sum_{j=1}^n\widehat{w}_j(u)\mathbf{Z}_j(u)e^{\widehat{\boldsymbol{\beta}}_C^T \mathbf{Z}_j(u)}}{\sum_{j=1}^n\widehat{w}_j(u)e^{\widehat{\boldsymbol{\beta}}_C^T \mathbf{Z}_j(u)}}\right\}I(u \geq t > X_i)d\tilde{M}_i(u),$$

$$\tilde{M}_i(t) = \int_0^t \widehat{w}_i(u)\{dN_i(u) - e^{\widehat{\boldsymbol{\beta}}_C^T \mathbf{z}_i(u)} d\tilde{\mu}_0^C(u)\}, \quad \tilde{M}_i^C(t) = N_i^C(t) - \int_0^t Y_i(u) d\tilde{\Lambda}_0^C(u),$$

$$\tilde{\mu}_0^C(t) = \sum_{i=1}^n \int_0^t \frac{\widehat{w}_i(u) dN_i(u)}{\sum_{j=1}^n \widehat{w}_j(u) e^{\widehat{\boldsymbol{\beta}}_C^T \mathbf{z}_j(u)}}, \quad \tilde{\Lambda}_0^C(t) = \sum_{i=1}^n \int_0^t \frac{dN_i^C(u)}{\sum_{j=1}^n Y_j(u)}.$$

The proof for this result is similar to, but simpler than, the proof of Theorem 1.

We describe below the asymptotic properties of $\widehat{\mu}_0^C(t)$.

Theorem 2. *The process $n^{1/2}\{\widehat{\mu}_0^C(t) - \mu_0(t)\}$, $0 \leq t \leq \tau$, converges weakly to a mean-zero Gaussian process whose covariance function at (s,t) can be consistently estimated by $\widehat{\xi}_C(s,t) \equiv n^{-1} \sum_{i=1}^n \widehat{\phi}_i^C(s)\widehat{\phi}_i^C(t)$, where*

$$\widehat{\phi}_i^C(t) = \int_0^t \frac{d\widehat{M}_i(u)}{\tilde{S}^{(0)}(\widehat{\boldsymbol{\beta}}_C, u)} + \int_0^\tau \frac{\widehat{p}_1^C(u,t)}{\widehat{R}^{(0)}(\widehat{\boldsymbol{\gamma}}_C, u)} d\widehat{M}_i^C(u)$$

$$+ \int_0^\tau \widehat{\mathbf{p}}_2^C(t)^T \left\{ \mathbf{Z}_i(u) - \frac{\widehat{\mathbf{R}}^{(1)}(\widehat{\boldsymbol{\gamma}}_C, u)}{\widehat{R}^{(0)}(\widehat{\boldsymbol{\gamma}}_C, u)} \right\} d\widehat{M}_i^C(u) + \mathbf{H}_C^T(\widehat{\boldsymbol{\beta}}_C, t)\widehat{\mathbf{A}}_C^{-1} n^{-1/2} \sum_{j=1}^n (\widehat{\boldsymbol{\eta}}_j^C + \widehat{\boldsymbol{\psi}}_j^C),$$

$$\mathbf{H}_C(\boldsymbol{\beta}, t) = -\int_0^t \overline{\mathbf{Z}}^C(\boldsymbol{\beta}, u) d\widehat{\mu}_0^C(u),$$

$$\widehat{p}_1^C(u,t) = -n^{-1} \sum_{i=1}^n \int_0^t \frac{I(s \geq u > X_i) e^{\widehat{\boldsymbol{\gamma}}_C^T \mathbf{z}_i(u)}}{\tilde{S}^{(0)}(\widehat{\boldsymbol{\beta}}_C, s)} d\widehat{M}_i(s),$$

$$\widehat{\mathbf{p}}_2^C(t) = -n^{-1} \sum_{i=1}^n \int_0^t \frac{I(u > X_i)\widehat{\mathbf{g}}^C(X_i, u; \mathbf{Z}_i)^T \widehat{\boldsymbol{\Omega}}_C^{-1}}{\tilde{S}^{(0)}(\widehat{\boldsymbol{\beta}}_C, u)} d\widehat{M}_i(u).$$

The asymptotic normality of $\widehat{\mu}_0^C(t)$, together with the consistent variance estimator $\widehat{\xi}_C(t,t)$, allows one to construct confidence intervals for $\mu_0(t)$. Since $\mu_0(t)$ is nonnegative, we consider the transformed variable $n^{1/2}[\log\{\widehat{\mu}_0^C(t)\} - \log\{\mu_0(t)\}]$, which has a distribution asymptotically equivalent to $n^{1/2}\{\widehat{\mu}_0^C(t) - \mu_0(t)\}/\mu_0(t)$ provided $\mu_0(t) > 0$. With the log-transformation, an approximate $(1-\alpha)$ confidence interval for $\mu_0(t)$ is given by $\widehat{\mu}_0^C(t)e^{\pm n^{-1/2}z_{\alpha/2}\widehat{\xi}_C^{1/2}(t,t)/\widehat{\mu}_0^C(t)}$, where $z_{\alpha/2}$ denotes the $100(1-\alpha/2)$ percentile of the standard normal distribution.

2.6. Asymptotic results for the IPSW method

We again impose regularity conditions A–E given in Section 2.5. Since $w_i^C(t)$ and $w_i^D(t)$ have the same expectation, $\overline{z}(\boldsymbol{\beta}, t)$ is also the limit of $\overline{\mathbf{Z}}^D(\boldsymbol{\beta}, t)$. For $i = 1, \ldots, n$, let $N_i^P(t) = I(X_i \leq t, \delta_i = 1)$, $M_i^P(t) = N_i^P(t) - \int_0^t Y_i(u)e^{\boldsymbol{\gamma}_D^T \mathbf{z}_i(u)} d\Lambda_0^D(u)$ and $M_i^\dagger(t) = \int_0^t w_i^D(u)\{dN_i(u) - e^{\boldsymbol{\beta}_0^T \mathbf{z}_i(u)} d\mu_0(u)\}$. Also, let $\widehat{M}_i^\dagger(t) = \int_0^t \widehat{w}_i^D(u)\{dN_i(u) - e^{\widehat{\boldsymbol{\beta}}_D^T \mathbf{z}_i(u)} d\widehat{\mu}_0^D(u)\}$, and $\widehat{M}_i^P(t) = N_i^P(t) - \int_0^t Y_i(u)e^{\widehat{\boldsymbol{\gamma}}_D^T \mathbf{z}_i(u)} d\widehat{\Lambda}_0^D(u)$. The asymptotic properties for $\widehat{\boldsymbol{\beta}}_D$ and $\widehat{\mu}_0^D(\cdot)$ are stated in the following theorems.

Theorem 3. *The estimator $\widehat{\boldsymbol{\beta}}_D$ is strongly consistent. The random vector $n^{1/2}(\widehat{\boldsymbol{\beta}}_D - \boldsymbol{\beta}_0)$ converges in distribution to a zero-mean normal random vector with a covariance matrix that can be consistently estimated by $\widehat{\mathbf{A}}_D^{-1}\widehat{\boldsymbol{\Sigma}}_D\widehat{\mathbf{A}}_D^{-1}$, where $\widehat{\mathbf{A}}_D = -n^{-1}\partial \mathbf{U}^D(\widehat{\boldsymbol{\beta}}_D)/\partial\boldsymbol{\beta}$, $\widehat{\boldsymbol{\Sigma}}_D = n^{-1}\sum_{i=1}^n(\widehat{\boldsymbol{\eta}}_i^D + \widehat{\boldsymbol{\psi}}_i^D)^{\otimes 2}$, $\widehat{\boldsymbol{\eta}}_i^D = \int_0^\tau \{\mathbf{Z}_i(t) - \overline{\mathbf{Z}}^D(\widehat{\boldsymbol{\beta}}_D,t)\}d\widehat{M}_i^\dagger(t),$*

$$\widehat{\boldsymbol{\psi}}_i^D = \int_0^\tau \widehat{\mathbf{B}}_D\left\{\mathbf{Z}_i(t) - \frac{\widehat{\mathbf{R}}^{(1)}(\widehat{\boldsymbol{\gamma}}_D,t)}{\widehat{R}^{(0)}(\widehat{\boldsymbol{\gamma}}_D,t)}\right\}d\widehat{M}_i^D(t) + \int_0^\tau \frac{\widehat{\mathbf{q}}^D(t)}{\widehat{R}^{(0)}(\widehat{\boldsymbol{\gamma}}_D,t)}d\widehat{M}_i^D(t),$$

$$\widehat{\mathbf{B}}_D = n^{-1}\sum_{i=1}^n\int_0^\tau\{\mathbf{Z}_i(t) - \overline{\mathbf{Z}}^D(\widehat{\boldsymbol{\beta}}_D,t)\}\widehat{\mathbf{g}}^D(t;\mathbf{Z}_i)^T\widehat{\boldsymbol{\Omega}}_D^{-1}d\widehat{M}_i^\dagger(t),$$

$$\widehat{\mathbf{g}}^D(t;\mathbf{Z}_i) = \int_0^t e^{\widehat{\boldsymbol{\gamma}}_D^T\mathbf{Z}_i(u)}\left\{\mathbf{Z}_i(u) - \frac{\widehat{\mathbf{R}}^{(1)}(\widehat{\boldsymbol{\gamma}}_D,u)}{\widehat{R}^{(0)}(\widehat{\boldsymbol{\gamma}}_D,u)}\right\}d\widehat{\Lambda}_0^D(u),$$

$$\widehat{\mathbf{q}}^D(t) = n^{-1}\sum_{i=1}^n\int_0^\tau\{\mathbf{Z}_i(u) - \overline{\mathbf{Z}}^D(\widehat{\boldsymbol{\beta}}_D,u)\}e^{\widehat{\boldsymbol{\gamma}}_D^T\mathbf{Z}_i(t)}I(u \geq t)d\widehat{M}_i^\dagger(u),$$

$$\widehat{\boldsymbol{\Omega}}_D = n^{-1}\sum_{i=1}^n\int_0^\tau\left[\frac{\mathbf{R}^{(2)}(\widehat{\boldsymbol{\gamma}}_D,t)}{R^{(0)}(\widehat{\boldsymbol{\gamma}}_D,t)} - \left\{\frac{\mathbf{R}^{(1)}(\widehat{\boldsymbol{\gamma}}_D,t)}{R^{(0)}(\widehat{\boldsymbol{\gamma}}_D,t)}\right\}^{\otimes 2}\right]dN_i^D(t).$$

Theorem 4. *The process $n^{1/2}\{\widehat{\mu}_0^D(t) - \mu_0(t)\}$, $0 \leq t \leq \tau$, converges weakly to a mean-zero Gaussian process whose covariance function at (s,t) can be consistently estimated by $\widehat{\xi}_D(s,t) \equiv n^{-1}\sum_{i=1}^n\widehat{\phi}_i^D(s)\widehat{\phi}_i^D(t)$, where*

$$\widehat{\phi}_i^D(t) = \int_0^t \frac{d\widehat{M}_i^\dagger(u)}{\widehat{S}^{(0)}(\widehat{\boldsymbol{\beta}}_D,u)} + \int_0^\tau \frac{\widehat{p}_1^D(u,t)}{\widehat{R}^{(0)}(\widehat{\boldsymbol{\gamma}}_D,u)}d\widehat{M}_i^D(u)$$

$$+ \int_0^\tau\widehat{\mathbf{p}}_2^D(t)^T\left\{\mathbf{Z}_i(u) - \frac{\widehat{\mathbf{R}}^{(1)}(\widehat{\boldsymbol{\gamma}}_D,u)}{\widehat{R}^{(0)}(\widehat{\boldsymbol{\gamma}}_D,u)}\right\}d\widehat{M}_i^D(u) + \mathbf{H}_D^T(\widehat{\boldsymbol{\beta}}_D,t)\widehat{\mathbf{A}}_D^{-1}n^{-1/2}\sum_{j=1}^n(\widehat{\boldsymbol{\eta}}_j^D + \widehat{\boldsymbol{\psi}}_j^D),$$

$$\mathbf{H}_D(\boldsymbol{\beta},t) = -\int_0^t\overline{\mathbf{Z}}^D(\boldsymbol{\beta},u)d\widehat{\mu}_0^D(u),$$

$$\widehat{p}_1^D(u,t) = n^{-1}\sum_{i=1}^n\int_0^t\frac{e^{\widehat{\boldsymbol{\gamma}}_D^T\mathbf{Z}_i(u)}}{\widehat{S}^{(0)}(\widehat{\boldsymbol{\beta}}_D,s)}I(s\geq u)d\widehat{M}_i^\dagger(s), \quad \widehat{\mathbf{p}}_2^D(t) = n^{-1}\sum_{i=1}^n\int_0^t\frac{\widehat{\mathbf{g}}^D(u;\mathbf{Z}_i)^T\widehat{\boldsymbol{\Omega}}_D^{-1}}{\widehat{S}^{(0)}(\widehat{\boldsymbol{\beta}}_D,u)}d\widehat{M}_i^\dagger(u).$$

Confidence intervals for $\mu_0(t)$ based on $\widehat{\mu}_0^D(t)$ and $\widehat{\xi}_D^{1/2}(t,t)$ can be constructed in a manner similar to that described in Section 2.5. In many applications, it is desirable to estimate the mean function $\mu_{\mathbf{z}}(t)$ for subjects with specific covariate values $\mathbf{z}$. If all the covariates are centered at $\mathbf{z}$, then $\mu_0(t)$ corresponds to $\mu_{\mathbf{z}}(t)$. Thus, one can obtain point and interval estimates for $\mu_{\mathbf{z}}(t)$ by using the formulae for $\widehat{\mu}_0^C(t)$ or $\widehat{\mu}_0^D(t)$, upon replacing $(\mathbf{Z}_1,\ldots,\mathbf{Z}_n)$ with $(\mathbf{Z}_1 - \mathbf{z},\ldots,\mathbf{Z}_n - \mathbf{z})$ in the dataset.

2.7. Model checking techniques

The IPCW and IPSW methods involve fitting models (4) and (7) for the censoring and survival distributions, respectively. Existing goodness-of-fit methods for the proportional hazards model with univariate right-censored data (e.g., Schoenfeld (1982); Therneau, Grambsch and Fleming (1990); Lin, Wei and Ying (1993)) can be used to check the adequacy of these models.

Here we develop goodness-of-fit techniques for models (1) and (2). For simplicity of description, we assume that the IPCW method is used, although the techniques apply to the IPSW method as well. Because $M_i(t)$, $i = 1, \ldots, n; 0 \leq t \leq \tau$, are zero-mean processes representing the differences between the observed and expected values of $N_i^*(t)$, it is natural to use $\widehat{M}_i(t)$'s as the goodness-of-fit measures. Let $\widehat{M}_i \equiv \widehat{M}_i(\tau)$, $i = 1, \ldots, n$, and assume that covariates are all time-invariant. To check the functional form of the jth component of $\mathbf{Z}$, we plot $\widehat{M}_i$ versus Z_{ji}, and compute a smoothed estimate, say using locally weighted least squares (Cleveland (1979)). If the functional form is appropriate, then the smoothed line should be close to zero for all values of Z_{ji}; otherwise, one would expect a systematic trend. Likewise, to check the exponential link function, we plot $\widehat{M}_i$ versus $\widehat{\boldsymbol{\beta}}_C^T \mathbf{Z}_i$.

To check the proportional rates/means assumption with respect to the jth component of $\mathbf{Z}$, we plot $\Delta U_j^C(\widehat{\boldsymbol{\beta}}_C, t)$ versus t, where $\Delta U_j^C(\boldsymbol{\beta}, t)$ is the increment in $U_j^C(\boldsymbol{\beta}, t)$, the jth component of

$$\mathbf{U}^C(\boldsymbol{\beta}, t) \equiv \sum_{i=1}^{n} \int_0^t \{\mathbf{Z}_i(u) - \overline{\mathbf{Z}}^C(\boldsymbol{\beta}, u)\} \widehat{w}_i^C(u) dN_i(u).$$

A lowess smooth based on the scatter plot is computed. If the estimated smooth is centered around zero for all t, then the assumption of proportionality is deemed reasonable. This procedure is similar in spirit to that of Schoenfeld (1982) for checking the proportional hazards assumption.

2.8. Joint inferences on covariate effects for recurrent events and death

There are two major reasons for simultaneously assessing the effects of covariates on recurrent events and death. First, survival time is of key interest in medical studies. Second, the marginal mean function of recurrent events is affected by the survival distribution. The survival distribution and the marginal mean function of recurrent events jointly characterize the subject's clinical experience.

In this section, we offer two strategies for performing joint inferences under models (2) and (7). We assume that the IPSW method is used for model (2), although the ideas presented here also apply to the IPCW method. To simplify

our presentation, we assume that Z consists of a single covariate. Let $\boldsymbol{\theta} = (\beta_0, \gamma_D)^T$ and $\widehat{\boldsymbol{\theta}} = (\widehat{\beta}_D, \widehat{\gamma}_D)^T$.

For our first strategy we assume, for simplicity, that $\beta_0 \leq 0$ and $\gamma_D \leq 0$. Define the hypotheses $H_{10} : \beta_0 = 0$ and $H_{20} : \gamma_D = 0$. We can test the hypotheses sequentially against one-sided alternatives along the lines of Wei, Lin and Weissfeld (1989). Let $\tilde{\beta}_D$ and $\tilde{\gamma}_D$ denote the standardized values of $\widehat{\beta}_D$ and $\widehat{\gamma}_D$. We suppose without loss of generality that $\tilde{\beta}_D < \tilde{\gamma}_D$. Let (V_1, V_2) be a bivariate zero-mean normal vector with unit variances and with a correlation equal to the estimated covariance between $\tilde{\beta}_D$ and $\tilde{\gamma}_D$. Then we reject H_{10} if $\Pr\{\min(V_1, V_2) \leq \tilde{\beta}_D\} \leq \alpha$; if H_{10} is rejected, we reject H_{20} if $\Pr(V_2 \leq \tilde{\gamma}_D) \leq \alpha$. It can be shown that the overall type I error for this multiple testing procedure is α. A similar procedure can be developed for two-sided alternatives.

For the second inference strategy, suppose that $\beta_0 = \gamma_D = \eta$. Then it is natural to estimate η by a linear combination of $\widehat{\beta}_D$ and $\widehat{\gamma}_D$, i.e., $\widehat{\eta} = c_1 \widehat{\beta}_D + c_2 \widehat{\gamma}_D$, where $c_1 + c_2 = 1$. Let $\widehat{\mathbf{V}}$ denote the estimated covariance matrix between $\widehat{\beta}_D$ and $\widehat{\gamma}_D$, and $\mathbf{e} = (1,1)^T$. It can be shown that the choice of $(c_1, c_2)^T \equiv (\mathbf{e}^T \widehat{\mathbf{V}}^{-1} \mathbf{e})^{-1} \widehat{\mathbf{V}}^{-1} \mathbf{e}$ yields an estimator of η that has the smallest asymptotic variance among all linear combinations of $\widehat{\beta}_D$ and $\widehat{\gamma}_D$ (Wei and Johnson (1985)). Although it might be unrealistic to expect $\beta_0 = \gamma_D$ exactly, the Wald statistic based on $\widehat{\eta}$ is always valid and potentially more powerful than separate tests in testing the null hypothesis of $\beta_0 = \gamma_D = 0$.

While we consider joint estimation procedures here, the interpretations of covariate effects on recurrences and death are based on the marginal models we are fitting. For example, if $\beta_0 < 0$ and $\gamma_D < 0$, then treatment decreases the mean number of recurrences and increases survival.

3. Simulation Studies

Extensive simulation studies were performed to assess the finite-sample behavior of the proposed inference procedures. In the ones reported here, Z is a 0/1 treatment indicator. Note that

$$\mu_Z(t) = \int_0^t S(u|Z) dR(u|Z), \tag{11}$$

where $S(t|Z) = \Pr(D \geq t|Z)$ and $dR(t|Z) = E\{dN^*(t)|D \geq t, Z\}$. The specification of $S(t|Z)$ and $dR(t|Z)$ induces a regression model for $\mu_Z(t)$. We considered the following models:

$$\lambda^D(t|Z, v) = v e^{\gamma_D Z} \lambda_0^D(t) \tag{12}$$

$$dR(t|Z, v) = v e^{\beta_R Z} dR_0(t), \tag{13}$$

where γ_D and β_R are regression parameters, and v is a frailty term that induces dependence among recurrences and death. For the cases considered in this section, γ_D in (12) will have the same meaning as it does in (7).

If $\gamma_D = 0$ in (12), then the induced model for $\mu_Z(t)$ is given by

$$\mu_Z(t) = e^{\beta_0 Z}\mu_0(t), \tag{14}$$

where $\beta_0 = \beta_R$ and $\mu_0(t)$ is a function of $\lambda_0^D(t)$, $dR_0(t)$ and the density of v. Clearly, (14) is a special case of (1).

In our first set of simulation studies, we generated data from (12) and (13) with $\gamma_D = 0$, $\lambda_0^D(t) = 0.25$, $R_0(t) = t$, $\beta_R = 0.5$, and v a gamma variable with mean 1 and variance σ^2. The parameter σ^2 controls the correlation among recurrences and death, the correlation being 0 under $\sigma^2 = 0$. We considered $\sigma^2 = 0, 0.25, 0.50, 1.0$. Censoring was generated from (4) with $\lambda_0^C(t) = 0.25$ and $\gamma_C = 0$ or 0.2, yielding approximately two observed recurrences per subject. We considered sample sizes $n = 50, 100, 200$. For each setting, 1000 simulation samples were generated. The results are presented in Table 1.

Table 1. Simulation results for the IPCW method.

		Based on $\widehat{w}_i(t)$ $(i=1,\ldots,n)$				Based on $\widehat{w}_i^C(t)$ $(i=1,\ldots,n)$							
		$\gamma_C = 0$				$\gamma_C = 0$				$\gamma_C = 0.2$			
n	σ^2	Bias	SE	SEE	CP	Bias	SE	SEE	CP	Bias	SE	SEE	CP
50	0	0.00	0.230	0.217	0.938	-0.02	0.214	0.201	0.940	-0.01	0.364	0.344	0.930
50	0.25	-0.01	0.281	0.267	0.937	-0.01	0.245	0.232	0.939	-0.03	0.597	0.561	0.930
50	0.5	0.02	0.307	0.311	0.940	0.01	0.260	0.245	0.938	0.01	0.577	0.541	0.929
50	1	-0.01	0.398	0.374	0.934	0.00	0.315	0.297	0.939	0.00	0.397	0.385	0.931
100	0	0.00	0.157	0.154	0.945	0.01	0.150	0.147	0.941	-0.01	0.254	0.240	0.942
100	0.25	0.00	0.192	0.191	0.943	0.00	0.165	0.159	0.943	0.01	0.419	0.401	0.941
100	0.5	0.00	0.222	0.220	0.942	0.01	0.183	0.178	0.945	-0.01	0.388	0.362	0.940
100	1	0.01	0.273	0.269	0.940	-0.01	0.222	0.215	0.944	-0.02	0.330	0.320	0.939
200	0	0.00	0.105	0.104	0.947	0.00	0.105	0.104	0.948	0.00	0.182	0.177	0.944
200	0.25	0.00	0.138	0.137	0.949	0.01	0.122	0.119	0.947	0.00	0.290	0.284	0.945
200	0.5	0.00	0.163	0.161	0.948	0.00	0.131	0.130	0.949	-0.02	0.284	0.281	0.946
200	1	0.01	0.194	0.195	0.950	0.00	0.158	0.155	0.948	0.00	0.196	0.190	0.944

Note: Bias is the mean of the estimator of β_0 minus β_0; SE is the standard error of the estimator of β_0; SEE is the mean of the standard error estimator; CP is the coverage probability of the 95% Wald confidence interval.

The results indicate that the IPCW estimators are virtually unbiased. The standard error estimators reflect well the true variabilities of the parameter estimators, and corresponding confidence intervals have reasonable coverage probabilities, at least for $n \geq 100$. The accuracy of the asymptotic approximation does not appear to depend appreciably on the amount of correlation between the terminal and recurrent event processes. When $\gamma_C = 0$, it is valid to use (5) with either $\widehat{w}_i(t)$ or $\widehat{w}_i^C(t)$, $i = 1,\ldots,n$. The results of Table 1 show that the use of

$\widehat{w}_i^C(t)$, $i = 1, \ldots, n$, in this situation leads to greater efficiency relative to the use of $\widehat{w}_i(t)$, $i = 1, \ldots, n$.

We also evaluated the IPSW method in our simulation studies. The implementation of this method requires that both (7) and (12) hold. This can be achieved by generating v from the positive stable distribution with Laplace transform $\exp(-v^\rho)$, $\rho \in (0, 1]$. The gap time between any two successive events and the survival time have a Kendall's tau correlation of $1 - \rho$.

Model (12) implies that $S_1(t|v) = S_0(t|v)^{\exp(\gamma_D)}$, where $S_k(t|v) = \exp\{-\int_0^t \lambda^D (u|v, Z = k)du\}$, $k = 0, 1$. If

$$dR(t|v, Z = 1) = vS_0(t|v)^{1-\exp(\gamma_D)}e^{\beta_0}dR_0(t), \qquad (15)$$

then the induced model for μ_Z is again in the form of (14).

In this set of simulation studies, we generated survival times from (12) with $\gamma_D = 0.2$ and $\lambda_0^D(t) = 0.25$, and conditional recurrence rate from (15) with $R_0(t) = t$ and $\beta_0 = 0.5$ or 0.2. We considered sample sizes $n = 50, 100, 200$ and correlations, in terms of Kendall's tau, of $0, 0.15, 0.30$ and 0.5. The censoring times were generated from an independent uniform $(0,5)$ variable, resulting in about two observed recurrences per subject. For each simulation setting, 1000 samples were obtained. The results are shown in Table 2.

Table 2. Simulation results for the IPSW method.

n	KT	$\gamma_D = 0.2, \beta_0 = 0.5$				$\gamma_D = \beta_0 = 0.2$			
		Bias	SE	SEE	CP	Bias	SE	SEE	CP
50	0	0.03	0.247	0.231	0.928	0.01	0.191	0.180	0.937
50	0.15	0.00	0.244	0.239	0.937	0.01	0.188	0.178	0.935
50	0.30	-0.01	0.249	0.242	0.929	0.01	0.193	0.184	0.933
50	0.5	0.00	0.353	0.339	0.933	-0.02	0.220	0.206	0.937
100	0	0.03	0.194	0.184	0.939	0.00	0.139	0.136	0.940
100	0.15	0.01	0.186	0.180	0.940	0.02	0.155	0.149	0.941
100	0.30	-0.01	0.194	0.188	0.940	0.01	0.163	0.159	0.942
100	0.5	0.01	0.255	0.245	0.940	-0.01	0.182	0.177	0.942
200	0	0.01	0.120	0.118	0.947	0.00	0.096	0.095	0.947
200	0.15	0.00	0.119	0.114	0.944	0.00	0.098	0.094	0.946
200	0.30	0.00	0.124	0.122	0.946	0.01	0.101	0.098	0.946
200	0.5	0.00	0.175	0.171	0.946	0.00	0.120	0.119	0.947

Note: KT represents Kendall's tau. Under $\gamma_D = 0.2$ and $\beta_0 = 0.5$, the method of §2.6 is used; under $\gamma_D = \beta_0 = 0.2$, the method of §2.8 is used. Bias is the mean of the estimator of β_0 minus β_0; SE is the standard error of the estimator of β_0; SEE is the mean of the standard error estimator; CP is the coverage probability of the Wald 95% confidence interval.

Based on these results, the IPSW method appears to perform similarly to the IPCW method with $\widehat{w}_i^C(t)$, $i = 1, \ldots, n$. There is a substantial decrease in the standard error if information on recurrences and deaths can be pooled.

Upon a referee's suggestion, we compared the methods proposed here with the two-sample procedures of Ghosh and Lin (2000). Specifically, we considered the Wald and score statistics from the IPCW procedure with $\widehat{w}_i(t)$, $i = 1, \ldots, n$, and the test statistic Q_{LR} from Ghosh and Lin (2000). Data were generated using $\gamma_D = 0$ and $\lambda_0^D(t) = 0.25$ in (12), and $R_0(t) = t$ in (13). We considered $\beta_0 = 0$ and 0.7. Censoring was generated using an independent uniform (0,5) random variable. This led to approximately 2.1 and 1.9 observed recurrences per subject under the two scenarios. We again took v to be a gamma variable with mean 1 and variance σ^2; we considered $\sigma^2 = 0$ and $\sigma^2 = 1$. Sample sizes $n = 50$ and $n = 100$ were examined. For each simulation setting, 1000 samples were obtained. The power results are given in Table 3. There appears to be good correspondence between the performance of the three statistics. In all settings examined, the concordance between the three statistics was greater than 95%.

As was mentioned in Section 2, the IPCW and IPSW methods of estimation require use of a truncation time τ. In the simulations, we set τ to be the last observed event time.

Table 3. Empirical powers of IPCW and Q_{LR}.

β_0	n	σ^2	IPCW		Q_{LR}
			Wald	Score	Q_{LR}
0	50	0	0.054	0.053	0.055
		1	0.051	0.056	0.055
	100	0	0.051	0.047	0.049
		1	0.050	0.050	0.048
0.7	50	0	0.721	0.701	0.732
		1	0.786	0.752	0.796
	100	0	0.956	0.942	0.951
		1	0.990	0.983	0.985

Note: Wald is the Wald statistic corresponding to the IPCW estimation method, while Score is the associated score test. Q_{LR} is the two-sample log-rank statistic from Ghosh and Lin (2000) with the usual log-rank weight function.

4. A Real Example

We now apply the methods developed in Section 2 to data from a cancer clinical trial conducted by the Veterans Administration Cooperative Urological

Research Group (Byar (1980)). These data have been analyzed extensively in the statistical literature. In this trial, 117 patients with stage I bladder cancer were randomized to placebo, pyridoxine or intravesical thiotepa and followed for recurrences of superficial bladder tumors. Following previous authors, we focus our attention on the comparison between thiotepa and placebo. In addition to the treatment assignment, two other covariates at baseline were measured: number of tumors and size of the largest tumor. Since all the covariates are time-invariant, models (1) and (2) are identical. Summary statistics for the two treatment arms of interest are given in Table 4.

Table 4. Recurrent and survival experiences for placebo and thiotepa groups in Stage I bladder cancer clinical trial.

| Treatment | Recurrences | | | | | | | Deaths |
	0	1	2	3	4	5	> 5	
Placebo	19	10	4	6	2	4	3	11
Thiotepa	20	8	3	2	2	2	1	12

As shown in this table, 23 of the 86 patients (26.7%) died during the study. In the analyses conducted by previous authors, death was treated as a censoring variable for cancer recurrences. Under the proportional rates model, this approach pertains to the cause-specific rate function, which is analogous to the cause-specific hazard function (Kalbfleisch and Prentice (1980, p.167)). If survival is independent of the recurrent events process, then the cause-specific rate function is the same as the marginal rate function for the recurrences. The results for this approach are provided in Table 5.

Table 5. Regression analysis for tumor recurrences.

| | | | | Mean function | | | | | |
| | Cause-specific rate | | | IPCW method | | | IPSW method | | |
Variable	Estimate	SE	P-value	Estimate	SE	P-value	Estimate	SE	P-value
Treatment	-0.540	0.270	0.046	-0.560	0.267	0.036	-0.556	0.288	0.053
Number	-0.199	0.043	0.002	-0.170	0.060	0.004	-0.266	0.070	<0.001
Size	0.041	0.065	0.600	0.003	0.044	0.969	0.071	0.090	0.432

Note: Treatment is coded as 1 (thiotepa) vs. 0 (placebo); Estimate denotes estimated regression parameter; SE represents estimated standard error; P-value represents the two-sided p-value for testing no covariate effect.

Table 5 also displays the results of the IPCW and IPSW methods under model (1). The three methods yield similar conclusions. As shown by Ghosh and Lin (2000), however, the use of the cause-specific rate method would yield

overestimation of the marginal mean function. Based on the proposed methods, thiotepa reduces the mean frequency of tumor recurrences by approximately 40% (after adjusting for number of tumors and size of the largest tumor), and this reduction is statistically significant at the 0.05 level (at least marginally).

As mentioned in Section 2, the IPCW and IPSW methods involve modeling the censoring and survival times with models (4) and (7), respectively. The results from fitting these two models are given in Table 6. None of the baseline covariates turn out to be significant predictors of survival or censoring, although there is slight evidence that number of tumors and size of the largest tumor might be predictive of survival. While thiotepa was seen to reduce tumor recurrences, it appears to be associated with increased mortality; however, the latter association is not significant.

Table 6. Proportional hazards regression for survival and censoring distributions.

Variable	Survival distribution			Censoring distribution		
	Estimate	SE	P-value	Estimate	SE	P-value
Treatment	0.290	0.425	0.495	0.002	0.264	0.994
Number	-0.149	0.107	0.165	0.077	0.099	0.437
Size	0.327	0.212	0.122	0.039	0.089	0.661

Note: See Note to Table 5.

To assess jointly the effects of treatment on recurrences and death, we employ the sequential test described in Section 2.8 with the IPSW method. Based on the results in Tables 4 and 5, the standardized estimates of treatment effects on recurrences and death are -1.93 and 0.68, respectively. The estimated correlation between the two estimators is -0.038. By numerical integration, $\Pr(\min(V_1, V_2) \leq -1.93) = 0.053$, and $\Pr(V_2 \leq 0.68) = 0.75$. Thus, there is evidence to suggest that thiotepa is effective in reducing recurrences but not in reducing mortality.

The application of the standard goodness-of-fit methods (e.g., Therneau, Grambsch and Fleming (1990)) did not reveal violation of the proportional hazards model for the survival or censoring distribution. Figure 1 displays the residual plots for checking the functional forms for number of tumors and size of largest tumor for model (1). The plots under both the IPCW and IPSW procedures are given. There is no clear systematic trend in any of the plots so that no transformations are needed. The plots of the Schoenfeld-type residuals based on the IPCW method are given in Figure 2; the plots based on the IPSW method are similar. These plots do not reveal systematic deviations, suggesting that the proportional means assumption is appropriate.
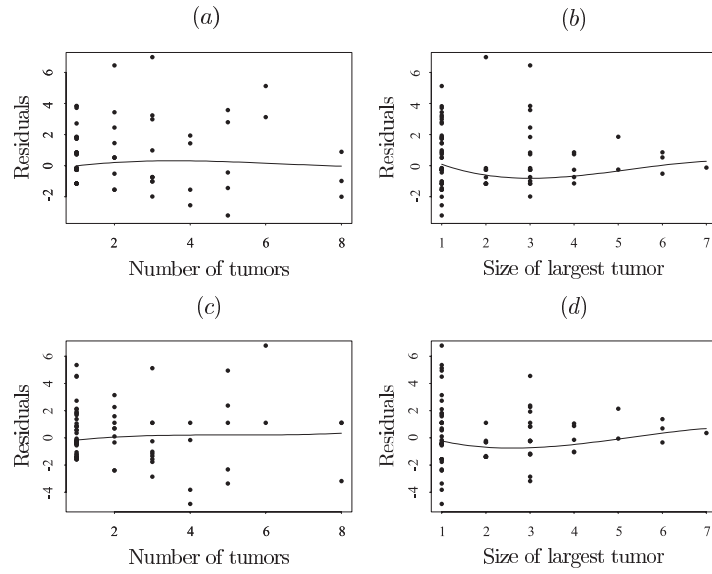
Figure 1. Plots of $\widehat{M}_i$'s for assessing the functional forms for number of tumors and size of largest tumor: figures (a) and (b) are based on IPCW method; (c) and (d) are based on the IPSW method; the solid line represents a locally weighted regression smooth with span of 0.5.
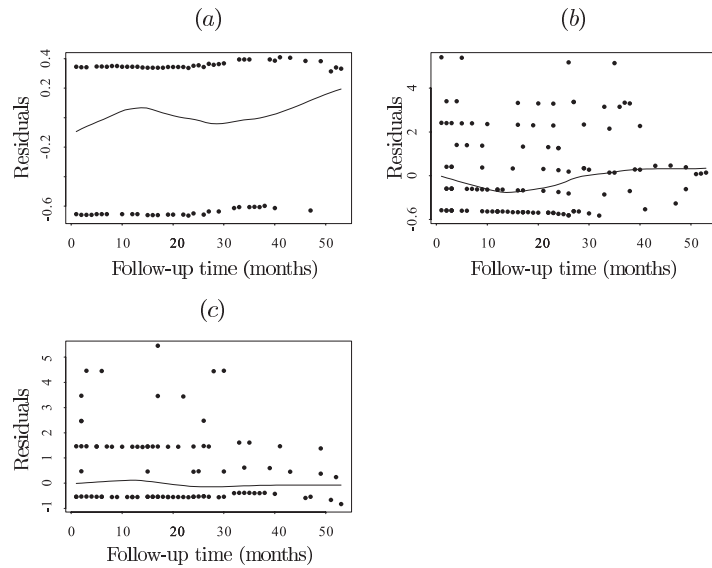

Figure 2. Plots of residuals $\Delta U_j^C(\widehat{\boldsymbol{\beta}}_C, \cdot)$ for assessing the proportional means assumption: the solid line represents a locally weighted regression smooth with span of 0.5; (a), (b) and (c) pertain to treatment, number of tumors and size of largest tumor, respectively.

The estimated regression model enables one to estimate/predict the cancer recurrences for subjects with certain covariate values. For example, Figure 3 displays the estimated mean function, along with the 95% confidence intervals, for thiotepa patients who have two tumors at baseline and whose largest tumors are 2 centimeters in diameter.
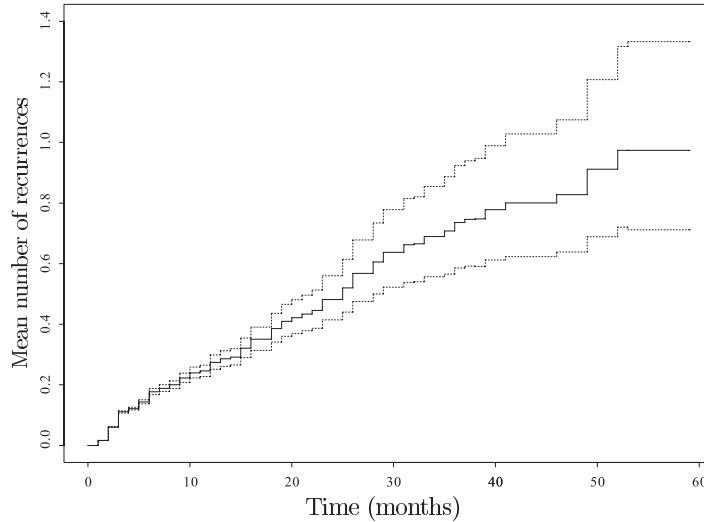

Figure 3. Estimated mean number of recurrences (solid line) and associated 95% confidence limits (dashed lines) for patients on thiotepa who have two tumors at baseline and whose largest tumors are 2 centimeters in diameter. The IPCW method is used. The figure was constructed in the manner described at the end of Section 2.6 using $\mathbf{z} = (1, 2, 2)$.

5. Discussion

The marginal mean function of recurrent events studied in this paper is analogous to the cumulative incidence function (Kalbfleisch and Prentice (1980, p.169); Pepe and Mori (1993); Fine and Gray (1999)) in the competing risks literature. This quantity is of clinical interest because it pertains to the frequency of recurrent events the subject actually experiences in the presence of death. As mentioned in Section 2.8, this quantity is affected by the survival distribution: the number of recurrences tends to be higher if the subject lives longer. Thus, the effects of covariates on death and disease recurrences should be examined simultaneously. If a new treatment reduces both disease recurrences and death or, as in the bladder tumor study, reduces disease recurrences but has no appreciable impact on survival, then the treatment is clearly preferred. If the treatment reduces disease recurrences but increases mortality, then it is more delicate to make a judgment on the treatment.

Two estimation procedures, IPSW and IPCW, have been proposed in the paper. The former seems more natural than the latter when survival is also of interest. However, if the marginal mean function of recurrent events is the primary interest and censoring is independent of covariates, then it is more attractive to use the IPCW procedure with a nonparametric estimator of the censoring distribution.

The estimation of models (1) and (2) requires modeling either the survival or censoring distributions. This is not very appealing as such models may be misspecified, but seems unavoidable. Models (1) and (2) may also be misspecified. It would be worthwhile to investigate the potential bias due to misspecification for each of these models.

The approach we have taken here is to formulate regression models in order to provide direct summary measures for covariate effects. However, if we were interested in prediction of the marginal mean function, then a more flexible approach would be to model $S(t|\mathbf{Z})$ and $R(t|\mathbf{Z})$ separately from (11) . Such an approach was taken in the competing risks setting by Cheng, Fine and Wei (1998).

The proposed estimators, although simple and intuitive, are not semiparametrically efficient. While it might be possible to develop an efficient estimation method based on nonparametric maximum likelihood, such a procedure is likely to be much more computationally intensive. An alternative approach is to apply results from locally efficient estimation theory (van der Laan, Robins and Gill (2000)). Further investigations are warranted.

Models (1) and (2) specify multiplicative covariate effects on the marginal mean/rate function of recurrent events. Another approach is to specify multiplicative covariate effects on the recurrent event times, which correspond to the accelerated time model: $\mu_{\mathbf{Z}}(t) = \mu_0(e^{\boldsymbol{\beta}_0^T \mathbf{z}}t)$. Inference procedures for this model can be developed by combining the ideas of this paper with those of Lin, Wei and Ying (1998).

Acknowledgements

Appendix: Proofs of Theorems

In this appendix, we outline the main steps in proving the theorems stated in Section 2. The interested readers are referred to Ghosh (2000) for further detail.

Proof of Theorem 1. Consider

$$X^C(\boldsymbol{\beta}) = n^{-1} \sum_{i=1}^{n} \int_0^\tau \left[(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^T \mathbf{Z}_i(t) - \log \left\{ \frac{\tilde{S}^{(0)}(\boldsymbol{\beta}, t)}{\tilde{S}^{(0)}(\boldsymbol{\beta}_0, t)} \right\} \right] \widehat{w}_i^C(t) dN_i(t).$$

The Strong Law of Large Numbers, together with the consistency of $\widehat{\boldsymbol{\gamma}}_C$ and $\widehat{\Lambda}_0^C(\cdot)$, implies that $X^C(\boldsymbol{\beta})$ converges almost surely to

$$E\left[\int_0^\tau (\boldsymbol{\beta} - \boldsymbol{\beta}_0)^T \mathbf{Z}(t) G(t|\mathbf{Z}) dN^*(t) - \int_0^\tau \log\{s^{(0)}(\boldsymbol{\beta}, t)/s^{(0)}(\boldsymbol{\beta}_0, t)\} G(t|\mathbf{Z}) dN^*(t)\right]$$

for every $\boldsymbol{\beta}$, where $\mathbf{s}^{(k)}(\boldsymbol{\beta}, t) = E\{G(t|\mathbf{Z})\mathbf{Z}(t)^{\otimes k} e^{\boldsymbol{\beta}^T \mathbf{Z}(t)}\}$, $k = 0, 1, 2$. The consistency of $\widehat{\boldsymbol{\beta}}_C$ now follows from the arguments in Appendix A.1 of Lin et al. (2000).

The Taylor series expansion, together with the Law of Large Numbers and the consistency of $\widehat{\boldsymbol{\beta}}_C$, yields

$$n^{1/2}(\widehat{\boldsymbol{\beta}}_C - \boldsymbol{\beta}_0) = \mathbf{A}^{-1} n^{-1/2} \mathbf{U}^C(\boldsymbol{\beta}_0) + o_P(1). \tag{16}$$

It remains to determine the asymptotic distribution of $n^{-1/2}\mathbf{U}^C(\boldsymbol{\beta}_0)$. Clearly,

$$n^{-1/2}\mathbf{U}^C(\boldsymbol{\beta}_0) = n^{-1/2} \sum_{i=1}^n \int_0^\tau \{\mathbf{Z}_i(t) - \overline{\mathbf{Z}}^C(\boldsymbol{\beta}_0, t)\} dM_i(t)$$

$$+ n^{-1/2} \sum_{i=1}^n \int_0^\tau \{\mathbf{Z}_i(t) - \overline{\mathbf{Z}}^C(\boldsymbol{\beta}_0, t)\} \left\{\frac{\widehat{G}(t|\mathbf{Z}_i)}{\widehat{G}(X_i \wedge t|\mathbf{Z}_i)} - \frac{G(t|\mathbf{Z}_i)}{G(X_i \wedge t|\mathbf{Z}_i)}\right\}$$

$$\times I(C_i \geq D_i \wedge t)\{dN_i(t) - e^{\boldsymbol{\beta}_0^T \mathbf{Z}_i(t)} d\mu_0(t)\}. \tag{17}$$

It follows from the functional delta method (Andersen, Borgan, Gill and Keiding (1993, p.111)), the $n^{1/2}$-consistency of $\widehat{G}$, formula (2.1) of Lin et al. (1994) and the Martingale Central Limit Theorem (Fleming and Harrington (1991, p.227)) that

$$n^{1/2}\left\{\frac{\widehat{G}(t|\mathbf{Z}_i)}{\widehat{G}(X_i \wedge t|\mathbf{Z}_i)} - \frac{G(t|\mathbf{Z}_i)}{G(X_i \wedge t|\mathbf{Z}_i)}\right\}$$

$$= -\frac{I(X_i < t)G(t|\mathbf{Z}_i)}{G(X_i \wedge t|\mathbf{Z}_i)}\left[n^{-1/2}\sum_{j=1}^n \int_{X_i}^t \frac{e^{\boldsymbol{\gamma}_C^T \mathbf{Z}_i(u)} dM_j^C(u)}{r^{(0)}(\boldsymbol{\gamma}_C, u)}\right.$$

$$+ \mathbf{g}^C(X_i, t; \mathbf{Z}_i)^T \boldsymbol{\Omega}_C^{-1} n^{-1/2}\sum_{j=1}^n \int_0^\tau \left\{\mathbf{Z}_j(u) - \frac{\mathbf{r}^{(1)}(\boldsymbol{\gamma}_C, u)}{r^{(0)}(\boldsymbol{\gamma}_C, u)}\right\} dM_j^C(u)\right] + o_P(1), \tag{18}$$

where $\mathbf{g}^C(X_i, t; \mathbf{Z}_i) = \int_{X_i}^t e^{\boldsymbol{\gamma}_C^T \mathbf{Z}_i(u)} \{\mathbf{Z}_i(u) - \mathbf{r}^{(1)}(\boldsymbol{\gamma}_C, u) / r^{(0)}(\boldsymbol{\gamma}_C, u)\} d\Lambda_0^C(u)$, $\mathbf{r}^{(k)}(\boldsymbol{\gamma}_C, t)$ is the limit of $\widehat{\mathbf{R}}^{(k)}(\boldsymbol{\gamma}_C, t)$, and $\boldsymbol{\Omega}_C$ the limit of $\widehat{\boldsymbol{\Omega}}_C$. Plugging (18) into (17) and interchanging integrals, we get

$$n^{-1/2}\mathbf{U}^C(\boldsymbol{\beta}_0) = n^{-1/2} \sum_{i=1}^n \int_0^\tau \{\mathbf{Z}_i(t) - \overline{\mathbf{Z}}^C(\boldsymbol{\beta}_0, t)\} dM_i(t)$$

$$+n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\tilde{\mathbf{B}}_C\left\{\mathbf{Z}_i(t)-\frac{\mathbf{r}^{(1)}(\boldsymbol{\gamma}_C,t)}{r^{(0)}(\boldsymbol{\gamma}_C,t)}\right\}dM_i^C(t)$$

$$+n^{-1/2}\sum_{i=1}^{n}\int_0^{\tau}\frac{\tilde{\mathbf{q}}^C(t)}{r^{(0)}(\boldsymbol{\gamma}_C,t)}dM_i^C(t)+o_P(1),$$

where $\tilde{\mathbf{B}}_C=-n^{-1}\sum_{i=1}^{n}\int_0^{\tau}\{\mathbf{Z}_i(t)-\overline{\mathbf{Z}}^C(\boldsymbol{\beta}_0,t)\}\mathbf{g}^C(X_i,t;\mathbf{Z}_i)^T\boldsymbol{\Omega}_C^{-1}I(t>X_i)dM_i(t)$, and $\tilde{\mathbf{q}}^C(t)=-n^{-1}\sum_{i=1}^{n}\int_0^{\tau}\{\mathbf{Z}_i(u)-\overline{\mathbf{Z}}^C(\boldsymbol{\beta}_0,u)\}e^{\boldsymbol{\gamma}_C^T\mathbf{Z}_i(t)}I(u\geq t>X_i)dM_i(u)$. By the Martingale Central Limit Theorem, $\tilde{\mathbf{B}}_C$ and $\tilde{\mathbf{q}}^C$ may be replaced by their limits, $\mathbf{B}_C$ and $\mathbf{q}^C(t)$ say, without altering the asymptotic distributions of the last two terms on the right-hand side of (19). In addition, using arguments from empirical process theory as given in Appendix A.2 of Lin et al. (2000), we can replace $\overline{\mathbf{Z}}^C(\boldsymbol{\beta}_0,t)\equiv\tilde{\mathbf{S}}^{(1)}(\boldsymbol{\beta}_0,t)/\tilde{S}^{(0)}(\boldsymbol{\beta}_0,t)$ in the first integral of (19) with its limit $\overline{\mathbf{z}}(\boldsymbol{\beta}_0,t)\equiv\mathbf{s}^{(1)}(\boldsymbol{\beta}_0,t)/s^{(0)}(\boldsymbol{\beta}_0,t)$. Thus, we have

$$n^{-1/2}\mathbf{U}^C(\boldsymbol{\beta}_0)=n^{-1/2}\sum_{i=1}^{n}(\boldsymbol{\eta}_i^C+\boldsymbol{\psi}_i^C)+o_P(1), \tag{20}$$

where $\boldsymbol{\eta}_i^C=\int_0^{\tau}\{\mathbf{Z}_i(t)-\overline{\mathbf{z}}(\boldsymbol{\beta}_0,t)\}dM_i(t)$, and

$$\boldsymbol{\psi}_i^C=\int_0^{\tau}\mathbf{B}_C\left\{\mathbf{Z}_i(t)-\frac{\mathbf{r}^{(1)}(\boldsymbol{\gamma}_C,t)}{r^{(0)}(\boldsymbol{\gamma}_C,t)}\right\}dM_i^C(t)+\int_0^{\tau}\frac{\mathbf{q}^C(t)}{r^{(0)}(\boldsymbol{\gamma}_C,t)}dM_i^C(t).$$

The right-hand side of (20) is a sum of n i.i.d. terms, so the Multivariate Central Limit Theorem implies that $n^{-1/2}\mathbf{U}^C(\boldsymbol{\beta}_0)\to_d N(\mathbf{0},\boldsymbol{\Sigma}_C)$, where $\boldsymbol{\Sigma}_C=E\{(\boldsymbol{\eta}_1^C+\boldsymbol{\psi}_1^C)^{\otimes2}\}$. Combing this result with (16), we have $n^{1/2}(\widehat{\boldsymbol{\beta}}_C-\boldsymbol{\beta}_0)\to_d N(\mathbf{0},\mathbf{A}^{-1}\boldsymbol{\Sigma}_C\mathbf{A}^{-1})$.

By replacing all the unknown quantities in $\mathbf{A}$ and $\boldsymbol{\Sigma}_C$ with their empirical counterparts, we obtain the covariance matrix estimator given in the statement of Theorem 1. By extending the arguments in Appendix A.3 of Lin et al. (2000), we can show that $\widehat{\mu}_0^C(t)$ is strongly consistent for $\mu_0(t)$. The consistency of $\widehat{\boldsymbol{\Sigma}}_C$ for $\boldsymbol{\Sigma}_C$ then follows from the strong consistency of $\widehat{\boldsymbol{\beta}}_C$, $\widehat{\mu}_0^C(t)$, $\widehat{\boldsymbol{\gamma}}_C$, $\widehat{\Lambda}_0^C(t)$ and repeated applications of the Uniform Strong Law of Large Numbers (Pollard (1990, p.41)).

Proof of Theorem 2. Algebraic manipulations yield

$$n^{1/2}\{\widehat{\mu}_0^C(t)-\mu_0(t)\}=n^{-1/2}\sum_{i=1}^{n}\int_0^t\frac{dM_i(u)}{\tilde{S}^{(0)}(\boldsymbol{\beta}_0,u)}$$

$$+n^{-1/2}\sum_{i=1}^{n}\int_0^t\frac{\widehat{w}_i^C(u)-w_i^C(u)}{\tilde{S}^{(0)}(\boldsymbol{\beta}_0,u)}\{dN_i(u)-e^{\boldsymbol{\beta}_0^T\mathbf{z}_i(u)}d\mu_0(u)\}$$

$$+n^{-1/2}\left\{\sum_{i=1}^{n}\int_0^t\frac{\widehat{w}_i^C(u)dN_i(u)}{\tilde{S}^{(0)}(\widehat{\boldsymbol{\beta}}_C,u)}-\sum_{i=1}^{n}\int_0^t\frac{\widehat{w}_i^C(u)dN_i(u)}{\tilde{S}^{(0)}(\boldsymbol{\beta}_0,u)}\right\}. \tag{21}$$

By the Taylor series expansion, along with the Uniform Strong Law of Large Numbers and the strong consistency of $\widehat{\boldsymbol{\beta}}_C$, $\widehat{\mu}_0^C(t)$, $\widehat{\boldsymbol{\gamma}}_C$ and $\widehat{\Lambda}_0^C(t)$, the third term on the right-hand side of (21) is asymptotically equal to $\mathbf{h}^T(\boldsymbol{\beta}_0,t)n^{1/2}(\widehat{\boldsymbol{\beta}}_C-\boldsymbol{\beta}_0)$, where $\mathbf{h}(\boldsymbol{\beta}_0,t) = -\int_0^t \overline{\mathbf{z}}(\boldsymbol{\beta}_0,u)d\mu_0(u)$. Thus,

$$
\begin{aligned}
n^{1/2}\{\widehat{\mu}_0^C(t)-\mu_0(t)\}={}&n^{-1/2}\sum_{i=1}^n\int_0^t\frac{dM_i(u)}{\tilde{S}^{(0)}(\boldsymbol{\beta}_0,u)}\\
&+n^{-1/2}\sum_{i=1}^n\int_0^t\frac{I(C_i\geq D_i\wedge u)}{\tilde{S}^{(0)}(\boldsymbol{\beta}_0,u)}\left\{\frac{\widehat{G}(u|\mathbf{Z}_i)}{\widehat{G}(X_i\wedge u|\mathbf{Z}_i)}-\frac{G(u|\mathbf{Z}_i)}{G(X_i\wedge u|\mathbf{Z}_i)}\right\}\\
&\times\{dN_i(u)-e^{\boldsymbol{\beta}_0^T\mathbf{Z}_i(u)}d\mu_0(u)\}\\
&+\mathbf{h}^T(\boldsymbol{\beta}_0,t)\mathbf{A}^{-1}n^{-1/2}\sum_{j=1}^n(\boldsymbol{\eta}_j^C+\boldsymbol{\psi}_j^C)+o_P(1).
\end{aligned}
$$

By manipulations similar to those in the proof of Theorem 1,

$$
\begin{aligned}
n^{1/2}\{\widehat{\mu}_0^C(t)-\mu_0(t)\} = {}&n^{-1/2}\sum_{i=1}^n\left[\int_0^t\frac{dM_i(u)}{\tilde{S}^{(0)}(\boldsymbol{\beta}_0,u)}+\int_0^\tau\frac{\tilde{p}_1^C(u,t)}{r^{(0)}(\boldsymbol{\gamma}_C,u)}dM_i^C(u)\right.\\
&\left.+\int_0^\tau\tilde{\mathbf{p}}_2^C(t)^T\left\{\mathbf{Z}_i(u)-\frac{\mathbf{r}^{(1)}(\boldsymbol{\gamma}_C,u)}{r^{(0)}(\boldsymbol{\gamma}_C,u)}\right\}dM_i^C(u)\right]\\
&+\mathbf{h}^T(\boldsymbol{\beta}_0,t)\mathbf{A}^{-1}n^{-1/2}\sum_{j=1}^n(\boldsymbol{\eta}_j^C+\boldsymbol{\psi}_j^C)+o_P(1), \qquad (22)
\end{aligned}
$$

where

$$
\tilde{p}_1^C(u,t) = -n^{-1}\sum_{i=1}^n\int_0^t\frac{I(s\geq u>X_i)e^{\boldsymbol{\gamma}_C^T\mathbf{Z}_i(u)}}{\tilde{S}^{(0)}(\boldsymbol{\beta}_0,s)}dM_i(s),
$$

$$
\tilde{\mathbf{p}}_2^C(t) = -n^{-1}\sum_{i=1}^n\int_0^t\frac{I(u>X_i)\mathbf{g}^C(X_i,u;\mathbf{Z}_i)^T\boldsymbol{\Omega}_C^{-1}}{\tilde{S}^{(0)}(\boldsymbol{\beta}_0,u)}dM_i(u).
$$

Using the fact that $M_i^C(t)$, $i=1,\ldots,n$, are martingales and the empirical process arguments in Lin et al. (2000), we can replace $\tilde{p}_1^C(u,t)$, $\tilde{\mathbf{p}}_2^C(t)$ and $\tilde{S}^{(0)}(\boldsymbol{\beta}_0,u)$ in (22) by their limits. Hence, $n^{1/2}\{\widehat{\mu}_0^C(t)-\mu_0(t)\} = n^{-1/2}\sum_{i=1}^n\phi_i^C(t)+o_P(1)$, where

$$
\begin{aligned}
\phi_i^C(t)={}&\int_0^t\frac{dM_i(u)}{s^{(0)}(\boldsymbol{\beta}_0,u)}+\int_0^\tau\frac{p_1^C(u,t)}{r^{(0)}(\boldsymbol{\gamma}_C,u)}dM_i^C(u)\\
&+\int_0^\tau\mathbf{p}_2^C(t)^T\left\{\mathbf{Z}_i(u)-\frac{\mathbf{r}^{(1)}(\boldsymbol{\gamma}_C,u)}{r^{(0)}(\boldsymbol{\gamma}_C,u)}\right\}dM_i^C(u)+\mathbf{h}^T(\boldsymbol{\beta}_0,t)\mathbf{A}^{-1}n^{-1/2}\sum_{j=1}^n(\boldsymbol{\eta}_j^C+\boldsymbol{\psi}_j^C),
\end{aligned}
$$

and $p_1^C(u,t)$ and $\mathbf{p}_2^C(t)$ are the limits of $\tilde{p}_1^C(u,t)$ and $\tilde{\mathbf{p}}_2^C(t)$. By the Multivariate Central Limit Theorem, the finite-dimensional distributions of $n^{1/2}\{\widehat{\mu}_0^C(t) - \mu_0(t)\}$ are asymptotically normal. Because $\phi_i^C(t)$ consists of monotone functions, $n^{-1/2}\sum_{i=1}^n \phi_i^C(t)$ is tight (Van der Vaart and Wellner (1996), p.215). Thus, the desired weak convergence is obtained. The consistency of the covariance function estimator follows from the consistency of $\widehat{\boldsymbol{\beta}}_C$, $\widehat{\mu}_0^C(t)$, $\widehat{\boldsymbol{\gamma}}_C$ and $\widehat{\Lambda}_0^C(t)$, and repeated applications of the Uniform Strong Law of Large Numbers.

Proof of Theorem 3. We derive the asymptotic distribution of $n^{-1/2}\mathbf{U}^D(\boldsymbol{\beta}_0)$; the rest of the proof is similar to that of Theorem 1. Clearly,

$$
n^{-1/2}\mathbf{U}^D(\boldsymbol{\beta}_0) = n^{-1/2}\sum_{i=1}^n \int_0^\tau \{\mathbf{Z}_i(t) - \overline{\mathbf{Z}}^D(\boldsymbol{\beta}_0,t)\}dM_i^\dagger(t)
$$
$$
+n^{-1/2}\sum_{i=1}^n \int_0^\tau \{\mathbf{Z}_i(t) - \overline{\mathbf{Z}}^D(\boldsymbol{\beta}_0,t)\}\left\{\frac{1}{\widehat{S}(t|\mathbf{Z}_i)} - \frac{1}{S(t|\mathbf{Z}_i)}\right\}
$$
$$
\times I(X_i \geq t)\{dN_i(t) - e^{\boldsymbol{\beta}_0^T\mathbf{Z}_i(t)}d\mu_0(t)\}. \tag{23}
$$

We can derive a representation for $n^{1/2}\{\widehat{S}^{-1}(t|\mathbf{Z}_i) - S^{-1}(t|\mathbf{Z}_i)\}$ similar to (18). Plugging that representation into (23) and interchanging integrals, we obtain

$$
n^{-1/2}\mathbf{U}^D(\boldsymbol{\beta}_0) = n^{-1/2}\sum_{i=1}^n \int_0^\tau \{\mathbf{Z}_i(t) - \overline{\mathbf{Z}}^D(\boldsymbol{\beta}_0,t)\}dM_i^\dagger(t)
$$
$$
+n^{-1/2}\sum_{i=1}^n \int_0^\tau \tilde{\mathbf{B}}_D\left\{\mathbf{Z}_i(t) - \frac{\mathbf{r}^{(1)}(\boldsymbol{\gamma}_D,t)}{r^{(0)}(\boldsymbol{\gamma}_D,t)}\right\}dM_i^D(t)
$$
$$
+n^{-1/2}\sum_{i=1}^n \int_0^\tau \frac{\tilde{\mathbf{q}}^D(t)}{\widehat{R}^{(0)}(\boldsymbol{\gamma}_D,t)}dM_i^D(t) + o_P(1),
$$

where $\tilde{\mathbf{B}}_D = n^{-1}\sum_{i=1}^n \int_0^\tau \{\mathbf{Z}_i(t) - \overline{\mathbf{Z}}^D(\boldsymbol{\beta}_0,t)\}\mathbf{g}^D(t;\mathbf{Z}_i)^T\boldsymbol{\Omega}_D^{-1}dM_i^\dagger(t)$,

$$
\mathbf{g}^D(t;\mathbf{Z}_i) = \int_0^t e^{\boldsymbol{\gamma}_D^T\mathbf{Z}_i(u)}\left\{\mathbf{Z}_i(u) - \frac{\mathbf{r}^{(1)}(\boldsymbol{\gamma}_D,u)}{r^{(0)}(\boldsymbol{\gamma}_D,u)}\right\}d\Lambda_0^D(u),
$$

$\tilde{\mathbf{q}}^D(t) = n^{-1}\sum_{i=1}^n \int_0^\tau \{\mathbf{Z}_i(u) - \overline{\mathbf{Z}}^D(\boldsymbol{\beta}_0,u)\}e^{\boldsymbol{\gamma}_D^T\mathbf{Z}_i(t)}I(u \geq t)dM_i^\dagger(u)$, and $\boldsymbol{\Omega}_D$ is the limit of $\widehat{\boldsymbol{\Omega}}_D$.

By the arguments leading to (20), we have $n^{-1/2}\mathbf{U}^D(\boldsymbol{\beta}_0) = n^{-1/2}\sum_{i=1}^n(\boldsymbol{\eta}_i^D + \boldsymbol{\psi}_i^D) + o_P(1)$, where $\boldsymbol{\eta}_i^D = \int_0^\tau \{\mathbf{Z}_i(t) - \overline{\mathbf{z}}(\boldsymbol{\beta}_0,t)\}dM_i^\dagger(t)$,

$$
\boldsymbol{\psi}_i^D = \int_0^\tau \left[\mathbf{B}_D\left\{\mathbf{Z}_i(t) - \frac{\mathbf{r}^{(1)}(\boldsymbol{\gamma}_D,t)}{r^{(0)}(\boldsymbol{\gamma}_D,t)}\right\} + \frac{\mathbf{q}^D(t)}{r^{(0)}(\boldsymbol{\gamma}_D,t)}\right]dM_i^D(t),
$$

$\mathbf{q}^D(t) = \lim_{n\to\infty} \tilde{\mathbf{q}}^D(t)$ and $\mathbf{B}_D = \lim_{n\to\infty} \tilde{\mathbf{B}}_D$. It then follows from the Multivariate Central Limit Theorem that $n^{-1/2}\mathbf{U}^D(\boldsymbol{\beta}_0) \to_d N(\mathbf{0}, \boldsymbol{\Sigma}_D)$, where $\boldsymbol{\Sigma}_D = \mathrm{E}\{(\boldsymbol{\eta}_1^D + \boldsymbol{\psi}_1^D)^{\otimes 2}\}$.

Proof of Theorem 4. Using the ideas in the proofs of Theorems 2 and 3, we can show that the process $n^{1/2}\{\hat{\mu}_0^D(t) - \mu_0(t)\}$, $0 \leq t \leq \tau$, converges weakly to a mean-zero Gaussian process with covariance function $\xi_D(s,t) \equiv E\{\phi_1^D(s)\phi_1^D(t)\}$, where

$$
\begin{aligned}
\phi_i^D(t) = & \int_0^t \frac{dM_i^\dagger(u)}{s^{(0)}(\boldsymbol{\beta}_0, u)} + \int_0^\tau \frac{p_1^D(u,t)}{r^{(0)}(\boldsymbol{\gamma}_D, u)} dM_i^D(u) \\
& + \int_0^\tau \mathbf{p}_2^D(t)^T \left\{ \mathbf{Z}_i(u) - \frac{\mathbf{r}^{(1)}(\boldsymbol{\gamma}_D, u)}{r^{(0)}(\boldsymbol{\gamma}_D, u)} \right\} dM_i^D(u) + \mathbf{h}^T(\boldsymbol{\beta}_0, t)\mathbf{A}^{-1} n^{-1/2} \sum_{j=1}^n (\boldsymbol{\eta}_j^D + \boldsymbol{\psi}_j^D),
\end{aligned}
$$

$$
p_1^D(u,t) = \lim_{n\to\infty} n^{-1} \sum_{i=1}^n \int_0^t \frac{I(s \geq u)e^{\boldsymbol{\gamma}_D^T \mathbf{Z}_i(u)}}{\widehat{S}^{(0)}(\boldsymbol{\beta}_0, s)} dM_i^\dagger(s),
$$

$$
\mathbf{p}_2^D(t) = \lim_{n\to\infty} n^{-1} \sum_{i=1}^n \int_0^t \frac{\mathbf{g}^D(u; \mathbf{Z}_i)^T \boldsymbol{\Omega}_D^{-1}}{\widehat{S}^{(0)}(\boldsymbol{\beta}_0, u)} dM_i^\dagger(u).
$$

The consistency of the covariance function estimator follows from the consistency of $\widehat{\boldsymbol{\beta}}_D$, $\hat{\mu}_0^D(t)$, $\widehat{\boldsymbol{\gamma}}_D$ and $\widehat{\Lambda}_0^D(t)$, along with repeated applications of the Uniform Strong Law of Large Numbers.

References

Albert, P. S. (1991). A two-state Markov mixture model for a time series of epileptic seizure counts. *Biometrics* **47**, 1371-1381.

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

Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study. *Ann. Statist.* **10**, 1100-1120.

Byar, D. P. (1980). The Veterans Administration study of chemoprophylaxis for recurrent stage I bladder tumors: comparisons of placebo, pyridoxine and topical thiotepa. In *Bladder Tumors and Other Topics in Urological Oncology*, (Edited by M. Pavone-Macaluso, P. H. Smith and F. Edsmyr), Plenum, 363-370, New Year.

Breslow, N. E. (1974). Covariance analysis of censored survival data. *Biometrics* **30**, 89-99.

Cheng, S. C., Fine, J. P. and Wei, L. J. (1998). Prediction of cumulative incidence function under the proportional hazards model. *Biometrics* **54**, 219-228.

Cheng, S. C., Wei, L. J. and Ying, Z. (1995). Analysis of transformation models with censored data. *Biometrika* **82**, 835-845.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *J. Amer. Statist. Assoc.* **74**, 829-836.

Cook, R. J. and Lawless, J. F. (1979). Marginal analysis of recurrent events and a terminating event. *Statist. in Medicine* **16** 911-924.

Cox, D. R. (1972). Regression models and life tables (with discussion). *J. Roy. Statist. Soc. Ser. B* **34**, 187-220.

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

Fine, J. P. and Gray, R. J. (1999). A Proportional hazards model for the subdistribution of a competing risk. *J. Amer. Statist. Assoc.* **94**, 496-509.

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

Ghosh, D. (2000). Nanparanmetric and semiparametric analysi of recurrent events in the presence of terminal events and dependent gensoring. Ph. D. dissertation, Department of Biostatistics, University of Washington.

Ghosh, D. and Lin, D. Y. (2000). Nonparametric analysis of recurrent events and death. *Biometrics* **56**, 554-562.

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

Lawless, J. F., Nadeau, C. and Cook, R. J. (1997). Analysis of mean and rate functions for recurrent events. In *Proceedings of the First Seattle Survival Analysis Symposium*, (Edited by D. Y. Lin and T. R. Fleming), 137-149. Springer-Verlag, New Yark.

Li, Q. and Lagakos, S. (1997). Use of the Wei-Lin-Weissfeld method for the analysis of a recurring and a terminating event. *Statist. in Medicine* **16**, 925-940.

Lin, D. Y., Fleming, T. R. and Wei, L. J. (1994). Confidence bands for survival curves under the proportional hazards model. *Biometrika* **81**, 73-81.

Lin, D. Y., Wei, L. J., Yang, I. and Ying, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *J. Roy. Statist. Soc. Ser. B* **62**, 711-730.

Lin, D. Y., Wei, L. J. and Ying, Z. (1993). Checking the Cox model with cumulative sums of martingale-based residuals. *Biometrika* **80**, 557-572.

Lin, D. Y., Wei. L. J. and Ying, Z. (1998). Accelerated failure time models for counting processes. *Biometrika* **85**, 605-618.

Lin, D. Y. and Ying, Z. (1993). A simple nonparametric estimator of the bivariate survival function under univariate censoring. *Biometrika* **80**, 573-581.

Pepe, M. S. and Cai, J. (1993). Some graphical displays and marginal regression analyses for recurrent failure times and time dependent covariates. *J. Amer. Statist. Assoc.* **88**, 811-820.

Pepe, M. S. and Mori, M. (1993). Kaplan-Meier, marginal or conditional probability curves in summarizing competing risks failure time data ? *Statist. in Medicine* **12**, 737-751.

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

Prentice R. L., Williams B. J. and Peterson A. V. (1981). On the regression analysis of multivariate failure time data. *Biometrika* **68**, 373-379.

Robins, J. and Rotinizky, A. (1992). Recovery of information and adjustment for dependent censoring using surrogate markers. In *AIDS Epidemiology-Methodological Issues* (Edited by N. Jewell, K. Dietz and V. Farewell), 297-331. Birkhauser, Bosion.

Schoenfeld, D. (1982). Partial residuals for the proportional hazards regression model. *Biometrika* **69**, 239-241.

Strawderman, R. (2000). Estimating the mean of an increasing stochastic process at a censored stopping time. *J. Amer. Statist. Assoc.* **95**, 1192-1208.

Therneau, T. M., Grambsch, P. M. and Fleming, T. R. (1990). Martingale-based residuals for survival models. *Biometrika* **77**, 147-160.

van der Laan, M., Robins, J. M. and Gill, R. D. (2000). Locally efficient estimation in censored data models: theory and examples. Technical report, University of California, Department of Biostatistics.

van der Vaart, A. W. and Wellner, J. A. (1993). *Weak Convergence and Empirical Processes.* Springer, New York.

Wei, L. J. and Johnson, W. E. (1985). Combining dependent tests with incomplete repeated measurements. *Biometrika* **72**, 359-364.

Wei, L. J., Lin D. Y. and Weissfeld L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *J. Amer. Statist. Assoc.* **84**, 1065-1073.

Department of Biostatistics, University of Michigan, 1420 Washington Heights, Ann Arbor, MI, 48109-2029, U.S.A.

E-mail: ghoshd@umich.edu

Department of Biostatistics, University of North Carolina, 3101E McGavran-Greenberg Hall, CB#7420, Chapel Hill, North Carolina 27599-7420, U.S.A.

E-mail: lin@bios.unc.edu