10-28-2005

Analyzing Panel Count Data with Informative Observation Times

Chiung-Yu Huang
Division of Biostatistics, School of Public Health, University of Minnesota, cyhuang@biostat.umn.edu

Mei-Cheng Wang
Johns Hopkins Bloomberg School of Public Health, Department of Biostatistics, mcwang@jhsph.edu

Ying Zhang
University of Iowa, Department of Biostatistics, ying-j-zhang@uiowa.edu

Analyzing panel count data with informative observation times

By CHIUNG-YU HUANG

Biostatistics Research Branch, National Institute of Allergy and Infectious Diseases, NIH, Bethesda, MD 20892, U.S.A.

cyhuang@biostat.umn.edu

MEI-CHENG WANG

Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland 21205, U.S.A.

mcwang@jhsph.edu

and YING ZHANG

Department of Biostatistics, University of Iowa, Iowa City, IA 52242, U.S.A.

ying-j-zhang@uiowa.edu

Summary

In this paper, we study panel count data with informative observation times. We assume nonparametric and semiparametric proportional rate models for the underlying recurrent event process, where the form of the baseline rate function is left unspecified and a subject-specific frailty variable inflates or deflates the rate function multiplicatively. The proposed models allow the recurrent event processes and observation times to be correlated through their connections with the unobserved frailty; moreover, the distributions of both the frailty variable and observation times are considered as nuisance parameters. The baseline rate function and the regression parameters are estimated by maximizing a conditional likelihood function of observed event counts and solving estimation equations. Large sample properties of the proposed estimators are studied. Numerical studies demonstrate that the proposed estimation procedures perform well for moderate sample sizes. An application to a bladder tumor study is presented to illustrate the use of the proposed methods.

Some key words: Dependent censoring; Frailty; Poisson process; Rate function; Recurrent events.

1 Introduction

Recurrent event data arise in longitudinal studies where each subject is at risk of experiencing serial events such as repeated tumor occurrences, or repeated graft rejection episodes. Often, the observations of recurrent events are taken at several distinct and random time points, and, instead of recording the exact times when the events occurred, only the number of events that have occurred prior to each observation time is recorded. Data of this type are commonly referred to as panel count data; see Thall & Lachin (1988) and Balshaw and Dean (2002).

The development of statistical methods for panel count data has progressed slowly. Because the exact event occurrence times are not observed, panel count data provide less information than recurrent event data about the underlying recurrent event process. For one-sample estimation, Sun & Kalbfleisch (1995) derived a nonparametric maximum pseudolikelihood estimator of the rate function for the recurrent event process. Wellner & Zhang (2000) studied the asymptotics of the nonparametric maximum pseudolikelihood estimator and showed that it is less efficient than the nonparametric maximum likelihood estimator through some simulation studies. For semiparametric modelling, the derivation of semiparametric maximum likelihood estimator is computationally intensive. To overcome the computational challenge, Zhang (2002) proposed an inference procedure based on a semiparametric pseudolikelihood function. Wellner et al. (2004) compared the large-sample properties of the semiparametric maximum pseudolikelihood estimator with the semiparametric maximum likelihood estimator, and showed that the former can be very inefficient when the distribution of the number of observation times is heavily tailed. Sun & Wei (2000) formulated estimation equations for regression parameters in the semiparametric proportional rate models. The Sun-Wei estimator, however, is inefficient as it ignores correlations among event counts in the estimation equations, and its validity relies heavily on correct modelling of the observation pattern.

Most proposed statistical models for panel count data assume that the observation times are independent of the recurrent events, conditioning on *observed* covariates such as treatment assignments. Such an assumption, however, can be easily violated in many applications. For example, patients with rapid disease progression may tend to visit clinics more often because they need more medical attention; moreover, these patients may have a shorter follow-up period due to a higher risk of failure. No existing methods handle panel count data with informative observation times. Motivated by Wang et al. (2001), we studied nonparametric and semiparametric models that allow observation times to be correlated with the recurrent event process with the correlation induced by a frailty variable. Estimation procedures that require no parametric assumptions on the distributions of the frailty variable and the observation time process are proposed for nonparametric and semiparametric models.

The paper is organized as follows: In Section 2, we introduce nonparametric and semiparametric rate models for the recurrent event process of a subject. In Section 3, we develop an estimation procedure for the baseline cumulative rate function in the nonparametric model. In Section 4, we present an estimation procedure for the regression parameters and the baseline cumulative rate function in the semiparametric model. The strong consistency properties of the proposed estimators and the rate of convergence are stated in Sections 3 and 4, with proofs given in the Appendix. We report numerical studies and the application to a bladder tumor study in Section 5 and conclude with discussions in Section 6.

2 Notations and Models

This paper focuses on the statistical inference of the rate function for the underlying counting process in a fixed time interval $[0, \tau]$. Let $N(t)$ denote the number of recurrent events that have occurred at or before time t, and assume that observations on a subject are collected at K random time points $0 < t_1 < \ldots < t_K \leq \tau$, where K is a random variable that takes values on positive integers and $y = t_K$ is the last observation time (i.e. censoring time). Let $m_j = N(t_j) - N(t_{j-1})$ be the number of recurrent events in the time interval $(t_{j-1}, t_j]$ and $m = N(y)$ the total number of recurrent events observed in $[0, \tau]$. We denote the observed data by $D = \{t_1, t_2, \ldots, t_K,\ K, y;\ m_1, m_2, \ldots, m_K, m\}$.

We consider the following nonparametric model for the recurrent event process $N(\cdot)$:

Model A. Let Z be a nonnegative latent variable with $E[Z] = 1$, so that, given $Z = z$, $N(\cdot)$ is a nonhomogeneous Poisson process with intensity function

$$\lambda(t|z) = z\lambda_0(t), \quad t \in [0, \tau],$$

where $\lambda_0(t)$ is an unspecified function. Given Z, the recurrent event process $N(\cdot)$ is independent of K and the random observation times $\{t_1, \ldots, t_K\}$.

Define the function $\Lambda_0(t) = \int_0^t \lambda_0(t)$. Model A implies that the cumulative rate function of recurrent events in the disease population is given by $E[Z] \cdot \Lambda_0(t) = \Lambda_0(t)$. Under Model

A, the recurrent event process $N(\cdot)$ and the observation times $\{t_1, \ldots, t_K\}$ are correlated through their connections with the frailty variable Z. In statistical literature, when a frailty variable is used to induce association among random variables, parametric assumptions on the distribution of the frailty variable are usually required for making inference. Model A, on the contrary, makes no parametric assumptions on the distribution of Z.

Let x be a $1 \times p$ vector of covariates. When the effects of x on the rate function of the recurrent event process are of interest, a semiparametric extension of Model A for the recurrent event process $N(\cdot)$ is given below:

Model B: There exists a nonnegative latent variable Z with $E[Z|x] = 1$ so that, conditioning on x and $Z = z$, $N(\cdot)$ is a nonhomogeneous Poisson process with intensity function

$$\lambda(t|x, z) = z e^{x\beta} \lambda_0(t), \quad t \in [0, \tau],$$

where $\lambda_0(t)$ is unspecified. Moreover, given x and z, the recurrent event process $N(\cdot)$ is independent of the number of observation time points, K, and the observation times $\{t_1, \ldots, t_K\}$.

Model B implies that the cumulative rate function of the subgroup with covariate x in the disease population is given by $e^{x\beta}\Lambda_0(t)$. Also note that under Model B the effect of x on the expected value of Z is through an exponential transformation. The regression coefficient β should be viewed as the joint effect on the rate function for x and Z. Model B allows the recurrent event process and the observation pattern of a subject to be correlated through their association with the observed covariates x and the unobserved frailty Z. Moreover, the distribution of the observation times and the distribution of the frailty variable are left unspecified.

3 Estimation Procedure for Model A

We use subscript i for a subject, $i = 1, \ldots, n$. Let z_i be the individual frailty value, k_i the number of observation times, and t_{ij} the j^{th} observation time for the i^{th} subject, where $j = 1, \ldots, k_i$ and $0 \equiv t_{i0} < \ldots < t_{ik_i} \leq \tau$. Let y_i denote the last observation time point, that is, $y_i = t_{ik_i}$. Let N_i be the underlying individual counting process and $m_{ij} = N_i(t_{ij}) - N_i(t_{ij-1})$ be the number of recurrent event in the time interval $(t_{ij-1}, t_{ij}]$. Finally, let $m_i = N(y_i)$ be the total number of recurrent events occurring during follow-up. For ease of notation we use m_{ij} and m_i to represent both random variables and realizations. We denote the observed data of the i^{th} subject by $D_i = \{t_{i1}, t_{i2}, \ldots, t_{ik_i}, k_i, y_i; m_{i1}, m_{i2}, \ldots, m_{ik_i}, m_i\}$, $i = 1, 2, \ldots, n$, and assume that $D_1, \ldots, D_n$ are independent identically distributed copies of D.

Model A implies that, given m_i and y_i, the m_i event times are order statistics of independent identically distributed (iid) random variables with density function $z_i\lambda_0(t)/z_i\Lambda_0(y_i)$. The likelihood of the event times is proportional to the truncation likelihood given in Wang et al. (2001). By further conditioning on $\{t_{ij}, j = 1, \ldots, k_i\}$, the conditional likelihood function can be derived by integrating out the the probability density function of the order statistics. Assuming that $\Lambda_0(\tau)$ is bounded, we define the shape function for the recurrent event process $N(\cdot)$ on $[0, \tau]$ as $F(t) = \frac{\Lambda_0(t)}{\Lambda_0(\tau)}$, $t \le \tau$. Thus F defines a proper cumulative distribution function on $[0, \tau]$ with $F(\tau) = 1$. The conditional likelihood function conditioning on z_i, k_i, m_i and $\{t_{ij}, j = 1, \ldots, k_i\}$ can be expressed as

$$Q \propto \prod_{i=1}^{n}\prod_{j=1}^{k_i}\left(\frac{\Lambda_0(t_{ij}) - \Lambda_0(t_{ij-1})}{\Lambda_0(y_i)}\right)^{m_{ij}} = \prod_{i=1}^{n}\prod_{j=1}^{k_i}\left(\frac{F(t_{ij}) - F(t_{ij-1})}{F(y_i)}\right)^{m_{ij}}. \quad (1)$$

Interestingly, no information from the frailty variable Z is required to form (1). Note that if $\sum_{j=1}^{k_i} m_{ij} = 1$, the right hand side of (1) is exactly the likelihood function of a set of independent interval censored and right-truncated data. Therefore, the estimation of $F(t)$ in (1) can be implemented by the self-consistency (EM) algorithm proposed by Turnbull (1976).

Let $0 \equiv t_0^* < t_1^* < \ldots < t_L^* \le \tau$ be the ordered and distinct observation times from $\{t_{ij}; k_i > 1, 1 \le i \le n, 1 \le j \le k_i\}$. For $1 \le l \le L$, define $p_k = F(t_k^*) - F(t_{k-1}^*)$. We maximize Q subject to the constraint $\sum_{k=1}^{L} p_k = 1$. Define $a_{ijk} = 1$ if $[t_{k-1}^*, t_k^*] \subseteq [t_{ij-1}, t_{ij}]$ and 0 otherwise. Additionally, we define $b_{ik} = 1$ if $t_k^* \le y_i$ and 0 otherwise. Given the estimates $p_k^{(l)}$, $k = 1, \ldots, L$, in the l^{th} iteration, the E-step is simply to compute

$$d_k^{(l)} = \sum_{i=1}^{n}\sum_{j=1}^{k_i} m_{ij}\left\{\frac{a_{ijk}p_k^{(l)}}{\sum_h a_{ijh}p_h^{(l)}} + \frac{(1 - b_{ik})p_k^{(l)}}{\sum_h b_{ih}p_h^{(l)}}\right\},$$

where $\sum_{h=1}^{L} b_{ih}p_h^{(l)} = \hat{F}_n^{(l)}(y_i)$ in the l^{th} iteration. Given the updated $d_k^{(l)}$, in the M-step we maximize the complete likelihood of $\{v_{ijk}\ ;\ i = 1, \ldots, n, j = 1, \ldots, k_i, k = 1, \ldots, m_{ij}\}$ and update estimate of p_k with $p_k^{(l+1)} = d_k^{(l)} / \sum_{h=1}^{L} d_h^{(l)}$. Note that $d_k^{(l)}$ is the expected number of events in the time interval $[t_{k-1}^*, t_k^*]$ and $\sum_{h=1}^{L} d_h^{(l)} = \sum_{i=1}^{n} \frac{m_i}{\hat{F}_n^{(l)}(y_i)}$ is the projected total number of recurrent events in the time interval $[0, \tau]$. Finally, the estimate of $F(t)$ is updated with $\hat{F}_n^{(l+1)}(t) = \sum_{t_h \le t} d_h^{(l)}$. We alternate between the E-step and M-step until convergence to obtain the estimate $\hat{F}_n$ of F.

Let $0 \equiv t_0^* < t_1^* < \ldots < t_L^* \le \tau$ be the ordered and distinct observation times from $\{t_{ij}; k_i > 1, 1 \le i \le n, 1 \le j \le k_i\}$. For $1 \le l \le L$, define $p_k = F(t_k^*) - F(t_{k-1}^*)$. We maximize Q subject to the constraint $\sum_{k=1}^{L} p_k = 1$. Define $a_{ijk} = 1$ if $[t_{k-1}^*, t_k^*] \subseteq [t_{ij-1}, t_{ij}]$ and 0 otherwise. Additionally, we define $b_{ik} = 1$ if $t_k^* \leq y_i$ and 0 otherwise. Given the estimates $p_k^{(l)}$, $k = 1, \ldots, L$, in the l^{th} iteration, the E-step is simply to compute

$$d_k^{(l)} = \sum_{i=1}^{n} \sum_{j=1}^{k_i} m_{ij} \left\{ \frac{a_{ijk} p_k^{(l)}}{\sum_h a_{ijh} p_h^{(l)}} + \frac{(1 - b_{ik}) p_k^{(l)}}{\sum_h b_{ih} p_h^{(l)}} \right\},$$

where $\sum_{h=1}^{L} b_{ih} p_h^{(l)} = \hat{F}_n^{(l)}(y_i)$ in the l^{th} iteration. Given the updated $d_k^{(l)}$, in the M-step we maximize the complete likelihood of $\{v_{ijk}\ ;\ i = 1, \ldots, n, j = 1, \ldots, k_i, k = 1, \ldots, m_{ij}\}$ and update estimate of p_k with $p_k^{(l+1)} = d_k^{(l)} / \sum_{h=1}^{L} d_h^{(l)}$. Note that $d_k^{(l)}$ is the expected number of events in the time interval $[t_{k-1}^*, t_k^*]$ and $\sum_{h=1}^{L} d_h^{(l)} = \sum_{i=1}^{n} \frac{m_i}{\hat{F}_n^{(l)}(y_i)}$ is the projected total number of recurrent events in the time interval $[0, \tau]$. Finally, the estimate of $F(t)$ is updated with $\hat{F}_n^{(l+1)}(t) = \sum_{t_h \leq t} d_h^{(l)}$. We alternate between the E-step and M-step until convergence to obtain the estimate $\hat{F}_n$ of F.

The cumulative rate function $\Lambda_0(t)$ is related to F through the equation $\Lambda_0(t) = F(t)\Lambda_0(\tau)$, where $\Lambda_0(\tau)$ is interpreted as the expected number of recurrent events occurring in the time interval $[0, \tau]$. Conditioning on z_i and y_i, m_i has the expected value $E[m_i|z_i, y_i] = z_i \Lambda_0(y_i) = z_i F(y_i) \Lambda_0(\tau)$. Thus we have $E[m_i F(y_i)^{-1}] = \Lambda_0(\tau)$, provided $E[Z] = 1$; that is, the ratio of m_i to $F(y_i)$ projects the number of events in $[0, \tau]$. Substituting F with $\hat{F}_n$, an estimator of $\Lambda_0(\tau)$ is given by $\hat{\Lambda}_n(\tau) = \frac{1}{n} \sum_{i=1}^{n} m_i \hat{F}_n(y_i)^{-1}$. Hence $\Lambda_0(t)$ can be estimated by $\hat{\Lambda}_n(t) = \hat{F}_n(t)\ \hat{\Lambda}_n(\tau)$.

Let $\mathcal{B}$ denote the Borel sets in $\mathcal{R}$ and let $\mathcal{B}_{[0,\tau]} = \{B \cap [0, \tau]\}$. Define the measures ν and ν_1 on $([0, \tau], \mathcal{B}_{[0,\tau]})$ by $\nu(B) = E\Big[\ E[\sum_{j=1}^{K} I(t_j \in B)|K]\ \Big]$ and $\nu_1(B_1 \times B_2) = E\Big[\ E[\sum_{j=1}^{K} I(t_{j-1} \in B_1, t_j \in B_2)|K]\ \Big]$ for $B, B_1, B_2 \in \mathcal{B}_{[0,\tau]}$. Note that ν and ν_1 are finite measures if $E[K] < \infty$. Let $\mathcal{F}$ be the class of functions defined by

$$\mathcal{F}_\tau = \{F : [0, \tau] \to [0, 1] | F \text{ is nondecreasing, } F(0) = 0, \text{ and } F(\tau) = 1\}$$

Then the $L_2(\nu)$ metric d on $\mathcal{F}$ is defined as

$$d^2(F_1, F_2) \;=\; \int |F_1(t) - F_2(t)|^2 d\nu(t) = E\Big[E[\sum_{j=1}^{K} (F_1(t_j) - F_2(t_j))^2 |K]\Big].$$

The strong consistency property of $\hat{\Lambda}_0$ is stated in Theorem 1 with the following conditions:

(C1) There exists an integer $k_0 < \infty$ such that the number of observation times, K, satisfies $Pr(K \leq k_0) = 1$ and $Pr(K > 1) > 0$.

(C2) The cumulative rate function Λ_0 satisfies $\Lambda_0(\tau) \leq M$ for some $M \in (0, \infty)$.

(C3) The random function $M_0 = \sum_{j=1}^{k} m_j \log(m_j)$ satisfies $E[M_0] < \infty$.

(C4) There exists a $\tau_1 > 0$ such that $Pr(Y \geq \tau_1) = 1$ and $\Lambda_0(\tau_1) \geq C^*$ for some $C^* > 0$

Theorem 1. Assume that (C1)$\sim$(C4) hold. Define $\tau_2 = \sup\{t : P(Y \geq t) > 0\}$. Then for every t such that $t \leq \tau_2$, $d(\hat{\Lambda}_n 1_{[0,t]}, \Lambda_0 1_{[0,t]}) \to 0$ almost surely when $n \to \infty$.

Because the estimation of Λ_0 shares similarities with the estimation of a distribution function under random interval censoring and truncation, the convergence rate of $\hat{\Lambda}_n(t)$ is expected to be non-regular, i.e. not of $n^{1/2}$-convergence rate. For the purpose of systematically studying the convergence rate of $\hat{\Lambda}_n(t)$, we consider the following technical conditions:

(C5) There exists a constant $\eta > 0$ such that adjacent observation times are separated by η, i.e. $t_j - t_{j-1} \geq \eta$ for $j = 1, 2, \ldots, K$.

(C6) The baseline cumulative rate function $\Lambda_0 \in C^1[0, \tau]$ and there exists a constant $\gamma > 0$ such that $\Lambda_0'(t) \geq \gamma$ for $t \in [0, \tau]$.

(C7) For any $\alpha = o_P(1)$, there exists a constant C^{**} such that $E\left(z^i e^{\alpha z}\right) \leq C^{**}$ for $i = 0, 1, 2$.

Theorem 2. In addition to (C1)$\sim$(C4), we further assume that (C5)$\sim$(C7) hold. Also we suppress the indicator,$1_{[0,t]}$ in our expression by assuming that the metric d is defined with $t \leq \tau_2$. Then we have $n^{1/3} d(\hat{\Lambda}_n, \Lambda_0) = O_p(1)$.

The proofs of the theorems are sketched in the Appendix using the modern empirical process theory. We leave the study of the asymptotic distribution of $\hat{\Lambda}$ to future research.

Remark. The conditions given above are sufficient for proving the theorems, though might not be necessary. Conditions (C1)$\sim$(C6) are often satisfied in practice, which warrants the usefulness of the theorems in applications. For the proof of Theorem 2, we need to characterize the unobserved frailty variable Z. Condition (C7) may be stronger than necessary, but it holds for the Gamma frailty variable, the most common choice for frailty models in practice.

4 Estimation Procedure for Model B

Under Model B the conditional likelihood for the i^{th} individual, given z_i, x_i, k_i, m_i and observation times $\{t_{i1}, \ldots, t_{ik_i}\}$, is proportional to

$$\prod_{j=1}^{k_i} \left(\frac{z_i e^{\beta x_i} \Lambda_0(t_{ij}) - z_i e^{\beta x_i} \Lambda_0(t_{ij-1})}{z_i e^{\beta x_i} \Lambda_0(y_i)} \right)^{m_{ij}} = \prod_{j=1}^{k_i} \left(\frac{F(t_{ij}) - F(t_{ij-1})}{F(y_i)} \right)^{m_{ij}},$$

where $F(t) = \Lambda_0(t)/\Lambda_0(\tau)$. Hence the baseline cumulative rate function can be estimated in the same way as that in Model A.

Note that $E[m_i F^{-1}(y_i)|x_i, y_i, z_i] = z_i \Lambda_0(\tau) e^{\beta x_i}$. Following $E[z_i|x_i] = 1$ we have $E[m_i F^{-1}(y_i)|x_i] = \Lambda_0(\tau) e^{\beta x_i}$. A class of unbiased estimating equations can be given by

$$n^{-1} \sum_{i=1}^{n} w_i x_i^{*\prime} (m_i F^{-1}(y_i) - e^{x_i^* \gamma}) = 0, \qquad (2)$$

where $x_i^* = (1, x_i)$, $\gamma = (\eta, \beta')'$, $\eta = \ln \Lambda_0(\tau)$, and w_i is a weight function depending on (x_i, β, Λ_0). In the case where Λ_0 is a known function, the optimal weight is given by $e^{x_i^* \gamma}/E[(m_i F^{-1}(y_i) - e^{x_i^* \gamma})^2]$ (Godambe, 1960). In real practice, however, F is estimated with a convergence rate $n^{1/3}$, hence the efficiency gain is unknown when $\hat{F}_n$ is used to replace F in the optimal weight function.

We denote the solutions of (2), with F replaced by $\hat{F}_n$, and γ by $\hat{\gamma} = (\hat{\eta}_n, \hat{\beta}_n')'$. In the Appendix we show that, under (C1)$\sim$(C4), $|\hat{\beta}_n - \beta|^2 \to 0$ almost surely as $n \to \infty$, where $|\cdot|$ represents the regular Euclidean L_2-norm. Moreover, using the estimator obtained by solving (2), we estimate the baseline cumulative rate function $\Lambda_0(t) = F(t)\Lambda_0(\tau)$ by $\hat{\Lambda}_n(t) = \hat{F}_n(t) e^{\hat{\eta}_n}$. The estimator $\hat{\Lambda}_n$ satisfies the following strong consistency property: $d(\hat{\Lambda}_n 1_{[0,t]}, \Lambda_0 1_{[0,t]}) \to 0$ almost surely for all $t \in [0, \tau_2]$ as $n \to 0$. The derivation of the asymptotic distribution of $\hat{\beta}_n$ and $\hat{\Lambda}_n(t)$ is a challenging problem and left for future research.

5 Simulations and Data Analysis

5.1 Monte-Carlo Simulations

Four sets of simulation studies with practical ($n = 100$) and large ($n = 1000$) sample sizes were conducted to evaluate the performance of the proposed nonparametric and semiparametric estimators. We used $\Lambda_0(t) = 2t$ for $t \in [0, 10]$ and conducted the simulations using

1000 replications. The first simulation study compared the efficiency of the proposed nonparametric estimator to that of the nonparametric maximum likelihood estimator (Wellner and Zhang, 2000) and the nonparametric maximum pseudolikelihood estimator (Sun and Kalbfleisch, 1995) under the assumption of independent observation process. The second set of simulation studies examined the bias in these three nonparametric estimators when the independence assumption is violated. Specifically, we assumed that subjects with $Z > 1$ have a higher event rate and tend to be observed more frequently than patients with $Z \leq 1$. We summarize in Tables 1 and 2 the Monte-Carlo bias and standard error estimates at selected time points. Table 1 shows that the bias in these three nonparametric estimators is very small when observation times are independent of the recurrent event process. The proposed estimator $\hat{\Lambda}_n(t)$ is more efficient, with smaller Monte-Carlo standard errors, than the nonparametric maximum pseudolikelihood estimator, and is slightly less efficient than the nonparametric maximum likelihood estimator. When sample size is large, the proposed estimator is highly efficient relative to the nonparametric maximum pseudolikelihood estimator. In Table 2, where the pattern of observation is correlated with the distribution of recurrent events, the nonparametric maximum likelihood estimator and the nonparametric maximum pseudolikelihood estimator are substantially biased, while the proposed estimator still gives valid results.

We evaluated the performance of the proposed semiparametric estimator in the last two sets of simulation studies. The covariate x was generated from a Bernoulli random variable with success probability 0.5, and Z was from a gamma distribution with mean 1 and standard deviation 0.25. We set the cumulative rate function to be $Ze^{\beta x}\Lambda_0(t)$ with $\beta = -1$. In the third simulation study we compared the efficiency of the proposed semiparametric estimator to that of the Sun-Wei estimator when the observation pattern depends only on observed covariates but not the subject's risk of recurrent events. The proposed semiparametric estimation procedure, with unit weights ($w_i = 1$) in the estimating equations (2), and the Sun-Wei estimator, with and without assuming that the observation process follows a proportional rate model, were applied to each simulated data set. As shown in Tables 3 and 4, both estimators have small biases; moreover, the proposed semiparametric estimator outperforms the Sun-Wei estimators in that it gives smaller Monte-Carlo standard errors. The last simulation study examined the validity of two semiparametric estimators in a setting where both the recurrent event process and the observation pattern are correlated with Z. Tables 5 and 6 show that bias in the proposed estimator is almost ignorable, while the Sun-Wei estimators yield substantial bias in estimating regression parameters.

5.2 Data Analysis

We used a subset of data from the bladder tumor study conducted by the Veterans Administration Cooperative Urological Research Group (Byar, 1980) to illustrate the proposed methods. All the recruited patients had superficial bladder tumors before entering the study, and were randomly allocated into one of the three treatment groups: placebo, thiotepa, and pyridoxine. Many patients experienced multiple tumor occurrences after enrollment, and new tumors were removed at follow-up clinic visits. We set $\tau = 30$ (month) and compared the thiotepa group with the placebo group in tumor occurrence rate during the first 30 months.

Figure 1 shows the estimated the cumulative rate function for placebo and thiotepa groups using the proposed nonparametric method, the nonparametric maximum likelihood estimator, and the nonparametric maximum pseudolikelihood estimator. Patients treated with thiotepa had a lower tumor occurrence rate, indicating the effectiveness of thiotepa in the first 30 months. Next, we applied the proposed semiparametric method and the Sun-Wei estimators to the bladder tumor data, with x an indicator of whether a patient was in the thiotepa group. With the proposed method, the estimate of the regression coefficient of the treatment indicator is -0.62 with a bootstrap standard error 0.43, yielding an estimated tumor occurrence rate in the thiotepa group of 0.54 ($= e^{-.62}$) times the placebo group during the first 30 months of follow-up. The estimated baseline cumulative rate function with 95% pointwise bootstrapped confidence interval at selected time points is given in Figure 2. Using the Sun-Wei estimators, the estimated coefficient of the treatment indicator is -0.88 with a bootstrap standard error 0.41 assuming the observation pattern is the same for both treatment groups, and is -1.48 with a bootstrap standard error 0.40 assuming that the observation process follows a proportional rate model. The proposed method estimates a smaller treatment effect in tumor occurrence rate than the Sun-Wei estimators do.

6 Final Remarks

As pointed out by Wellner, Zhang & Liu (2004), maximizing the "full" likelihood function of the recurrent event data under a semiparametric model is heavily computationally intensive. Instead of maximizing the likelihood function that involves both the regression parameters and the unspecified baseline cumulative rate function, the proposed estimation procedure maximizes a nonparametric likelihood function that only involves the nonparametric component F and has the advantage of computational simplicity. The convergence of the EM algorithm for estimating F is known to be slow, but can be improved using the gradient projection (GP) method (Polak, 1971) or a hybrid algorithm alternating between EM and GP described in Pan & Chappell (1998) and Zhang & Jamshidian (2004).

We have also applied our method to data generated outside of the working Poisson process, and concluded that the inferential results (not shown here) are still valid. Moreover, the Poisson process assumption is not required in our proof for the strong consistency. This indicates that the proposed methods have the same robust property as those proposed by Wellner & Zhang (2000) and Zhang (2002), namely, the validity of the proposed methods does not depend on the underlying counting process conditioning on the frailty variable under Model A or conditioning on the frailty variables and covariates under Model B.

Only a few papers, including Sun and Wei (2000), have studied panel counts with dependent observation times. The validity of the Sun-Wei estimators heavily depends on correct modelling of the observation time process. Even with correct specification of the observation pattern, the efficiency of the Sun-Wei estimator is still in doubt because the correlation among event counts of the same individual is ignored in the construction of the estimating functions. In this paper we propose models that allow the observation time process to be correlated with the event counts through the observed covariates as well as through the unobserved frailty variable. Our estimation procedures, in contrast, do not require the specification of observation time process; in fact, the distributions of observation times and latent variables are considered nuisance parameters. The relative efficiency of the proposed estimators, compared to the semiparametric maximum likelihood estimator, are expected to improve when K is large. The proposed estimators have the advantage of simplicity and robustness, and will likely to yield high efficiency when the number of observation times increases.

Acknowledgement

The authors thank Professor Wei Pan for useful discussions and Yan Zheng for the computing support.

References

Balshaw, R. F. & Dean, C. B. (2002). A semiparametric model for the analysis of recurrent-event panel data. *Biometrics*, **58**, 324–331.

Byar, D. P. (1980). The veterans administration study of chemoprophylaxis for recurrent stage I bladder tumors: comparisons of placebo, pyridoxine, and topical thiotepa. *Bladder Tumors and Other Topics in Urological Oncology*, New York: Plenum, pp. 363-370.

Godambe, V. P. (1960). An optimum property of regular maximum likelihood estimation. *The Annals of Mathematical Statistics*, **31**, 1208–1212

Pan, W. & Chappell, R. (1998). A nonparametric estimator of survival functions for arbitrary truncated and censored data. Lifetime Data Analysis, 187-202.

Polak, E. (1971). *Computational Methods in Optimization.* Academic Press, New York.

Sun, J. & Kalbfleisch, J. D. (1995). Estimation of the mean function of point processes based on panel count data. *Statistica Sinica*, **5** , 279-289

Sun, J. & Wei, L. J. (2000). Regression analysis of panel count data with covariate-dependent observation and censoring times. *Journal of the Royal Statistical Society, Series B*, **62**, 293-302.

Thall, P. F. & Lachin, J. M. (1988). Analysis of recurrent events: nonparametric methods for random-interval count data. *Journal of the American Statistical Association*, **83**, 339-347.

Turnbull, B.W. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data. *Journal of the Royal Statistical Society, Series B*, **38** , 290-295.

van der Vaart, A. W. & Wellner J. A. (1996). *Weak Convergence and Empirical Processes.* Springer-Verlag: New York.

Wang, M.-C, Qin, J. & Chiang, C.-T. (2001). Analyzing recurrent event data with informative censoring. *Journal of the American Statistical Association*, **96**, 1057–1065.

Wellner, J. A. & Zhang, Y. (2000). Two estimators of the mean of a counting process with panel count data. *The Annals of Statistics*, **28**, 779-814.

Wellner, J. A., Zhang, Y. & Liu, H (2004). A semiparametric regression model for panel count data: when do pseudo-likelihood estimators become badly inefficient? *Proceedings of the Second Seattle Biostatistical Symposium: Analysis of Correlated Data.* (Eds. Lin, D.Y. and P.J., Heagerty). Springer-Verlag: New York.

Zhang, Y. & Jamshidian, M. (2003). The gamma-frailty Poisson model for the nonparametric estimation of panel count data. *Biometrics*, **59**, 1099-1106.

Zhang, Y. & Jamshidian, M. (2004). On Algorithms for NPMLE of the Failure Function with Censored Data. *The Journal of Computational and Graphical Statistics*, **13**, 123-140.

Zhang, Y. (2002). A semiparametric pseudolikelihood estimation method for panel count data. *Biometrika*, **89**, 39-48.

Appendix

Sketch proof of Theorem 1: We start with proving the strong consistency of the estimator $\hat{F} \equiv \hat{F}_n$ of F. Following (C2) the function F is a proper distribution function with $0 \leq F(t) \leq 1$. Our proof of strong consistency closely follows the proof of Theorem 4.2 in Wellner and Zhang (2000). Note that $\hat{F}_n$, the maximizer of Q, is a non-decreasing step function which jumps only at the observation times $\{t_1^*, \ldots, t_L^*\}$ and $0 \leq \hat{F}_n(t) \leq 1$ for $t \in [0, \tau]$. We define the function

$$q(F; D) = \sum_{j=1}^{k} m_j \log(F(t_j) - F(t_{j-1})) - m \log(F(y)),$$

and we use the empirical process notations defined in Van der Vaart and Wellner (1996), $\mathbb{P}_n(F) = \frac{1}{n} \sum_{i=1}^{n} q(F; D_i)$ and $P(F) = E[q(F; D)]$.

Note that for any constant k and an arbitrary vector of nonnegative numbers $x = (x_1, \ldots, x_k)$ the function $g(x) = \sum_{j=1}^{k} a_j \log(x_j) - (\sum_{j=1}^{k} a_j) \log(\sum_{j=1}^{k} x_j)$ has the maximum $\sum_{j=1}^{k} a_j \log(a_j) - \left(\sum_{j=1}^{k} a_j\right) \log\left(\sum_{j=1}^{k} a_j\right)$ at $x_i = ca_i$, $i = 1, \ldots, k$ for all $c > 0$. Thus the function $q(F; D)$ satisfies

$$q(F; D) \leq \sum_{j=1}^{k} m_j \log(m_j) - m \log(m) \leq \sum_{j=1}^{k} m_j \log(m_j) = M_0, \quad \forall F \in \mathcal{F}_\tau, \tag{3}$$

with $0 \log 0 \equiv 0$. It follows (3) and (C3) that M_0 is the upper envelope for the class of functions $\mathcal{M} = \{q(F; D); F \in \mathcal{F}_\tau\}$. Moreover, $\mathcal{F}_\tau$ is compact with the metric d, and the function $F \to q(F; D)$ is upper semi-continuous in F for P almost all possible observations. It follows from the one-sided Glivenko-Cantelli Theorem that

$$\limsup_{n\to\infty} \sup_{F \in \mathcal{F}_\tau} (\mathbb{P}_n - P)(F) \leq 0 \ \text{ almost surely} \tag{4}$$

Because $\hat{F}_n$ is the maximizer of $\mathbb{P}_n(F)$,

$$\mathbb{P}_n(\hat{F}_n) = n^{-1} \sum_{i=1}^{n} q(\hat{F}_n; D_i) \geq n^{-1} \sum_{i=1}^{n} q(F; D_i) = \mathbb{P}_n(F).$$

It follows the law of large numbers that $\mathbb{P}_n(F) \to P(F)$ almost surely, hence we have

$$P(F) \leq \lim \inf_{n\to\infty} \mathbb{P}_n(\hat{F}_n) \quad \text{almost surely} \tag{5}$$

Because the sequence of functions $\{\hat{F}_n(t, \omega), t \in [0, \tau]\}$ is uniformly bounded, it follows Helly's selection theorem that, for any sequence $\hat{F}_n(\cdot, \omega)$ there exists a subsequence $n' = n(\omega)'$ so that $\hat{F}_{n'}(\cdot, \omega) \to F^*(\cdot, \omega)$, where F^* is a nondecreasing function on $[0, \tau]$.

By noting that $\mathbb{P}_n(F) = (\mathbb{P}_n - P)(F) + P(F)$ for any F, it follows from (4) that

$$\lim \sup_{n'\to\infty} \mathbb{P}_n(\hat{F}_{n'}) \leq 0 + P(F^*) \tag{6}$$

for any subsequence $\hat{F}_{n'}$ that converges almost surely to F^* on $[0, \tau]$. Combining (5) and (6) yields

$$0 \leq P(F^*) - P(F) = -E[\, q(F; D) - q(F^*; D)]$$

$$= \quad -E\Big[\sum_{j=1}^{k} m_j \log\left(\frac{F(t_j) - F(t_{j-1})}{F^*(t_j) - F^*(t_{j-1})}\right) - m \log\left(\frac{F(y)}{F^*(y)}\right)\Big]$$

$$= \quad -E\Big[z\Lambda_0(\tau) \sum_{j=1}^{k} (F(t_j) - F(t_{j-1})) \log\left(\frac{F(t_j) - F(t_{j-1})}{F^*(t_j) - F^*(t_{j-1})}\right)$$

$$-z\Lambda_0(\tau)F(y) \log\left(\frac{F(y)}{F^*(y)}\right)\}\Big] \leq 0$$

where the right inequality is due to the aforementioned property of g, and the equality holds if and only if $F^*(v) - F^*(u) = c[F(v) - F(u)]$ a.e. ν_1 for some constant $c > 0$. Arguing as the proof of Theorem 4.2 in Wellner and Zhang (2000), we can show that $F^*(t) = cF(t)$ a.e. ν. Because the result holds for any convergent subsequence, $cF(t)$ is the limit of any subsequence of $\{\hat{F}_n\}$. Hence $\lim_{n\to\infty} \hat{F}_n(t) = cF(t)$ almost surely for ν-almost all $t \leq \tau_2$, where $\tau_2 = \sup\{t : P(Y \geq t) > 0\}$. Because $\hat{F}_n$ is uniformly bounded and ν is a finite measure by (C1), it follows the dominated convergence theorem that $d(\hat{F}_n 1_{[0,t]}, cF 1_{[0,t]}) \to 0$ almost surely for any $t \in [0, \tau_2]$.

Now we prove that $\hat{\Lambda}_n(t)$ is a consistent estimator of $\Lambda_0(t)$ for t in $[0, \tau_2]$. Note that $\hat{\Lambda}_n(t) - \Lambda_0(t)$ can be reexpressed as

$$\hat{\Lambda}_n(t) - \Lambda_0(t) = \hat{F}_n(t)(\hat{\Lambda}_n(\tau) - c^{-1}\Lambda_0(\tau)) + c^{-1}\Lambda_0(\tau)(\hat{F}_n(t) - cF(t)).$$

Hence we have

$$\begin{aligned}\hat{\Lambda}_n(\tau) - c^{-1}\Lambda_0(\tau) &= n^{-1}\sum_{i=1}^{n} m_i\Big(\frac{1}{\hat{F}_n(y_i)} - \frac{1}{cF(y_i)}\Big) - c^{-1}n^{-1}\sum_{i=1}^{n}\Big(\frac{m_i}{F(y_i)} - \Lambda_0(\tau)\Big)\\ &= I + II.\end{aligned}$$

We define a new measure ν_2 on $([\tau_1, \tau], \mathcal{B}_{[\tau_1,\tau]})$ by $\nu_2(B) = E1_{[y_i \in B]}$. Obviously, ν_2 is dominated by the measure ν. For a $\delta_n > 0$ with $\delta_n \to 0$ as $n \to \infty$, we define a class $\mathcal{F} = \{f : f(t) = N(t)[g^{-1}(t) - c^{-1}F^{-1}(t)]$, where g is nondecreasing and non-negative with positive lower bound in $[\tau_1, \tau]$ and $d(g1_{[0,\tau]}, cF1_{[0,\tau]}) \leq \delta_n\}$. For a sufficient large n,

$$\begin{aligned}\left|n^{-1}\sum_{i=1}^{n} m_i\left[\hat{F}_n^{-1}(y_i) - c^{-1}F^{-1}(y_i)\right]\right| &\leq \sup_{g\in\mathcal{F}}\left|n^{-1}\sum_{i=1}^{n} m_i\left[g^{-1}(y_i) - c^{-1}F^{-1}(y_i)\right]\right|\\ &\leq \sup_{f\in\mathcal{F}}|Pf| + \|\mathbb{P}_n - P\|_{\mathcal{F}}.\end{aligned}$$

Under (C4) and applying Theorems 2.7.5 and 2.4.1 in Van der Vaart and Wellner (1996), we can show that $\mathcal{F}$ is a Glivenko-Catelli class. Thus $\|\mathbb{P}_n - P\|_{\mathcal{F}} \to 0$ almost surely Moreover,

$$\sup_{f\in\mathcal{F}}|Pf| = \left|\ E\left\{\Lambda_0(\tau)F(y_i)[g^{-1}(y_i) - c^{-1}F^{-1}(y_i)]\right\}\ \right| \leq c\delta_n$$

for some $c > 0$ following the fact that ν_2 is dominated by ν, (C4), and the Hölder inequality. This yields that $I \to 0$ almost surely. II converges to 0 almost surely following the fact that $E[m_i/F(y_i)] = \Lambda_0(\tau)$ and the law of large numbers. Thus we show that $\hat{\Lambda}_n(\tau) - c^{-1}\Lambda_0(\tau)$ converges to 0 almost surely. It is easy to see that $\hat{\Lambda}_n(t) - \Lambda_0(t) \to 0$ almost surely for $\nu-$ almost all $t \in [0, \tau_2]$ and it follows from the dominated convergence theorem (with dominating functions $\Lambda_0(\tau_2)$ since ν is a finite measure) that $d(\hat{\Lambda}_n 1_{[0,t]}, \Lambda_0 1_{[0,t]}) \to 0$ almost surely for any $t \in [0, \tau_2]$.

Sketch proof of Theorem 2: We apply Theorem 3.2.5 of Van der Vaart & Wellner (1996) to derive the rate of convergence. To do so, we verify that the conditions of that theorem hold in our problem with (C1)∼(C7).

First we rewrite $q(\Lambda; D) = \sum_{j=1}^{k} m_j \log\left[\Lambda^*(t_j) - \Lambda^*(t_{j-1})\right]$, where $\Lambda^*(t_j) = \Lambda(t_j)/\Lambda(y)$ for $j = 1, 2, \ldots, k$. We define

$$\mathbb{M}(\Lambda) \;=\; Pq(\Lambda; D) = E\left[\sum_{j=1}^{k} \Delta\Lambda_0(t_j)\log\left(\Delta\Lambda^*(t_j)\right)\right], \tag{7}$$

where $\Delta\Lambda_0(t_j) = \Lambda_0(t_j) - \Lambda_0(t_{j-1})$ and $\Delta\Lambda^*(t_j) = \Lambda^*(t_j) - \Lambda^*(t_{j-1})$.

First, we show that performing Taylor expansion on the right hand side of (7) along with (C5) and (C6) yields $\mathbb{M}(\Lambda_0) - \mathbb{M}(\Lambda) \geq CE\left\{\sum_{j=1}^{k}\left[\Lambda(t_j) - \Lambda_0(t_j)\right]^2\right\} = Cd^2(\Lambda, \Lambda_0)$ for any Λ in a neighborhood of Λ_0. C represents a constant. In a sequel, C may represent a different constant at different places in our proof withou futher specification.

Next, we consider a class $\mathcal{M}_\delta = \{q(\Lambda; D) - q(\Lambda_0; D) : \; d(\Lambda, \Lambda_0) < \delta\}$ for some $\delta > 0$ and $\delta = o(1)$. For any $f = q(\Lambda; D) - q(\Lambda_0; D) \in \mathcal{M}_\delta$, using (C1) and (C7), we can get $\|f\|_{P,B} \leq C\delta$, where $\|\cdot\|_{P,B}$ is the "Bernstein" norm defined as $\|f\|_{P,B} = \left(2P\left(e^{|f|} - 1 - |f|\right)\right)^{1/2}$. Hence by Lemma 3.4.3 of Van der Vaart & Wellnel (1996),

$$E_P\|\mathbb{G}_n\|_{\mathcal{M}_\delta} \leq C\tilde{J}_{[]}\left(\delta, \mathcal{M}_\delta, \|\cdot\|_{P,B}\right)\left(1 + \frac{\tilde{J}_{[]}\left(\delta, \mathcal{M}_\delta, \|\cdot\|_{P,B}\right)}{\delta^2\sqrt{n}}\right),$$

where $\tilde{J}_{[]}\left(\delta, \mathcal{M}_\delta, \|\cdot\|_{P,B}\right)$ is the bracketing integral of the class of functions $\mathcal{M}_\delta$ and is defined by

$$\tilde{J}_{[]}\left(\delta, \mathcal{M}_\delta, \|\cdot\|_{P,B}\right) = \int_0^\delta \sqrt{1 + \log N_{[]}(\epsilon, \mathcal{M}_\delta, \|\cdot\|_{P,B})}d\epsilon.$$

Finally, using (C5)-(C7), we can argue that the ϵ-bracketing number of class $\mathcal{M}_\delta$ with "Bernstein" norm is controlled by $e^{1/\epsilon}$, i.e. $N_{[]}\left(\epsilon, \mathcal{M}_\delta, \|\cdot\|_{P,B}\right) = O\left(e^{1/\epsilon}\right)$. Hence

$$\tilde{J}_{[]}\left(\delta, \mathcal{M}_\delta, \|\cdot\|_{P,B}\right) \;\leq\; C\int_0^\delta \sqrt{1 + \log(1/\epsilon)}d\epsilon \leq C\int_0^\delta \epsilon^{-1/2}d\epsilon \leq C\delta^{1/2}$$

This implies that the function $\phi_n(\delta)$, which is critical for the rate of convergence based on Theorem 3.2.5 of Van der Vaart & Wellner (1996) is given by

$$\phi_n(\delta) = \delta^{1/2}\left(1 + \frac{\delta^{1/2}}{\delta^2\sqrt{n}}\right) = \delta^{1/2} + \delta^{-1}/\sqrt{n}.$$

It can be easily verified that $\phi_n(\delta)/\delta$ is a decreasing function of δ and $n^{2/3}\phi_n\left(n^{-1/3}\right) = 2\sqrt{n}$. So $n^{1/3}d(\hat{\Lambda}_n, \Lambda_0) = O_P(1)$ due to Theorem 3.2.5 of Van der Vaart & Wellner (1996).

Consistency of $\hat{\beta}_n$:

The consistency of $\hat{F}_n$ under Model B can be established by arguing in the same way as described above, except for replacing z_i with $z_i \exp(\beta x_i)$. We now examine the consistency of $\hat{\beta}_n$ obtained by solving the estimating function (2). The consistency property of the estimator obtained from the alternative estimating function can be proven using a similar argument. Define the function $U(\gamma) = \frac{1}{n} \sum_{i=1}^{n} w_i x_i^{*\prime} \left[m_i \hat{F}_n(y_i)^{-1} - e^{x_i^* \gamma}\right]$. It can be shown that the function U converges to 0 almost surely when evaluated at $\gamma = (\log(\Lambda_0(\tau)/c), \beta)$. Further it is easy to see that the derivative of U evaluated at $(\log(\Lambda_0(\tau)/c), \beta)$ is negative definite. Applying Taylor expansion on $U(\gamma)$, one can show that the solution of (2), i.e. $\hat{\gamma} = (\hat{\eta}_n, \hat{\beta}_n)$, converges to $\gamma = (\log(\Lambda_0(\tau)/c), \beta)$ almost surely. Thus we prove that $\hat{\beta}_n$ converges to β almost surely.

Based on above (sketch) proof, $\hat{\eta}_n$ converges to $\log(\Lambda_0(\tau)/c)$ almost surely. Along the fact that $d(\hat{F}_n 1_{[0,t]}, cF1_{[0,t]}) \to 0$ almost surely for any $t \in [0, \tau_2]$, it can be shown that $d(\hat{\Lambda}_n 1_{[0,t]}, \Lambda_0 1_{[0,t]}) = d(\hat{F}_n 1_{[0,t]} e^{\hat{\eta}_n}, \Lambda_0 1_{[0,t]}) \to 0$ for any $t \in [0, \tau_2]$.

Table 1: Simulation results for nonparametric estimators under the assumption of independent observation times. Let $\Lambda_0(t) = 2t$ and $Z \equiv 1$. K was generated from a discrete uniform distribution on $\{1, 2, \ldots, 6\}$. The K distinct observation times $t_1, \ldots, t_K$ were order statistics of iid uniform random variables on $[0, 10]$. Observation times were rounded to the second decimal points.

		$n = 100$						$n = 1000$					
		Proposed		NPMLE		NPMPLE		Proposed		NPMLE		NPMPLE	
t	$\Lambda_0(t)$	Bias	SE	Bias	SE	Bias	SE	Bias	SE	Bias	SE	Bias	SE
1.0	2	-0.031	0.397	-0.030	0.392	-0.060	0.418	0.011	0.156	0.010	0.154	0.007	0.200
2.0	4	-0.009	0.449	-0.022	0.443	-0.067	0.539	0.007	0.187	0.002	0.186	-0.011	0.239
3.0	6	-0.006	0.488	-0.019	0.476	-0.058	0.633	0.005	0.202	-0.002	0.189	-0.017	0.276
4.0	8	-0.008	0.526	-0.025	0.504	-0.066	0.703	0.004	0.200	-0.008	0.190	-0.003	0.300
5.0	10	-0.024	0.542	-0.047	0.505	-0.063	0.723	0.014	0.210	-0.002	0.193	-0.030	0.336
6.0	12	0.015	0.627	-0.022	0.562	-0.064	0.822	0.012	0.217	-0.006	0.193	-0.038	0.341
7.0	14	-0.015	0.654	-0.055	0.559	-0.065	0.844	0.011	0.241	-0.009	0.211	-0.019	0.374
8.0	16	-0.027	0.717	-0.066	0.577	-0.117	0.885	0.012	0.255	-0.014	0.207	-0.033	0.406
9.0	18	0.004	0.804	-0.051	0.643	-0.045	0.904	0.025	0.277	-0.003	0.218	-0.023	0.408

Bias and SE are the Monte-Carlo sample mean and standard deviation of the 1000 estimates of $\Lambda_0(t)$.

Table 2: Simulation results for nonparametric estimators under the assumption of informative observation times. Let $\Lambda_0(t) = 2t$ and $Z \sim \text{gamma}(2, 1/2)$. For $z > 1$, K was generated from a discrete uniform distribution on $\{1, 2, \ldots, 8\}$ and $t_1, \ldots, t_K$ were order statistics of K iid exponential random variables with mean 2; for $z \leq 1$, K was generated from a discrete uniform distribution on $\{1, \ldots, 6\}$ and $t_1, \ldots, t_K$ were order statistics of K iid uniform random variables on $[0, 10]$. Observation times were rounded to the second decimal points.

		$n = 100$						$n = 1000$					
		Proposed		NPMLE		NPMPLE		Proposed		NPMLE		NPMPLE	
t	$\Lambda_0(t)$	Bias	SE	Bias	SE	Bias	SE	Bias	SE	Bias	SE	Bias	SE
1.0	2	-0.009	0.271	0.015	0.271	0.447	0.454	0.000	0.104	0.021	0.101	0.484	0.194
2.0	4	-0.002	0.426	0.019	0.407	0.582	0.631	0.013	0.159	0.031	0.145	0.680	0.276
3.0	6	-0.007	0.553	-0.040	0.512	0.361	0.795	0.015	0.214	-0.021	0.182	0.431	0.345
4.0	8	-0.008	0.735	-0.235	0.651	-0.291	0.884	0.029	0.278	-0.180	0.230	-0.207	0.392
5.0	10	0.049	0.874	-0.543	0.722	-1.183	0.949	0.031	0.343	-0.509	0.258	-1.157	0.407
6.0	12	0.053	1.021	-1.075	0.803	-2.273	1.026	0.033	0.391	-0.989	0.286	-2.233	0.438
7.0	14	0.008	1.198	-1.699	0.859	-3.395	1.067	0.035	0.459	-1.580	0.309	-3.321	0.459
8.0	16	0.094	1.349	-2.318	0.962	-4.455	1.180	0.050	0.493	-2.223	0.310	-4.418	0.488
9.0	18	0.127	1.532	-2.987	1.021	-5.324	1.349	0.038	0.564	-2.901	0.340	-5.448	0.508

Bias and SE are the Monte-Carlo sample mean and standard deviation of the 1000 estimates of $\Lambda_0(t)$.

Table 3: Simulation results of the semiparametric estimation where the observation time process is a nonhomogeneous Poisson process with cumulative intensity function given by $\log(1+2t)\exp(x/2)$. Let $\Lambda_0(t) = 2t$, $\beta = -1$, $x \sim$ Bernoulli(0.5), and $z \sim$ gamma$(2, 1/2)$. Observation times were rounded to the second decimal points.

	$n = 100$			$n = 1000$		
	Proposed	Sun-Wei[a]	Sun-Wei[b]	Proposed	Sun-Wei[a]	Sun-Wei[b]
Bias	-0.005	0.464	-0.036	-0.003	0.460	-0.040
SE	0.161	0.219	0.191	0.052	0.067	0.059

Bias and SE are Monte-Carlo sample mean and standard deviation for the 1000 estimates of β.
Sun-Wei[a] is the Sun-Wei estimator without modeling the observation pattern;
Sun-Wei[a] is the Sun-Wei estimator with modeling the observation pattern.

Table 4: *Bias and standard error of the proposed method under informative observation times.*

		$n = 100$		$n = 1000$	
t	$\Lambda_0(t)$	Bias	SE	Bias	SE
1.0	2	0.005	0.394	0.006	0.136
2.0	4	0.031	0.616	0.005	0.218
3.0	6	0.024	0.879	0.004	0.301
4.0	8	0.061	1.148	0.029	0.376
5.0	10	0.093	1.352	0.033	0.464
6.0	12	0.081	1.668	0.034	0.526
7.0	14	0.175	1.887	0.073	0.594
8.0	16	0.172	2.138	0.055	0.702
9.0	18	0.204	2.550	0.082	0.814

Bias and SE are the average and standard deviation of the 1000 estimates of $\Lambda_0(t)$.

Table 5: Simulation results of the semiparametric model under the assumption of informative observation times. Let $\Lambda_0(t) = 2t$, $\beta = -1$, $x \sim$ Bernoulli(0.5), and $z \sim$ gamma(2, 1/2). For $x = 1$ and $z > 1$, K was generated from a discrete uniform distribution on $\{1, 2, \ldots, 8\}$ and $t_1, \ldots, t_K$ were order statistics of K iid exponential random variables with mean 2; otherwise, K was generated from a discrete uniform distribution on $\{1, \ldots, 6\}$ and $t_1, \ldots, t_K$ were order statistics of K iid uniform random variables on $[0, 10]$. Observation times were rounded to the second decimal points.

	$n = 100$			$n = 1000$		
	Proposed	Sun-Wei^a	Sun-Wei^b	Proposed	Sun-Wei^a	Sun-Wei^b
Bias	-0.009	0.246	0.932	-0.001	0.245	0.928
SE	0.123	0.153	0.153	0.033	0.049	0.048

Bias and SE are Monte-Carlo sample mean and standard deviation for the 1000 estimates of β.
Sun-Wei^a is the Sun-Wei estimator without modeling the observation pattern;
Sun-Wei^a is the Sun-Wei estimator with modeling the observation pattern.

Table 6: *Bias and standard error of the proposed method under informative observation times.*

		$n = 100$		$n = 1000$	
t	$\Lambda_0(t)$	Bias	SE	Bias	SE
1.0	2	0.001	0.525	-0.005	0.132
2.0	4	0.020	1.137	0.007	0.181
3.0	6	0.045	1.698	0.003	0.220
4.0	8	0.071	2.050	0.020	0.267
5.0	10	0.088	2.560	0.036	0.296
6.0	12	0.166	3.304	0.029	0.334
7.0	14	0.142	3.663	0.010	0.375
8.0	16	0.222	4.164	0.035	0.421
9.0	18	0.217	4.529	0.033	0.463

Bias and SE are the average and standard deviation of the 1000 estimates of $\Lambda_0(t)$.

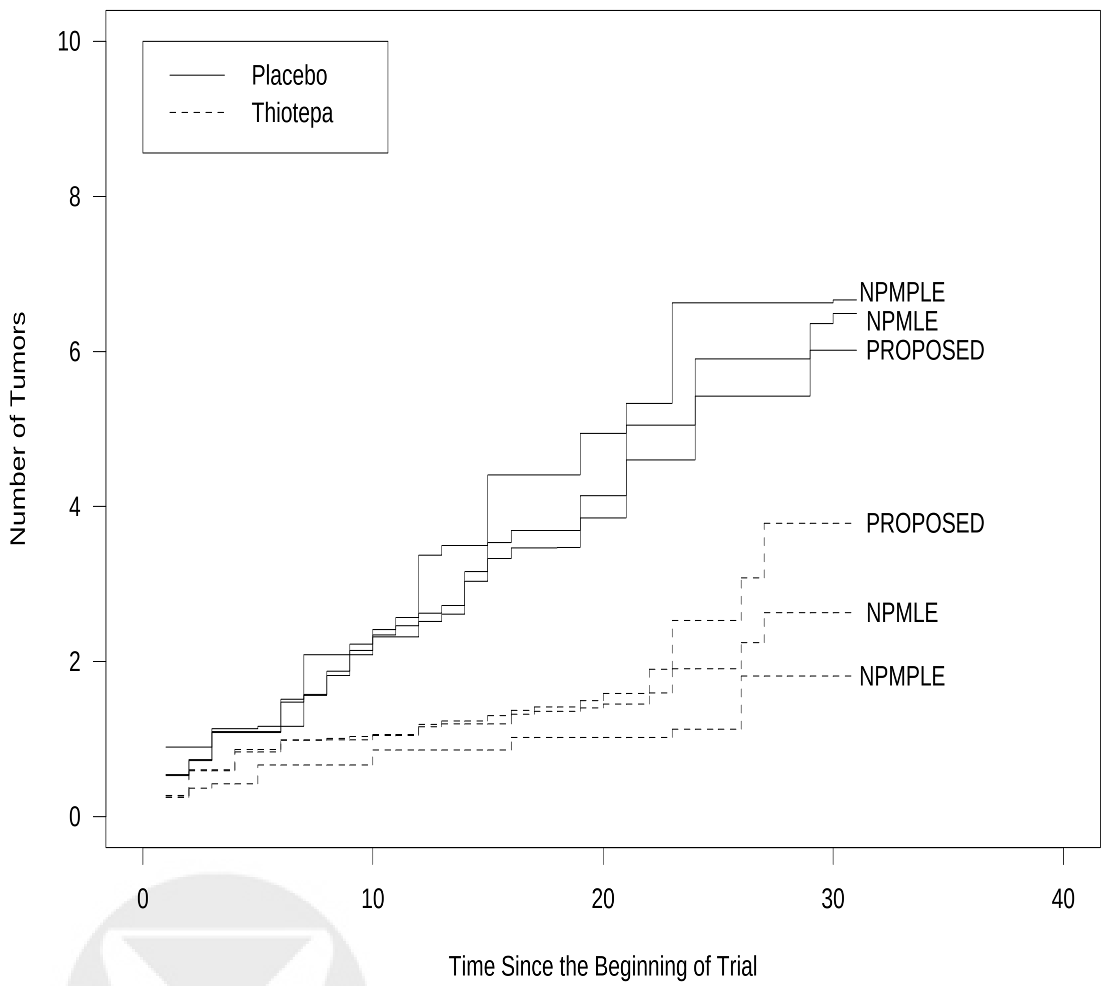

Figure 1: *Estimated cumulative rate function by treatment group.*

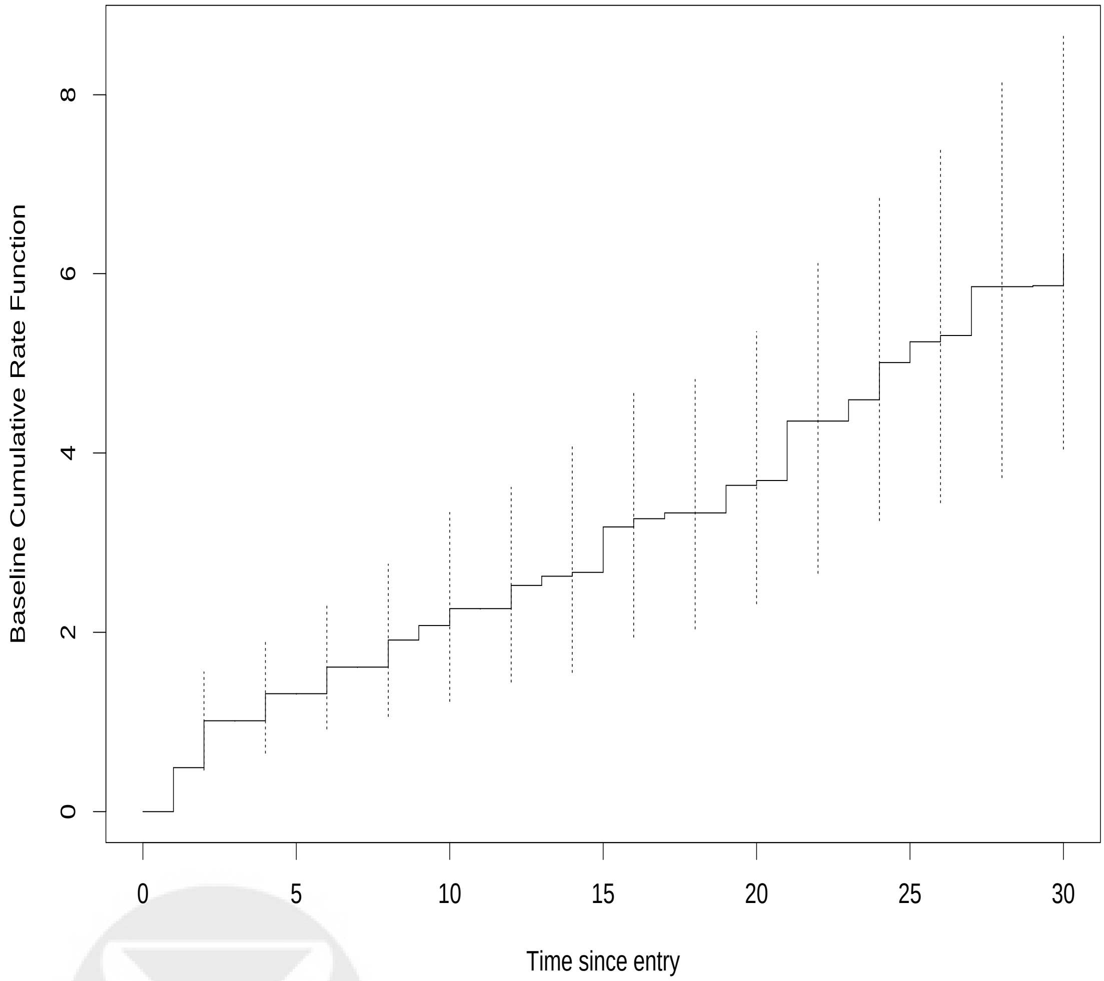

Figure 2: *Estimated baseline cumulative rate function with pointwise bootstrap 95% confidence intervals.*