Statistica Sinica **16**(2006), 919-932

THE ADDITIVE HAZARDS MODEL FOR RECURRENT GAP TIMES

Liuquan Sun[1], Do-Hwan Park[2] and Jianguo Sun[2]

[1] *Chinese Academy of Sciences and* [2] *University of Missouri*

Abstract: Recurrent event data and gap times between recurrent events are often the targets in the analysis of longitudinal follow-up or epidemiological studies. To analyze the gap times, Huang and Chen (2003), among others, proposed to fit the proportional hazards model. It is well-known, however, that the proportional hazards model might not fit the data well. To provide an alternative, this paper investigates the fit of the additive hazards model to gap time data, and an estimating equation approach is presented for inference about regression parameters. Both asymptotic and finite sample properties of the proposed parameter estimates are established. One major advantage of the use of the additive hazards model over the proportional hazards model is that the resulting parameter estimator has a closed form. The method is applied to a cancer study.

Key words and phrases: Additive hazards model, estimating equations, gap time, recurrent event data, regression analysis.

1. Introduction

Consider a longitudinal study that involves n independent subjects, each of which experiences recurrences of the same event (Chang and Wang (1999), Cook, Lawless and Nadeau (1996) and Prentice, Williams and Peterson (1981)). Here the event could be, for example, the occurrence of a certain disease, of hospitalization, or of a tumor. Suppose that one is interested in the gap times between recurrent events, and in making inference about effects of covariates such as age and treatment on the gap times. For subject i, let T_{ij} denote the time from the $(j-1)$th occurrence of the event to the jth occurrence of the event, $j = 1, 2, \ldots$, $i = 1, \ldots, n$. That is, $T_{i1} + \cdots + T_{ij}$ is the time at which the event occurs for the jth time. Also let Z_i denote the vector of time-independent covariates associated with subject i, and C_i the follow-up or censoring time.

Define $N_i = \{T_{ij} : j = 1, 2, \ldots\}$, and assume that $\{(N_i, Z_i, C_i); i = 1, \ldots n\}$ are n independent and identically distributed (i.i.d.) replicates of (N, Z, C). Also assume that N_i is independent of C_i given Z_i. Define M_i to be the number of observed gap times for subject i, the integer satisfying

$$\sum_{j=1}^{M_i-1} T_{ij} \le C_i \quad \text{and} \quad \sum_{j=1}^{M_i} T_{ij} > C_i .$$

Then observed data are $\{T_{i1}, \ldots, T_{i,M_i-1}, C_i, Z_i\}$. That is, the first $M_i - 1$ gap times are observed, but T_{i,M_i} is censored at

$$T^+_{i,M_i} = C_i - \sum_{j=1}^{M_i-1} T_{ij} \, .$$

A number of authors have discussed the analysis of recurrent event data in terms of the recurrence intensity or rate of the event (Cook and Lawless (1996), Cook et al. (1996), Lin et al. (2000), Wang and Chen (2000) and Wang, Qin and Chiang (2001)). In contrast, there exists limited research on recurrent gap times. Huang and Chen (2003) and Schaubel and Cai (2004) discussed regression analysis of recurrent gap times under the proportional hazards model. It is well-known, however, that the proportional hazards model might not fit the data well. To address this, in the following, we investigate the additive hazards model defined as

$$\lambda(t|Z_i) = \lambda_0(t) \, + \, \beta_0' Z_i \tag{1}$$

for T_{ij} given Z_i (Lin and Ying (1994)). In the above, β_0 denotes the vector of unknown regression parameters and $\lambda_0(t)$ is an unspecified baseline hazard function.

The additive hazards model describes a different aspect of the relationship between survival time and covariates than does the proportional hazards model. For the two-sample situation, for example, the additive hazards model addresses the risk difference, while the proportional hazards model concerns the risk ratio. Both models have sound biological and empirical bases, and which model should be used for a particular situation usually depends on factors such as the background, the quantity of interest and the availability of proper inference procedures. In tumorigenicity experiments that investigate the dose effect on tumor risk, for example, an additive hazards model may be more reasonable or often preferred since the excess risk is often the quantity of interest (Breslow and Day (1987, Chap.6)).

In the next section, for inference about regression parameters and the cumulative hazard function, an estimating equation approach is presented. It will be seen that one major advantage of the use of the additive hazards model over the proportional hazards model is that the resulting regression parameter estimator has a closed form. In addition, the asymptotic properties of the proposed estimators are established. Section 3 reports some results from a simulation study and Section 4 applies the proposed method to a bladder tumor study. Some concluding remarks are given in Section 5.

2. Statistical Methods

This section considers inference procedures for β_0 and the baseline cumulative hazard function $\Lambda_0(t) = \int_0^t \lambda_0(u)du$. For the purpose, as in Huang and Chen (2003), we assume that each individual recurrent event process is a renewal process. That is, for a given i, $\{T_{ij},\, j = 1, 2\ldots\}$ are i.i.d. For each i, define $\Delta_i = I(M_i > 1)$, $M_i^* = \max(M_i - 1, 1)$ and

$$X_{ij} = \begin{cases} T_{ij} & \text{if } \Delta_i = 1\,, \\ T_{ij}^{+} & \text{if } \Delta_i = 0\,, \end{cases}$$

$j = 1, \ldots, M_i^*$.

To present the inference procedure, note that given C_i, M_i and T^{+}_{i,M_i}, the observed complete gap times $\{T_{ij},\, j = 1, \ldots, M_i - 1\}$ are identically distributed (Huang and Chen (2003)). Assume that $\{(X_{ij},\, j = 1, \ldots, M_i^*, \Delta_i, Z_i);\, i = 1, \ldots, n\}$ are n i.i.d. replicates of $\{X_{(j)},\, j = 1, \ldots, M^*, \Delta, Z\}$. Since the first gap time is subject to independent censorship, the exchangeability of observed complete gap times then suggests that the subset $\{(X_{ij},\, j = 1, \ldots, M_i^*, \Delta_i, Z_i);\, i = 1, \ldots, n\}$ can be treated as clustered survival data. Of course the cluster size is informative, and the censored gap time needs to be removed for $M_i > 1$. Using this fact and following the ideas used in Lin and Ying (1994) and Huang and Chen (2003) we propose to use the estimating equation $U(\beta) = 0$ to estimate regression parameters β_0, where

$$U(\beta) = \int_0^{\tau} Q(t) \left[\hat{\mathcal{E}}_{ij}\{Z_i \Delta_i dI(X_{ij} \le t)\} - \frac{\hat{\mathcal{E}}_{ij}\{Z_i I(X_{ij} \ge t)\}}{\hat{\mathcal{E}}_{ij}\{I(X_{ij} \ge t)\}} d\hat{\mathcal{E}}_{ij}\{\Delta_i I(X_{ij} \le t)\} \right.$$
$$\left. - \left(\hat{\mathcal{E}}_{ij}\{Z_i^{\otimes 2} I(X_{ij} \ge t)\} - \frac{(\hat{\mathcal{E}}_{ij}\{Z_i I(X_{ij} \ge t)\})^{\otimes 2}}{\hat{\mathcal{E}}_{ij}\{I(X_{ij} \ge t)\}} \right) \beta dt \right].$$

In the above, $Q(t)$ is a weight process that may depend on data (see conditions given below), τ $(0 < \tau < \infty)$ is a prespecified constant such that $P(X_{(1)} \ge \tau) > 0$, $v^{\otimes 2} = v'v$ for a column vector v, $\hat{\mathcal{E}}_{ij} = \hat{\mathcal{E}}_i \hat{\mathcal{E}}_j$, where $\hat{\mathcal{E}}_i$ and $\hat{\mathcal{E}}_j$ denote empirical averages over $i = 1, \ldots, n$ and $j = 1, \ldots, M_i^*$, respectively. In practice, τ is usually taken as the longest follow-up time. In general, the choice of $Q(t)$ gives a class of estimates and in the numerical study and the example below, $Q(t) = 1$ is used; some comments follow.

Note that an alternative, but less efficient approach to the above method is to base the inference about β_0 on $\{(X_{i1}, \Delta_i, Z_i);\ i = 1, \ldots, n\}$, the data only on the time to the first occurrence of the event. In this case, $U(\beta)$ has the form

$$U_1(\beta) = \int_0^\tau Q(t) \Bigg[\hat{\mathcal{E}}_i\{Z_i \Delta_i dI(X_{i1} \le t)\} - \frac{\hat{\mathcal{E}}_i\{Z_i I(X_{i1} \ge t)\}}{\hat{\mathcal{E}}_i\{I(X_{i1} \ge t)\}} d\hat{\mathcal{E}}_i\{\Delta_i I(X_{i1} \le t)\} - \left(\hat{\mathcal{E}}_i\{Z_i^{\otimes 2} I(X_{i1} \ge t)\} - \frac{(\hat{\mathcal{E}}_i\{Z_i I(X_{i1} \ge t)\})^{\otimes 2}}{\hat{\mathcal{E}}_i\{I(X_{i1} \ge t)\}} \right) \beta dt \Bigg].$$

Define $\hat{K}(t) = \hat{\mathcal{E}}_{ij}\{\Delta_i I(X_{ij} \le t)\}$, $\hat{G}_0(t) = \hat{\mathcal{E}}_{ij}\{I(X_{ij} \ge t)\}$ and $\hat{G}_1(t) = \hat{\mathcal{E}}_{ij}\{Z_i I(X_{ij} \ge t)\}$. Let $\hat{\beta}$ denote the solution to $U(\beta) = 0$. Then

$$\hat{\beta} = \left[\int_0^\tau Q(t) \left(\hat{\mathcal{E}}_{ij}\{Z_i^{\otimes 2} I(X_{ij} \ge t)\} - \frac{(\hat{\mathcal{E}}_{ij}\{Z_i I(X_{ij} \ge t)\})^{\otimes 2}}{\hat{\mathcal{E}}_{ij}\{I(X_{ij} \ge t)\}} \right) dt \right]^{-1} \times \left[\int_0^\tau Q(t) \left(\hat{\mathcal{E}}_{ij}\{Z_i \Delta_i dI(X_{ij} \le t)\} - \frac{\hat{\mathcal{E}}_{ij}\{Z_i I(X_{ij} \ge t)\}}{\hat{\mathcal{E}}_{ij}\{I(X_{ij} \ge t)\}} d\hat{\mathcal{E}}_{ij}\{\Delta_i I(X_{ij} \le t)\} \right) \right].$$

It can be easily shown that $\hat{\beta}$ is a consistent estimator for β_0. For the asymptotic distribution of $\hat{\beta}$, it can be first shown that $n^{1/2} U(\beta_0)$ is asymptotically normally distributed with mean zero and covariance matrix that can be consistently estimated by $\hat{\Sigma} = \hat{\mathcal{E}}_i[(\hat{\mathcal{E}}_j\{\hat{\phi}(X_{ij}, \Delta_i, Z_i)\})^{\otimes 2}]$, where

$$\hat{\phi}(X_{ij}, \Delta_i, Z_i) = \int_0^\tau Q(t) \left(Z_i - \frac{\hat{G}_1(t)}{\hat{G}_0(t)\}} \right) \Bigg[\Delta_i dI(X_{ij} \le t) - \frac{I(X_{ij} \ge t)}{\hat{G}_0(t)} d\hat{K}(t) - I(X_{ij} \ge t) \hat{\beta}' \left(Z_i - \frac{\hat{G}_1(t)}{\hat{G}_0(t)} \right) dt \Bigg].$$

Then it follows from the Taylor series expansion of $U(\beta)$ that $n^{1/2}(\hat{\beta} - \beta_0)$ has an asymptotic normal distribution with zero mean and covariance matrix that can be consistently estimated by $\hat{\Omega}_F = \hat{A}^{-1} \hat{\Sigma} \hat{A}^{-1}$, where

$$\hat{A} = \int_0^\tau Q(t) \left[\hat{\mathcal{E}}_{ij}\{Z_i^{\otimes 2} I(X_{ij} \ge t)\} - \frac{\hat{G}_1(t)^{\otimes 2}}{\hat{G}_0(t)} \right] dt.$$

The proofs of the above results are sketched in Appendix A. Note that the variance estimator $\hat{\Omega}_F$ is model-free. Alternatively the asymptotic variance of $n^{1/2}(\hat{\beta} - \beta_0)$ can be estimated by $\hat{\Omega}_B = \hat{A}^{-1}(\hat{B}_1 - \hat{B}_2)\hat{A}^{-1}$, where

$$\hat{B}_1 = \hat{\mathcal{E}}_{ij} \left[\int_0^\tau Q(t) \Big(Z_i - \frac{\hat{G}_1(t)}{\hat{G}_0(t)} \Big)^{\otimes 2} \Delta_i dI(X_{ij} \le t) \right],$$

$$\hat{B}_2 = \hat{\mathcal{E}}_i[\hat{\mathcal{E}}_j\{\hat{\phi}(X_{ij}, \Delta_i, Z_i) - \hat{\mathcal{E}}_j\hat{\phi}(X_{ij}, \Delta_i, Z_i)\}^{\otimes 2}]\,.$$

In contrast to $\hat{\Omega}_F$, the validity of $\hat{\Omega}_B$ depends on (1). Let $\hat{\beta}_1$ denote the solution to $U_1(\beta) = 0$. It is shown in Appendix B that the asymptotic variance of $\hat{\beta}$ is smaller than that of $\hat{\beta}_1$. That is, $\hat{\beta}$ is more efficient than $\hat{\beta}_1$.

Sometimes one may be also interested in estimation of the baseline cumulative hazard function $\Lambda_0(t) = \int_0^t \lambda_0(u)du$. Following the same idea, and using the connection between gap data and clustered data discussed above, we propose the estimator

$$\hat{\Lambda}_0(t;\hat{\beta}) = \int_0^t \frac{d\hat{\mathcal{E}}_{ij}\{\Delta_i I(X_{ij} \le u)\} - \hat{\mathcal{E}}_{ij}\{I(X_{ij} \le u)\hat{\beta}' Z_i\}du}{\hat{\mathcal{E}}_{ij}\{I(X_{ij} \ge u)\}}$$

for $\Lambda_0(t)$. Note that this estimator may not always be monotone in t and, in this case, simple modifications as those discussed in Lin and Ying (1994) can be made to ensure monotonicity while preserving asymptotic properties. If one uses only the data about the time to the first occurrence of the event, the above estimator reduces to

$$\begin{aligned}\hat{\Lambda}_0^{(1)}(t;\hat{\beta}_1) &= \int_0^t \frac{\sum_{i=1}^n [\Delta_i dI(X_{i1} \le u) - I(X_{i1} \ge u)\hat{\beta}_1' Z_i du]}{\sum_{k=1}^n I(X_{k1} \ge u)} \\ &= \int_0^t \frac{d\hat{\mathcal{E}}_i\{\Delta_i I(X_{i1} \le u)\} - \hat{\mathcal{E}}_i\{I(X_{i1} \le u)\hat{\beta}_1' Z_i\}du}{\hat{\mathcal{E}}_i\{I(X_{i1} \ge u)\}}\,.\end{aligned}$$

As for $\hat{\beta}$, it can be shown that $\hat{\Lambda}_0(t;\hat{\beta})$ is consistent or, more specifically, $\sup_{0\le t\le\tau} |\hat{\Lambda}_0(t;\hat{\beta}) - \Lambda_0(t)| \to 0$ almost surely. Furthermore, we can show that $n^{1/2}(\hat{\Lambda}_0(t;\hat{\beta}) - \Lambda_0(t))$ converges weakly to a zero-mean Gaussian process whose covariance function at (s,t) can be consistently estimated by

$$\hat{\Gamma}(s,t) = \hat{\mathcal{E}}_i[\hat{\mathcal{E}}_j\{\hat{\psi}(s;X_{ij}, \Delta_i, Z_i)\}\hat{\mathcal{E}}_j\{\hat{\psi}(t;X_{ij}, \Delta_i, Z_i)\}]\,, \tag{2}$$

where

$$\begin{aligned}\hat{\psi}(t;X_{ij}, \Delta_i, Z_i) = \int_0^t \Bigg[&\frac{\Delta_i dI(X_{ij} \le u) - I(X_{ij} \ge u)\hat{\beta}' Z_i du}{\hat{G}_0(u)} \\ &- \frac{I(X_{ij} \ge u)d\hat{K}(u) - I(X_{ij} \ge u)\hat{\beta}'\hat{G}_1(u)du}{\hat{G}_0(u)^2}\Bigg] \\ &- \hat{C}'(t)\hat{A}^{-1}\hat{\phi}(X_{ij}, \Delta_i, Z_i)\end{aligned}$$

and $\hat{C}(t) = \int_0^t \hat{G}_1(u)du/\hat{G}_0(u)$. The sketch of the proof is given in Appendix C. Also similar to $\hat{\beta}$, $\hat{\Lambda}(t;\hat{\beta})$ is generally more efficient than $\hat{\Lambda}_0^{(1)}(t;\hat{\beta}_1)$.

Given $\hat{\Lambda}(t;\hat{\beta})$, one may also be interested in constructing confidence bands for $\Lambda_0(t)$. It can be seen by checking $\hat{\Gamma}(s,t)$ that the limiting Gaussian process of $\hat{\Lambda}(t;\hat{\beta})$ does not have independent increments, which makes the construction difficult. Corresponding to this, we propose to use the simulation approach discussed in Lin, Fleming and Wei (1994). The basic idea of the approach is to approximate $\hat{\mathcal{E}}_j\{\hat{\psi}(t;X_{ij},\Delta_i,Z_i)\}$ by $R_i\hat{\mathcal{E}}_j\{\hat{\psi}(t;X_{ij},\Delta_i,Z_i)\}$, where $\{R_i\,,\,i=1,\ldots,n\}$ are i.i.d. standard normal random variables independent of the observed data. Thus the construction of the simultaneous confidence bands for $\Lambda_0(t)$ can be carried out by repeatedly generating normal random samples $\{R_i\,,\,i=1,\ldots,n\}$ given the observed data. Note that since $\Lambda_0(t)$ is non-negative, one may first want to construct confidence bands for $\log\{\Lambda_0(t)\}$ using the above approach and then to transform them back to the confidence bands for $\Lambda_0(t)$. Specifically, define

$$\hat{W}(t)=\frac{n^{-\frac{1}{2}}\sum_{i=1}^{n}R_i\hat{\mathcal{E}}_j\{\hat{\psi}(t;X_{ij},\Delta_i,Z_i)\}}{\hat{\Lambda}_0(t;\hat{\beta})}\,.$$

Then by the functional delta-method, one can show that $n^{1/2}(\log\{\hat{\Lambda}_0(t;\hat{\beta})\}-\log\{\Lambda_0(t)\})$ is asymptotically equivalent to $\hat{W}(t)$. Let q_α be the boundary value given by $P\{\max_{t_1\le t\le t_2}|\hat{W}(t)|>q_\alpha\}=\alpha$ based on simulated $\hat{W}(t)$ for given observed data, where $0\le t_1\le t_2\le\tau$. Then an approximate $1-\alpha$ confidence band for $\Lambda_0(t)$ over $[t_1,t_2]$ is given by

$$[\hat{\Lambda}_0(t;\hat{\beta})\exp(-n^{-\frac{1}{2}}q_\alpha)\,,\,\hat{\Lambda}_0(t;\hat{\beta})\exp(n^{-\frac{1}{2}}q_\alpha)]\,.$$

3. Simulation Results

This section reports some of the simulation results obtained from a study conducted for investigating the finite sample performance of the statistical methods proposed in the previous section, with focus on covariate effects. To generate gap times, heterogeneous mixture renewal processes were used with the additive hazards model defined in (1). Specifically, in the simulation, the baseline gap time $T_{ij}^{(0)}$ was assumed to follow the standard exponential distribution, and given by $-\ln\{1-\Phi(A_i+B_{ij})\}$ where A_i and B_{ij} are independent normal random variables with mean zeros and variances ρ and $1-\rho$, respectively. Note that here ρ represents the heterogeneity among subjects and $1-\rho$ controls the heterogeneity among gap times for a given subject. Given the baseline gap times, general gap times were calculated as $T_{ij}=T_{ij}^{(0)}/(1+\beta_0 Z_i)$ with Z_i assumed to follow the uniform distribution over $(0,1)$.

In the simulation, we considered different configurations in terms of the heterogeneity characterized by ρ, different sample sizes, true values of β_0 and censoring distributions. Table 1 presents the means (BIAS) of the biases of point estimates $\hat{\beta}_1$ and $\hat{\beta}$ and their sample standard deviations (SSE) over 10,000 replications for the case of $\beta_0 = 0.5$, $n = 100$ or 200, and $\rho = 0.25$, 0.50 or 0.75, respectively. Note that $\hat{\beta}_1$ relies only on the first gap time and does not depend on ρ. The table also gives the average of observed gap times ($\bar{M}$), the means of two estimated standard deviations given by $\hat{\Omega}_F$ (SEE_F) and $\hat{\Omega}_B$ (SEE_B) and the estimated 95% empirical converge probabilities corresponding to $\hat{\Omega}_F$ (CP_F) and $\hat{\Omega}_B$ (CP_B), respectively. The top part of the table is for the case where the censoring time C is $U(0,1)$, while the bottom part is for the case where C is $U(0,2)$.

Table 1. Summary of the simulation study.

			Censoring time $\sim U(0,1)$						
	n	ρ	$\bar{M}$	BIAS	SSE	SEE_F	SEE_B	CP_F	CP_B
$\hat{\beta}_1$	100		1.0000	0.0055	0.6944	0.6636	0.6791	0.9420	0.9522
$\hat{\beta}$		0.25	1.3104	0.0142	0.6778	0.6579	0.6743	0.9422	0.9554
		0.50	1.5311	0.0110	0.6866	0.6598	0.6762	0.9396	0.9515
		0.75	1.9991	0.0054	0.6924	0.6625	0.6773	0.9427	0.9516
$\hat{\beta}_1$	200		1.0000	-0.0017	0.4816	0.4666	0.4734	0.9429	0.9488
$\hat{\beta}$		0.25	1.3097	0.0001	0.4711	0.4631	0.4702	0.9450	0.9506
		0.50	1.5319	0.0136	0.4791	0.4649	0.4717	0.9423	0.9491
		0.75	1.9996	-0.0010	0.4810	0.4656	0.4721	0.9425	0.9494

			Censoring time $\sim U(0,2)$						
	n	ρ	$\bar{M}$	BIAS	SSE	SEE_F	SEE_B	CP_F	CP_B
$\hat{\beta}_1$	100		1.0000	0.0067	0.5685	0.5400	0.5629	0.9318	0.9524
$\hat{\beta}$		0.25	1.8846	0.0123	0.5543	0.5210	0.5467	0.9266	0.9527
		0.50	2.3677	0.0156	0.5567	0.5276	0.5528	0.9320	0.9549
		0.75	3.2360	0.0081	0.5638	0.5333	0.5566	0.9278	0.9496
$\hat{\beta}_1$	200		1.0000	0.0053	0.3989	0.3810	0.3922	0.9368	0.9491
$\hat{\beta}$		0.25	1.8871	0.0055	0.3843	0.3685	0.3809	0.9374	0.9514
		0.50	2.3744	0.0062	0.3895	0.3720	0.3834	0.9352	0.9498
		0.75	3.2424	0.0052	0.3948	0.3763	0.3876	0.9360	0.9501

It can be seen from the table that both estimators $\hat{\beta}_1$ and $\hat{\beta}$ are approximately unbiased. As expected, $\hat{\beta}$ is more efficient than $\hat{\beta}_1$ and, when ρ is close to 1, the variances of the two estimators approach each other. Also as expected, the variances decrease when sample size increases and the follow-up period is longer. In terms of the two variance estimators, the model-based estimator $\hat{\Omega}_B$ seems better. These conclusions are similar to those obtained in Huang and Chen (2003) under the proportional hazards model. Other configurations gave similar results. For both the simulation here and the example in the next section, Fortran was used for programming.

4. An Application

In this section we apply the proposed methodology to a bladder cancer study conducted by the Veterans Administration Cooperative Urological Research Group (Andrews and Herzberg (1985), Byar (1980) and Wei, Lin and Weissfeld (1989)). The study consisted of 118 patients with superficial bladder tumors, a number of whom experienced recurrences of the tumors. The patients were randomly allocated to one of three treatments, placebo (48), pyridoxine (32) and thiotepa (38). In addition, for each patient, there were two potentially important baseline covariates: the number of initial tumors and the size of the largest initial tumor. The goal here is to assess the treatment effects on the gap time between recurrent tumors as well as baseline covariate effects. As mentioned in Huang and Chen (2003), the renewal process assumption seems reasonable.

To apply the method, define $Z_{i1} = 1$ if patient i received the pyridoxine treatment and 0 otherwise, $Z_{i2} = 1$ if patient i was given thiotepa and 0 otherwise, and Z_{i3} and Z_{i4} be the number of initial tumors and the size of the largest initial tumor, respectively. The results obtained by the application of the approach are given in Table 2, where the estimated standard errors are from the model-based method. For comparison, $\hat{\beta}_1$ and its estimated standard errors were also obtained and are presented in the table. The results suggest that the thiotepa treatment had a significant effect on delaying the recurrence of the bladder tumor, and a larger number of initial tumors implies a significantly shorter gap time. On the other hand, the gap time did not seem to be related to pyridoxine and the size of the largest initial tumor. The results based on model-free variance estimation are similar.

Table 2. Analysis results of the bladder cancer study.

		Pyridoxine	Thiotepa	Initial number	Initial size
$\hat{\beta}$	Estimate	0.0010	-0.0174	0.0084	-0.0029
	SEE_B	0.0072	0.0063	0.0033	0.0025
$\hat{\beta}_1$	Estimate	0.0010	-0.0167	0.0071	-0.0020
	SEE_B	0.0074	0.0066	0.0033	0.0025

Note that the above results were obtained under the additive hazards model (1). Huang and Chen (2003) gave similar conclusions using the proportional hazards model. To see graphically which model gives a better fit to the problem, we obtained separate estimators of the cumulative hazard functions corresponding to the three treatments. Figure 1 presents the differences of the three estimators in log scale, and Figure 2 gives the differences themselves. Note that under the proportional hazards model, one would expect the differences in log scale to be roughly a straight line with slope zero, while under the additive hazards model, the differences themselves would resemble a straight line with slope zero. It seems from Figures 1−2 that the additive hazards model fits a little better than the proportional hazards model.

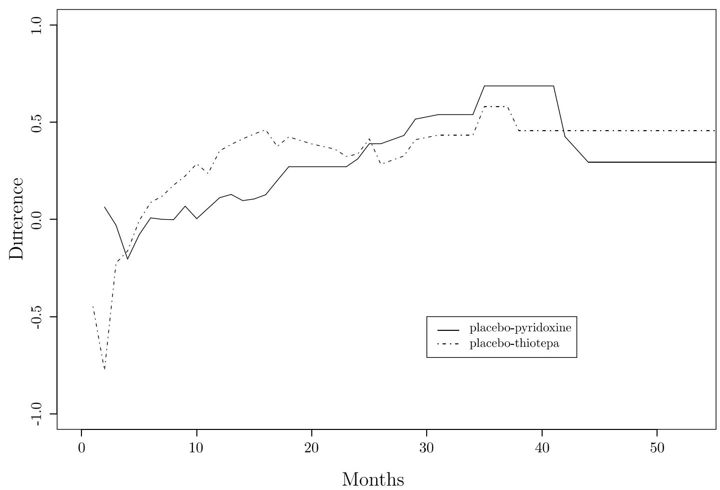

Figure 1. Differences of estimated cumulative hazard functions in log scale.

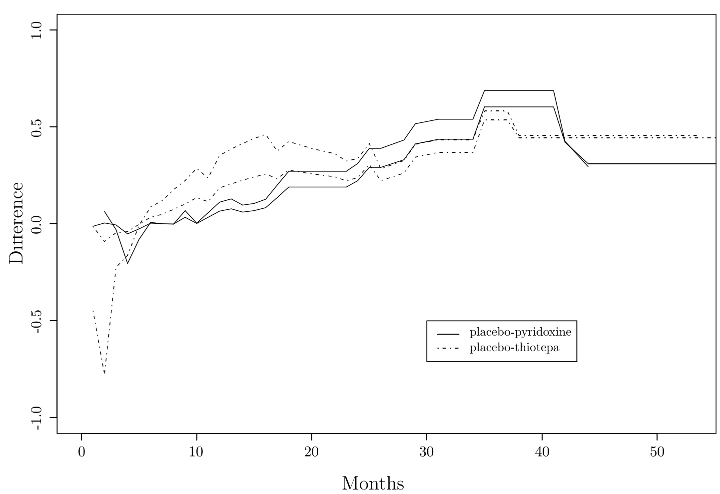

Figure 2. Differences of estimated cumulative hazard functions.

5. Concluding Remarks

An alternative to the proposed estimate $\hat{\beta}$ of the covariate effect is $\hat{\beta}_1$ based only on the time to the first occurrence of the event. Although the derivation of $\hat{\beta}_1$ may seem easier, it is always less efficient than $\hat{\beta}$ as proved in Appendix B and by the simulation for situations considered here. In general, however, this may not be true and needs more research. Another alternative is to replace $\{X_{i1}\ , i = 1, \ldots, n\}$ in $U_1(\beta)$ by each of $\prod_{i=1}^{n} M_i^*$ different sets of individual gap times and to base the inference on the sum of the resulting estimating functions. Specifically, let $j_i \in \{1, \ldots, M_i^*\}$ and

$$
\begin{aligned}
&U^{(j_1,\ldots,j_n)}(\beta) \\
&= \int_0^\tau Q(t) \left[\hat{\mathcal{E}}_i\{Z_i \Delta_i dI(X_{ij_i} \le t)\} - \frac{\hat{\mathcal{E}}_i\{Z_i I(X_{ij_i} \ge t)\}}{\hat{\mathcal{E}}_i\{I(X_{ij_i} \ge t)\}} d\hat{\mathcal{E}}_i\{\Delta_i I(X_{ij_i} \le t)\} \right. \\
&\quad \left. - \left(\hat{\mathcal{E}}_i\{Z_i^{\otimes 2} I(X_{ij_i} \ge t)\} - \frac{(\hat{\mathcal{E}}_i\{Z_i I(X_{ij_i} \ge t)\})^{\otimes 2}}{\hat{\mathcal{E}}_i\{I(X_{ij_i} \ge t)\}} \right) \beta dt \right] .
\end{aligned}
$$

Then one could use the estimating function

$$
U_2(\beta) = \sum_{j_1=1}^{M_1^*} \cdots \sum_{j_n=1}^{M_n^*} U^{(j_1,\ldots,j_n)}(\beta) = 0 .
$$

This approach may be more efficient, but it may be computationally overwhelming when n is large.

It should be noted that here we considered only time-independent covariates and that the proposed inference approach relies on the assumption that the observed complete gap times are identically distributed. In other words, we have the exchangeability of the gap times. For the case of time-dependent covariates, the proposed inference approach is still applicable as long as covariate effects are the same across gaps, which assures the exchangeability of the gap times. More discussions about the exchangeability can be found in Huang and Chen (2003) and Wang and Chang (1999), and recently Chen, Wang and Huang (2004) considered a situation where exchangeability is no longer available under the stratified proportional reverse-time hazards model.

A problem for future research is the development of appropriate tools for choosing a model that gives a good fit to a given problem. One approach is to use graphical tools as done in the last section, but they are often nonconclusive. Some procedures have been developed for the goodness-of-fit test for the proportional hazards model and the additive hazards model if regular right-censored failure time data are available. It seems, however, that there does not exist a formal procedure for either model for the situation considered here. Another problem that needs further study, but is beyond the scope of the current paper, is the selection of a weight process $Q(t)$ that gives the most efficient estimate of β for a particular situation. For the simulation study, in addition to $Q(t) = 1$, we also studied the process $Q(t) = n^{-1} \sum_{i=1}^{n} I(C_i \geq t)$ and obtained similar results. However, as in many situations, it is difficult to develop a guideline about the general selection of the weight process.

Appendix A. Asymptotic Properties of $U(\beta_0)$ and $\hat{\beta}$

Define $K(t) = E\{\Delta I(X_{(1)} \leq t)\}$, $G_0(t) = E\{I(X_{(1)} \geq t)\}$, and $G_1(t) = E\{ZI(X_{(1)} \geq t)\}$. To establish the asymptotic properties of $U(\beta_0)$ and $\hat{\beta}$, we need the following regularity conditions:

(C1) $Q(t)$ has bounded variation and converges almost surely to a deterministic function $q(t)$ uniformly over $t \in [0, \tau]$;

(C2) $E\|Z\|^2 < \infty$;

(C3) $A = \int_0^\tau q(t) \left[E\{Z^{\otimes 2} I(X_{(1)} \geq t)\} - G_1(t)^{\otimes 2}/G_0(t)\right] dt > 0$.

We first show the asymptotic normality of $U(\beta_0)$. Under (C1) and (C2), using the Functional Central Limit Theorem (Pollard (1990, p.53)) and the functional version of the Taylor expansion for the mapping, we have

$$\begin{aligned} n^{\frac{1}{2}} U(\beta_0) = n^{\frac{1}{2}} \int_0^\tau q(t) \Bigg[& \hat{\mathcal{E}}_{ij}\{Z_i \Delta_i dI(X_{ij} \leq t)\} - \frac{G_1(t)}{G_0(t)} d\hat{K}(t) - \frac{\hat{G}_1(t)}{G_0(t)} dK(t) \\ & + \frac{G_1(t)\hat{G}_0(t)}{G_0(t)^2} dK(t) - \hat{\mathcal{E}}_{ij}\{Z_i^{\otimes 2} I(X_{ij} \geq t)\}\beta_0 dt \\ & + \left(\frac{\hat{G}_1(t) G_1'(t)}{G_0(t)} + \frac{G_1(t)\hat{G}_1'(t)}{G_0(t)} - \frac{G_1(t)^{\otimes 2}\hat{G}_0(t)}{G_0(t)^2} \right) \beta_0 dt \Bigg] + o_p(1) \\ = n^{-\frac{1}{2}} \sum_{i=1}^{n} & \hat{\mathcal{E}}_j\{\phi(X_{ij}, \Delta_i, Z_i)\} + o_p(1), \end{aligned} \tag{A.1}$$

where

$$\begin{aligned} \phi(X_{ij}, \Delta_i, Z_i)\} = \int_0^\tau q(t) \left(Z_i - \frac{G_1(t)}{G_0(t)} \right) \Bigg[& \Delta_i dI(X_{ij} \leq t) - \frac{I(X_{ij} \geq t)}{G_0(t)} dK(t) \\ & - I(X_{ij} \geq t) \beta_0' \left(Z_i - \frac{G_1(t)}{G_0(t)} \right) dt \Bigg]. \end{aligned}$$

Note that $\hat{\mathcal{E}}_j\{\phi(X_{ij}, \Delta_i, Z_i)\}$ $(i = 1, \ldots, n)$ are i.i.d. zero-mean random vectors. By utilizing the Multivariate Central Limit Theorem, $n^{1/2} U(\beta_0)$

is asymptotically normal with mean zero and covariance matrix $\Sigma = E[(\hat{\mathcal{E}}_j\{\phi(X_{ij}, \Delta_i, Z_i)\})^{\otimes 2}]$, which can be consistently estimated by $\hat{\Sigma}$. Now we consider the asymptotic properties of $\hat{\beta}$. By applying the Uniform Strong Law of Large Numbers (Pollard (1990, p.41)), it can be first shown that $\hat{A} \to A$ and $\hat{U}(\beta_0) \to 0$ almost surely. The strong consistency of $\hat{\beta}$ thus follows from $\hat{\beta} - \beta_0 = \hat{A}^{-1}U(\beta_0)$ and $A > 0$. The asymptotic normality of $\hat{\beta}$ directly follows from the asymptotic normality of $U(\beta_0)$ and

$$n^{\frac{1}{2}}(\hat{\beta} - \beta_0) = A^{-1}n^{\frac{1}{2}}U(\beta_0) + o_p(1). \tag{A.2}$$

That is, $n^{1/2}(\hat{\beta} - \beta_0)$ is asymptotically normal with zero mean and covariance matrix $A^{-1}\Sigma A^{-1}$, which can be consistently estimated by $\hat{\Omega}$.

Appendix B. Comparison of Asymptotic Variances of $\hat{\beta}$ and $\hat{\beta}_1$

For $\hat{\beta}$, note that

$$\Sigma = E\{\phi(X_{(1)}, \Delta, Z)^{\otimes 2}\} - E[\hat{\mathcal{E}}_j\{\phi(X_{(j)}, \Delta, Z) - \hat{\mathcal{E}}_j\phi(X_{(j)}, \Delta, Z)\}^{\otimes 2}]\,. \tag{A.3}$$

Under the conditions (C1)$-$(C3), it can be shown that the asymptotic variance of $\hat{\beta}_1$ is given by

$$A^{-1}E\{\phi(X_{(1)}, \Delta, Z)^{\otimes 2}\}A^{-1}.$$

This, together with (A.3), shows that the asymptotic variance of $\hat{\beta}$ is smaller than or equal to that of $\hat{\beta}_1$.

Appendix C. Asymptotic Properties of $\hat{\Lambda}_0$

First note that under model (1), we have

$$\Lambda_0(t) = \int_0^t \frac{dK(u) - \beta_0' G_1(u)du}{G_0(u)}\,, \tag{A.4}$$

$$\hat{\Lambda}_0(t;\hat{\beta}) = \int_0^t \frac{d\hat{K}(u) - \hat{\beta}'\hat{G}_1(u)du}{\hat{G}_0(u)}. \tag{A.5}$$

The uniform strong consistency of $\hat{\Lambda}_0(t;\hat{\beta})$ thus follows directly from (A.4), (A.5), the Uniform Strong Law of Large Numbers and the strong consistency of $\hat{\beta}$. We now consider the asymptotic normality of $\hat{\Lambda}_0(t;\hat{\beta})$. In view of (A.1)$-$(A.2) and (A.4)$-$(A.5), using the Functional Central Limit Theorem and the functional version of the Taylor expansion for the mapping, we can show that

$$
\begin{aligned}
& n^{\frac{1}{2}}(\hat{\Lambda}_0(t;\hat{\beta}) - \Lambda_0(t)) \\
&= n^{\frac{1}{2}} \int_0^t \left[\frac{d\hat{K}(u) - \beta_0' \hat{G}_1(u)du}{G_0(u)} - \frac{\hat{G}_0(u)dK(u) - \beta_0' G_1(u)\hat{G}_0(u)du}{G_0(u)^2} \right] \\
&\quad - \int_0^t \frac{G_1'(u)du}{G_0(u)}(\hat{\beta} - \beta_0) + o_p(1) \\
&= n^{-\frac{1}{2}} \sum_{i=1}^{n} \hat{\mathcal{E}}_j \{\psi(t; X_{ij}, \Delta_i, Z_i)\} + o_p(1), \qquad \text{(A.6)}
\end{aligned}
$$

where

$$
\begin{aligned}
\psi(t; X_{ij}, \Delta_i, Z_i) &= \int_0^t \frac{\Delta_i dI(X_{ij} \leq u) - I(X_{ij} \geq u)\beta_0' Z_i du}{G_0(u)} \\
&\quad - \int_0^t \frac{I(X_{ij} \geq u)dK(u) - I(X_{ij} \geq u)\beta_0' G_1(u)du}{G_0(u)^2} \\
&\quad - C'(t)A^{-1}\phi(X_{ij}, \Delta_i, Z_i)
\end{aligned}
$$

and $C(t) = \int_0^t G_1(u)du/G_0(u)$. Thus, the finite-dimensional normality of $\hat{\Lambda}_0(t;\hat{\beta})$ follows from the Multivariate Central Limit Theorem. Note that $Z_i = \max\{Z_i, 0\} - \max\{-Z_i, 0\}$. Then $\int_0^t [\Delta_i dI(X_{ij} \leq u) - I(X_{ij} \geq u)\beta_0' Z du]/G_0(u)$ and $\int_0^t [I(X_{ij} \geq u)dK(u) - I(X_{ij} \geq u)\beta_0' G_1(u)du]/G_0(u)^2$ can be written as summations of monotone processes over $[0, \tau]$ and are therefore manageable (Pollard (1990, p.38)). It then follows from the Functional Central Limit Theorem that $n^{1/2}(\hat{\Lambda}_0(t;\hat{\beta}) - \Lambda_0(t))$ is tight and converges weakly to a zero-mean Gaussian process whose covariance function at (s,t) is given by $E\{\hat{\mathcal{E}}_j \psi(s; X_{(j)}, \Delta, Z)\hat{\mathcal{E}}_j \psi(t; X_{(j)}, \Delta, Z)\}$, which can be consistently estimated by (2).

Acknowledgements

The authors wish to thank the Editor, Professor Jane-Ling Wang, an associate editor and two referees for their many helpful comments and suggestions that greatly improved the paper. This work was partly supported by the National Natural Science Foundation of China grants 10471140 and 10571169 (to the first author) and a grant from the National Health of Institutes (to the third author).

References

Andrews, D. F. and Herzberg, A. M. (1985). *Data; A Collection of Problems from Many Fields for the Student and Research Worker*. Springer-Verlag, New York.

Breslow, N. E. and Day, N. E. (1987). *Statistical Methods in Cancer Research,* **2**, *The Design and Analysis of Cohort Studies*. Lyon: IARC.

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

Chang, S. H. and Wang, M. C. (1999). Conditional regression analysis for recurrence time data. *J. Amer. Statist. Assoc.* **94**, 1221-1230.

Chen, Y. Q., Wang, M. C. and Huang, Y. (2004). Semiparametric regression analysis on longitudinal pattern of recurrent gap times. *Biostatistics* **5**, 277-290.

Cook, R. J. and Lawless, J. F. (1996). Interim monitoring of longitudinal comparative studies with recurrent event responses. *Biometrics* **52**, 1311-1323.

Cook, R. J., Lawless, J. F. and Nadeau, J. C. (1996). Robust tests for treatment comparisons based on recurrent event responses. *Biometrics* **52**, 557-571.

Huang, Y. and Chen, Y. Q. (2003). Marginal regression of gaps between recurrent events. *Lifetime Data Anal.* **9**, 293-303.

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

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

Lin, D. Y. and Ying Z. (1994). Semiparametric analysis of the additive risk model. *Biometrika* **81**, 61-71.

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

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

Schaubel, D. E. and Cai, J. (2004). Regression analysis for gap time hazard functions of sequentially ordered multivariate failure time data. *Biometrika* **91**, 291-303.

Wang, M. C. and Chang, S. H. (1999). Nonparametric estimation of a recurrent survival function. *J. Amer. Statist. Assoc.* **94**, 146-153.

Wang, M. C. and Chen Y. Q. (2000). Nonparametric and semiparametric trend analysis of stratified recurrence time data. *Biometrics*, **56**, 789-794.

Wang, M. C., Qin, J. and Chiang, C. T. (2001). Analyzing recurrent event data with informative censoring. *J. Amer. Statist. Assoc.* **96**, 1057-1065.

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

Institute of Applied Mathematics, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100080, P.R. China.

E-mail: slq@amt.ac.cn

Department of Statistics, University of Missouri, 146 Middlebush Hall, Columbia, MO 65211, U.S.A.

E-mail: dp41e@mizzou.edu

Department of Statistics, University of Missouri, 146 Middlebush Hall, Columbia, MO 65211, U.S.A.

E-mail: sunj@missouri.edu

(Received June 2004; accepted October 2004)