ELSEVIER

Journal of Statistical Planning and Inference 44 (1995) 47–63

journal of statistical planning and inference

Semiparametric inference for the accelerated life model with time-dependent covariates

D.Y. Lin[a,*], Zhiliang Ying[b]

[a] Department of Biostatistics, SC-32, University of Washington, Seattle, WA 98195, USA
[b] Department of Statistics, University of Illinois, Champaign, IL 61820, USA

Received 1 April 1993; revised 13 January 1994

Abstract

The accelerated life model assumes that the failure time associated with a multi-dimensional covariate process is contracted or expanded relative to that of the zero-valued covariate process. In the present paper, the rate of contraction/expansion is formulated by a parametric function of the covariate process while the baseline failure time distribution is unspecified. Estimating functions for the vector of regression parameters are motivated by likelihood score functions and take the form of log rank statistics with time-dependent covariates. The resulting estimators are proven to be strongly consistent and asymptotically normal under suitable regularity conditions. Simple methods are derived for making inference about a subset of regression parameters while regarding others as nuisance quantities. Finite-sample properties of the estimation and testing procedures are investigated through Monte Carlo simulations. An illustration with the well-known Stanford heart transplant data is provided.

AMS Subject Classification: Primary 62J05; secondary 60F05, 62F12, 62G05

Key words: Accelerated failure time model; Censoring; Log rank statistic; Minimum dispersion statistic; Rank regression; Survival data; Time-varying covariates

1. Introduction

The accelerated life model specifies that the failure time variable T associated with a p-dimensional covariate process $\{Z(t),\ t>0\}$ is related to a baseline failure time variable T_0 by

$$T_0=\int_0^T \exp\{\beta' Z(u)\}\,\mathrm{d}u, \tag{1.1}$$

* Corresponding author. (206) 543-1044. Fax: (206) 543-3286. E-mail: danyu@biostat.washington.edu.

0378-3758/95/$09.50 © 1995—Elsevier Science B.V. All rights reserved
SSDI 0378-3758(94)00039-5

where T_0 refers to the condition $Z=0$, and β is a $p\times 1$ vector of regression parameters (Cox and Oakes, 1984, pp. 64–68). According to this model, any individual having failure time t under $Z(u)=z(u)$ $(0<u\leqslant t)$ would have failure time

$$\int_0^t \exp\{\beta' z(u)\}\,du$$

under $Z(u)=0$ $(0<u\leqslant t)$. The right-hand side of (1.1) may be replaced a different monotone and smooth transformation of T. In most instances, T is subject to right-censorship.

If Z is time-invariant, model (1.1) can be written as a linear regression model

$$\log T=-\beta' Z+\varepsilon, \tag{1.2}$$

where ε is a random variable independent of Z. When ε is assumed to come from a specific parametric family of distributions, the usual maximum likelihood method can be applied to model (1.2). On the other hand, if one is unwilling to make any parametric assumption on ε, then the rank regression approach (Prentice, 1978; Kalbfleisch and Prentice, 1980, pp. 143–62) is an appropriate choice. This semiparametric approach has received enormous recent attention (e.g., Ritov, 1990; Tsiatis, 1990; Wei et al., 1990; Lai and Ying, 1991; Ying, 1993).

To provide additional insights into the accelerated life model with *time-varying* covariates, suppose now that model (1.1) is employed to evaluate the effect of a test treatment on the survival time in a completed placebo-controlled clinical trial. The indicator covariate process $Z(\cdot)$ takes the value 1 when the patient is on the test treatment and takes the value 0 when he is off the test treatment or on the placebo. Imagine that a patient who was deceased at time t was on the test treatment from study entry to time $t^\dagger$ $(t^\dagger<t)$ and was off the treatment from $t^\dagger$ to t. According to model (1.1), this patient would have survived until $\{t^\dagger e^\beta+(t-t^\dagger)\}$ had he been on the placebo all along; his survival time was contracted or expanded relative to that of no treatment, depending on whether β is positive or negative, by a factor of e^β for the time period during which he was on the active treatment. This simple example reflects the essence of the accelerated life model with time-dependent covariates and demonstrates its immense potential in applications.

Formula (5.10) in Cox and Oakes (1984) provides the essential ingredients for the fully parametric analysis of model (1.1). Recently, Robins and Tsiatis (1992) proposed a class of rank-type estimating equations for accelerated life models with time-dependent covariates. They claimed that the techniques of Tsiatis (1990) could be used to establish an asymptotic linearity property of their estimating equations and thereby the existence of solutions that are asymptotically normal, but no formal theory or proof was given. As will be seen in the next section, the covariance matrices of these estimators involve the derivative of the unknown hazard function for T_0, and are therefore difficult to estimate well in practical applications.

In the present paper, we derive semiparametric procedures for making inference about β which bypass the variance–covariance estimation for the parameter estimators. The proposed methods are based on a rigorous large-sample theory which strengthens the results of Robins and Tsiatis (1992) in several key aspects. Simulation studies show that the rank-type estimators and our inference procedures perform satisfactorily for practical sample sizes. The familiar Stanford heart transplant data (Crowley and Hu, 1977) are used for illustration.

2. Inference procedures

Let T_i $(i=1,\dots,n)$ be a sequence of independent failure time variables and let $\{Z_i(t), t>0\}$ $(i=1,\dots,n)$ be the corresponding sequence of p-dimensional covariate processes. Assume that T_i $(i=1,\dots,n)$ are related to $Z_i(\cdot)$ $(i=1,\dots,n)$ through model (1.1). Under possible right-censorship, one can only observe $X_i=\min(T_i, C_i)$ and $\Delta_i=I\{T_i\leqslant C_i\}$ $(i=1,\dots,n)$, where C_i is the censoring time variable for the ith subject and $I\{\cdot\}$ is the indicator function. As usual, we assume that T_i and C_i are independent conditional on Z_i. Because all the following arguments will be conditional on Z_i, we shall regard Z_i nonrandom functions in our probability calculations.

It is convenient to introduce the notation:

$$h_i(t,b)=\int_0^t \exp\{b'Z_i(u)\}\,du, \qquad \tilde{X}_i(b)=h_i(X_i,b), \qquad \tilde{Z}_i(x,b)=Z_i\{h_i^{-1}(x,b)\},$$

$$N_i(x,b)=\Delta_i I\{\tilde{X}_i(b)\leqslant x\}, \qquad N(x,b)=\sum_{i=1}^n N_i(x,b),$$

$$N^Z(x,b)=\sum_{i=1}^n \int_0^x \tilde{Z}_i(u,b)\,dN_i(u,b),$$

$$Y_i(x,b)=I\{\tilde{X}_i(b)\geqslant x\}, \qquad Y(x,b)=\sum_{i=1}^n Y_i(x,b),$$

$$Y^Z(x,b)=\sum_{i=1}^n \tilde{Z}_i(x,b)\,Y_i(x,b).$$

Here, $h_i^{-1}(\cdot,b)$ is the inverse function of h_i with b being fixed. Note that $\tilde{Z}_i(x,\beta)$ is the ith covariate process on the baseline time scale.

2.1. Construction of parameter estimators

Under model (1.1), the likelihood score function for β is

$$\sum_{i=1}^n \int_0^\infty Q_i(x,\beta)\{dN_i(x,\beta)-Y_i(x,\beta)\lambda(x)\,dx\}, \tag{2.1}$$

where $\lambda(\cdot)$ is the hazard function for the baseline failure time variable T_0, and

$$Q_i(x,\beta)=\tilde{Z}_i(x,\beta)+\frac{\lambda'(x)}{\lambda(x)}\int_0^{h_i^{-1}(x,\beta)} Z_i(u)\exp\{\beta' Z_i(u)\}\,\mathrm{d}u.$$

Replacement of $\lambda(x)\,\mathrm{d}x$ in (2.1) by $\mathrm{d}N(x,\beta)/Y(x,\beta)$ results in

$$\sum_{i=1}^{n}\int_0^{\infty}\left\{Q_i(x,\beta)-\frac{\sum_{j=1}^{n} Q_j(x,\beta)\,Y_j(x,\beta)}{Y(x,\beta)}\right\}\mathrm{d}N_i(x,\beta). \tag{2.2}$$

In order to use (2.2) as an estimating function for β, we need to assume a specific form for the unknown hazard function $\lambda(\cdot)$ in $Q_i(\cdot,\beta)$. As in the usual rank regression, the choice of $\lambda(\cdot)$ for generating the estimating function will affect the efficiency, but not the validity of the resulting inference procedures. For simplicity, we shall replace $Q_i(\cdot,\beta)$ by $\tilde{Z}_i(\cdot,\beta)$ in (2.2), which will be optimal if the true $\lambda(\cdot)$ is constant. The corresponding estimating function becomes

$$\begin{aligned} U(b)&=\sum_{i=1}^{n}\int_0^{\infty}\left\{\tilde{Z}_i(x,b)-\frac{Y^Z(x,b)}{Y(x,b)}\,\mathrm{d}N_i(x,b)\right. \\ &=N^Z(\infty,b)-\int_0^{\infty}\frac{Y^Z(x,b)}{Y(x,b)}\,\mathrm{d}N(x,b). \end{aligned}$$

Note that $U(b)$ is the log rank statistic (Cox, 1972) based upon $\{\tilde{X}_i(b),\Delta_i,\tilde{Z}_i\}$ $(i=1,\ldots,n)$. Thus, the random vector $(nV_n)^{-1/2}U(\beta)$ is asymptotically p-variate standard normal (Andersen and Gill, 1982; Cuzick, 1985), where

$$V_n=n^{-1}\int_0^{\infty}\left[\frac{\mathscr{E}Y^{ZZ}(x,\beta)}{\mathscr{E}Y(x,\beta)}-\left\{\frac{\mathscr{E}Y^Z(x,\beta)}{\mathscr{E}Y(x,\beta)}\right\}^{\otimes 2}\right]\mathrm{d}\mathscr{E}N(x,\beta), \tag{2.3}$$

$$Y^{ZZ}(x,b)=\sum_{i=1}^{n}\tilde{Z}_i^{\otimes 2}(x,b)\,Y_i(x,b),$$

$\mathscr{E}$ denotes expectation, and $a^{\otimes 2}$ of a column vector a denotes aa'.

Unlike conventional estimating functions, $U(b)$ is discontinuous and generally nonmonotone. Thus the Newton-type algorithms for root-finding are inappropriate. In the present paper, the parameter estimator $\hat{\beta}$ is defined as a minimizer of $\|U(b)\|$ over a compact region. The minimization can be carried out by the method of simulated annealing (Kirkpatrick et al., 1983); such computational details have been provided by Lin and Geyer (1992).

2.2. Asymptotic properties of parameter estimators

Roughly speaking, in order to prove the consistency and asymptotic normality of $\hat{\beta}$, we need to establish the following two kinds of approximations for the estimating function $U(b)$.

(I) $U(b)=\bar{U}(b)+o_p(n^{(1/2)+\varepsilon})$ uniformly in $\|b\|\leqslant B$ for $\varepsilon>0$ and $B>0$, where $\bar{U}$ is a nonrandom function to be specified.

(II) $U(b)=U(\beta)-nA_n(b-\beta)+o_p(n^{1/2}+n\|b-\beta\|)$ as $n\to\infty$ and $b\to\beta$, where A_n is a nonrandom matrix to be specified.

Certain regularity conditions are required for making results (I) and (II) rigorous.

(C1) Each component process Z_{ik} has (uniformly) bounded total variation, i.e., there exists a constant D such that

$$Z_{ik}(0)+\int_0^\infty |\mathrm{d}Z_{ik}(x)|\leqslant D$$

for all $i=1,\dots,n$ and $k=1,\dots,p$.

From (C1), we obtain the unique Jordan decomposition (Rudin, 1974, p. 128)

$$Z_{ik}(x)=Z_{ik}(0)+Z_{ik}^{+}(x)-Z_{ik}^{-}(x),$$

where $Z_{ik}^{\pm}(\cdot)$ are increasing functions with $Z_{ik}^{\pm}(0)=0$. Write $\tilde{Z}_{ik}^{\pm}(x,b)=Z_{ik}^{\pm}\{h_i^{-1}(x,b)\}$.

(C2) There exist $\eta_0>0$ and $\kappa_0>0$ such that for $k=1,\dots,p$,

$$\sup_{|x-y|+\|b-a\|\leqslant n^{-\kappa_0}} n^{-1}\sum_{i=1}^{n}|\tilde{Z}_{ik}^{\pm}(x,b)-\tilde{Z}_{ik}^{\pm}(y,a)|=O(n^{-(1/2)-\eta_0}).$$

And for $0<d_n\to 0$, there exists $\varepsilon_0>0$ such that for $k=1,\dots,p$,

$$\sup_{|x-y|+\|b-a\|\leqslant d_n} n^{-1}\sum_{i=1}^{n}|\tilde{Z}_{ik}^{\pm}(x,b)-\tilde{Z}_{ik}^{\pm}(y-a)|=o(\max\{d_n^{\varepsilon_0},n^{-\varepsilon_0}\}).$$

(C3) The baseline density f and its derivative f' are bounded, and

$$\int_0^\infty\left\{\frac{f'(x)}{f(x)}\right\}^2 f(x)\,\mathrm{d}x<\infty,$$

$$\int_0^\infty x^{\theta_0}f(x)\,\mathrm{d}x<\infty \quad \text{for some } \theta_0>0.$$

(C4) The density functions g_i of C_i are uniformly bounded, i.e., $\sup_{t,i}g_i(t)<\infty$.

Remarks. Assumptions (C3) and (C4) were also made by Tsiatis (1990) and Ying (1993). They are essential in our proof of asymptotic linearity for the estimating function $U(b)$. The other two assumptions, (C1) and (C2) can be regarded as smoothness requirements on the covariate processes. They are satisfied when either the Z_i are smooth in the sense that they have uniformly bounded derivatives or when the Z_i are step functions with random between-jump times having uniformly bounded densities.

Our approach to justifying results (I) and (II) will be based upon approximations to a few basic weighted empirical processes with possibly time-varying weights.

Lemma 1 below is a main technical development in this direction, and it extends Theorem 1 of Lai and Ying (1988) to the setting of time-varying covariates.

Lemma 1. *Let $W(x,b)$ be any of the four processes $N(x,b)$, $N^Z(x,b)$, $Y(x,b)$ and $Y^Z(x,b)$. Similarly, let $\tilde{W}(x,b)$ be any of the four processes $\tilde{N}(x,b)$, $\tilde{N}^Z(x,b)$, $Y(x,b)$ and $Y^Z(x,b)$, where*

$$\tilde{N}(x,b)=\sum_{i=1}^{n}\int_{x}^{\infty}\mathrm{d}N_i(u,b),\quad \tilde{N}^Z(x,b)=\sum_{i=1}^{n}\int_{x}^{\infty}\tilde{Z}_i(u,b)\,\mathrm{d}N_i(u,b).$$

Then under (C1)–(C4),

(a) *for every $\varepsilon>0$ and $B>0$,*

$$\sup_{\|b\|\leqslant B}\|W(x,b)-\mathscr{E}W(x,b)\|=\mathrm{o}(n^{(1/2)+\varepsilon})\quad\textit{almost surely;}$$

(b) *for every $\gamma>0$, there exists $\eta>0$ such that*

$$\sup_{\|b\|\leqslant B,\,\mathscr{E}Y(x,b)\leqslant n^{1-\gamma}}\|\tilde{W}(x,b)-\mathscr{E}\tilde{W}(x,b)\|=\mathrm{o}(n^{(1/2)-\eta})\quad\textit{almost surely;}$$

(c) *for every $\gamma>0$, there exists $\eta>0$ such that*

$$\sup_{\|b-\beta\|\leqslant n^{-\gamma}}|W(x,b)-\mathscr{E}W(x,b)-W(x,\beta)+\mathscr{E}W(x,\beta)\|=\mathrm{o}(n^{(1/2)-\eta})\quad\textit{almost surely.}$$

The proof of Lemma 1 is relegated to the appendix.

Theorem 1. *Let*

$$\bar{U}(b)=\sum_{i=1}^{n}\int_{0}^{\infty}\left\{\tilde{Z}_i(x,b)-\frac{\mathscr{E}Y^Z(x,b)}{\mathscr{E}Y(x,b)}\right\}\mathrm{d}\mathscr{E}N_i(x,b),$$

$$A_n=n^{-1}\sum_{i=1}^{n}\int_{0}^{\infty}\left\{\tilde{Z}_i(x,\beta)-\frac{\mathscr{E}Y^Z(x,\beta)}{\mathscr{E}Y(x,\beta)}\right\}$$

$$\times\left\{\tilde{Z}_i(x,\beta)+\frac{\lambda'(x)}{\lambda(x)}\int_{0}^{h_i^{-1}(x,\beta)}Z_i(u)\exp\{\beta'Z_i(u)\}\,\mathrm{d}u\right\}'\{L_i(x,\beta)\}\,\mathrm{d}F(x),$$

where $L_i(\cdot,\beta)$ is the survival function of $\tilde{C}_i(\beta)=h_i(C_i,\beta)$. Then under (C1)–(C4),

(a) *for any $\varepsilon>0$ and $B>0$,*

$$\sup_{\|b\|\leqslant B}\|U(b)-\bar{U}(b)\|=\mathrm{o}\,(n^{(1/2)+\varepsilon})\quad\textit{almost surely;}$$

(b) *for any positive sequence $d_n\to 0$ almost surely (or in probability),*

$$\sup_{\|b-\beta\|\leqslant d_n}\{\|U(b)-U(\beta)+nA_n(b-\beta)\|/(n^{1/2}+n\|b-\beta\|)\}\to 0$$

almost surely (or in probability).

The proof of Theorem 1 is also given in the appendix.

Theorem 1 extends the main results, i.e., Theorems 1 and 3, of Ying (1993) to the setting of time-dependent covariates. Part (a) of Theorem 1 indicates that the (random) function U can be approximated by the nonrandom function $\bar{U}$ in any compact region, which leads to the consistency of $\hat{\beta}$. Part (b) shows that the function $U(b)$ is asymptotically linear with slope $-nA_n$ for b near β. The content of Theorem 1(b) is much stronger than result (4.1) of Robins and Tsiatis (1992). In particular, the asymptotic linearity for U holds in any shrinking neighborhood of β, rather than in the $n^{-1/2}$-neighborhood, which is cruial to our developments in the next subsection. An immediate consequence of Theorem 1 is the following theory on the consistency and asymptotic normality of $\hat{\beta}$.

Theorem 2. *Assume that* (C1)–(C4) *are satisfied and that* $\|A_n^{-1}\|$ *is bounded.*

(a) *There exists a fixed neighborhood* $\mathcal{N}$ *containing* β *as an interior point such that any minimizer* $\hat{\beta}\in\mathcal{N}$ *of* $\|U(b)\|$ *is strongly consistent and the random vector* $n^{1/2}V_n^{-1/2}A_n(\hat{\beta}-\beta)$ *converges to the p-variate standard normal distribution.*

(b) *If* $\liminf n^{-1}\|\bar{U}(b)\|>0$ *for every* $b\neq\beta$, *then the conclusions of* (a) *hold with* $\mathcal{N}$ *being any compact region containing* β *as its interior point.*

Remarks. (1) For time-invariant covariates, Ying (1993) verified that the condition for Theorem 2(b) holds in many important cases. (2) The results in Theorem 2 pertain to a *fixed* neighborhood whereas Robins and Tsiatis (1992) only studied the $n^{-1/2}$-neighborhood. The improvement made here is important since it is difficult, at least conceptually, to locate a neighborhood of β which shrinks to the unknown point β as $n\to\infty$.

Since A_n involves the unknown hazard function $\lambda(\cdot)$ and its derivative $\lambda'(\cdot)$, it is difficult to estimate the limiting covariance matrix of $\hat{\beta}$ well in practice. Making use of the fact that $n^{-1}U(b)$ is asymptotically linear with slope $-A_n$ in the $n^{-1/2}$-neighborhood of β, Robins and Tsiatis (1992) proposed to estimate A_n by a $p\times p$ matrix of numerical partial derivatives on $n^{-1}U(b)$ near the estimator $\hat{\beta}$ with step size of order $n^{-1/2}$. This approach, however, can yield rather different estimators for varying step sizes and will be unreliable in finite samples since $U(b)$ is neither continuous nor monotone.

In the next subsection, we provide alternative methods for making inference about β, which avoid estimating the covariance matrix of $\hat{\beta}$. In this context, we need a consistent estimator for the limit of V_n. In view of the definition of V_n given in (2.3), a natural choice is $n^{-1}V(\hat{\beta})$, where

$$V(b)=\int_0^{\infty}\left[\frac{Y^{ZZ}(x,b)}{Y(x,b)}-\left\{\frac{Y^{Z}(x,b)}{Y(x,b)}\right\}^{\otimes 2}\right]\mathrm{d}N(x,b).$$

Approximations given in Lemma 1 can be used to show that, under the assumptions for Theorem 2, $n^{-1}V(\hat{\beta})-V_n\to 0$ almost surely.

2.3. Minimum dispersion statistics

Suppose now that we are interested in $\beta^{(1)}$, a $q \times 1$ subvector of

$$\beta = \begin{bmatrix} \beta^{(1)} \\ \beta^{(2)} \end{bmatrix}.$$

The vectors b and $\hat{\beta}$ are similarly partitioned. Using the type of arguments given in Appendix 2 of Wei et al. (1990), we can prove that Theorem 3 follows from Theorems 1 and 2 of Section 2.2.

Theorem 3. *Suppose that all the assumptions for Theorem* 2 (b) *are satisfied. Then under* H_0: $\beta^{(1)} = \beta_0^{(1)}$, *the minimum dispersion statistic*

$$G(\beta_0^{(1)}) = \inf_{\substack{b^{(1)} = \beta_0^{(1)} \\ \|b^{(2)} - \hat{\beta}^{(2)}\| \leqslant B}} U'(b) V^{-1}(\hat{\beta}) U(b),$$

where $B > 0$ *is any given constant, is asymptotically distributed as* χ_q^2.

Theorem 3 is stronger than the corresponding result of Wei et al. (1990) in that Wei et al. were only able to prove the case of $B = n^{-1/3} d$ for some given $d > 0$. The annealing algorithm (Lin and Geyer, 1992) can also be used to calculate G. It follows immediately from Theorem 3 that $\{\psi: G(\psi) \leqslant \chi_q^2(\alpha)\}$ is an $(1-\alpha)$-level confidence region for $\beta^{(1)}$. Here $\chi_q^2(\alpha)$ is the upper α point of the χ_q^2 distribution.

3. Numerical results

3.1. Stanford heart transplant data

In this subsection, we illustrate the methods described in Section 2 with the familiar Stanford heart transplant data. Patients were accepted into the transplant program and then waited until a suitable donor could be found. Some patients were deceased before they received the transplantation. A total of 103 patients had been accepted to the program by the closing date of the study reported by Crowley and Hu (1977).

Crowley and Hu (1977) used Cox regression models with time-dependent covariates to evaluate the effectiveness of heart transplantation on the survival time measured from the date of acceptance. Three prognostic variables were considered in their final model. The first covariate, transplant status, is an indicator variable equal to zero before transplantation and one afterwards. Similarly, two other covariates, age at transplant and mismatch score, are zero before transplant and assume their measured

Table 1
Rergression analyses of the Stanford heart transplant data[a]

Covariate	Cox model[b]		Accelerated life model	
	$\hat{\beta}^c$	W^c	$\hat{\beta}$	G
Transplant status	−1.031	4.56	−1.986	4.85
Age at transplant −35	0.055	5.94	0.096	8.88
Mismatch score −0.5	0.445	2.52	0.930	2.02

[a]The data, kindly made available by Dr. John Crowley, contain a correction of the acceptance date for patient 71 from 08/20/71 in Table 1 of Crowley and Hu (1977) to 08/20/72.
[b]$\hat{\beta}^c$ denotes the maximum partial likelihood estimate, and W^c denotes the maximum partial likelihood statistic.

values after transplant. More specifically, if we let W_i denote the waiting time, from the date of acceptance to the date of transplant, for the ith patient, then

$$Z_{1i}(t)=\begin{cases}0 & \text{if } t<W_i,\\ 1 & \text{if } t\geqslant W_i,\end{cases}$$

$$Z_{2i}(t)=\begin{cases}0 & \text{if } t<W_i,\\ \text{age at transplant } -35 & \text{if } t\geqslant W_i,\end{cases}$$

$$Z_{3i}(t)=\begin{cases}0 & \text{if } t<W_i,\\ \text{mismatch score } -0.5 & \text{if } t\geqslant W_i.\end{cases}$$

In defining Z_2 and Z_3, we subtract age at transplant and mismatch score by their respective 15th sample percentiles so that the regression coefficient associated with Z_1 pertains to the effect of transplantation for a patient with 35 years of age and mismatch score of 0.5. Four patients who were not tissue-typed were excluded from the analysis. Out of the remaining 99 patients, 28 were censored as of the closing date. The results from the Cox regression analysis with the aforementioned covariates are displayed in Table 1. The maximum partial likelihood statistics (i.e., Wald statistics) indicate that transplant status and age at transplant are significant whereas mismatch score is not.

For comparison, let us use model (1.1) to evaluate the effects of the same set of covariates on the survival time. Because all three covariates are step functions, $\tilde{X}_i$ and $\tilde{Z}_i$ take simple forms:

$$\tilde{X}_i(b)=\begin{cases}X_i & \text{if } X_i\leqslant W_i,\\ W_i+(X_i-W_i)\exp\{b'Z_i(W_i)\} & \text{if } X_i>W_i,\end{cases}$$

$$\tilde{Z}_i(x,b)=\begin{cases}0 & \text{if } x<W_i,\\ Z_i(W_i) & \text{if } x\geqslant W_i.\end{cases}$$

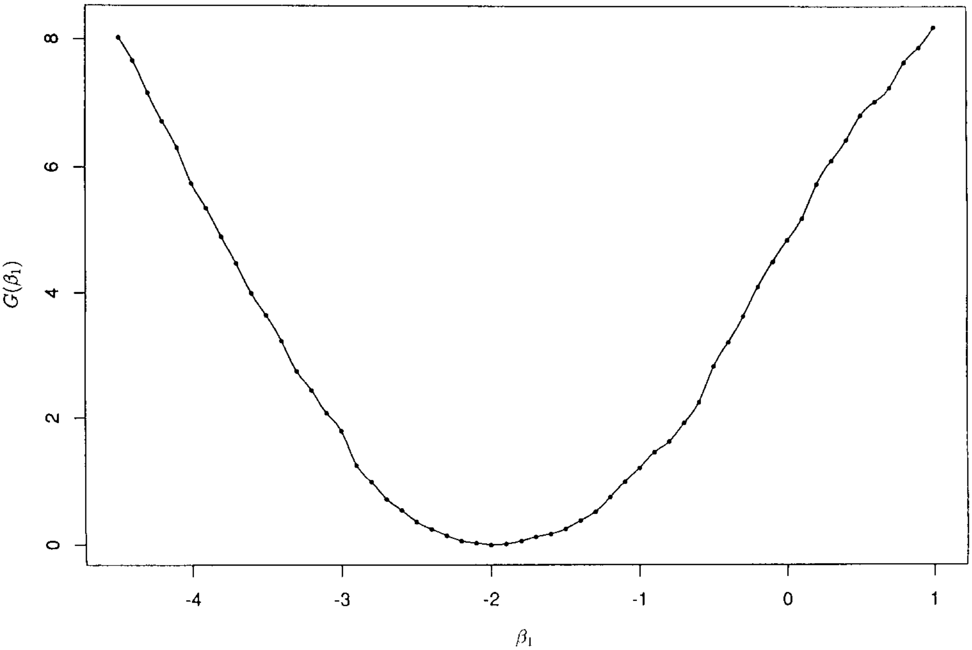

Fig. 1. Plot of $G(\beta_1)$ againist β_1 for the Stanford heart transplant data.

Consequently, calculations of $U(b)$ and $V(b)$ are fairly easy. The main results from the accelerated life model analysis are also shown in Table 1. The proposed estimates nearly doubled the maximum partial likelihood estimates under the Cox model, which is not surprising since the two models are defined on different scales. The minimum dispersion statistics appear to provide stronger evidence for the effects of transplant status and age at transplant than the maximum partial likelihood statistics.

Our results suggest that transplantation is beneficial for younger patients with lower mismatch scores. For a patient transplanted at the age of 35 with mismatch score of 0.5, $e^{\hat{\beta}'Z(W)} = e^{-1.986} \approx 0.137$, implying that this patient would have lived only 13.7% of his post-transplantation life had his heart not been transplanted. In contrast, older patients with higher mismatch scores are unlikely to benefit from transplantation. (For instance, the post-transplantation life time of a patient aged 53 with mismatch score of 1.8 would have been increased by about 160% had the operation not been done.) The foregoing conclusions confirm those of Crowley and Hu (1977) and others who used Cox models. (It is usually comforting to reach similar conclusions with different models.) Evidently, it is easier to visualize that transplantation expands (or shortens) the life time by a certain proportion than that the hazard rate is changed.

Figure 1 plots the values of $G(\beta_1)$ against β_1 for β_1 ranging from -4.5 to 1 0 with grid size 0.1. Confidence intervals for β_1 with various confidence levels can be read from the fitted curve shown in the figure. For example, the 95% confidence interval is found to be approximately $(-3.54, -0.26)$, corresponding to (3%, 77%) for the ratio of the nontransplanted survival time to the transplanted survival time (for a patient aged 35 with mismatch score of 0.5). As Fig. 1 indicates, $G(\cdot)$ is a fairly regular function. Thus, for most practical purposes, it suffices to evaluate G for only a few, say 10, values of the parameter and then obtain the confidence interval from a smooth curve fitted to those points (e.g., via a cubic spline).

3.2. *Simulation studies*

In order to get some ideas about how well the proposed asymptotic procedures behave in finite samples, we conducted a few simulation studies that mimicked the main features of the Stanford heart transplant data. We let the covariate processes be

$$Z_1(t)=\begin{cases}0 & \text{if } t<W,\\ 1 & \text{if } t\geqslant W,\end{cases}$$

$$Z_2(t)=\begin{cases}0 & \text{if } t<W,\\ H & \text{if } t\geqslant W,\end{cases}$$

where W is an exponential variable with 0.5 hazard rate, and H is an independent standard normal variable. The baseline failure time T_0 was generated independently from the standard exponential distribution, and the actual failure time associated with $\{Z_1(\cdot), Z_2(\cdot)\}$ was then determined by

$$T=\begin{cases}T_0 & \text{if } T_0\leqslant W,\\ W+(T_0-W)/\exp(\beta_1+\beta_2 H) & \text{if } T_0>W.\end{cases}$$

The censoring time C was a uniform random variable on $[0,\tau]$, where τ was so chosen that about 30% of the observations were censored. Because model (1.1) and the Cox regression model coincide when the baseline failure time T_0 has the standard exponential distribution, the partial-likelihood-based methods provide natural references for the simulation studies of the proposed statistics.

Table 2 displays the Monte Carlo estimates for the sampling means of the proposed estimator $\hat{\beta}$ and the maximum partial likelihood estimator $\hat{\beta}^c$. The former appears to be less biased than the latter. In small samples, $\hat{\beta}^c$ is unstable and may not be obtainable especially when the effect of a discrete covariate is strong. Related simulation results not shown in Table 2 indicate that $\hat{\beta}$ tends to have less sampling variability than $\hat{\beta}^c$.

The minimum dispersion statistic G and the maximum partial likelihood statistic W^c are compared in Table 3. Both tests maintain their sizes near the nominal level at least when n is not too small. In addition, the G test seems to be slightly more powerful than the W^c test, most notably for $\alpha=0.01$.

Table 2
Empirical means of the proposed estimator $\hat{\beta}$ and the maximum partial likelihood estimator $\hat{\beta}^c$ under the model $T_0=\int_0^T \exp\{\beta_1 Z_1(u)+\beta_2 Z_2(u)\}\,du$[a]

	$\{\beta_1=0, \beta_2=0.5\}$				$\{\beta_1=1, \beta_2=0.5\}$			
n	$\hat{\beta}_1$	$\hat{\beta}_1^c$	$\hat{\beta}_2$	$\hat{\beta}_2^c$	$\hat{\beta}_1$	$\hat{\beta}_1^c$	$\hat{\beta}_2$	$\hat{\beta}_2^c$
20	0.018	−0.148	0.553	0.631	1.034	1.045	0.542	0.600
50	0.0002	−0.034	0.519	0.536	1.005	1.002	0.512	0.527
100	0.005	−0.011	0.505	0.511	1.007	1.004	0.504	0.509

[a]Each entry is based on 2000 replications. The cases that $\hat{\beta}^c$ is not obtainable (19 cases for $n=20$ and 1 case for $n=50$ under $\{\beta_1=1, \beta_2=0.5\}$) are excluded from the calculations for $\hat{\beta}^c$.

Table 3
Empirical sizes/powers of the minimum dispersion statistic G and the maximum partial likelihood statistic W^c for testing H_0: $\beta_1=0$ under the model $T_0=\int_0^T \exp\{\beta_1 Z_1(u)+0.5Z_2(u)\}\,du$[a]

Sample size	Test statistic	True model: $\beta_1=0$			True model: $\beta_1=1$		
		$\alpha=0.01$	0.05	0.1	$\alpha=0.01$	0.05	0.1
20	G	0.014	0.046	0.085	0.160	0.321	0.419
	W^c	0.004	0.037	0.090	0.122	0.319	0.439
50	G	0.011	0.048	0.092	0.556	0.735	0.824
	W^c	0.009	0.045	0.095	0.488	0.720	0.814
100	G	0.013	0.056	0.101	0.898	0.962	0.978
	W^c	0.013	0.057	0.101	0.860	0.956	0.973

[a]Each entry is based on 2000 replications. The cases that $\hat{\beta}^c$ is not obtainable are excluded from the calculations for W^c.

4. Remarks

Model (1.1) is one special, perhaps the most important, form of accelerated life models with time-varying covariates. In some applications, alternative formulations may be more appropriate. The techniques presented in this paper can be used to derive asymptotic theories for other regression forms in similar fashions. Furthermore, the main conclusions of Section 2 would still hold if $Q_i(\cdot,\beta)$ in (2.2) were replaced by processes other than $\tilde{Z}_i(\cdot,\beta)$ that are also independent of T_0.

An application of the Schwarz inequality to the variance expression for $n^{1/2}(\hat{\beta}-\beta)$ indicates that the asymptotic variance for the parameter estimator will be minimized if the unknown hazard function $\lambda(\cdot)$ in $Q_i(\cdot,\beta)$ is correctly specified. In fact, such an estimator achieves the semiparametric efficiency bound (Begun et al., 1983), as was pointed out by Robins and Tsiatis (1992). It is therefore desirable to derive estimating functions that adaptively estimate $\lambda'(\cdot)/\lambda(\cdot)$. The sample-splitting technique of

Lai and Ying (1991) seems a promising approach, but its practically requires further investigations.

Model (1.1) provides a useful alternative to the Cox regression model and has a more straightforward interpretation since it deals with the failure time directly. These two classes of models intersect when the baseline failure time T_0 has a Weibull distribution. Even though the Cox model is the current method of choice in survival analysis, the partial likelihood inference may be criticized on the basis of robustness and efficiency (Kalbfleisch and Prentice, 1980, p. 144). In most applications, the choice between the Cox model and the accelerated life model will be an empirical matter, depending on which one fits the data better. We are currently developing techniques for discriminating between the two models.

The chief stumbling block in the practical use of the accelerated life model was the lack of efficient computing algorithms. Recent work by Lin and Geyer (1992) produced some useful numerical solutions. Using their annealing algorithm, we were able to conduct large simulations (e.g., sample size of 100 with 2000 replications) rather quickly. The results for the accelerated life model shown in Table 1 only consumed a few CPU minutes on a SPARC 4 station.

One must exercise care when utilizing time-dependent covariates. Kalbfleisch and Prentice (1980, Section 5.3) made a distinction between external and internal covariates. They cautioned that only the former are easy to interpret, though the latter can be useful in some applications (e.g., the surrogate marker problem). Robins and Tsiatis (1992) assumed that T_0 (denoted by U in their paper) is independent of Z and illustrated such independence using a conceptual example of an external covariate.

When using model (1.1) (or the Cox model) to compare the survival experience of transplanted and nontransplanted patients, it is essential that there be no selection bias in the assignment of hearts to individuals. Since assignments were not made at random in the Stanford heart transplant study, the possibility of selection bias exists and the results shown in Section 3.1 should be viewed with some caution. (This is always a problem in observational studies.) Recently, Robins (1992) discussed the estimation of the causal effect of a time-dependent treatment or exposure on the failure time in the presence of time-varying confounders.

Appendix

Proof of Lemma 1. For notational simplicity, we shall take $p=1$. Furthermore, we shall prove the three types of approximations for Y^Z only because the same kinds of arguments can be used for the other five processes. The corresponding approximations for time-invariant covariates were developed in Lai and Ying (1988). A key idea there was to use the monotonicity of $Y^Z(x,b)$ (in x and b) to get certain sandwich inequalities. For time-varying covariates, this is no longer true, and in fact the behavior of $Y^Z(x,b)$ is far more complicated. Our approach is to decompose Z

according to the Jordan decomposition so that similar sandwich inequalities can be established. Since Z_i $(i=1,\ldots,n)$ are uniformly bounded under Condition (C1) and since $U(b)$ is location-shift invariant, we may, without loss of generality, assume that the Z_i are positive. When the Z_{ik} are nonnegative, the $h_i(t,b)$ are increasing in t and b. Therefore, the $Y_i(x,b)$ are increasing in b and decreasing in x. By considering the Z_{ik}^+ and Z_{ik}^- separately, we may, without loss of generality, assume that the Z_{ik} are monotonically increasing.

Choose $x_k=k/n^{\kappa_0}$, $k=0,1,\ldots,n_x$ and $b_j=B_j/n^{\kappa_0+1}$, $j=0,\pm 1,\ldots,\pm n_b$, where n_x and n_b are the smallest integers greater than $n^{\kappa_0+\theta_0^{-1}+1}$ and n^{κ_0+2}, respectively. Observe that the $\tilde{Z}_i(x,b)$ are nonnegative and increasing in x and decreasing in b since the $h_i^{-1}(x,b)$ are increasing in x and decreasing in b. Thus, for $x\in[x_{k-1},x_k)$ and $b\in[b_{j-1},b_j)$,

$$\begin{aligned} Y^Z(x,b)-\mathscr{E}Y^Z(x,b)\leqslant{}& \sum_{i=1}^n \tilde{Z}_i(x_{k-1},b_j)\{Y_i(x_{k-1},b_j)-\mathscr{E}Y_i(x_{k-1},b_j)\} \\ &+\sum_{i=1}^n\{\tilde{Z}_i(x_k,b_{j-1})-\tilde{Z}_i(x_{k-1},b_j)\}\{Y_i(x_{k-1},b_j) \\ &-\mathscr{E}Y_i(x_{k-1},b_j)\}+\sum_{i=1}^n\{\tilde{Z}_i(x_k,b_{j-1})\mathscr{E}Y_i(x_{k-1},b_j) \\ &-\tilde{Z}_i(x_{k-1},b_j)\mathscr{E}Y_i(x_k,b_{j-1})\}, \end{aligned} \tag{A.1}$$

$$\begin{aligned} Y^Z(x,b)-\mathscr{E}Y^Z(x,b)\geqslant{}& \sum_{i=1}^n \tilde{Z}_i(x_k,b_{j-1})\{Y_i(x_k,b_{j-1})-\mathscr{E}Y_i(x_k,b_{j-1})\} \\ &+\sum_{i=1}^n\{\tilde{Z}_i(x_{k-1},b_j)-\tilde{Z}_i(x_k,b_{j-1})\}\{Y_i(x_k,b_{j-1}) \\ &-\mathscr{E}Y_i(x_k,b_{j-1})\}+\sum_{i=1}^n\{\tilde{Z}_i(x_{k-1},b_j)\mathscr{E}Y_i(x_k,b_{j-1}) \\ &-\tilde{Z}_i(x_k,b_{j-1})\mathscr{E}Y_i(x_{k-1},b_j)\}. \end{aligned} \tag{A.2}$$

Given assumptions (C2)–(C4), the last two terms in (A.1) and (A.2) are of the order $O(n^{(1/2)-\eta})$ for some $\eta>0$, uniformly in j and k. Therefore, it suffices to show that, with probability 1,

$$\sup_{j,k}\left|\sum_{i=1}^n \tilde{Z}_i(x_k,b_j)\{Y_i(x_k,b_j)-\mathscr{E}Y_i(x_k,b_j)\}\right|=o(n^{(1/2)+\varepsilon}), \tag{A.3}$$

$$\sup_{\substack{j,k \\ \mathscr{E}Y(x_k,b_j)\leqslant n^{1-\gamma}}}\left|\sum_{i=1}^n \tilde{Z}_i(x_k,b_j)\{Y_i(x_k,b_j)-\mathscr{E}Y_i(x_k,b_j)\}\right|=o(n^{(1/2)-\eta}), \tag{A.4}$$

$$\begin{aligned} &\sup_{j,k,|b_j-\beta|\leqslant n^{-\gamma}}\left|\sum_{i=1}^n \tilde{Z}_i(x_k,b_j)\{Y_i(x_k,b_j)-\mathscr{E}Y_i(x_k,b_j)-Y_i(x_k,\beta)+\mathscr{E}Y_i(x_k,\beta)\}\right| \\ &=o(n^{(1/2)-\eta}). \end{aligned} \tag{A.5}$$

In order to prove (A.3), we apply Bennett's inequality (Shorack and Wellner, 1986, pp. 851–852) to obtain

$$P\left[\left|\sum_{i=1}^{n} \tilde{Z}_i(x_k,b_j)\{Y_i(x_k,b_j)-\mathscr{E}\,Y_i(x_k,b_j)\}\right|\geqslant n^{(1/2)+\varepsilon/2}\right]\leqslant K\exp(-n^{\theta})$$

for some $\theta>0$, $K>0$ and all k and j. Since there are only polynomially many combinations of k and j, the forgoing exponential inequality implies (A.3). Moreover, on $\mathscr{E}\,Y(x_k,b_j)\leqslant n^{1-\gamma}$,

$$\mathrm{Var}\left[\sum_{i=1}^{n} \tilde{Z}_i(x_k,b_j)\{Y_i(x_k,b_j)-\mathscr{E}\,Y_i(x_k,b_j)\}\right]\leqslant D^2 n^{1-\gamma},$$

which coupled with Bennett's inequality implies (A.4) with $\eta<\gamma/2$. A similar use of Bennett's inequality results in (A.5). □

Proof of Theorem 1. Our approach is to use Lemma 1 and the arguments detailed in the proof of Theorem 1 of Ying (1993). From Lemma 1(a), with probability 1,

$$\sup_{\|b\|\leqslant B}\|N^Z(\infty,b)-\mathscr{E}N^Z(\infty,b)\|=o(n^{(1/2)+\varepsilon}). \tag{A.6}$$

In addition, Lemma 1(a), together with the integration by parts formula and the fact that the total variation

$$\sup_{\|b\|\leqslant B}\int_0^\infty\left\|\frac{dY^Z(x,b)}{Y^Z(x,b)}\right\|=O(\log n),$$

implies that with probability 1

$$\sup_{\|b\|\leqslant B}\left\|\int_0^\infty\frac{Y^Z(x,b)}{Y(x,b)}d[N(x,b)-\mathscr{E}N(x,b)]\right\|=o(n^{(1/2)+\varepsilon}). \tag{A.7}$$

Likewise, we can show that with probability 1

$$\sup_{\|b\|\leqslant B}\left\|\int_0^\infty\left\{\frac{Y^Z(x,b)}{Y(x,b)}-\frac{\mathscr{E}\,Y^Z(x,b)}{\mathscr{E}\,Y(x,b)}\right\}\mathrm{d}\mathscr{E}N(x,b)\right\|=o(n^{(1/2)+\varepsilon}). \tag{A.8}$$

From (A.6)–(A.8), we get part (a) of Theorem 1.

Because the proof of part (b) is rather involved, we shall only sketch its main steps here. Let $t_b(\alpha)=\inf\{x\colon \mathscr{E}\,Y(x,b)\leqslant n^{1-\alpha}\}$ for $\alpha\in(0,1)$. Analogous to (2.6) of Ying (1993), we can use Lemma 1(b) to show that with probability 1

$$\begin{aligned}\sup_{\|b-\beta\|\leqslant n^{-1/3},\,t\geqslant t_\beta(\alpha)}&\left\|\sum_{i=1}^{n}\int_{t_\beta(\alpha)}^{t}\left\{\tilde{Z}_i(x,b)-\frac{Y^Z(x,b)}{Y(x,b)}\right\}\mathrm{d}N_i(x,b)\right.\\ &\left.-\sum_{i=1}^{n}\int_{t_\beta(\alpha)}^{t}\left\{\tilde{Z}_i(x,b)-\frac{\mathscr{E}\,Y^Z(x,b)}{\mathscr{E}\,Y(x,b)}\right\}\mathrm{d}\mathscr{E}N_i(x,b)\right\|\\ &=o(n^{(1/2)-\gamma_1})\quad\text{for some }\gamma_1>0.\end{aligned} \tag{A.9}$$

Let $\alpha \in (0, \frac{1}{6})$ be fixed. As in the proof Lemma 3 of Ying (1993), we can use Lemma 1 to show that

$$\sup_{\|b-\beta\| \leqslant n^{-1/3}, t \leqslant t_\beta(\alpha)} \left\| \sum_{i=1}^{n} \int_0^t \left\{ \tilde{Z}_i(x,b) - \frac{Y^Z(x,b)}{Y(x,b)} \right\} dN_i(x,b) \right.$$
$$- \sum_{i=1}^{n} \int_0^t \left\{ \tilde{Z}_i(x,b) - \frac{\mathscr{E} Y^Z(x,b)}{\mathscr{E} Y(x,b)} \right\} d\mathscr{E} N_i(x,b)$$
$$\left. - \sum_{i=1}^{n} \int_0^t \left\{ \tilde{Z}_i(x,\beta) - \frac{Y^Z(x,\beta)}{Y(x,\beta)} \right\} dN_i(x,\beta) \right\|$$
$$= o(n^{(1/2)-\gamma_2}) \quad \text{for some } \gamma_2 > 0 \qquad \text{(A.10)}$$

by noting that $\sum_{i=1}^n \int_0^t \{\tilde{Z}_i(x,\beta) - \mathscr{E} Y^Z(x,\beta)/\mathscr{E} Y(x,\beta)\} \, d\mathscr{E} N_i(x,\beta) = 0$. It follows from (A.9) and (A.10) that, for some $\theta > 0$,

$$\sup_{\|b-\beta\| \leqslant n^{-1/3}} \| U(b) - \bar{U}(b) - U(\beta) \| = o(n^{(1/2)-\theta}) \quad \text{almost surely.} \qquad \text{(A.11)}$$

From (A.11) and Theorem 1(a), we have

$$\sup_{\|b-\beta\| \leqslant d_n} \| U(b) - \bar{U}(b) - U(\beta) \| / (n \|b-\beta\| + n^{1/2})$$
$$\leqslant \sup_{\|b-\beta\| \leqslant n^{-1/3}} \{ n^{-1/2} \| U(b) - \bar{U}(b) - U(\beta) \| \}$$
$$+ 2 \sup_{n^{-1/3} \leqslant \|b-\beta\|, \|b\| \leqslant B} \{ n^{-2/3} \| U(b) - \bar{U}(b) \| \}$$
$$\to 0 \quad \text{almost surely.} \qquad \text{(A.12)}$$

Moreover, analogous to Lemma 5 of Ying (1993),

$$\bar{U}(b) = n \{ A_n(b-\beta) + r_n(b) \}, \qquad \text{(A.13)}$$

where $r_n(b)$ satisfies

$$\sup_{\|b-\beta\| \leqslant d_n} \frac{\| r_n(b) \|}{\| b-\beta \| + n^{-1/2}} \to 0.$$

Combining (A.12) and (A.13), we obtain part (b) of the theorem. □

Acknowledgements

The authors are grateful to Raymond Carroll, an associate editor and two referees for their helpful comments and suggestions, and to John Crowley for providing an electronic file of the Stanford heart transplant data. This work was supported by the U.S. National Institutes of Health (for Lin) and the National Science Foundation (for Ying).

References

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

Begun, J.M., W.J. Hall, Huang, W.-M. and J.A. Wellner (1983). Information and asymptotic efficiency in parametric-nonparametric models. *Ann. Statist.* **11**, 432–452.

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

Cox, D.R. and D. Oakes (1984). *Analysis of Survival Data.* Chapman & Hall, London.

Crowley, J. and M. Hu (1977). Covariance analysis of heart transplant survival data. *J. Amer. Statist. Assoc.* **72**, 27–36.

Cuzick, J. (1985). Asymptotic properties of censored linear rank tests. *Ann. Statist.* **13**, 133–141.

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

Kirkpatrick, S., C.D. Gelatt and M.P. Vecchi (1983). Optimization by simulated annealing. *Science* **220**, 671–680.

Lai, T.L. and Z. Ying (1988). Stochastic integrals of empirical-type processes with applications to censored regression. *J. Multivariate Anal.* **27**, 334–358.

Lai, T.L. and Z. Ying (1991). Rank regression methods for left-truncated and right-censored data. *Ann. Statist.* **19**, 531–556.

Lin, D.Y. and C.J. Geyer (1992). Computational methods for semiparametric linear regression with censored data. *J. Comput. and Graph. Statist.* **1**, 77–90.

Prentice, R.L. (1978). Linear rank tests with right censored data. *Biometrika* **65**, 167–179.

Ritov, Y. (1990). Estimation in a linear regression model with censored data. *Ann. Statist.* **18**, 303–28.

Robins, J.M. (1992). Estimation of the time-dependent accelerated failure time model in the presence of confounding factors. *Biometrika* **79**, 313–334.

Robins, J.M. and A.A. Tsiatis (1992). Semiparametric estimation of an accelerated failure time model with time-dependent covariates. *Biometrika* **79**, 311–319.

Rudin, W. (1974). *Real and Complex Analysis*, 2nd ed. McGraw-Hill, New York.

Shorack, G.R. and J.A. Wellner (1986). *Empirical Processes with Applications to Statistics.* Wiley, New York.

Tsiatis, A.A. (1990). Estimating regression parameters using linear rank tests for censored data. *Ann. Statist.* **18**, 354–72.

Wei, L.J., Z. Ying and D.Y. Lin (1990). Linear regression analysis of censored survival data based on rank tests. *Biometrika* **77**, 845–51.

Ying, Z. (1993). A large sample study of rank estimation for censored regression data. *Ann. Statist.* **21**, 76–99.