Published in final edited form as:
Biometrika. 2006 December ; 93(4): 763–775. doi:10.1093/biomet/93.4.763.

Analysing panel count data with informative observation times

CHIUNG-YU HUANG,
Biostatistics Research Branch, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, U.S.A

MEI-CHENG WANG, and
Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland 21205, U.S.A

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

CHIUNG-YU HUANG: huangchi@niaid.nih.gov; MEI-CHENG WANG: ncwang@jhsph.edu; YING ZHANG: 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 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 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 maximising 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 tumour study is presented.

Keywords

Dependent censoring; Frailty; Poisson process; Rate function; Serial events

1. Introduction

In longitudinal studies of serial events such as repeated tumour occurrences or graft rejection episodes the cumulative number of these events experienced by each subject may be observed only at several distinct and random observation times, specific to each subject. Data of this type are commonly referred to as panel count data; see Thall & Lachin (1988) and Balshaw & Dean (2002). Statistical methodology for panel count data has developed slowly. Sun & Kalbfleisch (1995) derived a one-sample nonparametric maximum pseudolikelihood estimator of the rate function for the serial 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 the semiparametric maximum likelihood estimator is computationally intensive, and 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 heavy-tailed. Sun & Wei (2000) formulated estimation equations for regression parameters in the semiparametric proportional rate models. However, the Sun–Wei estimator 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 serial events, conditioning on observed covariates such as treatment assignments. However, such an assumption can be violated in many applications. No existing method can handle panel count data with informative observation times. Motivated by Wang et al. (2001), we study nonparametric and semiparametric models that allow observation times to be correlated with the event process, where the correlation is induced by a frailty variable. Estimation procedures that require no parametric assumption about the distributions of the frailty variable and the observation time process are proposed for nonparametric and semiparametric models.

2. Notation and models

This paper focuses on statistical inference for the rate function for the underlying event process in a fixed time interval $[0, \tau]$. Let $N(t)$ denote the number of serial 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 positive integer values and $y = t_K$ is the last observation time, i.e. the censoring time. Let $m_j= N(t_j)-N(t_{j-1})$ be the number of serial events in the time interval $(t_{j-1}, t_j]$ and $m = N(y)$ the total number of 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 event process $N(\cdot)$.

Model 1

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

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

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

Define the function $\Lambda_0(t)=\int_0^t\lambda_0(u)du$. Model 1 implies that the cumulative rate function of the event process in the disease population is given by $E(Z)\Lambda_0(t)=\Lambda_0(t)$. Under Model 1, the event process $N(\cdot)$ and the observation times $\{t_1, \ldots, t_K\}$ are correlated through the frailty variable Z. Unlike most frailty models in the literature, Model 1 makes no parametric assumption about the distribution of Z.

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

Model 2

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

$$\lambda(t|x,z)=ze^{x'\beta}\lambda_0(t),\qquad t\in[0,\tau],$$

where β is a p×1 vector of parameters and $\lambda_0(t)$ is unspecified. Moreover, given x and z, the event process $N(\cdot)$ is independent of the number of observation time points, K, and the observation times $\{t_1, \ldots, t_K\}$.

In our formulation the distribution of the frailty variables and the conditional distribution of the observation times given the frailty can be arbitrary and are left unspecified.

3. The estimators and their asymptotic properties

3·1. Estimation procedure for model 1

We use subscript i for a subject, i=1, …, n. Let z_i be the individual frailty value, k_i the number of observation times and t_{ij} the jth observation time for the ith subject, where j=1, …, 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 let $m_{ij}= N_i(t_{ij})-N_i(t_{ij-1})$ be the number of serial events in the time interval $(t_{ij-1}, t_{ij}]$. Finally, let $m_i = N(y_i)$ be the total number of events occurring during follow-up. For ease of notation we use m_{ij} and m_i to represent both random variables and realisations. We denote the observed data of the ith subject by $D_i=\{t_{i1}, t_{i2}, \ldots, t_{ik_i}, k_i, y_i; m_{i1}, m_{i2}, \ldots, m_{ik_i}, m_i\}$, for i=1, 2, …, n, and assume that $D_1, \ldots, D_n$ are independent and identically distributed copies of D.

Model 1 implies that, given m_i and y_i, the m_i event times are order statistics of independent and identically distributed 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). If we further condition on $\{t_{ij}, j=1, \ldots, k_i\}$, the conditional likelihood function can be derived by integrating out the probability density function of the order statistics. Assuming that $\Lambda_0(\tau)$ is bounded, we define the shape function for the event process $N(\cdot)$ on [0, τ] as $F(t)=\Lambda_0(t)/\Lambda_0(\tau)$, for $t\leq\tau$. Thus F defines a proper cumulative distribution function on [0, τ] with $F(\tau)$=1. The conditional likelihood function, conditioning on z_i, k_i, m_i and $\{t_{ij}, j=1, \ldots, k_i\}$, is

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

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 algorithm proposed by Turnbull (1976).

Turnbull's self-consistency algorithm is equivalent to the Expectation-Maximisation algorithm. When applied to the conditional likelihood function Q, the E-step and M-step have simple closed solutions as described below. Let $0\equiv t_0^*<t_1^*<\ldots<t_L^*\leq\tau$ be the ordered and distinct observation times from $\{t_{ij}; k_i>1, 1\leq i\leq n, 1\leq j\leq k_i\}$. For $1\leq l\leq L$, define $p_k=F(t_k^*)-F(t_{k-1}^*)$. We maximise 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 lth iteration, the E-step is to compute

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

where $\sum_{h=1}^{L}b_{ih}p_h^{(l)}=\widehat{F}_n^{(1)}(y_i)$ in the lth iteration. Given the updated $d_k^{(l)}$, in the M-step we update the 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 $\sum_{h=1}^{L}d_h^{(l)}=\sum_{i=1}^{n}m_i/\widehat{F}_n^{(l)}(y_i)$ is the projected total number of serial events in the time interval $[0, \tau]$. Finally, the estimate of $F(t)$ is updated with $\widehat{F}_n^{(l+1)}(t)=\sum_{t_h\le t}d_h^{(l)}$. We alternate between the E-step and M-step until convergence to obtain the estimate $\hat{F}_n$ of F.

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 serial events occurring in the time interval $[0, \tau]$. If we condition on z_i and y_i, m_i has the expected value $E(m_i|z_i, y_i) = z_i\Lambda_0(y_i) = z_iF(y_i)\Lambda_0(\tau)$. Thus we have $E\{m_iF(y_i)^{-1}\}=\Lambda_0(\tau)$, since $E(Z)=1$; that is, the ratio of m_i to $F(y_i)$ projects the number of events in $[0, \tau]$. If we substitute F with $\hat{F}_n$, an estimator of $\Lambda_0(\tau)$ is given by $\widehat{\Lambda}_n(\tau)=n^{-1}\sum_{i=1}^{n}m_i/\widehat{F}_n(y_i)$. Hence $\Lambda_0(t)$ can be estimated by $\hat{\Lambda}_n(t)=\hat{F}_n(t)\hat{\Lambda}_n(\tau)$.

Let $\mathscr{F}$ be the class of functions defined by

$$\mathscr{F}=\{\mathscr{F}:[0,\tau]\rightarrow[0,1]|F\text{ is nondecreasing, }F(0)=0\text{ and }F(\tau)=1\}.$$

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

$$d^2(F_1,F_2)=\int|F_1(t)-F_2(t)|^2dv(t)=E\left(E\left[\sum_{j=1}^{K}\{F_1(t_j)-F_2(t_j)\}^2|K\right]\right)$$

where

$$v(t)=\sum_{k=1}^{\infty}\text{pr}(K=k)\sum_{j=1}^{k}\text{pr}(t_{ij}\le t|K=k).$$

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

Condition 1. There exists an integer $k_0<\infty$ such that the number of observation times, K, satisfies $\text{pr}(K\le k_0) = 1$ and $\text{pr}(K>1)>0$.

Condition 2. The cumulative rate function Λ_0 satisfies $\Lambda_0(\tau)\le M$ for some $M\in(0, \infty)$.

Condition 3. The random function $M_0=\sum_{j=1}^{k}m_j\log(m_j)$ satisfies $E[M_0]<\infty$.

Condition 4. There exists a $\tau_1>0$ such that $\text{pr}(Y\ge\tau_1)=1$ and $\Lambda_0(\tau_1)\ge C^*$ for some $C^*>0$.

Theorem 1— *We assume that Conditions* 1–4 *hold. Define* $\tau_2=\sup\{t: \mathrm{pr}(Y\geq t)>0\}$. *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* $n\to\infty$.

Since 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 nonregular, 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.

Condition 5. There exists a constant $\eta>0$ such that adjacent observation times are separated by η, that is $t_j-t_{j-1}\geq\eta$ for $j=1, 2, \ldots, K$.

Condition 6. 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 $t\in[0, \tau]$.

Condition 7. For any $\alpha=o_P(1)$, there exists a constant C^{**} such that $E(z^i e^{\alpha z})\leq C^{**}$ for $i=0, 1, 2$.

Theorem 2— *We assume that Conditions* 1–7 *hold, and 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 that* $n^{1/2} d(\hat{\Lambda}_n, \Lambda_0)=O_p(1)$.

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

Remark: Conditions 1–7 are sufficient, but may not all be necessary. In particular, Condition 7 may be stronger than necessary, but it does hold for the Gamma frailty variable.

3·2. Estimation procedure for model 2

Under Model 2 the conditional likelihood for the ith individual, given z_i, x_i, k_i, m_i and observation times $\{t_{i1}, \ldots, t_{ik_i}\}$, is proportional to

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

where $F(t)=\Lambda_0(t)/\Lambda_0(\tau)$. Note that the unobserved frailty z_i and the observed covariates x_i are cancelled out in the formula, yielding the same conditional likelihood function given by (1) in § 3·1. Thus the baseline cumulative rate function can be estimated in the same way as that in Model 1. Intuitively, if all subjects are under observation up to time τ the total number of events of each subject contains all the information about β. Note that $E\{m_i F^{-1}(y_i)|x_i, y_i, z_i\}=z_i\Lambda_0(\tau)e^{x_i'\beta}$. Following $E(z_i|x_i)=1$ we have

$$E\{m_i F^{-1}(y_i)|x_i\}=\Lambda_0(\tau)e^{x_i'\beta};$$

that is, the ratio of m_i to $F(y_i)$ projects the number of events in $[0, \tau]$. We can derive the inferential results for β based on a class of unbiased estimating equations given by

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

where $x_i^*=(1, x_i')'$, $\gamma = (\eta,\beta')'$, η=log $\Lambda_0(\tau)$, and w_i is a weight function depending on (x_i, β, Λ_0). If 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 practice, however, F is estimated with a convergence rate of $n^{1/3}$, and 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 $\hat{F}_n$, by $\widehat{\gamma}=(\widehat{\eta}_n,\widehat{\beta}_n')'$. In the Appendix we show that, under Conditions 1–4, $|\hat{\beta}_n - \beta|^2 \rightarrow 0$ almost surely as $n \rightarrow \infty$, where $|\cdot|$ represents the usual 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 $\hat{\Lambda}_n(t)=\hat{F}_n(t)e^{\eta^n}$. The estimator $\hat{\Lambda}_n$ satisfies the following strong consistency property: $d(\hat{\Lambda}_n 1_{[0,t]}, \Lambda_0 1_{[0,t]})\rightarrow 0$ almost surely for all $t \in [0, \tau_2]$ as $n \rightarrow 0$. The derivation of the asymptotic distribution of $\hat{\beta}_n$ and $\hat{\Lambda}_n(t)$ is a challenging problem and is left for future research.

4. Simulations and data analysis

4·1. Monte Carlo simulations

Four sets of simulation studies with moderate, 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 & Zhang, 2000) and the nonparametric maximum pseudolikelihood estimator (Sun & Kalbfleisch, 1995) under the assumption of independent observation process. To be specific, we set $z\equiv 1$ and generated K from a discrete uniform distribution on {1, 2, …, 6}. The K distinct observation times $t_1, \ldots, t_K$ were order statistics of independent and identically distributed uniform random variables on [0, 10], and observation times were rounded to the second decimal place. The second set of simulation studies examined the bias in these three nonparametric estimators when the independence assumption is violated. Let $z\sim \text{Ga}(2, \frac{1}{2})$. For $z>1$, K was generated from a discrete uniform distribution on {1, 2, …, 8} and $t_1, \ldots, t_K$ were order statistics of K independent and identically distributed exponential random variables with mean 2; for $z\le 1$, K was generated from a discrete uniform distribution on {1, …, 6} and $t_1, \ldots, t_K$ were order statistics of K independent and identically distributed uniform random variables on [0, 10]. Thus, subjects with $z>1$ have a higher event rate and tend to be observed more frequently than patients with $z\le 1$.

Table 1 gives the Monte Carlo bias and standard error estimates of these three nonparametric estimators at selected time points. Table 1(a) shows that the bias in these three nonparametric estimators is very small when observation times are independent of the 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 the sample size is large, the proposed estimator is highly efficient relative to the nonparametric maximum pseudolikelihood estimator. In Table 1(b), where the pattern of observation is correlated with the distribution of serial 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 Ber(0·5) random variable, and z was from a Ga(2, 0·5) distribution. We set the cumulative intensity function to be $ze^{x\beta}\Lambda_0(t)$ with $\beta=-1$. In the third simulation study we compared the efficiency of the proposed semiparametric estimator to that of the Sun–Wei estimator under the assumption that the observation time process is a nonhomogeneous Poisson process with cumulative intensity function given by $\log(1+2t)e^{x/2}$. Thus, the observation pattern depends only on observed covariates but not on the subject's risk of serial 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 dataset. Table 2 gives the Monte Carlo bias and standard error of the estimated β, and Table 3 gives estimates of $\Lambda_0(t)$ at selected time points using the proposed semiparametric method. As shown in Table 2, 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 the two semiparametric estimators in a setting where both the event process and the observation pattern are correlated with z. For $x=1$ and $z>1$, K was generated from a discrete uniform distribution on {1, 2, …, 8} and $t_1, \ldots, t_K$ were order statistics of K independent and identically distributed exponential random variables with mean 2; otherwise, K was generated from a discrete uniform distribution on {1, …, 6} and $t_1, \ldots, t_K$ were order statistics of K independent and identically distributed uniform random variables on [0, 10]. Tables 4 and 5 show that bias in the proposed estimator is almost ignorable, while the Sun–Wei estimators yield substantial bias in estimating regression parameters.

4·2. Data analysis

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

Figure 1(a) shows the estimated 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 tumour 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 tumour data, with x an indicator of whether or not 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 of 0·43, yielding an estimated tumour occurrence rate in the thiotepa group of $0{\cdot}54=e^{-0{\cdot}62}$ times that of 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 Fig. 1(b). With the Sun–Wei estimators, the estimated coefficient of the treatment indicator is −0·88 with a bootstrap standard error of 0·41 under the assumption that the observation pattern is the same for both treatment groups, and is −1·48 with a bootstrap standard error of 0·40 under the assumption that the observation process follows a proportional rate model. The proposed method estimates a smaller treatment effect in the tumour occurrence rate than do the Sun–Wei estimators.

5. Final remarks

We have applied our method to data generated from other than 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 robustness property as those proposed by Wellner & Zhang (2000) and Zhang (2002), namely that the validity of the proposed methods does not depend on the underlying counting process conditioning on the frailty variable under Model 1 or conditioning on the frailty variables and covariates under Model 2. We have also considered in our simulation studies scenarios where the scheduled visits are fixed by design and the chance of missing a visit depends on the frailty z. The results, not shown, suggest that the proposed method gives valid results, while the nonparametric maximum likelihood estimator and the nonparametric pseudolikelihood estimator are substantially biased.

The standard asymptotic theory applies to the proposed method when the schedules of visits are fixed by study design: instead of maximising a nonparametric conditional likelihood with infinite-dimensional parameters, the proposed estimation procedure maximises a conditional likelihood with finite number of parameters. This asymptotic normality with a convergence rate of $n^{1/2}$ is expected for the proposed estimators.

It is important to indicate that the proposed methodology relies on the assumption that the effect of frailty on the intensity function is multiplicative. This assumption is widely used in modelling clustered survival times, where a parametric assumption for the frailty distribution is usually required for statistical inference. The estimation procedure proposed in this paper does not rely on the frailty distribution and hence is more robust against departure from the true frailty distribution. While the use of multiplicative frailty is crucial to our methodology, the technique for checking the multiplicative assumption needs to be developed in future research.

Acknowledgments

The authors thank Professor Wei Pan for useful discussions and Yan Zheng and Tai Feng for the computing support. The authors thanks the editor and the associate editor, whose comments greatly improved the paper.

References

Balshaw RF, Dean CB. A semiparametric model for the analysis of recurrent-event panel data. Biometrics. 2002; 58:324–31. [PubMed: 12071405]

Byar, DP. The veterans administration study of chemoprophylaxis for recurrent stage I bladder tumors: comparisons of placebo, pyridoxine, and topical thiotepa. In: Pavone-Macaluso, M.; Smith, PH.; Edsmyr, F., editors. Bladder Tumors and Other Topics in Urological Oncology. New York: Plenum; 1980. p. 363-70.

Godambe VP. An optimum property of regular maximum likelihood estimation. Ann Math Statist. 1960; 31:1208–12.

Sun J, Kalbfleisch JD. Estimation of the mean function of point processes based on panel count data. Statist Sinica. 1995; 5:279–89.

Sun J, Wei LJ. Regression analysis of panel count data with covariate-dependent observation and censoring times. J R Statist Soc B. 2000; 62:293–302.

Thall PF, Lachin JM. Analysis of recurrent events: nonparametric methods for random-interval count data. J Am Statist Assoc. 1988; 83:339–47.

Turnbull BW. The empirical distribution function with arbitrarily grouped, censored and truncated data. J R Statist Soc B. 1976; 38:290–5.

van der Vaart, AW.; Wellner, JA. Weak Convergence and Empirical Processes. Springer-Verlag; New York: 1996.

Wang MC, Qin J, Chiang CT. Analyzing recurrent event data with informative censoring. J Am Statist Assoc. 2001; 96:1057–65.

Wellner JA, Zhang Y. Two estimators of the mean of a counting process with panel count data. Ann Statist. 2000; 28:779–814.

Wellner, JA.; Zhang, Y.; Liu, H. A semiparametric regression model for panel count data: when do pseudo-likelihood estimators become badly inefficient?. In: Lin, DY.; Heagerty, PJ., editors. Proceedings of the Second Seattle Biostatistical Symposium: Analysis of Correlated Data. New York: Springer-Verlag; 2004. p. 143-74.

Zhang Y. A semiparametric pseudolikelihood estimation method for panel count data. Biometrika. 2002; 89:39–48.

Appendix

Proofs

Sketch proof of Theorem 1

We only state the main results for the proof of Theorem 1. Readers are referred to the technical report available at http://www.bepress.com/jhubiostat/paper90 for details. The proof of strong consistency of $\hat{F}_n$ closely follows Wellner & Zhang (2000). Arguing as in the proof of Theorem 4·2 in Wellner & Zhang (2000), we can show that $d(\hat{F}_n 1_{[0,t]}, cF1_{[0,t]}) \rightarrow 0$ almost surely for any $t \in [0, \tau_2]$, where $\tau_2 = \sup\{t : \text{pr}(Y \geq t) > 0\}$ and c is a fixed constant.

Now we prove that $\hat{\Lambda}_n(t)$ is a consistent estimator of $\Lambda_0(t)$ for t in $[0, \tau_2]$. We write

$$\widehat{\Lambda}_n(\tau) - \frac{1}{c}\Lambda_0(\tau) = \frac{1}{n}\sum_{i=1}^{n} m_i \left\{ \frac{1}{\widehat{F}_n(y_i)} - \frac{1}{cF(y_i)} \right\} - \frac{1}{cn}\sum_{i=1}^{n} \left\{ \frac{m_i}{F(y_i)} - \Lambda_0(\tau) \right\} = \text{I} + \text{II}.$$

Let $\mathscr{B}$ denote the Borel sets in $\mathscr{R}$. We define a new measure ν_2 on $([\tau_1, \tau], \mathscr{B}_{[\tau_1,\tau]})$ by $\nu_2(B) = E1_{[y_1 \in B]}$. Obviously, ν_2 is dominated by the measure ν. For a $\delta_n > 0$ with $\delta_n \rightarrow 0$ as $n \rightarrow \infty$, we define a class $\mathscr{F} = \{ f : f(t) = N(t)\{g^{-1}(t) - c^{-1}F^{-1}(t)\}$, where g is nondecreasing and nonnegative with positive lower bound in $[\tau_1, \tau]$ and $d(g1_{[0,tau;]}, cF1_{[0,tau;]}) \leq \delta_n$. For a sufficiently large n,

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

where $\mathbb{P}_n$ denotes the empirical measure and P denotes the probability measure.

Under Condition 4 and applying Theorems 2·7·5 and 2·4·1 in van der Vaart & Wellner (1996), we can show that $\mathscr{F}$ is a Glivenko–Cantelli class. Thus $\|\mathbb{P}_n - P\|_{\mathscr{F}} \to 0$ almost surely. Moreover,

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

for some $c>0$ following from the fact that ν_2 is dominated by ν and the Hölder inequality. This implies that I→0 almost surely. The quantity II converges to 0 almost surely because $E\{m_i/F(y_i)\}=\Lambda_0(\tau)$ and by 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 ν-almost-all $t \in [0, \tau_2]$, and it follows from the dominated convergence theroem, 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 Conditions 1–7.

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

$$\mathrm{M}(\Lambda) = Pq(\Lambda;D) = E\left\{\sum_{j=1}^{k} \Delta\Lambda_0(t_j) \log \Delta\Lambda*(t_j)\right\}, \quad \text{(A1)}$$

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 (A1) along with Conditions 5 and 6 yields $\mathrm{M}(\Lambda_0) - \mathrm{M}(\Lambda) \geq CE[\sum_{j=1}^{k} \{\Lambda(t_j) - \Lambda_0(t_j)\}^2] = Cd^2(\Lambda, \Lambda_0)$ for any Λ in a neighbourhood of Λ_0. Here C represents a generic constant.

Next, we consider a class $\mathscr{M}_\delta = \{q(\Lambda; D) - q(\Lambda_0; D) : d(\Lambda, \Lambda_0) < \delta\}$ for some $\delta>0$ and $\delta=o(1)$. For any $f=q(\Lambda; D) - q(\Lambda_0; D) \in \mathscr{M}_\delta$, using Conditions 1 and 7, we can obtain $\| f\|_{P,B} \leq C\delta$, where $\|\cdot\|_{P,B}$ is the Bernstein norm defined as $\| f\|_{P,B} = \{2P(e^{|f|} - 1 - | f|\}^{1/2}$. Hence, by Lemma 3·4·3 of van der Vaart & Wellner (1996),

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

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

$$\tilde{J}_{[\,]}(\delta, \mathscr{M}_\delta, \|\cdot\|_{P,B}) = \int_0^\delta \{1 + \log N_{[\,]}(\varepsilon, \mathscr{M}_\delta, \|\cdot\|_{P,B})\}^{1/2} d\varepsilon.$$

Finally, using Conditions 5–7, we can argue that the ε-bracketing number of class $\mathscr{M}_\delta$ with Bernstein norm is controlled by $e^{1/\varepsilon}$, that is $N_{[]}(\varepsilon, \mathscr{M}_\delta, \|\cdot\|_{P,B}) = O(e^{1/\varepsilon})$. Hence

$$\tilde{J}_{[\,]}(\delta, \mathscr{M}_\delta, \|\cdot\|_{P,B}) \le C\int_0^\delta \{1 + \log(1/\varepsilon)\}^{1/2} d\varepsilon \le C\int_0^\delta \varepsilon^{-1/2} d\varepsilon \le C\delta^{1/2}.$$

This implies that the function $\varphi_n(\delta)$, which is critical for the rate of convergence based on Theorem 3·2·5 of van der Vaart & Wellner (1996), is given by

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

It can be easily verified that $\varphi_n(\delta)/\delta$ is a decreasing function of δ and $n^{2/3}\varphi_n(n^{-1/3})=2n^{1/2}$, so that $n^{1/3}d(\hat{\Lambda}_n, \Lambda_0)=O_P(1)$ because of Theorem 3·2·5 of van der Vaart & Wellner (1996). %

Consistency of $\hat{\beta}_n$

The consistency of $\hat{F}_n$ under Model 2 can be established by arguing in the same way as described above, except for replacing z_i with $z_i\exp(x_i'\beta)$. We now examine the consistency of $\beta\,\hat{}_n$ 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 $U(\gamma)=n^{-1}\sum_{i=1}^n w_i x_i^{*'}\{m_i\widehat{F}_n(y_i)^{-1} - e^{x_i^{*'}\gamma}\}$. It can be shown that the function U converges to 0 almost surely when evaluated at $\gamma=(\log\{\Lambda_0(\tau)/c\}, \beta')'$. Furthermore, it is easy to see that the derivative of U evaluated at $(\log\{\Lambda_0(\tau)/c\}, \beta')'$ is negative definite. Applying Taylor expansion to $U(\gamma)$, one can show that the solution of (2), that is $\widehat{\gamma}=(\widehat{\eta}_n, \widehat{\beta}_n)'$, 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 the above sketch proof, $\hat{\eta}_n$ converges to $\log\{\Lambda_0(\tau)/c\}$ almost surely. Along with 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 $t \in [0, \tau_2]$.

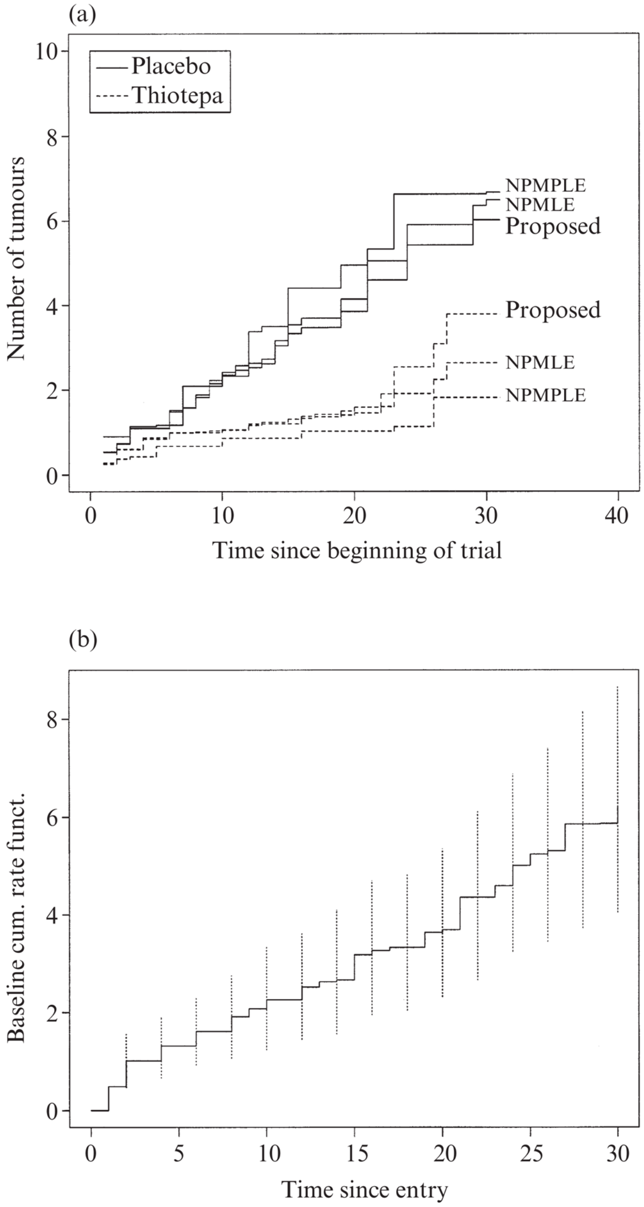

Fig. 1.

Bladder tumour study. (a) Nonparametric estimation of cumulative rate function by treatment group; (b) Semiparametric estimation of baseline cumulative rate function with pointwise bootstrap 95% confidence intervals.

Table 1

Simulation results for nonparametric estimators under the assumptions of independent and informative observation times

			(*a*) *Independent observation times*					
			Proposed		**NPMLE**		**NPMPLE**	
n	***t***	**$\Lambda_0(t)$**	**Bias**	**SE**	**Bias**	**SE**	**Bias**	**SE**
100	1·0	2	−0·031	0·397	−0·030	0·392	−0·060	0·418
	3·0	6	−0·006	0·488	−0·019	0·476	−0·058	0·633
	5·0	10	−0·024	0·542	−0·047	0·505	−0·063	0·723
	7·0	14	−0·015	0·654	−0·055	0·559	−0·065	0·844
	9·0	18	0·004	0·804	−0·051	0·643	−0·045	0·904
1000	1·0	2	0·011	0·156	0·010	0·154	0·007	0·200
	3·0	6	0·005	0·202	−0·002	0·189	−0·017	0·276
	5·0	10	0·014	0·210	−0·002	0·193	−0·030	0·336
	7·0	14	0·011	0·241	−0·009	0·211	−0·019	0·374
	9·0	18	0·025	0·277	−0·003	0·218	−0·023	0·408

			(*b*) *Informative observation times*					
			Proposed		**NPMLE**		**NPMPLE**	
n	***t***	**$\Lambda_0(t)$**	**Bias**	**SE**	**Bias**	**SE**	**Bias**	**SE**
100	1·0	2	−0·009	0·271	0·015	0·271	0·447	0·454
	3·0	6	−0·007	0·553	−0·040	0·512	0·361	0·795
	5·0	10	0·049	0·874	−0·543	0·722	−1·183	0·949
	7·0	14	0·008	1·198	−1·699	0·859	−3·395	1·067
	9·0	18	0·127	1·532	−2·987	1·021	−5·324	1·349
1000	1·0	2	0·000	0·104	0·021	0·101	0·484	0·194
	3·0	6	0·015	0·214	−0·021	0·182	0·431	0·345
	5·0	10	0·031	0·343	−0·509	0·258	−1·157	0·407
	7·0	14	0·035	0·459	−1·580	0·309	−3·321	0·459
	9·0	18	0·038	0·564	−2·901	0·340	−5·448	0·508

NPMLE, nonparametric maximum likelihood estimator; NPMPLE, nonparametric maximum pseudolikelihood estimator; Bias and SE are the Monte Carlo sample mean and standard deviation of the 1000 estimates of $\Lambda_0(t)$.

Table 2

Bias and standard error of $\hat{\beta}$ using the proposed semi-parametric method and the Sun–Wei estimators when the observation time process depends only on observed covariates

	***n*=100**			***n*=1000**		
	Proposed	**SW[a]**	**SW[b]**	**Proposed**	**SW[a]**	**SW[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

SW[a], the Sun–Wei estimator of β without modelling the observation pattern; SW[b], the Sun–Wei estimator with modelling the observation pattern; Bias and SE, Monte Carlo sample mean and standard deviation for the 1000 estimates.

Table 3

Bias and standard error of $\hat{\Lambda}_0(t)$ using the proposed semiparametric method when the observation time process depends only on observed covariates

		***n*=100**		***n*=1000**	
t	**$\Lambda_0(t)$**	**Bias**	**SE**	**Bias**	**SE**
1·0	2	0·005	0·394	0·006	0·136
3·0	6	0·024	0·879	0·004	0·301
5·0	10	0·093	1·352	0·033	0·464
7·0	14	0·175	1·887	0·073	0·594
9·0	18	0·204	2·550	0·082	0·814

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

Table 4

Bias and standard error of $\hat{\beta}$ from the proposed semiparametric method and the Sun–Wei estimators when the observation times are informative

	***n*=100**			***n*=1000**		
	Proposed	**SW^a**	**SW^b**	**Proposed**	**SW^a**	**SW^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

SW^a, the Sun–Wei estimator without modelling the observation pattern; SW^b, the Sun–Wei estimator with modelling the observation pattern; Bias and SE, Monte Carlo sample mean and standard deviation for the 1000 estimates of β.

Table 5

Bias and standard error of $\hat{\Lambda}_0(t)$ using the proposed semiparametric method when the observation times are informative

		***n*=100**		***n*=1000**	
t	**$\Lambda_0(t)$**	**Bias**	**SE**	**Bias**	**SE**
1·0	2	0·001	0·525	−0·005	0·132
3·0	6	0·045	1·698	0·003	0·220
5·0	10	0·088	2·560	0·036	0·296
7·0	14	0·142	3·663	0·010	0·375
9·0	18	0·217	4·529	0·033	0·463

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