

Semiparametric regression for the mean and rate functions of recurrent events

D. Y. Lin,

University of Washington, Seattle, USA

L. J. Wei,

Harvard University, Boston, USA

I. Yang

Schering-Plough Research Institute, Kenilworth, USA

and Z. Ying

Rutgers University, Piscataway, USA

[Received April 1999. Revised April 2000]

Summary. The counting process with the Cox-type intensity function has been commonly used to analyse recurrent event data. This model essentially assumes that the underlying counting process is a time-transformed Poisson process and that the covariates have multiplicative effects on the mean and rate functions of the counting process. Recently, Pepe and Cai, and Lawless and co-workers have proposed semiparametric procedures for making inferences about the mean and rate functions of the counting process without the Poisson-type assumption. In this paper, we provide a rigorous justification of such robust procedures through modern empirical process theory. Furthermore, we present an approach to constructing simultaneous confidence bands for the mean function and describe a class of graphical and numerical techniques for checking the adequacy of the fitted mean and rate models. The advantages of the robust procedures are demonstrated through simulation studies. An illustration with multiple-infection data taken from a clinical study on chronic granulomatous disease is also provided.

Keywords: Counting process; Empirical process; Intensity function; Martingale; Partial likelihood; Poisson process

1. Introduction

Andersen and Gill (1982) introduced a counting process model with the Cox (1972) type of intensity function for recurrent events. Specifically, let $N^*(t)$ be the number of events that occur over the interval $[0, t]$ and $Z(\cdot)$ be a p -dimensional covariate process. Also, let \mathcal{F}_t be the σ -field generated by $\{N^*(s), Z(s): 0 \leq s \leq t\}$ and $\lambda_Z(t)$ be the intensity function of $N^*(t)$ associated with \mathcal{F}_t , i.e.

$$E\{dN^*(t)|\mathcal{F}_{t-}\} = \lambda_Z(t) dt,$$

where $dN^*(t)$ is the increment $N^*\{(t + dt) -\} - N^*(t-)$ of N^* over the small interval $[t, t + dt)$. Then the Andersen–Gill intensity model takes the form

$$\lambda_Z(t) = \exp\{\beta_0^T Z(t)\} \lambda_0(t), \quad (1.1)$$

where $\lambda_0(\cdot)$ is an unspecified continuous function and β_0 is a $p \times 1$ vector of regression parameters. Under this model, $N^*(t)$ is a time-transformed Poisson process in that $N^*\{\Lambda_Z^{-1}(t)\}$ is a Poisson process, where $\Lambda_Z^{-1}(t)$ is the inverse function of

$$\Lambda_Z(t) \equiv \int_0^t \exp\{\beta_0^T Z(u)\} \lambda_0(u) du.$$

If Z is time invariant, then $N^*(t)$ is a non-homogeneous Poisson process. Using the powerful martingale theory, Andersen and Gill (1982) developed an elegant large sample theory for the partial likelihood (Cox, 1975) estimation of model (1.1). These inference procedures have been implemented in major software packages and are commonly used by practitioners.

Model (1.1) consists of two major components:

- (a) $E\{dN^*(t)|\mathcal{F}_{t-}\} = E\{dN^*(t)|Z(t)\};$
- (b) $E\{dN^*(t)|Z(t)\} = \exp\{\beta_0^T Z(t)\} \lambda_0(t) dt.$

Assumption (a) implies that all the influence of the prior events on the future recurrence, if there is any, is mediated through time-varying covariates at t , whereas assumption (b) specifies how the covariates affect the instantaneous rate of the counting process. If Z is time invariant, then assumption (a) corresponds to the independent increments structure of the Poisson process.

It would be desirable to relax assumption (a) because the dependence of the recurrent events may not be adequately captured by time-varying covariates and because no method is available to verify this assumption. Thus, we remove assumption (a) and take assumption (b) as the defining property of our model. Denoting $E\{dN^*(t)|Z(t)\}$ by $d\mu_Z(t)$, we express the resulting model as

$$d\mu_Z(t) = \exp\{\beta_0^T Z(t)\} d\mu_0(t), \quad (1.2)$$

or

$$\mu_Z(t) = \int_0^t \exp\{\beta_0^T Z(u)\} d\mu_0(u), \quad (1.3)$$

where $\mu_0(\cdot)$ is an unknown continuous function. If Z consists of external covariates (Kalbfleisch and Prentice (1980), page 123) only, then

$$\mu_Z(t) = E\{N^*(t)|Z(s): s \geq 0\}$$

so $\mu_Z(t)$ and $\mu_0(t)$ pertain to the mean functions of the recurrent events; otherwise, they can only be interpreted as the cumulative rates. If Z is time invariant, then model (1.3) simplifies to

$$\mu_Z(t) = \exp(\beta_0^T Z) \mu_0(t). \quad (1.4)$$

Models (1.2) and (1.4) are referred to as the proportional rates and proportional means models.

Model (1.2) characterizes the rate of the counting process under model (1.1). Clearly, model (1.1) implies model (1.2) with $d\mu_0(t) = \lambda_0(t) dt$, but not vice versa. Model (1.2) is more

versatile than model (1.1) in that it allows arbitrary dependence structures among recurrent events and is applicable to any counting process for recurrent events. To illustrate this point, suppose that some subjects are more prone to recurrent events than others and that this heterogeneity can be characterized through the random-effect intensity model

$$\lambda_Z(t|\eta) = \eta \exp\{\beta_0^\top Z(t)\} \lambda_0(t), \quad (1.5)$$

where η is an unobserved unit-mean positive random variable that is independent of Z . Given model (1.5), model (1.2) holds whereas model (1.1) does not.

Pepe and Cai (1993) studied models in the form of equation (1.2) and advocated the use of the rate function of recurrence after the first event. They established a large sample theory for the semiparametric estimation of the regression parameters; however, there are some gaps in their technical developments, especially in the proof of their lemma A.1. More recently, Lawless and Nadeau (1995) and Lawless *et al.* (1997) studied the estimation of β_0 and μ_0 for models (1.2) and (1.4), though the proofs of the asymptotic results were given for the case of discrete times only.

In this paper, we provide a rigorous justification for the important results of Pepe and Cai and Lawless and colleagues by appealing to modern empirical process theory. In fact, we study a class of estimators which is a modification of those of Pepe and Cai and Lawless and colleagues, as will be discussed at the end of Section 2. Furthermore, we show how to construct simultaneous confidence bands for $\mu_0(\cdot)$ and covariate-specific mean functions. We also develop numerical and graphical techniques for checking the adequacy of model (1.2). These theoretical and methodological developments are presented in Sections 2–4, most of the technical details being relegated to Appendix A. In Section 5, we report some simulation results and provide an illustration with multiple-infection data taken from a clinical trial on chronic granulomatous disease (CGD).

2. Inferences on the regression parameters

In most applications, the subject is followed for a limited amount of time so $N^*(\cdot)$ is not fully observed. Let C denote the follow-up or censoring time. The censoring mechanism is assumed to be independent in the sense that

$$E\{dN^*(t)|Z(t), C \geq t\} = E\{dN^*(t)|Z(t)\}$$

for all $t \geq 0$. Define $N(t) = N^*(t \wedge C)$ and $Y(t) = I(C \geq t)$, where $a \wedge b = \min(a, b)$, and $I(\cdot)$ is the indicator function. For a random sample of n subjects, the observable data consist of $\{N_i(\cdot), Y_i(\cdot), Z_i(\cdot)\}$ ($i = 1, \dots, n$).

Let

$$S^{(k)}(\beta, t) = n^{-1} \sum_{i=1}^n Y_i(t) Z_i(t)^{\otimes k} \exp\{\beta^\top Z_i(t)\}$$

($k = 0, 1, 2$), where $a^{\otimes 0} = 1$, $a^{\otimes 1} = a$ and $a^{\otimes 2} = aa^\top$. Also, let $\bar{Z}(\beta, t) = S^{(1)}(\beta, t)/S^{(0)}(\beta, t)$, and $\bar{z}(\beta, t)$ be the limit of $\bar{Z}(\beta, t)$. We impose the following regularity conditions:

- (a) $\{N_i(\cdot), Y_i(\cdot), Z_i(\cdot)\}$ ($i = 1, \dots, n$) are independent and identically distributed;
- (b) $\Pr(C_i \geq \tau) > 0$ ($i = 1, \dots, n$), where τ is a predetermined constant;
- (c) $N_i(\tau)$ ($i = 1, \dots, n$) are bounded by a constant;
- (d) $Z_i(\cdot)$ ($i = 1, \dots, n$) have bounded total variations, i.e. $|Z_{ji}(0)| + \int_0^\tau |dZ_{ji}(t)| \leq K$ for all $j = 1, \dots, p$ and $i = 1, \dots, n$, where Z_{ji} is the j th component of Z_i and K is a constant.

(e) $A \equiv E[\int_0^t \{Z(t) - \bar{z}(\beta_0, t)\}^{\otimes 2} Y(t) \exp\{\beta_0^\top Z(t)\} d\mu_0(t)]$ is positive definite, where E is the expectation.

These conditions are analogous to those of Andersen and Gill (1982), theorem 4.1. Condition (e) holds if, at least for some interval of t , the distribution of $Z(t)$ conditional on $Y(t) = 1$ does not concentrate on a $(p - 1)$ -dimensional hyperplane.

Under model (1.1), the partial likelihood score function for β_0 is $U(\beta, \tau)$, where

$$U(\beta, t) = \sum_{i=1}^n \int_0^t \{Z_i(u) - \bar{Z}(\beta, u)\} dN_i(u). \quad (2.1)$$

Denote the solution to $U(\beta, \tau) = 0$ by $\hat{\beta}$. Under conditions (a)–(e), the random vectors $n^{-1/2} U(\beta_0; \tau)$ and $n^{1/2}(\hat{\beta} - \beta_0)$ are asymptotically zero mean normal with covariance matrices A and A^{-1} respectively, provided that model (1.1) is true (Andersen and Gill, 1982).

As mentioned in Section 1, for arbitrary counting processes, we formulate the relationship between $Z(\cdot)$ and $N^*(\cdot)$ through model (1.2) rather than through model (1.1). As shown in Appendix A, $U(\beta_0, t)$ is still centred at zero asymptotically as long as equation (1.2) holds. Thus, we estimate β_0 of model (1.2) by $\hat{\beta}$ as well. This is analogous to the use of a generalized estimating equation with an independence working assumption for longitudinal data (Liang and Zeger, 1986). In Appendix A.1, we show that, under model (1.2), $\hat{\beta}$ converges almost surely to β_0 .

To establish the asymptotic distribution of $\hat{\beta}$ under model (1.2), we need to derive the corresponding distribution of $U(\beta_0, \tau)$. Simple algebraic manipulation yields

$$U(\beta_0, t) = \sum_{i=1}^n \int_0^t \{Z_i(u) - \bar{Z}(\beta_0, u)\} dM_i(u), \quad (2.2)$$

where

$$M_i(t) = N_i(t) - \int_0^t Y_i(u) \exp\{\beta_0^\top Z_i(u)\} d\mu_0(u).$$

If $\{N_1^*(\cdot), \dots, N_n^*(\cdot)\}$ satisfy the intensity model given in equation (1.1), then $\{M_1(\cdot), \dots, M_n(\cdot)\}$ are martingales, in which case the weak convergence of $U(\beta_0, t)$ follows directly from the martingale central limit theorem (Andersen and Gill, 1982). For other counting processes, the M_i are not martingales so the martingale central limit theorem is not applicable. However, $E\{dM_i(t)|Z_i(t)\} = 0$ under model (1.2) since

$$dM_i(t) = I(C_i \geq t)[dN_i^*(t) - \exp\{\beta_0^\top Z_i(t)\} d\mu_0(t)].$$

Then, using modern empirical process theory, we show in Appendix A.2 that the process $n^{-1/2} U(\beta_0, t)$ ($0 \leq t \leq \tau$) converges weakly to a continuous zero-mean Gaussian process with covariance function

$$\Sigma(s, t) = E \left[\int_0^s \{Z_1(u) - \bar{z}(\beta_0, u)\} dM_1(u) \int_0^t \{Z_1(v) - \bar{z}(\beta_0, v)\}^\top dM_1(v) \right], \quad 0 \leq s, t \leq \tau,$$

between time points s and t . We also show in Appendix A.2 that $n^{1/2}(\hat{\beta} - \beta_0)$ is asymptotically zero mean normal with covariance matrix $\Gamma \equiv A^{-1} \Sigma A^{-1}$, where $\Sigma = \Sigma(\tau, \tau)$.

In general, $\Sigma \neq A$, so $\Gamma \neq A^{-1}$. In other words, the limiting covariance matrix for $\hat{\beta}$ under model (1.2) generally differs from its counterpart under model (1.1). If model (1.1) holds, which implies that model (1.2) also holds, then $\Sigma = A$ and consequently $\Gamma = A^{-1}$.

The covariance matrix Γ involves $\mu_0(\cdot)$. It is natural to estimate $\mu_0(t)$ by the Aalen–Breslow-type estimator

$$\hat{\mu}_0(t) \equiv \int_0^t \frac{d\bar{N}(u)}{n\hat{S}^{(0)}(\hat{\beta}, u)}, \quad t \in [0, \tau], \quad (2.3)$$

where $\bar{N}(u) = \sum_{i=1}^n N_i(u)$. In Appendix A.3, we show that $\hat{\mu}_0(\cdot)$ converges almost surely to $\mu_0(\cdot)$ and that the covariance matrix Γ can be consistently estimated by $\hat{\Gamma} \equiv \hat{A}^{-1} \hat{\Sigma} \hat{A}^{-1}$, where

$$\begin{aligned} \hat{A} &\equiv -n^{-1} \partial U(\hat{\beta}, \tau) / \partial \beta^T = n^{-1} \sum_{i=1}^n \int_0^\tau \{Z_i(u) - \bar{Z}(\hat{\beta}, u)\}^{\otimes 2} Y_i(u) \exp\{\hat{\beta}^T Z_i(u)\} d\hat{\mu}_0(u), \\ \hat{\Sigma} &= n^{-1} \sum_{i=1}^n \int_0^\tau \{Z_i(u) - \bar{Z}(\hat{\beta}, u)\} d\hat{M}_i(u) \int_0^\tau \{Z_i(v) - \bar{Z}(\hat{\beta}, v)\}^T d\hat{M}_i(v), \\ \hat{M}_i(t) &= N_i(t) - \int_0^t Y_i(u) \exp\{\hat{\beta}^T Z_i(u)\} d\hat{\mu}_0(u). \end{aligned}$$

We shall refer to $\hat{\Gamma}$ and \hat{A}^{-1} as the robust and naïve covariance matrix estimators respectively. The former is always valid whether or not the dependence structure for recurrent events is modelled correctly, whereas the latter is not.

By incorporating a random weight function $Q(\hat{\beta}, t)$ into $U(\beta, \tau)$, we obtain the following class of weighted estimating functions for β_0 :

$$U_Q(\beta, \tau) = \sum_{i=1}^n \int_0^\tau Q(\hat{\beta}, u) \{Z_i(u) - \bar{Z}(\beta, u)\} dN_i(u). \quad (2.4)$$

We assume that $Q(\hat{\beta}, t)$ is non-negative, bounded and monotone in t , and converges almost surely to a continuous deterministic function $q(t)$ in $t \in [0, \tau]$. (The monotonicity assumption may be relaxed to that $Q(\hat{\beta}, \cdot)$ is of bounded total variation so that it can be written as a sum of two monotone functions, each of which is assumed to converge to a deterministic function.) Weighted partial likelihood score functions similar to formula (2.4) have been studied by Lin (1991) and Sasieni (1993) in the context of censored survival data. Let $\hat{\beta}_Q$ be the solution to $U_Q(\beta, \tau) = 0$. We show in Appendix A.2 that $n^{1/2}(\hat{\beta}_Q - \beta_0)$ is asymptotically zero mean normal with covariance matrix $A_Q^{-1} \Sigma_Q A_Q^{-1}$, where

$$A_Q = E \left[\int_0^\tau q(t) \{Z_1(t) - \bar{z}(\beta_0, t)\}^{\otimes 2} Y_1(t) \exp\{\beta_0^T Z_1(t)\} d\mu_0(t) \right]$$

and

$$\Sigma_Q = E \left[\int_0^\tau q(u) \{Z_1(u) - \bar{z}(\beta_0, u)\} dM_1(u) \int_0^\tau q(v) \{Z_1(v) - \bar{z}(\beta_0, v)\}^T dM_1(v) \right].$$

We may estimate A_Q and Σ_Q consistently by replacing the unknown parameters by their respective sample estimators, as in the case of A and Σ .

The estimator $\hat{\beta}$ always exists and is unique at least for large n because $U(\beta, \tau)$ is the derivative of a concave function. In our definition of the weighted estimating functions given in formula (2.4), the weight function $Q(\hat{\beta}, t)$ does not involve β , which ensures that $U_Q(\beta, \tau)$ remains concave. By contrast, Pepe and Cai (1993), Lawless and Nadeau (1995) and Lawless *et al.* (1997) used $Q(\beta, t)$ instead of $Q(\hat{\beta}, t)$ in their weighted estimating functions. Consequently, their estimating functions may have multiple roots even in the limit, and the asymptotic properties of the resulting estimators have yet to be rigorously studied.

3. Inferences on the mean function

In this section, we assume that the covariates are external so that μ_0 has the mean function interpretation. As mentioned in Section 2, we may estimate $\mu_0(t)$ by $\hat{\mu}_0(t)$ given in expression (2.3). More generally, we may estimate $\mu_0(t)$ by $\hat{\mu}_Q(t)$, which is obtained from expression (2.3) by replacing $\hat{\beta}$ with $\hat{\beta}_Q$. The weak convergence for $\hat{\mu}_Q$ in continuous time has not previously been investigated. We show in Appendix A.4 that the process $V(t) \equiv n^{1/2}\{\hat{\mu}_Q(t) - \mu_0(t)\}$ ($t \leq \tau$) is asymptotically equivalent to $n^{-1/2} \sum_{i=1}^n \Psi_i(t)$, where

$$\Psi_i(t) = \int_0^t \frac{dM_i(u)}{S^{(0)}(\beta_0, u)} - h^T(\beta_0, t) A_Q^{-1} \int_0^\tau q(u) \{Z_i(u) - \bar{z}(\beta_0, u)\} dM_i(u) \quad (3.1)$$

and

$$h(\beta, t) = \int_0^t \bar{z}(\beta, u) d\mu_0(u).$$

It then follows from the multivariate central limit theorem, together with a proof of tightness again given in Appendix A.4, that $V(t)$ converges weakly to a zero-mean Gaussian process with covariance function $\xi(s, t) \equiv E\{\Psi_1(s) \Psi_1(t)\}$ at (s, t) . A natural estimator for $\xi(s, t)$ is

$$\hat{\xi}(s, t) \equiv n^{-1} \sum_{i=1}^n \hat{\Psi}_i(s) \hat{\Psi}_i(t),$$

where

$$\hat{\Psi}_i(t) = \int_0^t \frac{d\hat{M}_i(u)}{S^{(0)}(\hat{\beta}_Q, u)} - H^T(\hat{\beta}_Q, t) \hat{A}_Q^{-1} \int_0^\tau Q(\hat{\beta}_Q, u) \{Z_i(u) - \bar{Z}(\hat{\beta}_Q, u)\} d\hat{M}_i(u),$$

$$H(\beta, t) = \int_0^t \frac{\bar{Z}(\beta, u)}{n S^{(0)}(\beta, u)} d\bar{N}(u),$$

and \hat{A}_Q is analogous to \hat{A} . The consistency of $\hat{\xi}$ is established in Appendix A.4.

The asymptotic normality for $\hat{\mu}_Q(t)$, together with the consistent variance estimator $\hat{\xi}(t, t)$, enables us to construct pointwise confidence intervals for $\mu_0(t)$. Since $\mu_0(t)$ is non-negative, we consider the transformed random variable $n^{1/2}[\log\{\hat{\mu}_Q(t)\} - \log\{\mu_0(t)\}]$, whose distribution is asymptotically equivalent to that of $V(t)/\mu_0(t)$ provided that $\mu_0(t) > 0$. With the log-transformation, an approximate $1 - \alpha$ pointwise confidence interval for $\mu_0(t)$ is

$$\hat{\mu}_Q(t) \exp\{\pm n^{-1/2} z_{\alpha/2} \hat{\xi}^{1/2}(t, t)/\hat{\mu}_Q(t)\}, \quad (3.2)$$

where $z_{\alpha/2}$ is the upper $100\alpha/2$ percentage point of the standard normal distribution.

To construct simultaneous confidence bands for $\mu_0(t)$ over a time interval of interest $[t_1, t_2]$ ($0 < t_1 < t_2 \leq \tau$), we need to evaluate the distribution of the supremum of the process $V(t)$ or a related process over $[t_1, t_2]$. It is not possible to evaluate such distributions analytically because the limiting process of $V(t)$ does not have an independent increments structure even when the $N_i^*(\cdot)$ are Poisson processes. We show in Appendix A.4 that the distribution of the process $V(t)$ can be approximated by that of the zero-mean Gaussian process

$$\tilde{V}(t) \equiv n^{-1/2} \sum_{i=1}^n \hat{\Psi}_i(t) G_i,$$

where (G_1, \dots, G_n) are independent standard normal variables which are independent of

$\{N_i(\cdot), Y_i(\cdot), Z_i(\cdot)\}$ ($i = 1, \dots, n$). To approximate the distribution of $V(t)$, we obtain a large number of realizations from $\tilde{V}(t)$ by repeatedly generating the normal random sample (G_1, \dots, G_n) while fixing the data $\{N_i(\cdot), Y_i(\cdot), Z_i(\cdot)\}$ ($i = 1, \dots, n$) at their observed values. Using this simulation method, we may determine an approximate value of $c_{\alpha/2}$ which satisfies

$$\Pr \left\{ \sup_{t_1 \leq t \leq t_2} \left| \frac{\tilde{V}(t)}{\hat{\xi}^{1/2}(t, t)} \right| < c_{\alpha/2} \right\} = 1 - \alpha.$$

Then, with the log-transformation, an approximate $1 - \alpha$ simultaneous confidence band for $\mu_0(t)$ over $[t_1, t_2]$ is

$$\hat{\mu}_Q(t) \exp \{ \pm n^{-1/2} c_{\alpha/2} \hat{\xi}^{1/2}(t, t) / \hat{\mu}_Q(t) \}. \quad (3.3)$$

In applications, we are often interested in estimating or predicting the mean function $\mu_z(t)$ for subjects with specific covariate value z . If all the covariates are centred at z , then μ_0 corresponds to μ_z . Thus, we may obtain a consistent estimator of μ_z and construct the pointwise confidence interval or simultaneous confidence bands for μ_z by using the above formulae for μ_0 after replacing (Z_1, \dots, Z_n) with $(Z_1 - z, \dots, Z_n - z)$ in the data set.

4. Model checking techniques

Since $\hat{M}_i(t)$ is the difference between the observed and model-predicted numbers of events on the i th subject by time t , it is natural to use these residuals to check the adequacy of model (1.2). Following Lin *et al.* (1993), we develop a class of graphical and numerical methods by using certain cumulative sums of the $\hat{M}_i(t)$. Because the basic ideas are similar to those of Lin *et al.* (1993), we keep our discussion fairly brief in this section. However, it is important to note that there is an additional technical challenge here because the recurrent event times are correlated so the $M_i(t)$ are not martingales.

We first consider the problem of checking the functional forms of the covariates. For this, we assume that the covariates are time invariant. Let $\tilde{M}_i = \hat{M}_i(\tau)$. To check the functional form for the j th component of Z , we may plot the \tilde{M}_i against the Z_{ji} . To construct more objective and formal procedures, we consider the cumulative sum of the \tilde{M}_i over the values of Z_{ji} :

$$W_j(x) = n^{-1/2} \sum_{i=1}^n I(Z_{ji} \leq x) \tilde{M}_i.$$

We show in Appendix A.5 that the null distribution of $W_j(x)$ can be approximated by the zero-mean Gaussian process

$$\begin{aligned} \tilde{W}_j(x) = n^{-1/2} \sum_{i=1}^n \int_0^\tau \left\{ I(Z_{ji} \leq x) - \frac{S(\hat{\beta}, u, x)}{S^{(0)}(\hat{\beta}, u)} \right\} d\hat{M}_i(u) G_i \\ - B^T(\hat{\beta}, \tau, x) \hat{A}^{-1} n^{-1/2} \sum_{i=1}^n \int_0^\tau \{Z_i - \bar{Z}(\hat{\beta}, u)\} d\hat{M}_i(u) G_i, \end{aligned} \quad (4.1)$$

where

$$S(\beta, u, x) = n^{-1} \sum_{i=1}^n Y_i(u) \exp(\beta^T Z_i) I(Z_{ji} \leq x)$$

and

$$B(\beta, t, x) = n^{-1} \sum_{i=1}^n \int_0^t Y_i(u) \exp(\beta^T Z_i) I(Z_{ji} \leq x) \{Z_i - \bar{Z}(\beta, u)\} d\hat{\mu}_0(u).$$

In this section, the null distribution pertains to the distribution of the goodness-of-fit process when model (1.2) holds. As in the case of $\tilde{V}(\cdot)$, the distribution of $\tilde{W}_j(\cdot)$ may be simulated. To assess how unusual the observed residual pattern is, we plot a few, say 20, realizations from $\tilde{W}_j(\cdot)$ along with the observed $W_j(\cdot)$. Furthermore, we may complement the graphical inspection with the supremum test $\sup_x |W_j(x)|$. The p -value of this test is obtained by generating a large number of, say 1000, realizations from $\sup_x |\tilde{W}_j(x)|$ and comparing them with the observed value of $\sup_x |W_j(x)|$.

To check the exponential link function of model (1.2), we consider the process

$$W_r(x) = n^{-1/2} \sum_{i=1}^n I(\hat{\beta}^T Z_i \leq x) \tilde{M}_i.$$

The null distribution of this process can be approximated by that of $\tilde{W}_r(x)$, which is obtained from expression (4.1) by replacing $I(Z_{ji} \leq x)$ with $I(\hat{\beta}^T Z_i \leq x)$. Graphical and numerical inspections can be conducted in the same fashion as for $W_j(\cdot)$.

To check the proportional rates or means assumption with respect to the j th covariate component, we consider the standardized ‘score’ process

$$U_j^*(t) = (\hat{\Sigma}_{jj}^{-1})^{1/2} n^{-1/2} U_j(\hat{\beta}, t),$$

where $U_j(\beta, t)$ is the j th component of $U(\beta, t)$ and $\hat{\Sigma}_{jj}^{-1}$ is the j th diagonal element of $\hat{\Sigma}^{-1}$. Clearly,

$$U(\hat{\beta}, t) = \sum_{i=1}^n \int_0^t Z_i(u) d\hat{M}_i(u),$$

which involves transformations of the $\hat{M}_i(t)$. The justification for using $U_j^*(\cdot)$ ($j = 1, \dots, p$) to check the proportional rates and means assumptions is similar to that of Lin *et al.* (1993) for checking the proportional hazards assumption. We show in Appendix A.5 that the null distribution of $U_j^*(t)$ can be approximated by that of the zero-mean Gaussian process

$$\begin{aligned} \tilde{U}_j^*(t) = & (\hat{\Sigma}_{jj}^{-1})^{1/2} \left[n^{-1/2} \sum_{i=1}^n \int_0^t \left\{ Z_{ji}(u) - \frac{S_j^{(1)}(\hat{\beta}, u)}{S^{(0)}(\hat{\beta}, u)} \right\} d\hat{M}_i(u) G_i \right. \\ & \left. - J^T(\hat{\beta}, t) \hat{A}^{-1} n^{-1/2} \sum_{i=1}^n \int_0^t \{Z_i(u) - \bar{Z}(\hat{\beta}, u)\} d\hat{M}_i(u) G_i \right], \end{aligned} \quad (4.2)$$

where

$$J(\beta, t) = n^{-1} \sum_{i=1}^n \int_0^t Y_i(u) \exp\{\beta^T Z_i(u)\} Z_{ji}(u) \{Z_i(u) - \bar{Z}(\beta, u)\} d\hat{\mu}_0(u)$$

and $S_j^{(1)}$ is the j th component of $S^{(1)}$. Again, graphical and numerical inspections can be performed by simulation.

An omnibus test for checking the overall fit of the model can be constructed from the process

$$W_o(t, z) = n^{-1/2} \sum_{i=1}^n I(Z_i \leq z) \hat{M}_i(t),$$

where the event $\{Z_i \leq z\}$ means that each of the p components of Z_i is no larger than the respective component of z . We show in Appendix A.5 that the null distribution of $W_o(t, z)$ can be approximated by the zero-mean Gaussian process $\tilde{W}_o(t, z)$, which is obtained from expression (4.1) by replacing $I(Z_{ji} \leq x)$ with $I(Z_i \leq z)$, τ in the first integral with t and $B(\hat{\beta}, \tau, x)$ with $B(\hat{\beta}, t, z)$. An omnibus test statistic is $\sup_{t,z} |W_o(t, z)|$.

5. Numerical results

5.1. Simulation studies

A series of simulation studies was conducted to assess the performance of the robust method described in Section 2 and the corresponding method of Andersen and Gill (1982). We considered randomized clinical trials with $m \equiv n/2$ patients assigned to each of the two treatment groups. We generated recurrent event times from model (1.5) with $\lambda_0(t) = 1$ and $\beta_0 = 0.5$, Z being the treatment indicator and η having a gamma distribution with mean 1 and variance σ^2 . The subject's follow-up time was uniform $[0, 3]$, which yielded an average of approximately two observed events per subject during the trial period. We set $m = 50, 100, 200$ and $\sigma^2 = 0, 0.25, 0.5, 1$. For each combination of m and σ^2 , 10000 data sets were simulated.

For each simulated data set, we estimated β_0 under two intensity models:

$$\lambda_Z(t) = \exp(\beta_0 Z) \lambda_0(t), \quad (5.1)$$

$$\lambda_Z(t) = \exp\{\beta_0 Z + \gamma_0 X(t)\} \lambda_0(t), \quad (5.2)$$

where Z is the treatment indicator and

$$X(t) = \begin{cases} 1 & \text{if there was an event within the interval } [t-1, t), \\ 0 & \text{otherwise.} \end{cases}$$

These models are similar to models I–III of Andersen and Gill (1982), pages 1113–1115, and have commonly been used to analyse recurrent events for clinical trials. Of course, fitting model (5.1) with the robust variance estimator is the same as fitting the proportional means model

$$\mu_Z(t) = \exp(\beta_0 Z) \mu_0(t). \quad (5.3)$$

As discussed in Section 1, model (5.3) holds whereas models (5.1) and (5.2) do not unless $\sigma^2 = 0$. We took τ as the largest recurrence time so that all the data were used in the analysis.

The results of the simulation studies are displayed in Table 1. Under model (5.1), $\hat{\beta}$ is virtually unbiased. The robust variance estimator provides a fairly accurate estimation of the true variance of $\hat{\beta}$, and the corresponding confidence intervals have reasonable coverage probabilities. The performance of the robust method tends to improve as m increases and as σ^2 decreases. For non-zero σ^2 , the naïve variance estimator underestimates the true variance of $\hat{\beta}$, and thus the corresponding confidence intervals do not have proper coverage probabilities. Under model (5.2), the estimator of β_0 is biased downwards unless $\sigma^2 = 0$. The performance of the intensity-based method is not acceptable even if $\sigma^2 = 0.25$ and worsens rapidly as σ^2 increases.

Simulation studies were also conducted to evaluate the inference procedures developed in Sections 3 and 4. The results indicate that the asymptotic approximations are sufficiently accurate for practical use.

Table 1. Summary statistics for the simulation studies†

m	σ^2	Results for model (5.1)						Results for model (5.2)			
		<i>Bias</i>	<i>SE</i>	<i>SEE(R)</i>	<i>CP(R)</i>	<i>SEE(N)</i>	<i>CP(N)</i>	<i>Bias</i>	<i>SE</i>	<i>SEE(N)</i>	<i>CP(N)</i>
50	0	0.002	0.149	0.145	0.942	0.148	0.949	0.004	0.153	0.151	0.951
50	0.25	0.004	0.189	0.183	0.940	0.148	0.878	−0.040	0.175	0.150	0.899
50	0.5	0.002	0.221	0.214	0.939	0.149	0.813	−0.072	0.191	0.151	0.843
50	1	0.003	0.275	0.263	0.935	0.150	0.718	−0.106	0.222	0.151	0.764
100	0	0.002	0.104	0.103	0.945	0.104	0.950	0.004	0.106	0.106	0.951
100	0.25	0.003	0.134	0.131	0.942	0.104	0.875	−0.043	0.123	0.105	0.888
100	0.5	0.001	0.157	0.153	0.943	0.104	0.807	−0.075	0.135	0.105	0.814
100	1	−0.002	0.195	0.190	0.939	0.105	0.713	−0.111	0.157	0.105	0.702
200	0	0.001	0.074	0.073	0.946	0.073	0.948	0.001	0.075	0.074	0.949
200	0.25	0.002	0.093	0.093	0.949	0.073	0.881	−0.045	0.086	0.074	0.862
200	0.5	0.001	0.111	0.109	0.945	0.073	0.805	−0.075	0.096	0.074	0.755
200	1	0.002	0.137	0.135	0.947	0.074	0.711	−0.108	0.110	0.074	0.622

†Bias is the mean of the estimates of β_0 minus β_0 , SE is the standard error of the estimates of β_0 ; SEE(R) and SEE(N) are the means of the robust and naïve standard error estimates respectively, and CP(R) and CP(N) are the coverage probabilities of the corresponding 95% confidence intervals. The robust and naïve standard error estimates are based on $\hat{\Gamma}$ and \hat{A}^{-1} respectively.

5.2. A real example

We now apply the methods proposed to the multiple-infection data taken from the CGD study presented by Fleming and Harrington (1991), section 4.4. CGD is a group of inherited rare disorders of the immune function characterized by recurrent pyogenic infections. To evaluate the ability of gamma interferon in reducing the rate of infection, a placebo-controlled randomized clinical trial was conducted between October 1988 and March 1989. A total of 128 patients were enrolled into the study. By the end of the trial, 14 of 63 treated patients and 30 of 65 untreated patients had experienced at least one infection. Of the 30 untreated patients with at least one infection, five had experienced two infections, four others had experienced three and three patients had four or more. Of the 14 treated patients with at least one infection, four experienced two and another had a third infection.

Let us first fit models (5.1) and (5.2) to these data. The treatment indicator Z takes the value 1 if the subject received gamma interferon. For model (5.2), we redefine $X(t)$ to be the indicator on whether or not there was an infection within the last 60 days. Model (5.1) was previously used by Fleming and Harrington (1991), page 163, to analyse an earlier version of the CGD data and is similar to models I and II of Andersen and Gill (1982), whereas model (5.2) is similar to model III of Andersen and Gill. Of course, the previous researchers only used the naïve variance estimator.

The results for models (5.1) and (5.2) are shown in Table 2. The treatment variable is highly significant, indicating that gamma interferon is effective in reducing the rate of infection. Under model (5.1), the robust standard error estimate is considerably larger than the naïve estimate. Under model (5.2), the time-varying covariate $X(\cdot)$ is highly significant, the standardized parameter estimate being 2.43; the difference between the naïve and robust standard error estimates becomes smaller, though still appreciable. This suggests that the time-varying covariate captures part of the dependence among recurrent infections. The estimate of the treatment effect is about 10% smaller under model (5.2). This is consistent with the simulation results shown in Table 1 in that the estimator of the treatment effect under model (5.2) has a negative bias if there is overdispersion. In the CGD study, as in many other clinical trials, some patients were more prone to recurrent events than others, which

Table 2. Estimation of the treatment effect for the CGD study†

Model	Parameter estimate	Results for the naïve method		Results for the robust method	
		SEE	Estimate/SEE	SEE	Estimate/SE
(5.1)	−1.097	0.261	−4.20	0.311	−3.53
(5.2)	−0.989	0.266	−3.72	0.294	−3.36

†SEE is the standard error estimate.

Table 3. Estimation of the effects of treatment and age for the CGD study†

Covariate	Parameter estimate	SEE		<i>p</i> -value	
		Robust	Naïve	Robust	Naïve
Treatment	−1.12	0.309	0.261	0.0003	0.00002
Age	−0.03	0.014	0.013	0.034	0.020

†SEE is the standard error estimate; *p*-value pertains to testing no covariate effect.

suggests that model (1.5) or model (5.3) is more appropriate than models (5.1) and (5.2). Note that the test for no treatment difference based on model (5.3) is purely nonparametric in that it does not model any aspect of the recurrent events.

For further illustration, we consider model (1.4) with $Z = (Z_1, Z_2)^T$, where Z_1 is the treatment indicator defined above and Z_2 is the patient’s age at enrolment. This model would be useful for predicting the experience of infection for patients with specific age and treatment assignments. The estimates of the regression parameters are presented in Table 3. Again, the naïve standard error estimate for the treatment effect is much smaller than the robust estimate, yielding a markedly smaller *p*-value for testing no treatment difference. Incidentally, in both Tables 2 and 3, we let $Q = 1$ and τ be the largest observed infection time.

Fig. 1 displays the estimates of the cumulative frequencies of recurrent infections for 14-year-old patients who received gamma interferon *versus* those who did not receive gamma interferon. The estimates are shown from day 4 to day 373, which are respectively the smallest and largest infection times observed in the data set. The simultaneous confidence bands are based on expression (3.3) with 1000 simulated realizations of $\tilde{V}(\cdot)$. Clearly, the treated patients tend to have fewer infection episodes over time.

Fig. 2 summarizes the graphical and numerical results for checking the adequacy of the assumed model: Figs 2(a) and 2(b) pertain to the functional form for age and the exponential link function respectively, whereas Figs 2(c) and 2(d) pertain to the proportional means assumptions with respect to treatment and age respectively. In all the plots, the observed residual processes appear to be within the normal ranges. Thus, there is no evidence against the model assumed.

6. Remarks

The intensity model requires the correct specification of the dependence of the recurrent events within the same subject. In many applications, especially in a medical context, the dependence structure is complex and unknown. Even if the dependence structure is known, it may not be possible to fit the corresponding intensity model. Under model (1.5), for instance, the induced intensity model for $\lambda_Z(t)$ has a very complicated form and cannot be analysed

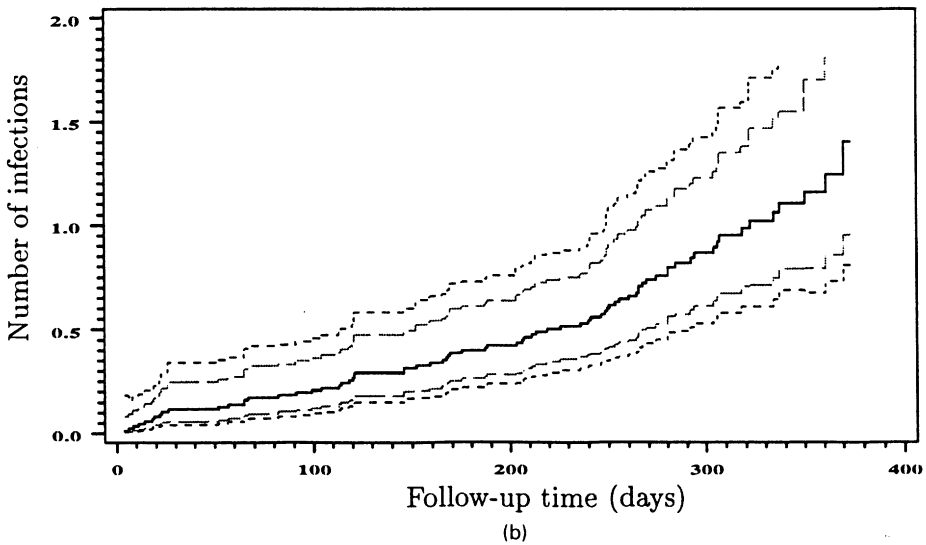
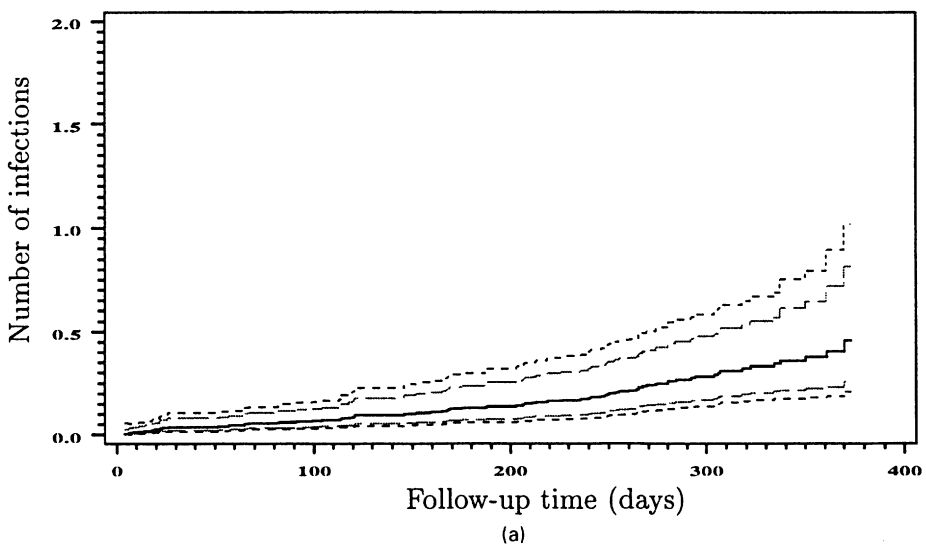


Fig. 1. Estimation of the cumulative frequencies of infections for 14-year-old CGD patients (a) receiving gamma interferon and (b) not receiving gamma interferon (—, point estimates; — —, 95% confidence limits; - - - -, 95% confidence bands)

by the usual partial likelihood approach. To accommodate the dependence, we may add time-varying covariates extracted from the counting process history into model (1.1). However, most dependence structures, such as the heterogeneity specified by model (1.5), cannot be adequately characterized by simple time-varying covariates. Furthermore, the inclusion of such time-varying covariates which are part of the response may result in biased estimation of the overall treatment effect for randomized trials, as elucidated by Kalbfleisch and Prentice (1980), pages 124–125, and demonstrated in Section 5.

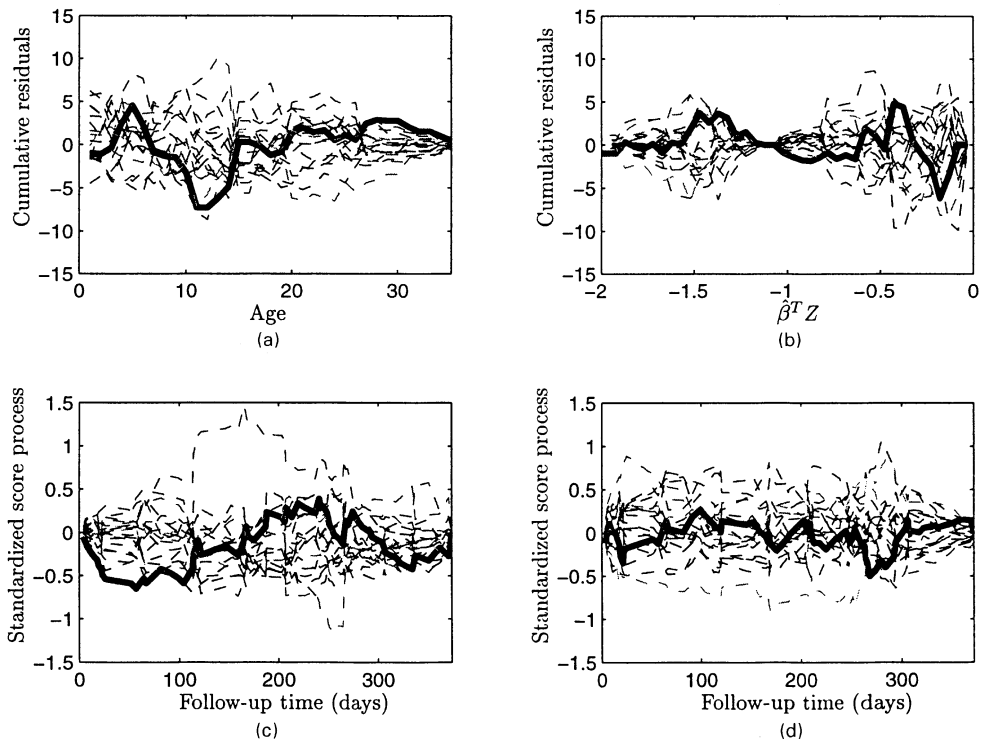


Fig. 2. Plots of residual processes for the CGD study: (a) cumulative residuals *versus* age (p -value 0.23); (b) cumulative residuals *versus* $\hat{\beta}^T Z$ (p -value 0.30); (c) standardized ‘score’ process *versus* follow-up time for treatment (p -value 0.42); (d) standardized ‘score’ process *versus* follow-up time for age (p -value 0.75) (—, observed processes; — —, simulated realizations from the null distributions; p -values pertain to the supremum goodness-of-fit tests)

By contrast, the proposed rate and mean models allow arbitrary dependence structures. As in the case of the intensity model, we may use the rate model with time-varying covariates which reflect the past history of the counting process to predict the probability of another event in the near future or to understand the dependence of the recurrent events. However, if we are interested in predicting the recurrence experience based on base-line covariate values, then model (1.4) can be used. Under model (1.4) or model (1.3) with external covariates only, μ_Z has the mean function interpretation. This quantity is particularly easy for non-statisticians to understand.

The approach of modelling the marginal mean and rate functions for recurrent events taken here is in the same vein as the approach of modelling the marginal hazard functions for multivariate failure times (Wei *et al.*, 1989). The latter uses a stratified proportional hazards model with a separate stratum depending on the number of previous events and can only handle a small number of recurrent events; the former treats all recurrent events of the same subject as a single counting process. Thus, the current approach is more efficient, flexible and parsimonious than the method of Wei *et al.* (1989) in handling recurrent events. Both approaches have been implemented in major software packages, although the asymptotic theory for the current approach was lacking.

Many applications are concerned with point processes that have arbitrary jump sizes. An

example of this kind is the accumulated cost of medical care, which has received tremendous recent interest. In such situations, we may formulate the rate or mean of the accumulation through model (1.2) or (1.4), whereas intensity- or hazard-based models no longer make sense. All the results stated in this paper continue to hold when $N^*(\cdot)$ is a point process with positive jumps of arbitrary sizes; the proofs given in Appendix A do not rely on the counting process feature of $N^*(\cdot)$ at all.

The results in this paper can be extended to other rate models, including those of Pepe and Cai (1993). In the presence of a terminal event, such as death, we may model the conditional rate function $E\{dN^*(t)|D \geq t; Z(t)\}$ by equation (1.2), where D is the time to the terminal event. Then, by redefining $N(t) = N^*(t \wedge D \wedge C)$ and $Y(t) = I(D \wedge C \geq t)$, we can show that the basic results in the paper hold for the conditional rate function provided that censoring by C is independent in that

$$E\{dN^*(t)|D \wedge C \geq t; Z(t)\} = E\{dN^*(t)|D \geq t; Z(t)\}.$$

The statistics encountered in this paper as well as in many other contexts are of the form

$$\sum_{i=1}^n \int H(t) dM_i(t),$$

where $M_i(\cdot)$ ($i = 1, \dots, n$) are zero-mean processes and $H(\cdot)$ involves the data from all the n subjects. If $M_i(\cdot)$ ($i = 1, \dots, n$) are martingales and $H(\cdot)$ is predictable, then the asymptotic properties for such statistics follow from martingale theory. The techniques developed in Appendix A do not require $M_i(\cdot)$ ($i = 1, \dots, n$) to be martingales or $H(\cdot)$ to be predictable and can be applied to a broad range of problems.

To simplify the proofs, we impose a truncation point τ in the estimating function for β_0 . Our simulation results indicated that it is appropriate to set τ to be the last observed recurrence time so that all the data are used in the inferences. In fact, it is possible to show that the desired asymptotic results hold without the tail restriction, though the proofs would be more tedious.

We assume that $N^*(\cdot)$ is a continuous time process. All the inference procedures proposed remain valid for the discrete time case. We then interpret $d\mu(t)$ as a mass function. The asymptotic proofs are similar, but simpler.

Appendix A: Proofs of asymptotic results

In this appendix, we prove the asymptotic properties of the inference procedures proposed in Sections 2–4. We shall repeatedly use the assumptions described in Section 2, especially conditions (a)–(e). We first state and prove a technical lemma which will be useful in proving the weak convergence of $U(\beta_0, t)$, $V(t)$ and other processes.

Lemma 1. Let f_n and g_n be two sequences of bounded functions such that, for some constant τ ,

- (a) $\sup_{0 \leq t \leq \tau} |f_n(t) - f(t)| \rightarrow 0$, where f is continuous on $[0, \tau]$,
- (b) $\{g_n\}$ are monotone on $[0, \tau]$ and
- (c) $\sup_{0 \leq t \leq \tau} |g_n(t) - g(t)| \rightarrow 0$ for some bounded function g . Then

$$\sup_{0 \leq t \leq \tau} \left| \int_0^t f_n(s) dg_n(s) - \int_0^t f(s) dg(s) \right| \rightarrow 0, \quad (\text{A.1})$$

$$\sup_{0 \leq t \leq \tau} \left| \int_0^t g_n(s) df_n(s) - \int_0^t g(s) df(s) \right| \rightarrow 0. \quad (\text{A.2})$$

Proof. It suffices to prove expression (A.1) because expression (A.2) follows from it through the integration by parts formula. Since $f_n = \max(f_n, 0) - \max(-f_n, 0)$, we assume without loss of generality that $\{f_n\}$ are all non-negative. Clearly,

$$\int_0^t f_n(s) dg_n(s) - \int_0^t f(s) dg(s) = \int_0^t \{f_n(s) - f(s)\} dg_n(s) + \left\{ \int_0^t f(s) dg_n(s) - \int_0^t f(s) dg(s) \right\}.$$

The first term on the right-hand side of this equation converges uniformly to 0 in view of assumption (a) and the fact that $\{g_n\}$ are monotone and uniformly bounded. Note that the uniform boundedness of $\{g_n\}$ follows directly from assumptions (b) and (c). Because the convergence of g_n to g is uniform, we can extend Helly's theorem (Serfling (1980), page 352) to show that

$$\int_0^{t\pm} f(s) dg_n(s) \rightarrow \int_0^{t\pm} f(s) dg(s)$$

for every t , where $t\pm = \lim_{\Delta t \rightarrow 0} (t \pm \Delta t)$. Thus, lemma 8.2.3 in Chow and Teicher (1988), page 265, can be used to conclude that the second term also converges to 0 uniformly in t .

A.1. Consistency of $\hat{\beta}$ and $\hat{\beta}_Q$

Define

$$X(\beta) = n^{-1} \left[\sum_{i=1}^n \int_0^\tau (\beta - \beta_0)^T Z_i(t) dN_i(t) - \int_0^\tau \log \left\{ \frac{S^{(0)}(\beta, t)}{S^{(0)}(\beta_0, t)} \right\} d\tilde{N}(t) \right].$$

In view of conditions (a)–(d) in Section 2, we can use the strong law of large numbers and the fact that $n^{-1} \tilde{N}(t)$ and $S^{(0)}(\beta, t)$ have bounded variations to show that $X(\beta)$ converges almost surely to

$$\mathcal{X}(\beta) \equiv E \left[\int_0^\tau (\beta - \beta_0)^T Z_1(t) dN_1(t) - \int_0^\tau \log \left\{ \frac{s^{(0)}(\beta, t)}{s^{(0)}(\beta_0, t)} \right\} dN_1(t) \right] \quad (\text{A.3})$$

for every β , where

$$s^{(k)}(\beta, t) = E[Y_1(t) \exp\{\beta^T Z_1(t)\} Z_1(t)^{\otimes k}]$$

($k = 0, 1, 2$). Clearly,

$$\frac{\partial^2 X(\beta)}{\partial \beta^2} = -n^{-1} \sum_{i=1}^n \int_0^\tau \{Z_i(t) - \bar{Z}(\beta, t)\}^{\otimes 2} Y_i(t) \exp\{\beta^T Z_i(t)\} \frac{d\tilde{N}(t)}{S^{(0)}(\beta, t)},$$

which is negative semidefinite. Thus, $X(\beta)$ is concave. This implies that the convergence of $X(\beta)$ to $\mathcal{X}(\beta)$ is uniform on any compact set of β (Rockafellar (1970), theorem 10.8). In particular, letting $\mathcal{B}_r = \{\beta: \|\beta - \beta_0\| \leq r\}$, we have

$$\sup_{\beta \in \mathcal{B}_r} \|X(\beta) - \mathcal{X}(\beta)\| \rightarrow 0 \quad (\text{A.4})$$

almost surely. It is easy to show that \mathcal{X} is concave with $\partial \mathcal{X}(\beta_0)/\partial \beta = 0$ and $\partial^2 \mathcal{X}(\beta_0)/\partial \beta^2 = -A$ under model (1.2). Since A is positive definite by condition (e), \mathcal{X} has a unique maximizer β_0 . In particular, $\sup_{\beta \in \partial \mathcal{B}_r} \{\mathcal{X}(\beta)\} < \mathcal{X}(\beta_0)$, where $\partial \mathcal{B}_r = \{\beta: \|\beta - \beta_0\| = r\}$. This fact, together with expression (A.4), implies that $X(\beta) < \mathcal{X}(\beta_0)$ for all $\beta \in \partial \mathcal{B}_r$ and all large n . Therefore, there must be a maximizer of $X(\beta)$, i.e. a solution to $\partial X(\beta)/\partial \beta = 0$, say $\hat{\beta}$, in the interior of \mathcal{B}_r . Because of condition (e), the argument of Jacobsen (1989), page 338, can be used to show the (global) uniqueness of $\hat{\beta}$ for all large n . Since r can be chosen arbitrarily small, $\hat{\beta}$ must converge to β_0 almost surely. The consistency of $\hat{\beta}_Q$ can be proved in the same manner.

A.2. Weak convergence of $U(\beta_0, t)$, $U_Q(\beta_0, t)$, $\hat{\beta}$ and $\hat{\beta}_Q$

In view of equation (2.2),

$$U(\beta_0, t) = \bar{M}_Z(t) - \int_0^t \bar{Z}(\beta_0, u) d\bar{M}(u),$$

where $\bar{M}(t) = \sum_{i=1}^n M_i(t)$ and

$$\bar{M}_Z(t) = \sum_{i=1}^n \int_0^t Z_i(u) dM_i(u),$$

both of which are sums of independent and identically distributed zero-mean terms for fixed t . By the multivariate central limit theorem, $(n^{-1/2} \bar{M}, n^{-1/2} \bar{M}_Z)$ converges in finite dimensional distributions to a zero-mean Gaussian process, say $(\mathcal{W}_M, \mathcal{W}_{M_Z})$. Obviously, $M_i(t)$ is the difference of two monotone functions in t . Since condition (d) in Section 2 implies that $Z_i(\cdot)$ is bounded, we may assume without loss of generality that $Z_i(\cdot) \geq 0$. Thus, each of the p components of $\int_0^t Z_i(u) dM_i(u)$ is also a difference of two monotone functions in t . Because monotone functions have pseudodimension 1 (Pollard (1990), page 15, and Biliias *et al.* (1997), lemma A.2), the processes $\{M_i(t); i = 1, \dots, n\}$ and $\{\int_0^t Z_i(u) dM_i(u); i = 1, \dots, n\}$ are manageable (Pollard (1990), page 38, and Biliias *et al.* (1997), lemma A.1). It then follows from the functional central limit theorem (Pollard (1990), page 53) that $(n^{-1/2} \bar{M}, n^{-1/2} \bar{M}_Z)$ is tight and thus converges weakly to $(\mathcal{W}_M, \mathcal{W}_{M_Z})$. This weak convergence also follows from example 2.11.16 of van der Vaart and Wellner (1996), page 215. Furthermore, it can be shown that $E\{\mathcal{W}_M(t) - \mathcal{W}_M(s)\}^4 \leq \tilde{K}\{\mu_0(t) - \mu_0(s)\}^2$ for some constant $\tilde{K} > 0$. It then follows from the Kolmogorov–Centsov theorem (Karatzas and Shreve (1988), page 53) that \mathcal{W}_M has continuous sample paths under the Euclidean distance. Likewise, \mathcal{W}_{M_Z} also has continuous sample paths.

By the strong embedding theorem (Shorack and Wellner (1986), pages 47–48), we can obtain in a new probability space almost sure convergence of $(n^{-1/2} \bar{M}, n^{-1/2} \bar{M}_Z, S^{(1)}, S^{(0)})$ to $(\mathcal{W}_M, \mathcal{W}_{M_Z}, s^{(1)}, s^{(0)})$. Clearly, $S^{(0)}(\beta_0, t)$ is a monotone function in t . Since $Z_i(\cdot) \geq 0$ ($i = 1, \dots, n$), each component of $S^{(1)}(\beta_0, t)$ is also a monotone function in t . It then follows from lemma 1 that

$$n^{-1/2} \int_0^t \frac{d\bar{M}(u)}{S^{(0)}(\beta_0, u)} \rightarrow \int_0^t \frac{d\mathcal{W}_M(u)}{s^{(0)}(\beta_0, u)}$$

uniformly in t almost surely. Applying lemma 1 once more, we obtain

$$n^{-1/2} \int_0^t \frac{S^{(1)}(\beta_0, u)}{S^{(0)}(\beta_0, u)} d\bar{M}(u) \rightarrow \int_0^t \frac{s^{(1)}(\beta_0, u)}{s^{(0)}(\beta_0, u)} d\mathcal{W}_M(u),$$

again uniformly in t almost surely. This convergence, coupled with the convergence of $n^{-1/2} \bar{M}_Z$ to \mathcal{W}_{M_Z} , yields the uniform convergence of $n^{-1/2} U(\beta_0, t)$ to

$$\mathcal{W}_{M_Z}(t) - \int_0^t \bar{z}(\beta_0, u) d\mathcal{W}_M(u)$$

almost surely in the new probability space and thus weakly in the original probability space. The limiting covariance function $\Sigma(s, t)$ given in Section 2 follows from a straightforward calculation.

By Taylor series expansion,

$$n^{1/2}(\hat{\beta} - \beta_0) = \hat{A}^{-1}(\beta^*) n^{-1/2} U(\beta_0, \tau),$$

where $\hat{A}(\beta) = -n^{-1} \partial U(\beta, \tau) / \partial \beta$, and β^* is on the line segment between $\hat{\beta}$ and β_0 . The consistency of $\hat{\beta}$ and $\hat{A}(\beta_0)$ for β_0 and A , together with the weak convergence of $n^{-1/2} U(\beta_0, \tau)$, implies that $n^{1/2}(\hat{\beta} - \beta_0)$ converges in distribution to a zero-mean normal random vector with covariance matrix $\Gamma \equiv A^{-1} \Sigma A^{-1}$. For future reference, we display the asymptotic approximation

$$n^{1/2}(\hat{\beta} - \beta_0) = A^{-1} n^{-1/2} \sum_{i=1}^n \int_0^\tau \{Z_i(u) - \bar{z}(\beta_0, u)\} dM_i(u) + o_p(1). \quad (\text{A.5})$$

For the weighted estimators, $U_Q(\beta_0, \tau) = \int_0^\tau Q(\hat{\beta}, t) dU(\beta_0, t)$. Since $Q(\hat{\beta}, t)$ is monotone with limit $q(t)$ and $n^{-1/2} U(\beta_0, \cdot)$ converges weakly, the strong embedding theorem can again be used to show the weak convergence of $n^{-1/2} U_Q(\beta_0, \tau)$. By Taylor series expansion and the consistency of $\hat{\beta}_Q$, we have

$$n^{1/2}(\hat{\beta}_Q - \beta_0) = A_Q^{-1} n^{-1/2} U_Q(\beta_0, \tau) + o_p(1),$$

which, in view of the convergence of $n^{-1/2} U_Q(\beta_0, \tau)$, is asymptotically zero mean normal with covariance matrix $A_Q^{-1} \Sigma_Q A_Q^{-1}$.

A.3. Consistency of $\hat{\mu}_0(\cdot)$ and $\hat{\Sigma}$

By the uniform strong law of large numbers (Pollard (1990), page 41), $n^{-1} \bar{N}(t) \rightarrow E\{N_1(t)\}$ and $S^{(0)}(\beta, t) \rightarrow s^{(0)}(\beta, t)$ uniformly in t and β . This entails uniform convergence of

$$\hat{\mu}_0(\beta, t) \equiv \int_0^t \frac{d\bar{N}(u)}{n S^{(0)}(\beta, u)}$$

to

$$\int_0^t \frac{s^{(0)}(\beta_0, u)}{s^{(0)}(\beta, u)} d\mu_0(u)$$

under model (1.2). The derivative of $\hat{\mu}_0(\beta, t)$ with respect to β is uniformly bounded for all large n and β in a bounded region. Therefore, the strong consistency of $\hat{\beta}$ implies that $\hat{\mu}_0(t) \equiv \hat{\mu}_0(\hat{\beta}, t)$ converges almost surely to $\mu_0(t)$ uniformly in t . This convergence, together with the almost sure convergence of $\hat{\beta}$ and $\bar{Z}(\beta_0, t)$ to β_0 and $\bar{z}(\beta_0, t)$, entails that

$$n^{-1} \sum_{i=1}^n \left\| \int_0^\tau \{Z_i(t) - \bar{Z}(\hat{\beta}, t)\} d\hat{M}_i(t) - \int_0^\tau \{Z_i(t) - \bar{z}(\beta_0, t)\} dM_i(t) \right\|^2 \rightarrow 0$$

almost surely. Thus, to prove that $\hat{\Sigma} \rightarrow \Sigma$ almost surely, it suffices to establish that

$$n^{-1} \sum_{i=1}^n \left[\int_0^\tau \{Z_i(t) - \bar{z}(\beta, t)\} dM_i(t) \right]^{\otimes 2} \rightarrow \Sigma$$

almost surely. The latter convergence follows from the strong law of large numbers. In addition, the almost sure convergence of $\hat{\beta}$ and $\hat{A}(\beta_0)$ to β_0 and A implies the almost sure convergence of \hat{A} to A . Hence, $\hat{\Gamma}$ converges almost surely to Γ .

A.4. Weak convergence of $V(t)$ and $\tilde{V}(t)$

We make the simple decomposition

$$V(t) = n^{1/2} \left\{ \int_0^t \frac{d\bar{N}(u)}{n S^{(0)}(\beta_0, u)} - \mu_0(t) \right\} + n^{1/2} \left\{ \int_0^t \frac{d\bar{N}(u)}{n S^{(0)}(\hat{\beta}_Q, u)} - \int_0^t \frac{d\bar{N}(u)}{n S^{(0)}(\beta_0, u)} \right\}. \quad (\text{A.6})$$

The first term on the right-hand side of equation (A.6) can be written as

$$n^{-1/2} \sum_{i=1}^n \int_0^t \frac{dM_i(u)}{S^{(0)}(\beta_0, u)}$$

for $t \leq \max_i(C_i)$. By the arguments of Appendix A.2, this term is tight and equals

$$n^{-1/2} \sum_{i=1}^n \int_0^t \frac{dM_i(u)}{s^{(0)}(\beta_0, u)} + o_p(1).$$

Taylor series expansion shows that the second term on the right-hand side of equation (A.6) equals $-H^T(\beta^\dagger, t) n^{1/2}(\hat{\beta}_Q - \beta_0)$, where β^\dagger is on the line segment between $\hat{\beta}_Q$ and β_0 . By lemma 1 and the uniform strong law of large numbers (Pollard (1990), page 41), $H(\beta_0, t)$ converges almost surely to $h(\beta_0, t)$ uniformly in t . Furthermore, analogously to equation (A.5),

$$n^{1/2}(\hat{\beta}_Q - \beta_0) = A_Q^{-1} n^{-1/2} \sum_{i=1}^n \int_0^\tau q(u) \{Z_i(u) - \bar{z}(\beta_0, u)\} dM_i(u) + o_p(1).$$

Thus, the second term on the right-hand side of equation (A.6) is tight and equals

$$-h^T(\beta_0, t)A_Q^{-1}n^{-1/2}\sum_{i=1}^n\int_0^T q(u)\{Z_i(u) - \bar{z}(\beta_0, u)\}dM_i(u) + o_p(1).$$

Hence, $V(t) = n^{-1/2}\sum_{i=1}^n\Psi_i(t) + o_p(1)$, which converges weakly to a zero-mean Gaussian process with covariance function ξ . By the same arguments as those of Appendix A.3, $\hat{\xi}(s, t) \rightarrow \xi(s, t)$ almost surely uniformly in t and s .

Conditionally on the data $\{N_i(\cdot), Y_i(\cdot), Z_i(\cdot); i = 1, \dots, n\}$, the only random components in $\tilde{V}(t)$ are (G_1, \dots, G_n) . Thus, it follows from the multivariate central limit theorem and a straightforward covariance calculation that, conditionally on the data, $\tilde{V}(t)$ converges in finite dimensional distributions to a zero-mean Gaussian process with covariance function $\hat{\xi}$. As mentioned above, $\hat{\xi} \rightarrow \xi$ almost surely. Therefore, $\tilde{V}(t)$ converges to the same limiting distribution as $V(t)$ provided that $\tilde{V}(t)$ is tight. The tightness of $\tilde{V}(t)$ again follows from the functional central limit theorem (Pollard (1990), page 53) because $\tilde{V}(t)$ comprises monotone functions in t , which are manageable.

A.5. Weak convergence of W_f , W_r , U^* and W_o

The processes W_j , U^* and W_o are all special cases of the multiparameter process

$$W(t, z) = n^{-1/2}\sum_{i=1}^n\int_0^t f(Z_i)I(Z_i \leq z)d\hat{M}_i(u),$$

where f is a smooth function. We shall establish the weak convergence of W under model (1.2). For simplicity, we assume that the covariates are time invariant. By Taylor series expansion and some simple algebra,

$$W(t, z) = n^{-1/2}\sum_{i=1}^n\int_0^t \left\{ f(Z_i)I(Z_i \leq z) - \frac{S_f(\beta_0, u, z)}{S^{(0)}(\beta_0, u)} \right\} dM_i(u) - B_f^T(\beta^*, t, z)n^{1/2}(\hat{\beta} - \beta_0), \quad (\text{A.7})$$

where

$$S_f(\beta, u, z) = n^{-1}\sum_{i=1}^n Y_i(u) \exp(\beta^T Z_i) f(Z_i) I(Z_i \leq z),$$

$$B_f(\beta, t, z) = n^{-1}\sum_{i=1}^n\int_0^t Y_i(u) \exp(\beta^T Z_i) f(Z_i) I(Z_i \leq z) \{Z_i - \bar{Z}(\beta, u)\} d\hat{\mu}_0(u)$$

and β^* is on the line segment between $\hat{\beta}$ and β_0 .

By the strong consistency of $\hat{\beta}$ and $\hat{\mu}_0$ and the uniform strong law of large numbers, $S_f(\beta_0, u, z)$ and $B_f(\beta^*, t, z)$ converge almost surely to some deterministic functions, $s_f(\beta_0, u, z)$ and $b_f(\beta_0, t, z)$ say. Because the first term on the right-hand side of equation (A.7) takes a similar form to $U(\beta_0, t)$, its tightness follows from the arguments given in Appendix A.2. In addition, the second term is tight since $n^{1/2}(\hat{\beta} - \beta_0)$ converges in distribution and $B_f(\beta^*, t, z)$ converges uniformly to $b_f(\beta_0, t, z)$. Therefore, $W(t, z)$ is tight.

It follows from lemma 1, equation (A.5) and the convergence of $\hat{\beta}$, $S^{(0)}$, S_f and B_f that

$$W(t, z) = n^{-1/2}\sum_{i=1}^n \Upsilon_i(t, z) + o_p(1),$$

where

$$\Upsilon_i(t, z) = \int_0^t \left\{ f(Z_i)I(Z_i \leq z) - \frac{s_f(\beta_0, u, z)}{s^{(0)}(\beta_0, u)} \right\} dM_i(u) - b_f^T(\beta_0, t, z)A^{-1}\int_0^t \{Z_i - \bar{z}(\beta_0, u)\} dM_i(u). \quad (\text{A.8})$$

The multivariate central limit theorem, together with the tightness of W , then implies that $W(t, z)$ converges weakly to a zero-mean Gaussian process with covariance function $E\{\Upsilon_1(t, z)\Upsilon_1(t^\dagger, z^\dagger)\}$ at (t, z) and (t^\dagger, z^\dagger) . By the arguments of Appendices A.3 and A.4, this covariance function can be consistently estimated by

$$n^{-1} \sum_{i=1}^n \hat{\Upsilon}_i(t, z) \hat{\Upsilon}_i(t^\dagger, z^\dagger),$$

where $\hat{\Upsilon}_i(t, z)$ are obtained from expression (A.8) by replacing all the unknown parameters by their respective sample estimators.

To establish the weak convergence of W_r , we let $B_\epsilon(\beta_0) = \{\beta: \|\beta - \beta_0\| \leq \epsilon\}$ and suppose that, for some $\epsilon > 0$, the function $\Pr(\beta^\top Z \leq x)$ is continuous in $(\beta, x) \in B_\epsilon(\beta_0) \times [a, b]$. It follows from the earlier arguments for $W(t, z)$ that $W_r(x) = W_r^*(\hat{\beta}, x) + o_p(1)$, where

$$W_r^*(\beta, x) = n^{-1/2} \sum_{i=1}^n \int_0^\tau \left[\left\{ I(\beta^\top Z_i \leq x) - \frac{s_r(\beta, u, x)}{s^{(0)}(\beta_0, u)} \right\} - b_r^\top(\beta, x) A^{-1} \{Z_i - \bar{z}(\beta_0, u)\} \right] dM_i(u), \quad (\text{A.9})$$

$$s_r(\beta, u, x) = E\{Y_1(u) \exp(\beta_0^\top Z_1) I(\beta^\top Z_1 \leq x)\}$$

and

$$b_r(\beta, x) = E \left[\int_0^\tau Y_1(u) \exp(\beta_0^\top Z_1) I(\beta^\top Z_1 \leq x) \{Z_1 - \bar{z}(\beta_0, u)\} d\mu_0(u) \right].$$

Since the right-hand side of equation (A.9) is a sum of independent zero-mean terms, the earlier arguments for $W(t, z)$ can again be used to verify the conditions including the manageability for the functional central limit theorem (Pollard (1990), page 53). Therefore, $W_r^*(\beta, x)$ converges weakly on $B_\epsilon(\beta_0) \times [a, b]$ to a Gaussian process and is stochastically equicontinuous (Pollard (1990), pages 52–53). In particular, $W_r^*(\hat{\beta}, x)$ and $W_r^*(\beta_0, x)$ are asymptotically equivalent and thus converge to the same limiting Gaussian process.

In view of equation (A.8) and by the arguments given in the second paragraph of Appendix A.4, the distribution of $W(t, z)$ can be approximated by the zero-mean Gaussian process

$$\tilde{W}(t, z) \equiv n^{-1/2} \sum_{i=1}^n \hat{\Upsilon}_i(t, z) G_i,$$

which contains $\tilde{W}_j(x)$ and $\tilde{U}_j^*(t)$ as special cases. Likewise, the distribution of $W_r(x)$ can be approximated by $\tilde{W}_r(x)$.

References

- Andersen, P. K. and Gill, R. D. (1982) Cox's regression model for counting processes: a large sample study. *Ann. Statist.*, **10**, 1100–1120.
- Biliyas, Y., Gu, M. and Ying, Z. (1997) Towards a general asymptotic theory for Cox model with staggered entry. *Ann. Statist.*, **25**, 662–682.
- Chow, Y. S. and Teicher, H. (1988) *Probability Theory: Independence, Interchangeability, Martingales*, 2nd edn. New York: Springer.
- Cox, D. R. (1972) Regression models and life-tables (with discussion). *J. R. Statist. Soc. B*, **34**, 187–220.
- (1975) Partial likelihood. *Biometrika*, **62**, 269–276.
- Fleming, T. R. and Harrington, D. P. (1991) *Counting Processes and Survival Analysis*. New York: Wiley.
- Jacobsen, M. (1989) Existence and unicity of MLEs in discrete exponential family distributions. *Scand. J. Statist.*, **16**, 335–349.
- Kalbfleisch, J. D. and Prentice, R. L. (1980) *The Statistical Analysis of Failure Time Data*. New York: Wiley.
- Karatzas, I. and Shreve, S. E. (1988) *Brownian Motion and Stochastic Calculus*. New York: Springer.
- Lawless, J. F. and Nadeau, C. (1995) Some simple robust methods for the analysis of recurrent events. *Technometrics*, **37**, 158–168.
- Lawless, J. F., Nadeau, C. and Cook, R. J. (1997) Analysis of mean and rate functions for recurrent events. In *Proc. 1st Seattle Symp. Biostatistics: Survival Analysis* (eds D. Y. Lin and T. R. Fleming), pp. 37–49. New York: Springer.
- Liang, K.-Y. and Zeger, S. L. (1986) Longitudinal data analysis using generalized linear models. *Biometrika*, **73**, 13–22.
- Lin, D. Y. (1991) Goodness-of-fit analysis for the Cox regression model based on a class of parameter estimators. *J. Am. Statist. Ass.*, **86**, 725–728.
- 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.

- Pepe, M. S. and Cai, J. (1993) Some graphical displays and marginal regression analyses for recurrent failure times and time dependent covariates. *J. Am. Statist. Ass.*, **88**, 811–820.
- Pollard, D. (1990) *Empirical Processes: Theory and Applications*. Hayward: Institute of Mathematical Statistics.
- Rockafellar, R. T. (1970) *Convex Analysis*. Princeton: Princeton University Press.
- Sasieni, P. (1993) Some new estimators for Cox regression. *Ann. Statist.*, **21**, 1721–1759.
- Serfling, R. J. (1980) *Approximation Theorems of Mathematical Statistics*. New York: Wiley.
- Shorack, G. R. and Wellner, J. A. (1986) *Empirical Processes with Applications to Statistics*. New York: Wiley.
- van der Vaart, A. W. and Wellner, J. A. (1996) *Weak Convergence and Empirical Processes*. New York: Springer.
- Wei, L. J., Lin, D. Y. and Weissfeld, L. (1989) Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *J. Am. Statist. Ass.*, **84**, 1065–1073.