Modeling event count data in the presence of informative dropout with application to bleeding and transfusion events in myelodysplastic syndrome

Guoqing Diao,[a*†] **Donglin Zeng,**[b] **Kuolung Hu**[c] **and Joseph G. Ibrahim**[b]

In many biomedical studies, it is often of interest to model event count data over the study period. For some patients, we may not follow up them for the entire study period owing to informative dropout. The dropout time can potentially provide valuable insight on the rate of the events. We propose a joint semiparametric model for event count data and informative dropout time that allows for correlation through a Gamma frailty. We develop efficient likelihood-based estimation and inference procedures. The proposed nonparametric maximum likelihood estimators are shown to be consistent and asymptotically normal. Furthermore, the asymptotic covariances of the finite-dimensional parameter estimates attain the semiparametric efficiency bound. Extensive simulation studies demonstrate that the proposed methods perform well in practice. We illustrate the proposed methods through an application to a clinical trial for bleeding and transfusion events in myelodysplastic syndrome.

Keywords: Cox model; informative dropout; nonparametric maximum likelihood estimators; Poisson regression model; semiparametric efficiency

1. Introduction

Recurrent events data refer to a subject experiencing repeated occurrences of the same or related types of events during the study time. The number of the event count during the time is clearly the interest in medical and epidemiological researches. Many clinical trials are designed to demonstrate treatment efficacy or safety based on the frequency of the events over time, defined as the event rate. Although the exposure period may be different from subject to subject, the frequency of events is usually summarized as a rate or number per year or per 100 subject years of exposure. Examples include occurrence of serious infections in clinical trials of AIDS prophylaxis [1], relapses of multiple sclerosis [2], the frequency of secondary infections during myelosuppressive chemotherapy treatment [3], or the occurrence of red blood transfusions [4].

The Poisson regression model is straightforward in analyzing this type of data, and it has been illustrated in many papers; for instance, see [5–9]. Typically, the event count data are assumed to follow a Poisson distribution conditional on the follow-up time, and the logarithm of the follow-up time is treated as a fixed intercept in the Poisson regression model. However, the number of events can potentially impact the dropout time, and additional information on the informative dropout time can provide more insight in the estimation and inference of the Poisson rates. It is therefore very important to account for potential correlation between the event count data and informative dropout time.

[a]*Department of Statistics, George Mason University, Fairfax, VA, U.S.A.*
[b]*Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC, U.S.A.*
[c]*Amgen Inc., One Amgen Center Drive, Thousand Oaks, CA, U.S.A.*
[*]*Correspondence to: Guoqing Diao, Department of Statistics, George Mason University, Fairfax, VA 22030, U.S.A.*
[†]*E-mail: gdiao@gmu.edu*

A motivating example involves a phase 2, multicenter, randomized, double-blind, placebo-controlled clinical trial with low or intermediate-1 risk myelodysplastic syndrome (MDS) patients that have severe thrombocytopenia (platelet counts $< 50 \times 109$/L). The key study objective is to evaluate the treatment effect on reduction of both platelet transfusion incidences and bleeding adverse events reported during the study efficacy follow-up (duration of 26 weeks). Living with abnormal platelet function, severe thrombocytopenia patients have an increased risk of bleeding incidence. Platelet transfusion interventions are often required therapeutically (given to patients who are actively bleeding) and prophylactically (to prevent future bleeding). Therefore, the incidence of platelet transfusion intervention and bleeding event occurrence are highly correlated. In addition, thrombocytopenia in the MDS population is associated with shortened survival and an increased risk of evolution to acute myeloid leukemia, part of the natural progression of MDS. During this period, patients may obtain disease progression and then discontinue the treatment and dropout of the study in order to receive the MDS disease-modifying treatment. There were also patients that were informatively censored owing to death or administrative withdrawal due to other types of adverse events.

Several authors have proposed nonparametric or semiparametric methods to jointly model recurrent event data and informative follow-up time. In particular, Wang *et al.* [10] and Huang and Wang [11] model the association between the intensity of the recurrent event process and the hazard of the failure time through a common subject-specific latent variable. They propose a 'borrow-strength estimation procedure' by first estimating the value of the latent variable from the recurrent event data and then use the estimated value in the failure time model. Alternatively, several authors have proposed to use frailty models to account for the correlation between the recurrent event process and terminal event process [12–17]. In these papers, the recurrent event times are assumed to be observed before the dropout time. A comprehensive review of statistical analysis of recurrent event data was provided in [18].

In many applications, we do not observe the event times directly. Instead, only the number of recurrent events during the follow-up period is available. Such data are often referred to as the panel count data. For correlated panel count data, Zhang and Jamshidian [19] and Yao *et al.* [20] proposed Gamma frailty Poisson models but assumed conditional independent observation processes. Among others, several studies [21–27] proposed semiparametric regression analysis of panel count data with dependent observation processes. Essentially all of these existing methods are based on estimating equations, and it is not clear whether the estimators are asymptotically efficient. Additionally the bootstrap method is often needed to estimate the standard errors of the parameter estimates. Furthermore, these papers focus on parameter estimation, and little has been investigated in comparing the event rates of two treatment groups. Sun and Zhao [28] gave a thorough review of the statistical analysis of panel count data.

In this paper, we develop likelihood-based methods to compare the event rates in the treatment group and placebo group adjusting for covariates and meanwhile accounting for informative dropout. In Section 2, we propose a joint model of the event count data and hazard of the dropout time and allow for correlation by using a Gamma frailty. Unlike most existing joint models (e.g., [21,23,24]), we model the cumulative mean count at each time instead of assuming a Poisson process for the recurrent data and modeling the whole intensity function. To achieve more flexibility and robustness, we consider a stratified Cox model for the informative dropout time, allowing for unspecified (conditional) baseline hazard in each treatment group. We derive the nonparametric likelihood function of the unknown parameters based on the observed data. In addition, we describe several tests for comparing the incidence rates between the treatment group and the placebo group. In Section 3, we establish the asymptotic properties of the proposed nonparametric maximum likelihood estimators (NPMLEs). The proposed NPMLEs are shown to be consistent and asymptotically normal. Furthermore, the asymptotic covariances of the finite-dimensional parameter estimates attain the semiparametric efficiency bound. Extensive simulation results are provided in Section 4. We illustrate the proposed methodology through the MDS clinical trial in Section 5. We conclude the paper with a few brief remarks in Section 6. All technical details of the proofs of the asymptotic results are relegated to the Appendix.

2. Methods

Suppose that we have n subjects in the data set, among which n_0 subjects were allocated to the control arm and n_1 subjects are allocated to the treatment arm. For the ith subject, $i = 1, \ldots, n$, let A_i denote the treatment indicator, which takes the value 0 for the control arm and 1 for the treatment arm; C_i is the administrating censoring time; $\widetilde{T}_i$ is the time to adverse event or death; $N_i(t)$ is the event count by time

point t; and $\mathbf{Z}_i$ is a $d \times 1$ vector of covariates at baseline. Therefore, the observed data are

$$\left\{ A_i, \mathbf{Z}_i, X_i = N_i(T_i), T_i \equiv \tilde{T}_i \wedge C_i, \Delta_i = I(\tilde{T}_i \leqslant C_i) \right\}, \quad i = 1, \ldots, n,$$

where the T_i's are the observed follow-up times, and Δ_i are the corresponding censoring indicators.

Denote the end of study by τ. For subject i, we assume that $N_i(t), t \in [0, \tau]$ follows a Poisson distribution with

$$E\{N_i(t)|\xi_i, A_i, \mathbf{Z}_i\} = \lambda_{A_i} \xi_i t \exp(\boldsymbol{\beta}^T \mathbf{Z}_i), \tag{1}$$

where λ_{A_i} is the conditional baseline event rate in group A_i, $\xi_i \sim Gamma(\theta^{-1}, \theta^{-1})$ represents the patient's heterogeneity due to other characteristics, and $\boldsymbol{\beta}$ is a set of regression coefficients. The gamma frailties ξ_i have mean 1 and variance θ and are mutually independent. It is worth to note that unconditional on ξ_i, the rate function satisfies $E\{N_i(t)/t|A_i, \mathbf{Z}_i\} = \lambda_{A_i} \exp(\boldsymbol{\beta}^T \mathbf{Z}_i)$.

While model (1) is a parametric model, we assume a semiparametric model for the informative dropout time. Specifically, we further assume that $\widetilde{T}_i$ is independent of $N_i(t)$ given ξ_i and the conditional hazard function of $\tilde{T}_i$ is

$$h(t|\xi_i, A_i, \mathbf{Z}_i) = \xi_i h_{A_i}(t) \exp(\boldsymbol{\zeta}^T \mathbf{Z}_i), \quad t \in [0, \tau], \tag{2}$$

where $\xi_i h_{A_i}(t)$ is the conditional baseline hazard function of $\tilde{T}_i$ given ξ_i and $\mathbf{Z}_i$, and $\boldsymbol{\zeta}$ is a set of regression coefficients. The functions $h_0(t)$ and $h_1(t)$ are unspecified. Model (2) is referred to as the Cox model with a shared Gamma frailty [29,30]. This model allows one to model the positive correlation between the event rate of the Poisson process and the hazard rate of the time to adverse event or death. Additionally, model (2) is also a stratified model allowing the (conditional) baseline hazards to be different in the control and treatment groups. Note that we may allow different covariates in models (1) and (2).

Write $H_a(t) = \int_0^t h_a(s)ds, a = 0, 1$. From model (2), we can show by using the Bayes rule that conditional on $\{A_i, \mathbf{Z}_i, T_i, \Delta_i\}$, the distribution of ξ_i is $\Gamma(\theta^{-1} + \Delta_i, \theta^{-1} + H_{A_i}(T_i)e^{\boldsymbol{\zeta}^T \mathbf{Z}_i})$. Then we can show that conditional on $\{A_i, \mathbf{Z}_i, T_i, \Delta_i\}$, the distribution of X_i is negative binomial with parameters $\theta^{-1} + \Delta_i$ and success probability

$$\frac{\lambda_{A_i} T_i e^{\boldsymbol{\beta}^T \mathbf{Z}_i}}{\theta^{-1} + H_{A_i}(T_i)e^{\boldsymbol{\zeta}^T \mathbf{Z}_i} + \lambda_{A_i} T_i e^{\boldsymbol{\beta}^T \mathbf{Z}_i}}.$$

The conditional mean and variance of X_i given $\{A_i, \mathbf{Z}_i, T_i, \Delta_i\}$ are given by

$$(\theta^{-1} + \Delta_i) \frac{\lambda_{A_i} T_i e^{\boldsymbol{\beta}^T \mathbf{Z}_i}}{\theta^{-1} + H_{A_i}(T_i)e^{\boldsymbol{\zeta}^T \mathbf{Z}_i}}, \tag{3}$$

and

$$(\theta^{-1} + H_{A_i}(T_i)e^{\boldsymbol{\zeta}^T \mathbf{Z}_i} + \lambda_{A_i} T_i e^{\boldsymbol{\beta}^T \mathbf{Z}_i})(\theta^{-1} + \Delta_i) \frac{\lambda_{A_i} T_i e^{\boldsymbol{\beta}^T \mathbf{Z}_i}}{(\theta^{-1} + H_{A_i}(T_i)e^{\boldsymbol{\zeta}^T \mathbf{Z}_i})^2}, \tag{4}$$

respectively. Therefore, by introducing a Gamma frailty, we allow for overdispersion in the recurrent events data. Furthermore, the Gamma frailty allows for correlation between the recurrent events and the follow-up time, and consequently, the distribution of X_i depends on the parameters not only in the Poisson model (1) but also in the survival model (2).

Write $\boldsymbol{\phi} = (\lambda_0, \lambda_1, \theta, \boldsymbol{\beta}, \boldsymbol{\zeta})$. Assume that, conditional on A and $\mathbf{Z}$, the censoring time C is independent of the failure time $\tilde{T}$ and the event count X. Based on the observed data, the likelihood function for the unknown parameters $(\boldsymbol{\phi}, H_0, H_1)$ is

$$\begin{aligned} L(\boldsymbol{\phi}, H_0, H_1) = \prod_{i=1}^{n} \int_{\xi_i} & \frac{(\lambda_{A_i} \xi_i T_i e^{\boldsymbol{\beta}^T \mathbf{Z}_i})^{X_i} \exp\{-\lambda_{A_i} \xi_i T_i e^{\boldsymbol{\beta}^T \mathbf{Z}_i}\}}{X_i!} \\ & \times \{h_{A_i}(T_i) \xi_i e^{\boldsymbol{\zeta}^T \mathbf{Z}_i}\}^{\Delta_i} \exp\{-H_{A_i}(T_i) \xi_i e^{\boldsymbol{\zeta}^T \mathbf{Z}_i}\} f(\xi_i) d\xi_i, \end{aligned} \tag{5}$$

where $f(\cdot)$ is the $Gamma(\theta^{-1}, \theta^{-1})$ density given by

$$f(\xi) = \frac{\theta^{-\theta^{-1}}}{\Gamma(\theta^{-1})}\xi^{\theta^{-1}-1}e^{-\xi/\theta}.$$

The likelihood (5) is equivalent to

$$L(\boldsymbol{\phi}, H_0, H_1) = \prod_{i=1}^{n} \frac{T_i^{X_i}}{X_i!}\left(\lambda_{A_i}e^{\boldsymbol{\beta}^T\mathbf{Z}_i}\right)^{X_i}\left\{h_{A_i}(T_i)e^{\boldsymbol{\zeta}^T\mathbf{Z}_i}\right\}^{\Delta_i}\frac{\Gamma(\theta^{-1}+X_i+\Delta_i)}{\Gamma(\theta^{-1})}$$
$$\times \quad \frac{\theta^{X_i+\Delta_i}}{\left[1+\theta\{\lambda_{A_i}T_ie^{\boldsymbol{\beta}^T\mathbf{Z}_i}+H_{A_i}(T_i)e^{\boldsymbol{\zeta}^T\mathbf{Z}_i}\}\right]^{\theta^{-1}+X_i+\Delta_i}}.$$

We propose to estimate $H_a(\cdot), a = 0, 1$ nonparametrically by leaving both functions unspecified. By using the nonparametric maximum likelihood approach and allowing $H_a(\cdot), a = 0, 1$ to be right continuous, we obtain the nonparametric likelihood function, still denoted by $L(\boldsymbol{\phi}, H_0, H_1)$, for ease of exposition,

$$L(\boldsymbol{\phi}, H_0, H_1) = \prod_{i=1}^{n} \frac{T_i^{X_i}}{X_i!}\left(\lambda_{A_i}e^{\boldsymbol{\beta}^T\mathbf{Z}_i}\right)^{X_i}\left(H_{A_i}\{T_i\}e^{\boldsymbol{\zeta}^T\mathbf{Z}_i}\right)^{\Delta_i}\frac{\Gamma(\theta^{-1}+X_i+\Delta_i)}{\Gamma(\theta^{-1})}$$
$$\times \quad \frac{\theta^{X_i+\Delta_i}}{\left[1+\theta\{\lambda_{A_i}T_ie^{\boldsymbol{\beta}^T\mathbf{Z}_i}+H_{A_i}(T_i)e^{\boldsymbol{\zeta}^T\mathbf{Z}_i}\}\right]^{\theta^{-1}+X_i+\Delta_i}},$$

where $H_a\{t\}, a = 0, 1$, is the jump size of $H_a(\cdot)$ at t. After some simple algebra, we can show that the log-likelihood function of $(\boldsymbol{\phi}, H_0, H_1)$ is given by

$$l(\boldsymbol{\phi}, H_0, H_1) = c + \sum_{i=1}^{n}\left[X_i(\log\lambda_{A_i} + \boldsymbol{\beta}^T\mathbf{Z}_i) + \Delta_i[\log H_{A_i}\{T_i\} + \boldsymbol{\zeta}^T\mathbf{Z}_i] + \sum_{k=0}^{X_i+\Delta_i-1}\log(\theta^{-1}+k)\right.$$
$$\left. + (X_i+\Delta_i)\log\theta - (\theta^{-1}+X_i+\Delta_i)\log\{1+\theta\lambda_{A_i}T_ie^{\boldsymbol{\beta}^T\mathbf{Z}_i}+\theta H_{A_i}(T_i)e^{\boldsymbol{\zeta}^T\mathbf{Z}_i}\}\right],$$

where c is a constant. We maximize $l(\boldsymbol{\phi}, H_0, H_1)$ and obtain the NPMLEs of $(\boldsymbol{\phi}, H_0, H_1)$, denoted by $(\hat{\boldsymbol{\phi}}_n, \hat{H}_{0n}, \hat{H}_{1n})$. The maximization can be accomplished by using an iterative procedure such as the quasi-Newton algorithm Press *et al.* [31]. The quasi-Newton algorithm has been successfully applied in optimization problems with a large number of parameters [32, 33] and has been implemented in software packages such as SAS, R, and MATLAB. Section 3 establishes the consistency and asymptotic normality of the NPMLEs.

Note that our main interest is to test $\lambda_0 = \lambda_1$. We consider several different tests: (i) likelihood ratio test; (ii) Wald test based on $\hat{\lambda}_0 - \hat{\lambda}_1$; (iii) Wald test based on $\hat{\lambda}_0/\hat{\lambda}_1$; and (iv) Wald test based on $\log\hat{\lambda}_0 - \log\hat{\lambda}_1$. It can be shown from the asymptotic results in Section 3 that under the null hypothesis, all four test statistics are asymptotically chi-square with 1 degree of freedom.

Remark 2.1

When the follow-up time is short, we can assume $h_0(t) = h_0$ and $h_1(t) = h_1$. Furthermore, if we assume that there are no covariates effects and reparameterize

$$p_a = \lambda_a/(\lambda_a + h_a), \quad \nu_a = \theta(\lambda_a + h_a),$$

for $a = 0, 1$, then the likelihood function for the unknown parameters $\boldsymbol{\phi} \equiv (p_0, \nu_0, p_1, \nu_1, \theta)$ takes the form

$$L(\boldsymbol{\phi}) = \prod_{i=1}^{n}\binom{X_i+\Delta_i}{X_i}p_{A_i}^{X_i}(1-p_{A_i})^{\Delta_i}\frac{\Gamma(\theta^{-1}+X_i+\Delta_i)}{(X_i+\Delta_i)!\Gamma(\theta^{-1})}\frac{(\nu_{A_i}T_i)^{X_i+\Delta_i}}{\left[1+\nu_{A_i}T_i\right]^{\theta^{-1}+X_i+\Delta_i}}.$$

Concerning the inference for the parameters, we can assume that the T's are fixed and that the data are generated from the following mechanism: in arm $A = 0$, conditional on $T_i, X_i+\Delta_i$ follows a negative binomial distribution $NB(\theta^{-1}, \nu_0 T_i/(1 + \nu_0 T_i))$ and conditional on $X_i + \Delta_i$, X_i follows a binomial distribution, $Binomial(X_i + \Delta_i, p_0)$. This also holds for $A = 1$.

3. Asymptotic results

In this section, we establish the asymptotic properties of the proposed NPMLEs. In particular, we will prove the following three theorems.

Theorem 1
Under the conditional independent censoring assumption and conditions (C1)–(C4) in Appendix A, the unknown parameters $(\boldsymbol{\phi}, H_0, H_1)$ are identifiable.

The proof of Theorem 1 is given in Appendix B.

Theorem 2
Under the conditional independent censoring assumption and conditions (C1)–(C4) in Appendix A, with probability tending to 1,

$$||\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_0|| + \sup_{t\in[0,\tau_a]} |\hat{H}_{0n}(t) - H_{00}(t)| + \sup_{t\in[0,\tau_a]} |\hat{H}_{1n}(t) - H_{01}(t)| \to 0,$$

where $||\cdot||$ is the Euclidean norm.

Remark 3.1
Theorem 2 states the consistency of the NPMLEs. The basic idea for proving Theorem 2 is as follows. We first show that neither $\hat{H}_{0n}(\tau)$ nor $\hat{H}_{1n}(\tau)$ is allowed to diverge by using the partitioning argument described in [29]. Once the boundedness of $\hat{H}_{0n}(\tau)$ and $\hat{H}_{1n}(\tau)$ is established, a subsequence of $\hat{H}_{an}, a = 0, 1$ can be found to converge pointwise to a bounded monotone function H_a^* in $[0, \tau]$ and the same subsequence of $\hat{\boldsymbol{\phi}}_n$ to some $\boldsymbol{\phi}^*$. We construct a step function $\bar{H}_{an}$ with jumps at the observed failure times converging to H_{a0}. Then, as $l_n(\hat{\boldsymbol{\phi}}_n, \hat{H}_{0n}, \hat{H}_{1n}) \geqslant l_n(\boldsymbol{\phi}_0, \bar{H}_0 n, \bar{H}_{1n})$, by taking the limit, we prove that the Kullback–Leibler information between the true density and the density indexed by $(\boldsymbol{\phi}^*, H_0^*, H_1^*)$ is non-positive. Therefore, the true density must be equal to the density indexed by $(\boldsymbol{\phi}^*, H_0^*, H_1^*)$. Consistency will then follow from the identifiability result. The details of the proofs are given in Appendix C.

Theorem 3
Under the conditional independent censoring assumption and conditions (C1)–(C4) in Appendix A,

$$n^{1/2}(\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_0, \hat{H}_{0n} - H_{00}, \hat{H}_{1n} - H_{01}) \to_d \mathcal{G},$$

where $\mathcal{G}$ is a continuous zero-mean Gaussian process in the metric space $l^\infty(\mathcal{H})$ and

$$\begin{aligned}\mathcal{H} =&\{(\mathbf{h}_1, h_2, h_3) : \mathbf{h}_1 \in R^{3+2d}, h_2 \text{ and } h_3 \text{ are functions on } [0, \tau];\\ &||h_1|| \leqslant 1, |h_2|_V \leqslant 1, |h_3|_V \leqslant 1\}.\end{aligned}$$

Here $|h|_V$ denotes the total variation of h in $[0, \tau]$ and $l^\infty(\mathcal{H})$ is the collection of all bounded functions on $\mathcal{H}$. Additionally, $\hat{\boldsymbol{\phi}}_n$ is asymptotically efficient.

Remark 3.2
In the statement of Theorem 3, asymptotically efficient estimators mean that the asymptotic covariances attain the semiparametric efficiency bound [34, Chapter 3]. Once the consistency of the NPMLEs is established, the asymptotic distribution of the NPMLEs stated in Theorem 3 can be derived by verifying the four conditions in Theorem 3.3.1 of [35]. The proof of Theorem 3 is given in Appendix D.

Remark 3.3
Theorem 3 implies that for any $(\mathbf{h}_1, h_2, h_3) \in \mathcal{H}$, $\sqrt{n}(\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_0)^T\mathbf{h}_1 + \sqrt{n}\int_0^\tau h_2(t)d(\hat{H}_{0n} - H_{00}) + \sqrt{n}\int_0^\tau h_3(t)d(\hat{H}_{1n} - H_{10})$ is asymptotically normal with mean zero. To estimate the covariance matrix of $(\hat{\boldsymbol{\phi}}_n, \hat{H}_{0n}, \hat{H}_{1n})$, we view (5) as a parametric likelihood with $\boldsymbol{\phi}$ and the jump sizes of H_0 and H_1 at the

observed failure times as parameters. Following the arguments in Theorem 2 of [36], we can then estimate the asymptotic covariance matrix of the unknown parameters by inverting the observed information matrix according to parametric likelihood theory.

Remark 3.4
Using the arguments in [37], we can prove that under the conditional independent censoring assumption and conditions (C1)–(C4) in Appendix A, the asymptotic null distribution of the likelihood ratio test statistic for testing $\theta = 0$ is a 50:50 mixture of χ_0^2 and χ_1^2.

4. Simulation studies

We conduct extensive simulation studies to evaluate the finite-sample performance of the MLEs of $\boldsymbol{\phi}$ under the proposed model. We generate data under the proposed joint model and set the true parameters as $(\lambda_0, \lambda_1) = (1, 1)$. We generate the follow-up time from a Weibull distribution with $H_a(t) = 0.5t^b$ and vary b among $(0.7, 1.0, 1.3)$. The variance of the Gamma frailty varies from 0.5 to 1. We generate 100 subjects for each treatment group, that is, $n_0 = n_1 = 100$. We use the quasi-Newton algorithm to maximize the nonparametric maximum likelihood function. The covariance matrix of the NPMLEs is estimated by inverting the observed Fisher information matrix, that is, the negative second derivatives of the nonparametric log-likelihood with respect to the finite-dimensional parameters and the jump sizes of $H_0(\cdot)$ and $H_1(\cdot)$ at the observed dropout times. The quasi-Newton algorithm converges quickly and is robust to the choices of the initial values of the unknown parameters. It takes about 0.01 s to analyze one simulated data set (including covariance matrix estimation) under the aforementioned simulation setting on a MacBook Pro with 2.8-GHz Intel Core i7 processors.

Tables I presents the summary statistics of the NPMLEs of the unknown parameters based on 1000 replicates for $b = 0.7$. Throughout this section, $\gamma \equiv \theta^{-1}$. The proposed NPMLEs have small biases, the estimated standard errors reflect the actual variation of the estimators, and the coverage probabilities of the

Table I. Summary statistics of the NPMLEs with $n_0 = n_1 = 100$ and $H_a(t) = 0.5t^{0.7}$.

Parameter	True	Bias	SE	SEE	CP
		$\theta = 0.5$			
γ	2	0.140	0.438	0.426	0.942
λ_0	1	−0.003	0.120	0.120	0.954
λ_1	1	−0.007	0.125	0.120	0.930
$H_0(0.2)$	0.162	0.000	0.043	0.044	0.958
$H_0(0.3)$	0.215	0.000	0.051	0.052	0.966
$H_0(0.4)$	0.263	−0.002	0.058	0.059	0.961
$H_0(0.5)$	0.308	−0.002	0.065	0.066	0.950
$H_1(0.2)$	0.162	0.001	0.046	0.044	0.946
$H_1(0.3)$	0.215	0.001	0.056	0.052	0.953
$H_1(0.4)$	0.263	0.000	0.063	0.059	0.950
$H_1(0.5)$	0.308	0.000	0.070	0.066	0.941
		$\theta = 1.0$			
γ	1	0.055	0.172	0.168	0.941
λ_0	1	−0.012	0.155	0.152	0.934
λ_1	1	−0.007	0.158	0.153	0.944
$H_0(0.2)$	0.162	−0.001	0.047	0.046	0.950
$H_0(0.3)$	0.215	−0.001	0.059	0.056	0.938
$H_0(0.4)$	0.263	0.000	0.067	0.064	0.941
$H_0(0.5)$	0.308	0.000	0.075	0.072	0.945
$H_1(0.2)$	0.162	0.001	0.048	0.046	0.936
$H_1(0.3)$	0.215	0.001	0.059	0.056	0.948
$H_1(0.4)$	0.263	0.000	0.068	0.064	0.945
$H_1(0.5)$	0.308	0.001	0.075	0.072	0.947

Bias and SE denote the bias and standard deviation of the nonparametric maximum likelihood estimators (NPMLEs), SEE is the average of the standard error estimates, and CP is the coverage probability of 95% confidence interval.

95% confidence intervals are close to the nominal level. Results for the scenarios when $b = 1$ and $b = 1.3$ are presented in Tables S1 and S2, and similar results are obtained. We conducted additional simulations by increasing the sample size from 100 to 200 in each treatment group. The results are summarized in Tables S3–S5. Biases and standard errors of the NPMLEs, and the coverage probabilities of the 95% confidence intervals improve as sample size increases. The standard errors decrease by a factor of $\sqrt{2}$ suggesting root-n convergence rate of the NPMLEs.

For comparison, we also consider the parametric method described in Section 2 in which $H_a(t), a = 0, 1$ are assumed to be constant. Table II presents the results based on 1000 replicates with $n_0 = n_1 = 100$ comparing the performance of the NPMLEs and the parametric MLEs. The NPMLEs are robust to departures from the parametric assumption on the baseline hazards and has little loss of efficiency compared with the parametric MLEs when the parametric assumption is true (i.e., when $b = 1$). On the other hand, the parametric method tends to have biased estimators under model mis-specification especially for the case of $b = 1.3$. When $b = 0.7$, the parametric MLEs of λ_0 and λ_1 have larger standard errors than the NPMLEs, whereas the standard error of parametric MLE of γ is larger when $b = 1.3$. It appears that the value of θ does not impact the relative efficiency much.

In the second set of simulations, we compare the performance of different tests for testing $\lambda_0 = \lambda_1$ under different scenarios. We also consider a naïve method by fitting a Poisson model to the event count data directly. Specifically, we fix the true values of (θ, h_0, h_1) at $(0.5, 0.2, 0.2)$ and vary the values of (λ_0, λ_1), as follows:

(a) $\lambda_0 = 0.1, \quad \lambda_1 = 0.1, 0.125, 0.15, 0.175, 0.2;$

(b) $\lambda_0 = 0.5, \quad \lambda_1 = 0.5, 0.625, 0.75, 0.875, 1.0;$ and

(c) $\lambda_0 = 1, \quad \lambda_1 = 1, 1.25, 1.5, 1.75, 2.$

Scenario (a) corresponds to the case with rare event, whereas the event of interest is more common under the other two scenarios. The censoring time was set at $C = 10$ yielding an approximate 25%

Table II. Comparison of the NPMLEs and parametric MLEs.

		Semiparametric			Parametric			
Parameter	True	Bias	SE	MSE	Bias	SE	MSE	RE
				$\theta = 0.5, H_a(t) = 0.5t^{0.7}$				
γ	2	0.140	0.438	0.211	−0.246	0.330	0.169	0.799
λ_0	1	−0.003	0.120	0.014	0.083	0.135	0.025	1.753
λ_1	1	−0.007	0.125	0.016	0.081	0.143	0.027	1.731
				$\theta = 0.5, H_a(t) = 0.5t$				
γ	2	0.133	0.456	0.225	0.092	0.413	0.179	0.796
λ_0	1	−0.006	0.121	0.015	0.001	0.120	0.014	0.983
λ_1	1	−0.001	0.119	0.014	0.005	0.119	0.014	0.994
				$\theta = 0.5, H_a(t) = 0.5t^{1.3}$				
γ	2	0.126	0.433	0.203	0.621	0.563	0.702	3.459
λ_0	1	−0.007	0.121	0.015	−0.064	0.112	0.017	1.137
λ_1	1	−0.009	0.118	0.014	−0.065	0.110	0.016	1.172
				$\theta = 1, H_a(t) = 0.5t^{0.7}$				
γ	1	0.055	0.172	0.032	−0.106	0.138	0.030	0.932
λ_0	1	−0.012	0.155	0.024	0.116	0.188	0.049	2.002
λ_1	1	−0.007	0.158	0.025	0.121	0.194	0.052	2.075
				$\theta = 1, H_a(t) = 0.5t$				
γ	1	0.040	0.169	0.030	0.027	0.160	0.026	0.874
λ_0	1	−0.007	0.153	0.023	0.001	0.152	0.023	0.984
λ_1	1	0.000	0.157	0.025	0.010	0.155	0.024	0.990
				$\theta = 1, H_a(t) = 0.5t^{1.3}$				
γ	1	0.050	0.157	0.027	0.212	0.184	0.079	2.924
λ_0	1	−0.009	0.153	0.024	−0.076	0.142	0.026	1.101
λ_1	1	−0.007	0.153	0.024	−0.075	0.140	0.025	1.077

Bias and SE denote the bias and standard deviation of the estimators, MSE is the mean squared error of the estimators, and RE is MSE relative efficiency of the proposed non-parametric maximum likelihood estimators (NPMLEs) compared with the parametric MLEs.

Table III. Type I error rate and powers of different tests for testing $\lambda_0 = \lambda_1$ based on 1000 replicates with $n_0 = n_1 = 100$.

λ_1/λ_0	SPLRT	SPWD	SPWR	SPWLR	PLRT	PWD	PWR	PWLR
				$\lambda_0 = 0.1$				
1.00	0.050	0.045	0.215	0.047	0.077	0.074	0.205	0.074
1.25	0.158	0.144	0.332	0.155	0.201	0.2	0.353	0.199
1.50	0.399	0.387	0.658	0.396	0.503	0.498	0.683	0.492
1.75	0.648	0.638	0.846	0.648	0.759	0.756	0.870	0.751
2.00	0.854	0.854	0.950	0.854	0.913	0.911	0.967	0.910
				$\lambda_0 = 0.5$				
1.00	0.061	0.058	0.141	0.062	0.216	0.214	0.361	0.214
1.25	0.327	0.324	0.430	0.332	0.599	0.600	0.696	0.599
1.50	0.788	0.788	0.837	0.792	0.932	0.932	0.953	0.931
1.75	0.952	0.953	0.968	0.954	0.996	0.996	0.998	0.996
2.00	0.997	0.998	0.999	0.998	0.999	0.999	0.999	0.999
				$\lambda_0 = 1.0$				
1.00	0.049	0.045	0.048	0.049	0.284	0.283	0.338	0.283
1.25	0.393	0.397	0.308	0.403	0.760	0.760	0.777	0.760
1.50	0.849	0.854	0.749	0.859	0.980	0.980	0.980	0.980
1.75	0.987	0.988	0.961	0.987	1.000	1.000	0.999	1.000
2.00	0.999	0.999	0.995	0.999	1.000	1.000	1.000	1.000

SPLRT refers to the semiparametric likelihood ratio test; SPWD refers to the semiparametric Wald test based on difference of rates; SPWR refers to the semiparametric Wald test based on ratio of rates; SPWLR refers to the semiparametric Wald test based on log-ratio of rates; PLRT refers to the parametric likelihood ratio test; PWD refers to the parametric Wald test based on difference of rates; PWR refers to the parametric Wald test based on ratio of rates; and PWLR refers to the parametric Wald test based on log-ratio of rates.

censoring rate. Table III presents the type I error rates and powers of various tests for testing $H_0 : \lambda_0 = \lambda_1$ at significance level of 0.05 based on 1000 replicates with $n_0 = n_1 = 100$ under scenarios (a)–(c), respectively. In particular, we consider four tests—namely, the likelihood ratio test, the Wald test based on difference of rates, the Wald test based on ratio of rates, and the Wald test based on log-ratio of rates—under both the proposed semiparametric model and the naïve parametric Poisson model. The naïve tests lead to inflated type I error rates especially when the rates are large. The Wald test based on the rate ratio tends to yield inflated type I error rates for rare events and loses power compared with other tests when the event rate is high. All other three tests perform similarly under the semiparametric model.

Finally, we conducted simulation studies to examine the performance of the likelihood ratio test statistics for testing $\theta = 0$, that is, whether there is correlation between the two processes. Figure S1 presents the type I error rates and powers of the likelihood ratio test at significance level of 0.05 based on 1000 replicates with $n_0 = n_1 = 200$ at three different values of $\lambda = \lambda_0 = \lambda_1$, 0.1, 0.5, and 1.0. The likelihood ratio tests have correct control of the type I error rate and have reasonable powers under the alternative hypotheses.

5. Myelodysplastic syndrome trial

We apply the proposed methodology to the motivating example described in Section 1. The trial was conducted in multiple countries, and a 2 to 1 randomization ratio (treatment vs. placebo) was adopted. Furthermore, the randomization stratifications were based on baseline platelet counts (50–25 × 109/L and below 25 × 109/L) and disease risk status (low and intermediate-1). One hundred fifty patients (100 treatment and 50 placebo) with at least 1 week of exposure are further summarized. Thirty-six treatment patients (36%) and 24 placebo patients (48%) dropped out early and during the efficacy follow-up phase. The major reason for this is due to alternative therapy. The median treatment duration is 113 days, and the total number of bleeding or platelet transfusion events ranges from 0 to 45 per patient. Table IV presents the frequency and relative frequency of the number of events in the control group and treatment group. Figure 1 presents the Kaplan–Meier survival curves of the failure time in each group and their 95% confidence bands. It appears that there is little difference between the two survival curves (p-value of log-rank test = 0.677).

Table IV. Frequencies and relative frequencies of number of events in the MDS trial.

	Control		Treatment	
Count	Frequency	Proportion	Frequency	Proportion
0	13	0.26	49	0.49
1	11	0.22	17	0.17
2	3	0.06	2	0.02
3	4	0.08	7	0.07
5	1	0.02	2	0.02
6	1	0.02	4	0.04
7	2	0.04	3	0.03
9	1	0.02	3	0.03
11	1	0.02	3	0.03
12	2	0.04	2	0.02
>12	11	0.22	8	0.08

MDS, myelodysplastic syndrome.

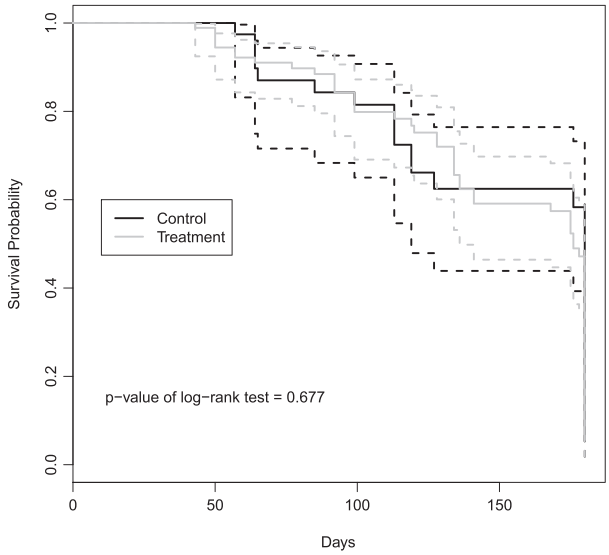

Figure 1. Kaplan–Meier survival curves of the failure time in each treatment group and their 95% confidence bands.

The main objective of the MDS trial is to compare the event rates in the treat group and the placebo group. We now apply the proposed methodology to this data set. For comparison, we also fit a standard Poisson model to the data directly. Table V presents the parameter estimates and their standard errors. The variance of the Gamma frailty is estimated as 1.44 with a standard error of 0.186. There is strong evidence that the event count data and the hazard rate of the dropout time are positively correlated (p-value $< 1.0\times10^{-10}$). The naïve method, which does not account for such a correlation, leads to underestimated standard errors of the parameter estimates. Consequently, all four tests described in Section 4 based on the naïve method for testing $H_0 : \lambda_0 = \lambda_1$ lead to p-values smaller than 0.0001. On the other hand, we obtain non-significant results at $\alpha = 0.05$ from the likelihood ratio test, the Wald test based on the rate difference, and the Wald test based on the log-ratio of the rates by using the proposed method, with p-values 0.074, 0.081, and 0.052, respectively. The Wald test based on the ratio of the rates has a p-value smaller than 0.0001.

For comparison, we analyzed the MDS trial data using two existing methods, HWZ [21] and YWH [20]. The HWZ method, implemented in R package *spef*, accounts for dependent observation processes, whereas the second one, implemented in R package *PCDSpline*, assumes conditional independent observations. In both methods, treatment indicator was included in the models as the only covariate, and Wald test statistics were for testing the regression coefficient (i.e., $\log \lambda_0/\lambda_1 = 0$). Note that for the method

Table V. Results of MDS trial.

	Naïve		New	
Parameter	Estimate	SE	Estimate	SE
θ			1.443558	0.186276
λ_0	0.058131	0.003270	0.057941	0.010688
λ_1	0.033841	0.001711	0.037557	0.004722

MDS, myelodysplastic syndrome; SE, standard error.

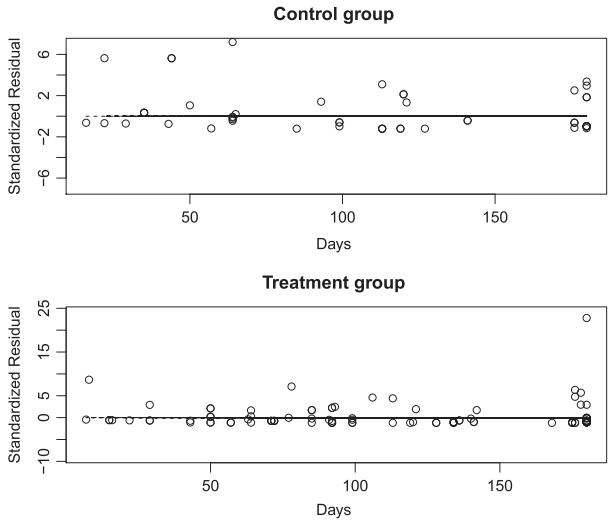

Figure 2. Plot of standardized residuals against follow-up times for the MDS trial.

of HWZ, bootstrapping is used to calculate the standard errors. The method of HWZ yields a p-value of 0.026 for testing the treatment effect. On the other hand, the method of YWH yields a p-value of 2.2×10^{-10} for testing the treatment effect. Although the truth is unknown here, as is evident from the simulation studies, ignoring the correlation between the event count data and dropout time can lead to false positive results, and the Wald statistic based on the ratio of the rates can also lead to false positive results when the event rates are small, which is exactly the case in this application.

Finally, we use a graphical procedure to check the goodness-of-fit of the proposed model. Specifically, we plot the standardized residuals against the follow-up times. As there are no additional covariates in the model, based on equations (3) and (4), the standardized residual for the ith subject is calculated as

$$\hat{\epsilon}_{ni} = (X_i - \hat{\mu}_{ni})/\sqrt{\hat{v}_{ni}},$$

where

$$\hat{\mu}_i = (\hat{\theta}_n^{-1} + \Delta_i)\frac{\hat{\lambda}_{A_i n} T_i}{\hat{\theta}_n^{-1} + \hat{H}_{A_i n}(T_i)},$$

and

$$\hat{v}_{ni} = (\hat{\theta}_n^{-1} + \hat{H}_{A_i n}(T_i) + \hat{\lambda}_{A_i n} T_i)(\hat{\theta}_n^{-1} + \Delta_i)\frac{\hat{\lambda}_{A_i n} T_i}{\{\hat{\theta}_n^{-1} + \hat{H}_{A_i n}(T_i)\}^2}$$

are the conditional mean and variance of X_i given $\{A_i, T_i, \Delta_i\}$, respectively, evaluated at the NPMLEs.

Figure 2 presents the plot of standardized residuals against the follow-up times for each treatment group. The majority of the residuals appear to be scattered around zero in both groups except for one possible outlier in the treatment group, which was from a subject with a very large event count ($X_i = 45$), and $\tilde{T}_i$ was censored at 180 days. After removing this outlier, the proposed model appears to fit the data reasonably well (Figure 3); however, we do observe smaller variations for the negative residuals, which were likely due to an excessive number of zero events in both groups. We address this issue with more detail in Section 6.

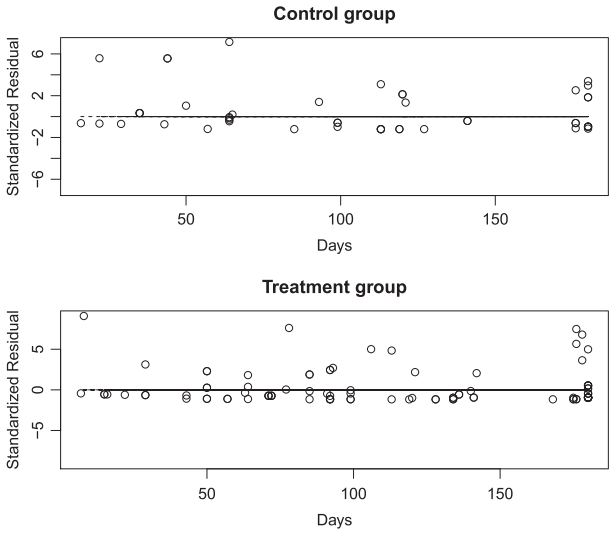

Figure 3. Plot of standardized residuals against follow-up times for the MDS trial after removing an outlier in the treatment group with $X_i = 45, T_i = 180$ (days), and $\Delta_i = 0$.

6. Discussion

We have developed a joint semiparametric model for recurrent events data and informative dropout time. This joint model allows one to model these two types of data simultaneously, and the correlation between them is modeled through a Gamma frailty. The proposed method allows one to utilize the additional information in the informative dropout time data to make inference on the event rate. Ignoring such information can lead to invalid inference and inflated type I error rates when comparing the event rates from two groups.

We assumed a gamma shared frailty to account for the correlation between the event count data and the informative dropout time. One advantage of by using the Gamma frailty is that there is a closed form for the observed data likelihood. The joint Gamma frailty model can be easily generalized to different frailty distributions such as the log-normal distribution and the positive stable distribution. However, under these frailty models, the numerical approximations or Monte Carlo approximations are needed to compute the likelihood functions, and hence maximization of the likelihood function is computationally intensive. One drawback of using the shared frailty in models (1) and (2) is that it induces positive correlations, which can be restrictive in practice. Following the idea of [12], we may modify model (2) such that

$$\lambda(t|\xi_i, A_i, \mathbf{Z}_i) = \xi_i^{\alpha} h_{A_i}(t) \exp(\boldsymbol{\zeta}^T \mathbf{Z}_i), \quad t \in [0, \tau],$$

where α is an unknown shape parameter. With this modeling, both positive and negative correlations are allowed. However, the likelihood function does not have a closed form except when α is fixed at 1; therefore, it will be computationally more intensive to maximize the likelihood.

In the application, a large portion of the patients had zero events in both the treatment group and the control group. While the proposed method allows for overdispersion in modeling the event count data, it does not handle count data that have an excess of zero counts. We describe a graphical procedure to check the goodness-of-fit of the proposed model through residual plots. It would be desirable to develop formal model diagnostic procedures to examine the goodness-of-fit of the proposed model and further extend the proposed model to event count data with an excess of zero counts. Future research is warranted in this direction.

In model (1), the event rate is proportional to the length of the follow-up time. To improve the flexibility and robustness of the model, we may consider the following semiparametric proportional rate model,

$$E[N_i(t)|\xi_i, A_i, \mathbf{Z}_i] = \lambda_{A_i} \xi_i G(t) \exp(\boldsymbol{\beta}^T \mathbf{Z}_i), \tag{6}$$

where $G(t)$ is an unknown increasing function in $[0, \tau]$ with $G(0) = 0$ and λ_0 is set to be 1 in order to ensure identifiability of the model. Several authors, including [20, 38–41], proposed spline-based methods to

estimate the unspecified baseline mean function $G(\cdot)$ assuming non-informative dropout. Splines provide a smooth estimate of the unknown functions, but they require selection of knots and splines orders. On the other hand, such selection is not needed in the nonparametric maximum likelihood approach. It would be interesting to develop efficient likelihood-based inference procedures for the joint models of (2) and (6). This is a topic of current research.

Appendix A. Regularity conditions

Besides the conditional independent censoring assumption described in Section 2, we impose the following regularity conditions needed for proving Theorems 1–1:

(C1) If there exist c_0 and $\mathbf{c}_1$ such that $\mathbf{c}_1^T\mathbf{Z} = c_0$ with probability one, then $c_0 = 0$ and $\mathbf{c}_1 = \mathbf{0}$.
(C2) There exists some positive constant number δ_0 such that $P(C \geqslant \tau|A, \mathbf{Z}) = P(C = \tau|A, \mathbf{Z}) \geqslant \delta_0$ almost surely, where τ is a constant denoting the end of the study.
(C3) The true parameter values of $\lambda_0, \lambda_1, \theta, \boldsymbol{\beta}$, and $\boldsymbol{\zeta}$ belong to a known compact set $\mathcal{B}_0$ in $(R^+)^3 \times R^{2d}$.
(C4) The true baseline cumulative distribution functions $H_{0a}, a = 0, 1$ belong to the following class:

$$\begin{aligned}\mathcal{A}_0 =&\{\Lambda : \Lambda \text{ is a strictly increasing function in } [0, \tau] \text{ and is continuously differentiable} \\ &\text{with } \Lambda(0) = 0, \Lambda'(0) > 0 \text{ and } \Lambda(\tau) < \infty\}.\end{aligned}$$

Appendix B. Proof of Theorem 1

Suppose that two sets of parameters, $(\lambda_0, H_0, \lambda_1, H_1, \theta, \boldsymbol{\beta}, \boldsymbol{\zeta})$ and $(\tilde{\lambda}_0, \tilde{H}_0, \tilde{\lambda}_1, \tilde{H}_1, \tilde{\theta}, \tilde{\boldsymbol{\beta}}, \tilde{\boldsymbol{\zeta}})$, give the same likelihood function for the observed data, that is,

$$\begin{aligned}&\left(\lambda_A e^{\boldsymbol{\beta}^T\mathbf{Z}}\right)^X \left\{h_A(T)e^{\boldsymbol{\zeta}^T\mathbf{Z}}\right\}^\Delta \frac{\Gamma(\theta^{-1} + X + \Delta)}{\Gamma(\theta^{-1})} \frac{\theta^{X+\Delta}}{\left[1 + \theta\{\lambda_A T e^{\boldsymbol{\beta}^T\mathbf{Z}} + H_A(T)e^{\boldsymbol{\zeta}^T\mathbf{Z}}\}\right]^{\theta^{-1}+X+\Delta}} \\ &= \left(\tilde{\lambda}_A e^{\tilde{\boldsymbol{\beta}}^T\mathbf{Z}}\right)^X \left\{\tilde{h}_A(T)e^{\tilde{\boldsymbol{\zeta}}^T\mathbf{Z}}\right\}^\Delta \frac{\Gamma(\tilde{\theta}^{-1} + X + \Delta)}{\Gamma(\tilde{\theta}^{-1})} \frac{\tilde{\theta}^{X+\Delta}}{\left[1 + \tilde{\theta}\{\tilde{\lambda}_A T e^{\tilde{\boldsymbol{\beta}}^T\mathbf{Z}} + \tilde{H}_A(T)e^{\tilde{\boldsymbol{\zeta}}^T\mathbf{Z}}\}\right]^{\tilde{\theta}^{-1}+X+\Delta}}.\end{aligned} \tag{B.1}$$

Equation (B.1) holds also for an event count of $X + 1$. It follows that

$$\lambda_A e^{\boldsymbol{\beta}^T\mathbf{Z}} \frac{1 + \theta(X + \Delta)}{1 + \theta\{\lambda_A T e^{\boldsymbol{\beta}^T\mathbf{Z}} + H_A(T)e^{\boldsymbol{\zeta}^T\mathbf{Z}}\}} = \tilde{\lambda}_A e^{\tilde{\boldsymbol{\beta}}^T\mathbf{Z}} \frac{1 + \tilde{\theta}(X + \Delta)}{1 + \tilde{\theta}\{\tilde{\lambda}_A T e^{\tilde{\boldsymbol{\beta}}^T\mathbf{Z}} + \tilde{H}_A(T)e^{\tilde{\boldsymbol{\zeta}}^T\mathbf{Z}}\}}. \tag{B.2}$$

The preceding equation holds for both $\Delta = 0$ and $\Delta = 1$. Therefore,

$$\frac{1 + \theta(X + 1)}{1 + \theta X} = \frac{1 + \tilde{\theta}(X + 1)}{1 + \tilde{\theta}X}.$$

Simple algebra yields $\theta = \tilde{\theta}$. Now let $T = 0$. From Equation (B.2), immediately, we obtain $\lambda_A e^{\boldsymbol{\beta}^T\mathbf{Z}} = \tilde{\lambda}_A e^{\tilde{\boldsymbol{\beta}}^T\mathbf{Z}}$. As this equation holds for any $\mathbf{Z}$ and $\mathbf{Z}^*$ in the support of the covariates, we obtain $\boldsymbol{\beta}^T(\mathbf{Z} - \mathbf{Z}^*) = \tilde{\boldsymbol{\beta}}^T(\mathbf{Z} - \mathbf{Z}^*)$. It follows from condition (C1) that $\boldsymbol{\beta} = \tilde{\boldsymbol{\beta}}$. Then we obtain $\lambda_A = \tilde{\lambda}_A$ for both $A = 0$ and $A = 1$. From equation (B.2), we can show that $H_A(t)e^{\boldsymbol{\zeta}^T\mathbf{Z}} = \tilde{H}_A(t)e^{\tilde{\boldsymbol{\zeta}}^T\mathbf{Z}}$ for any $t \in [0, \tau]$ and $\mathbf{Z}$ in its support. Similar techniques and condition (C1) yield $\boldsymbol{\zeta} = \tilde{\boldsymbol{\zeta}}$ and $H_A(t) = \tilde{H}_A(t)$, for any $t \in [0, \tau]$. The identifiability of $(\lambda_0, H_0, \lambda_1, H_1, \theta, \boldsymbol{\beta}, \boldsymbol{\zeta})$ is thus established.

Appendix C. Proof of Theorem 2

We introduce some notation that will be used throughout the proof of Theorems 2 and 3. Let $\mathbf{O}_i$ denote the observations for the ith subject consisting of $(A_i, X_i, T_i, \Delta_i, \mathbf{Z}_i)$. Let $\mathbf{P}_n$ and $\mathbf{P}$ be the empirical measure and the expectation of n i.i.d. observations $\mathbf{O}_1, \ldots, \mathbf{O}_n$. That is, for any measurable function $g(\mathbf{O})$,

$$\mathbf{P}_n[g(\mathbf{O})] = \frac{1}{n}\sum_{i=1}^n g(\mathbf{O}_i), \quad \mathbf{P}[g(\mathbf{O})] = E[g(\mathbf{O})].$$

The proof of consistency consists of two major steps. In the first step, we prove that $\limsup_n \hat{H}_{an}(\tau), a = 0, 1$ has an upper bound with probability one. Therefore, there exists a subsequence of $(\hat{\boldsymbol{\phi}}_n, \hat{H}_{0n}, \hat{H}_{1n})$ that converges to $(\boldsymbol{\phi}^*, H_0^*, H_1^*)$. In the second step, we prove that $(\boldsymbol{\phi}^*, H_0^*, H_1^*) = (\boldsymbol{\phi}_0, H_{00}, H_{10})$.

Step 1. We will prove the uniform boundedness of $\hat{H}_{an}(\tau), a = 0, 1$ by contradiction. Suppose that $\hat{H}_{an}(\tau) \to \infty, a = 0, 1$ in some sample space Ω with positive probability. For each sample in Ω, by selecting a subsequence still indexed by n, we assume that $\hat{\boldsymbol{\phi}}_n \to \boldsymbol{\phi}^*$ and $\hat{H}_{an}(\tau) \to \infty, a = 0, 1$. The idea of obtaining a contradiction is as follows: we first construct a step function $\bar{H}_{an}$ with jumps only at the observed T_i in the ath group such that $\bar{H}_{an}$ is close to the true function H_{a0}; then because $(\hat{\boldsymbol{\phi}}_n, \hat{H}_{0n}, \hat{H}_{1n})$ maximizes $l_n(\boldsymbol{\phi}, H_0, H_1)$, it holds that $\{l_n(\hat{\boldsymbol{\phi}}_n, \hat{H}_{0n}, \hat{H}_{1n}) - l_n(\boldsymbol{\phi}_0, \bar{H}_{0n}, \bar{H}_{1n})\}/n \geqslant 0$; finally, we prove that if $\hat{H}_{0n}(\tau) \to \infty$ and/or $\hat{H}_{1n}(\tau) \to \infty$, the left-hand side of the preceding inequality will eventually be negative, which yields the contradiction.

Recall that the nonparametric log-likelihood takes the form

$$l_n(\boldsymbol{\phi}, H_0, H_1) = n\mathbf{P}_n[R(\mathbf{O}; \boldsymbol{\phi}, H_0, H_1) + \Delta \log H_A\{T\}],$$

where

$$\begin{aligned} R(\mathbf{O}; \boldsymbol{\phi}, H_0, H_1) =& X(\log \lambda_A + \boldsymbol{\beta}^T\mathbf{Z}) + \Delta\boldsymbol{\zeta}^T\mathbf{Z} + \sum_{k=0}^{X+\Delta-1} \log(\theta^{-1} + k) \\ &+ (X + \Delta)\log\theta - (\theta^{-1} + X + \Delta)\log\{1 + \theta\lambda_A T e^{\boldsymbol{\beta}^T\mathbf{Z}} + \theta H_A(T) e^{\boldsymbol{\zeta}^T\mathbf{Z}}\}. \end{aligned}$$

By differentiating $l_n(\boldsymbol{\phi}, H_0, H_1)$ with respect to $H_a\{T_i\}$ and setting it to zero, we can see that $\hat{H}_{an}\{T_i\}$ satisfies the following equation.

$$\hat{H}_{an}\{T_i\} = \left.\frac{I(A_i = a)\Delta_i}{n\mathbf{P}_n[I(T \geqslant t, A = a)Q(\mathbf{O}; \hat{\boldsymbol{\phi}}_n, \hat{H}_{0n}, \hat{H}_{1n})]}\right|_{t=T_i}, \tag{C.1}$$

where

$$Q(\mathbf{O}; \hat{\boldsymbol{\phi}}_n, \hat{H}_{0n}, \hat{H}_{1n}) = \frac{(1 + \hat{\theta}_n X + \hat{\theta}_n\Delta)e^{\hat{\boldsymbol{\zeta}}_n^T\mathbf{Z}}}{1 + \hat{\theta}_n\hat{\lambda}_{An}e^{\hat{\boldsymbol{\beta}}_n^T\mathbf{Z}} + \hat{\theta}_n\hat{H}_{An}(T)e^{\hat{\boldsymbol{\zeta}}_n^T\mathbf{Z}}}.$$

In view of (C.1), we construct another step function $\bar{H}_{an}(t)$ with jumps only at the observed T_i in the ath group and the jump size satisfies that

$$\bar{H}_{an}\{T_i\} = \left.\frac{I(A_i = a)\Delta_i}{n\mathbf{P}_n[I(T \geqslant t, A = a)Q(\mathbf{O}; \boldsymbol{\phi}_0, H_{00}, H_{01})]}\right|_{t=T_i}. \tag{C.2}$$

We verify that $\bar{H}_{an}(t)$ converges to H_{0a} uniformly in $t \in [0, \tau]$ with probability one. It can be shown that the class

$$\mathcal{F}_a = \{I(T \geqslant t, A = a)Q(t, \mathbf{O}; \boldsymbol{\phi}, H_0, H_1) : t \in [0, \tau], (\lambda_0, \lambda_1, \theta, \boldsymbol{\beta}, \boldsymbol{\zeta}) \in \mathcal{B}_0, H_a \in \mathcal{A}, H_a(0) = 0\}$$

is a bounded and P-Donsker class, where

$$\mathcal{A} = \{g : g \text{ is a nondecreasing function in } [0, \tau], g(\tau) \leqslant B_0\}$$

and B_0 is a positive constant based on condition (C5).

As a P-Donsker class is also a Glivenko–Cantelli class, by the Glivenko–Cantelli theorem in [35], $\bar{H}_{an}(t)$ uniformly converges to

$$E[I(T \leqslant t, A = a)\Delta/\mu_a(T)],$$

where $\mu_a(t) = E[I(T \geqslant t, A = a)Q(\mathbf{O}; \boldsymbol{\phi}_0, H_{00}, H_{01})]$.

By the conditional independent censoring assumption (C1) and with some tedious algebra, we can prove that

$$E[I(T \leqslant t, A = a)\Delta/\mu_a(T)] = H_{0a}(t).$$

As $(\hat{\boldsymbol{\phi}}_n, \hat{H}_{0n}, \hat{H}_{1n})$ maximizes $l_n(\boldsymbol{\phi}, H_0, H_1)$,

$$0 \leqslant \frac{1}{n} l_n(\hat{\boldsymbol{\phi}}_n, \hat{H}_{0n}, \hat{H}_{1n}) - \frac{1}{n} l_n(\boldsymbol{\phi}_0, \bar{H}_{0n}, \bar{H}_{1n}).$$

By plugging $\hat{H}_{an}\{T_i\}$ of equation (C.1) and $\bar{H}_{an}\{T_i\}$ of equation (C.2) into $l_n(\hat{\boldsymbol{\phi}}_n, \hat{H}_{0n}, \hat{H}_{1n})$ and $l_n(\boldsymbol{\phi}_0, \bar{H}_{0n}, \bar{H}_{1n})$, respectively, we obtain

$$\begin{aligned} 0 \leqslant & \frac{1}{n} \sum_{i=1}^{n} \sum_{a=0}^{1} I(A_i = a) \Bigg[- \Delta_i \log \left\{ \frac{1}{n} \sum_{j=1}^{n} I(T_j \geqslant T_i, A_j = a) Q(\mathbf{O}_j; \hat{\boldsymbol{\phi}}_n, \hat{H}_{0n}, \hat{H}_{1n}) \right\} \\ & + R(\mathbf{O}_i; \hat{\boldsymbol{\phi}}_n, \hat{H}_{0n}, \hat{H}_{1n}) - R(\mathbf{O}_i; \boldsymbol{\phi}_0, \bar{H}_{0n}, \bar{H}_{1n}) \\ & + \Delta_i \log \left\{ \frac{1}{n} \sum_{j=1}^{n} I(T_j \geqslant T_i, A_j = a) Q(\mathbf{O}_j; \boldsymbol{\phi}_0, H_{00}, H_{10}) \right\} \Bigg] \\ \leqslant & O(1) - \frac{1}{n} \sum_{i=1}^{n} \sum_{a=0}^{1} I(A_i = a) \Bigg[\Delta_i \log \left\{ \frac{1}{n} \sum_{j=1}^{n} \frac{I(T_j \geqslant T_i, A_j = a)}{g_1 + \hat{H}_{an}(T_j)} \right\} \\ & + (\hat{\theta}_n^{-1} + X_i + \Delta_i) \log \left\{ 1 + \hat{\theta}_n \hat{\lambda}_{an} T_i e^{\hat{\boldsymbol{\beta}}_n^T \mathbf{Z}_i} + \hat{\theta}_n \hat{H}_{an}(T_i) e^{\hat{\boldsymbol{\xi}}_n^T \mathbf{Z}_i} \right\} \Bigg], \end{aligned} \tag{C.3}$$

where g_1 is some positive constant. The last inequality above is obtained from conditions (C4) and (C5).

Recall that $\hat{\theta}_n$ converges to θ^*. If $\theta^* = 0$, and we can show that the right-hand side of inequality (C.3) diverges to negative infinity as $\hat{H}_{0n}(\tau) \to \infty$ or $\hat{H}_{1n}(\tau) \to \infty$. We have the contradiction because the left-hand side is non-negative. Therefore, we assume that $\theta^* > 0$. It follows that

$$\begin{aligned} 0 \leqslant & O(1) - \frac{1}{n} \sum_{i=1}^{n} \sum_{a=0}^{1} I(A_i = a) \Bigg[\Delta_i \log \left\{ \frac{1}{n} \sum_{j=1}^{n} \frac{I(T_j \geqslant T_i, A_j = a)}{g_1 + \hat{H}_{an}(T_j)} \right\} \\ & + (g_2 + \Delta_i) \log \left\{ g_3 + \hat{H}_{an}(T_i) \right\} \Bigg], \end{aligned} \tag{C.4}$$

for some positive constants g_2 and g_3.

We will show that if $\hat{H}_{0n}(\tau) \to \infty$ or $\hat{H}_{1n}(\tau) \to \infty$, the right-hand side of (C.4) diverges to $-\infty$. We mimic the arguments of Murphy [29] to prove the divergence of the right-hand side. Specifically, for both $a = 0$ and $a = 1$, we choose a partition of $[0, \tau]$, as follows: with $s_0 = \tau$, choose $s_1 < s_0$ such that

$$\frac{1}{2} E\{(g_2 + \Delta_i) I(T_i = s_0, A_i = a)\} > E\{\Delta_i I(T_i \in [s_1, s_0), A_i = a)\}.$$

By conditions (C2) and (C3), such s_1 exists. We next define a constant $\epsilon \in (0, 1)$ such that

$$\frac{\epsilon}{g_2(1-\epsilon)} < \frac{E\{I(T_i \in [s_1, s_0), A_i = a)\}}{E\{\Delta_i I(T_i \in [0, \tau), A_i = a)\}}.$$

If $s_1 > 0$, we can choose $s_2 = \max(0, s)$ such that s is the minimum value less than s_1 satisfying that

$$(1-\epsilon) E\{(g_2 + \Delta_i) I(T_i \in [s_1, s_0), A_i = a)\} \geqslant E\{\Delta_i I(T_i \in [s, s_1), A_i = a)\}.$$

Clearly, s_2 exists under the regularity conditions, and $s_2 < s_1$. This process can continue so that we obtain a sequence $\tau = s_0 > s_1 > s_2 > \cdots \geqslant 0$ such that

$$\begin{aligned} \frac{1}{2} E\{(g_2 + \Delta_i) I(T_i = s_0, A_i = a)\} & \geqslant E\{\Delta_i I(T_i \in [s_1, s_0), A_i = a)\}, \\ (1-\epsilon) E\{(g_2 + \Delta_i) I(T_i \in [s_q, s_{q-1}), A_i = a)\} & \geqslant E\{\Delta_i I(T_i \in [s_{q+1}, s_q), A_i = a)\}, \quad q \geqslant 1. \end{aligned}$$

Such a sequence cannot be infinite, that is, there exists a finite N such that $s_{N+1} = 0$; otherwise, $s_q \to s^*$ for some $s^*[0, \tau)$. By the definition of s_q, it holds that

$$(1-\epsilon)E\{(g_2+\Delta_i)I(T_i \in [s_q, s_{q-1}), A_i = a)\} = E\{\Delta_i I(T_I \in [s_{q+1}, s_q), A_i = a)\}, \quad q \geqslant 1.$$

By the continuity of true densities, we sum over $q = 1, 2, \ldots$ and obtain

$$(1-\epsilon)E\{(g_2+\Delta_i)I(T_i \in [s^*, \tau), A_i = a)\} = E\{\Delta_i I(T_i \in [s^*, s_1), A_i = a)\}.$$

Thus,

$$g_2(1-\epsilon)E\{I(T_i \in [s^*, \tau), A_i = a)\} \leqslant \epsilon E\{\Delta_i I(T_i \in [s^*, s_1), A_i = a)\},$$

which contradicts with the choice of ϵ. Therefore, the sequence is finite: $\tau = s_0 > s_1 > \cdots > s_{N+1} = 0$. Therefore, the right-hand side of (C.4) can be bounded from above by

$$\begin{aligned}
&O(1) - \frac{1}{n}\sum_{i=1}^{n}\sum_{a=0}^{1} I(A_i = a)\Bigg[(g_2+\Delta_i)I(T_i = \tau)\log\{g_3 + \hat{H}_{an}(\tau)\} \\
&+ \sum_{q=0}^{N}(g_2+\Delta_i)I(T_i \in [s_{q+1}, s_q))\log\{g_3 + \hat{H}_{an}(s_{q+1})\} \\
&+ \sum_{q=0}^{N}\Delta_i I(T_i \in [s_{q+1}, s_q))\log\left\{\frac{1}{n}\sum_{j=1}^{n}\frac{I(T_j \geqslant T_i, A_j = a, T_j \in [s_{q+1}, s_q))}{g_1 + \hat{H}_{an}(s_q)}\right\}\Bigg] \\
&\leqslant O(1) - \frac{1}{n}\sum_{i=1}^{n}\sum_{a=0}^{1} I(A_i = a)\Bigg[\frac{1}{2}(g_2+\Delta_i)I(T_i = \tau)\log\{g_3 + \hat{H}_{an}(\tau)\} \\
&+ \left\{\frac{1}{2}(g_2+\Delta_i)I(T_i = \tau)\log\{g_3 + \hat{H}_{an}(\tau)\} - \Delta_i I(T_i \in [s_1, s_0))\log\{g_3 + \hat{H}_{an}(\tau)\}\right\} \\
&+ \sum_{q=1}^{N}\Bigg\{(g_2+\Delta_i)I(T_i \in [s_q, s_{q-1}))\log\{g_3 + \hat{H}_{an}(s_q)\} \\
&- \Delta_i I(T_i \in [s_{q+1}, s_q))\log\{g_3 + \hat{H}_{an}(s_q)\}\Bigg\} \\
&+ \sum_{q=0}^{N}\Delta_i I(T_i \in [s_{q+1}, s_q))\log\left\{\frac{1}{n}\sum_{j=1}^{n} I(T_j \geqslant T_i, T_j \in [s_{q+1}, s_q))\right\}\Bigg].
\end{aligned} \tag{C.5}$$

The first term on the right-hand side of (C.5) diverges to $-\infty$ as $\hat{H}_{0n}(\tau) \to \infty$ or $\hat{H}_{1n} \to \infty$. The second term is negative for large n owing to the choice of s_1. By the selection of $s_q, q = 1, \ldots, N$, the third term cannot diverge to ∞. Finally, the fourth term is bounded because of the Glivenko–Cantelli theorem. Hence, the right-hand side of (C.5) diverges to $-\infty$. This contradicts the fact that the left-hand side of (C.5) is non-negative. Therefore, we have shown that, with probability one, $\hat{H}_{an}(\tau), a = 0, 1$ is bounded for any sample size n.

Thus, by Helly's selection theorem, we can choose a further subsequence, still indexed by $\{n\}$, such that $(\hat{\lambda}_{0n}, \hat{H}_{0n}, \hat{\lambda}_{1n}, \hat{H}_{1n}, \hat{\theta}_n, \hat{\boldsymbol{\beta}}_n, \hat{\boldsymbol{\zeta}}_n)$ converges to $(\lambda_0^*, H_0^*, \lambda_1^*, H_1^*, \theta^*, \boldsymbol{\beta}^*, \boldsymbol{\zeta}^*)$ with probability one.

Step 2. In this step, we will show that

$$(\lambda_0^*, H_0^*, \lambda_1^*, H_1^*, \theta^*, \boldsymbol{\beta}^*, \boldsymbol{\zeta}^*) = (\lambda_{00}, H_{00}, \lambda_{10}, H_{10}, \theta_0, \boldsymbol{\beta}_0, \boldsymbol{\zeta}_0).$$

We use $\hat{H}_{an}(t)$ and $\bar{H}_{an}(t)$ in step 1. By the construction of $\hat{H}_{an}(t)$ and $\bar{H}_{an}(t)$, we can see that $\hat{H}_{an}(t)$ is absolutely continuous with respect to $\bar{H}_{an}(t)$ and

$$\hat{H}_{an}(t) = \int_0^t \frac{\mathbb{P}_n[I(T \geqslant y, A = a)Q(y, \mathbf{O}; \boldsymbol{\phi}_0, H_{00}, H_{01})]}{\mathbb{P}_n[I(T \geqslant y, A = a)Q(y, \mathbf{O}; \hat{\boldsymbol{\phi}}_n, \hat{H}_{0n}, \hat{H}_{1n})]} d\bar{H}_{an}(y). \tag{C.6}$$

By taking limits on both sides of (C.6), we obtain that

$$H_a^*(t) = \int_0^t \frac{\mathbb{P}[I(T \geqslant y, A = a)Q(y, \mathbf{O}; \boldsymbol{\phi}_0, H_{00}, H_{01})]}{\mathbb{P}[I(T \geqslant y, A = a)Q(y, \mathbf{O}; \boldsymbol{\phi}^*, H_0^*, H_1^*)]} dH_{0a}(y).$$

Therefore, $H_a^*(t)$ is differentiable with respect to $H_{0a}(t)$ so that $H_a^*(t)$ is differentiable with respect to t. It follows that $d\hat{H}_{an}(t)/d\bar{H}_{an}(t)$ converges to $dH_a^*(t)/dH_{0a}(t)$ uniformly in $t \in [0, \tau]$.

Note that

$$\begin{aligned} & n^{-1} l_n(\hat{\lambda}_{0n}, \hat{H}_{0n}, \hat{\lambda}_{1n}, \hat{H}_{1n}, \hat{\theta}_n) - n^{-1} l_n(\lambda_{00}, \bar{H}_{0n}, \lambda_{01}, \bar{H}_{1n}, \theta_0) \\ & = \mathbb{P}_n \left[\Delta \log \frac{\hat{H}_{An}\{T\}}{\bar{H}_{An}\{T\}} \right] + \mathbb{P}_n[R(\mathbf{O}; \hat{\lambda}_{0n}, \hat{H}_{0n}, \hat{\lambda}_{1n}, \hat{H}_{1n}, \hat{\theta}_n) - R(\mathbf{O}; \lambda_{00}, \bar{H}_{0n}, \lambda_{01}, \bar{H}_{1n}, \theta_0)] \\ & \geqslant 0. \end{aligned} \tag{C.7}$$

As $\mathcal{B}_0 \times \mathcal{A}$ is a Donsker class and the functionals $R(\mathbf{O}; \boldsymbol{\phi}, H_0, H_1)$ are bounded Lipschitz functionals with respect to $\mathcal{B}_0 \times \mathcal{A}$, the following class

$$\mathcal{F}_2 = \{R(\mathbf{O}; \boldsymbol{\phi}, H_0, H_1) : (\lambda_0, \lambda_1, \theta, \boldsymbol{\beta}, \boldsymbol{\zeta}) \in \mathcal{B}_0, H_a \in \mathcal{A}, H_a(0) = 0, H_a(\tau) \leqslant B_0, a = 0, 1\}$$

is P-Donsker and hence a Glivenko–Cantelli class. Therefore, by letting $n \to \infty$ in (C.7), we have

$$0 \leqslant \mathbb{P}\left[\log \left\{ \frac{h_A^*(T)^\Delta e^{R(\mathbf{O}; \boldsymbol{\phi}^*, H_0^*, H_1^*)}}{h_{A0}(T)^\Delta e^{R(\mathbf{O}; \boldsymbol{\phi}_0, H_{00}, H_{01})}} \right\} \right],$$

which is the negative Kullback–Leibler information. Then it follows that, with probability one,

$$h_A^*(T)^\Delta e^{R(\mathbf{O}; \boldsymbol{\phi}^*, H_0^*, H_1^*)} = h_{A0}(T)^\Delta e^{R(\mathbf{O}; \boldsymbol{\phi}_0, H_{00}, H_{01})}.$$

Therefore, from the identifiability result proved earlier, we obtain $(\boldsymbol{\phi}^*, H_0^*, H_1^*) = (\boldsymbol{\phi}_0, H_{00}, H_{01})$. This completes the proof of Theorem 2. Note that the uniform convergence of $\hat{H}_{an}$ to $H_{a0}, a = 0, 1$ follows from the fact that H_{a0} are continuous functions.

Appendix D. Proof of Theorem 3

We prove Theorem 3 by verifying the four conditions in Theorem 3.3.1 of [35]. For this purpose, we first define a neighborhood of the true parameters $(\boldsymbol{\phi}_0, H_0, H_1)$, denoted by

$$\mathcal{U} = \{(\boldsymbol{\phi}, H_0, H_1) : ||\boldsymbol{\phi} - \boldsymbol{\phi}_0|| + \sup_{t \in [0,\tau]} (|H_0(t) - H_{00}(t)| + |H_1(t) - H_{01}(t)|) < \epsilon_0\},$$

for a very small constant ϵ_0. Based on the consistency theorem, $(\hat{\boldsymbol{\phi}}_n, \hat{H}_{0n}, \hat{H}_{1n})$ belongs to $\mathcal{U}$ with probability close to 1 when the sample size n is large enough.

For any one-dimensional submodel given as $\{\lambda_0 + \epsilon h_1, \lambda_1 + \epsilon h_2, \theta + \epsilon h_3, \boldsymbol{\beta} + \epsilon \mathbf{h}_4, \boldsymbol{\zeta} + \epsilon \mathbf{h}_5, H_0 + \epsilon \int h_6 dH_0, H_1 + \epsilon \int h_7 dH_1\}, (\boldsymbol{\phi}, H_0, H_1) \in \mathcal{U}, \mathbf{H} \equiv (h_1, h_2, h_3, \mathbf{h}_4, \mathbf{h}_5, h_6, h_7) \in \mathcal{H}$, we can derive the score function for a single observation $\mathbf{O}$

$$\begin{aligned} W(\mathbf{O}; \boldsymbol{\phi}, H_0, H_1)[\mathbf{H}] = & l_{\lambda_0}(\boldsymbol{\phi}, H_0, H_1)h_1 + l_{\lambda_1}(\boldsymbol{\phi}, H_0, H_1)h_2 + l_\theta(\boldsymbol{\phi}, H_0, H_1)h_3 \\ & + l_\beta(\boldsymbol{\phi}, H_0, H_1)^T \mathbf{h}_4 + l_\zeta(\boldsymbol{\phi}, H_0, H_1)^T \mathbf{h}_5 \\ & + l_{H_0}(\boldsymbol{\phi}, H_0, H_1) \left[\int h_6 dH_0 \right] + l_{H_1}(\boldsymbol{\phi}, H_0, H_1) \left[\int h_7 dH_1 \right], \end{aligned} \tag{D.1}$$

where

$$l_{\lambda_a}(\mathbf{O}; \boldsymbol{\phi}, H_0, H_1) = I(A = a) \left\{ \frac{X}{\lambda_a} - \frac{(1 + \theta X + \theta \Delta) T e^{\beta^T \mathbf{Z}}}{1 + \theta \lambda_a T e^{\beta^T \mathbf{Z}} + \theta H_a(T) e^{\zeta^T \mathbf{Z}}} \right\}, \quad a = 0, 1,$$

$$l_\theta(\mathbf{O};\boldsymbol{\phi},H_0,H_1) = -\left(\sum_{k=0}^{X+\Delta-1}\frac{1}{\theta+\theta^2 k}\right) + \log\{1+\theta\lambda_a T e^{\beta^T\mathbf{Z}} + \theta H_a(T)e^{\zeta^T\mathbf{Z}}\}/\theta^2$$
$$+ \frac{X+\Delta}{\theta} - \frac{(\theta^{-1}+X+\Delta)(\lambda_A T e^{\beta^T\mathbf{Z}} + H_A(T)e^{\zeta^T\mathbf{Z}})}{1+\theta\lambda_a T e^{\beta^T\mathbf{Z}} + \theta H_a(T)e^{\zeta^T\mathbf{Z}}},$$

$$l_\beta(\mathbf{O};\boldsymbol{\phi},H_0,H_1) = \left\{X - \frac{(1+\theta X+\theta\Delta)\lambda_A T e^{\beta^T\mathbf{Z}}}{1+\theta\lambda_A T e^{\beta^T\mathbf{Z}} + \theta H_A(T)e^{\zeta^T\mathbf{Z}}}\right\}\mathbf{Z},$$

$$l_\zeta(\mathbf{O};\boldsymbol{\phi},H_0,H_1) = \left\{\Delta - \frac{(1+\theta X+\theta\Delta)H_A(T)e^{\zeta^T\mathbf{Z}}}{1+\theta\lambda_A T e^{\beta^T\mathbf{Z}} + \theta H_A(T)e^{\zeta^T\mathbf{Z}}}\right\}\mathbf{Z},$$

$$l_{H_0}(\boldsymbol{\phi},H_0,H_1)\left[\int h_6 dH_0\right] = I(A=0)\left\{\Delta h_6(T) - \frac{(1+\theta X+\theta\Delta)e^{\zeta^T\mathbf{Z}}}{1+\theta\lambda_A T e^{\beta^T\mathbf{Z}} + \theta H_A(T)e^{\zeta^T\mathbf{Z}}}\int_0^T h_6 dH_0\right\},$$

and

$$l_{H_1}(\boldsymbol{\phi},H_0,H_1)\left[\int h_7 dH_1\right] = I(A=1)\left\{\Delta h_7(T) - \frac{(1+\theta X+\theta\Delta)e^{\zeta^T\mathbf{Z}}}{1+\theta\lambda_A T e^{\beta^T\mathbf{Z}} + \theta H_A(T)e^{\zeta^T\mathbf{Z}}}\int_0^T h_7 dH_1\right\}.$$

We define

$$U_n(\boldsymbol{\phi},H_0,H_1)[\mathbf{H}] = \mathbb{P}_n\{W(\mathbf{O};\boldsymbol{\phi},H_0,H_1)[\mathbf{H}]\}$$

and

$$U(\boldsymbol{\phi},H_0,H_1)[\mathbf{H}] = \mathbb{P}\{W(\mathbf{O};\boldsymbol{\phi},H_0,H_1)[\mathbf{H}]\}.$$

Thus, it is easy to see that both $U_n(\boldsymbol{\phi},H_0,H_1)[\mathbf{H}]$ and $U(\boldsymbol{\phi},H_0,H_1)[\mathbf{H}]$ are maps from $\mathcal{U}$ to $l^\infty(\mathcal{H})$ and $\sqrt{n}\{U_n(\boldsymbol{\phi},H_0,H_1) - U(\boldsymbol{\phi},H_0,H_1)\}$ is an empirical process in the space $l^\infty(\mathcal{H})$. It is easy to see that $U_n(\hat{\boldsymbol{\phi}}_n,\hat{H}_{0n},\hat{H}_{1n}) = 0$ and $U(\boldsymbol{\phi}_0,H_{00},H_{01}) = 0$.

We shall prove the theorem by verifying the following four properties stated in Theorem 3.3.1 of [35].

(P1) $\sqrt{n}(U_n - U)(\hat{\boldsymbol{\phi}}_n,\hat{H}_{0n},\hat{H}_{1n}) - \sqrt{n}(U_n - U)(\boldsymbol{\phi}_0,H_{00},H_{01}) = o_P(1 + \sqrt{n}||\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_0|| + \sqrt{n}\sup_{t\in[0,\tau]}\{|\hat{H}_{0n}(t) - H_{00}(t)| + |\hat{H}_{1n}(t) - H_{01}(t)|\})$.

(P2) $\sqrt{n}(U_n - U)(\boldsymbol{\phi}_0,H_{00},H_{01})$ converges to a tight random element $\boldsymbol{\xi}$.

(P3) $U(\boldsymbol{\phi},H_0,H_1)$ is Frechet-differentiable at $(\boldsymbol{\phi}_0,H_{00},H_{01})$.

(P4) The derivative of $U(\boldsymbol{\phi},H_0,H_1)$ at $(\boldsymbol{\phi}_0,H_{00},H_{01})$, denoted by $U'(\boldsymbol{\phi}_0,H_{00},H_{01})$ is continuously invertible.

To prove property (P1), we make use of Lemma 3.3.5 of [35]. Based on the explicit expression in (D.1), $W(\mathbf{O};\boldsymbol{\phi},H_0,H_1)[\mathbf{H}]$ is continuously differentiable with respect to $\boldsymbol{\phi}$ and

$$\left\|\frac{dW(\mathbf{O};\boldsymbol{\phi},H_0,H_1)}{d\boldsymbol{\phi}}\right\| \leqslant g_3 + g_4 X,$$

where g_3 and g_4 are some positive constants. Furthermore,

$$|W(\mathbf{O};\boldsymbol{\phi},H_0,H_1)[\mathbf{H}] - W(\mathbf{O};\boldsymbol{\phi},\tilde{H}_0,H_1)[\mathbf{H}]| \leqslant (g_5+g_6X)\left\{|H_0(T)-\tilde{H}_0(T)| + \int_0^\tau |H_0(t)-\tilde{H}_0(t)|dt\right\}$$

and

$$|W(\mathbf{O};\boldsymbol{\phi},H_0,H_1)[\mathbf{H}] - W(\mathbf{O};\boldsymbol{\phi},H_0,\tilde{H}_1)[\mathbf{H}]| \leqslant (g_5+g_6X)\left\{|H_1(T)-\tilde{H}_1(T)| + \int_0^\tau |H_1(t)-\tilde{H}_1(t)|dt\right\}$$

for some positive constants g_5 and g_6. Therefore,

$$\sup_{\mathbf{H}\in\mathcal{H}} E\left[\left\{W(\mathbf{O};\boldsymbol{\phi},H_0,H_1)[\mathbf{H}] - W(\mathbf{O};\boldsymbol{\phi}_0,H_{00},H_{01})[\mathbf{H}]\right\}^2\right]$$

converges to zero if $||\boldsymbol{\phi}-\boldsymbol{\phi}_0||+\sup_{t\in[0,\tau]}\{|H_0(t)-H_{00}(t)|+|H_1(t)-H_{01}(t)|\}\to 0$. Additionally, we can show that the class

$$\mathcal{F}_3=\{W(\mathbf{O};\boldsymbol{\phi},H_0,H_1)[\mathbf{H}]-W(\mathbf{O};\boldsymbol{\phi}_0,H_{00},H_{01})[\mathbf{H}]:(\boldsymbol{\phi},H_0,H_1)\in\mathcal{U},\mathbf{H}\in\mathcal{H}\}$$

is P-Donsker. Therefore, according to Lemma 3.3.5 of [35], property (P1) holds.

Property (P2) holds again because of the P-Donsker property of the class

$$\{W(\mathbf{O};\boldsymbol{\phi},H_0,H_1)[\mathbf{H}]:\mathbf{H}\in\mathcal{H}\}.$$

Furthermore, the limit random elements $\boldsymbol{\xi}$ is a Gaussian process indexed by $\mathbf{H}\in\mathcal{H}$ and the covariance between $\boldsymbol{\xi}(\mathbf{H}_1)$ and $\boldsymbol{\xi}(\mathbf{H}_2)$ is equal to

$$E\left[W(\mathbf{O};\boldsymbol{\phi}_0,H_{00},H_{01})[\mathbf{H}_1]\times W(\mathbf{O};\boldsymbol{\phi}_0,H_{00},H_{01})[\mathbf{H}_2]\right].$$

The Frechet differentiability in (P3) can be directly verified by using the smoothness of $U(\boldsymbol{\phi},H_0,H_1)$. The derivative of $U(\boldsymbol{\phi},H_0,H_1)$ at $(\boldsymbol{\phi}_0,H_{00},H_{01})$, denoted by $U'(\boldsymbol{\phi}_0,H_{00},H_{01})$, is a map from the space

$$\{(\boldsymbol{\phi}-\boldsymbol{\phi}_0,H_0-H_{00},H_1-H_{01}):(\boldsymbol{\phi},H_0,H_1)\in\mathcal{U}\}$$

to $l^\infty(\mathcal{H})$.

It remains to show that U' is continuously invertible at $(\boldsymbol{\phi}_0,H_{00},H_{01})$. Following the argument in the Appendix of [42], it suffices to prove that for any one-dimensional submodel given as $\{\lambda_0+\epsilon h_1,\lambda_1+\epsilon h_2,\theta+\epsilon h_3,\boldsymbol{\beta}+\epsilon\mathbf{h}_4+\boldsymbol{\zeta}+\epsilon\mathbf{h}_5,H_0+\epsilon\int h_6dH_0,H_1+\epsilon\int h_7dH_1\},\mathbf{H}\in\mathcal{H}$, the Fisher information along this submodel is nonsingular. If the Fisher information along this submodel is singular, the score function along this submodel is zero with probability one. We will show that $W(\mathbf{O};\boldsymbol{\phi}_0,H_{00},H_{01})[\mathbf{H}]=0$ yields that $h_1=h_2=h_3=0$, $\mathbf{h}_4=\mathbf{h}_5=\mathbf{0}$, and $h_6=h_7=0$. We follow the ideas of proving the identifiability in the proof of Theorem 1. Let $\mathbf{O}=(A=0,X=0,T,\Delta,\mathbf{Z})$ and $\tilde{\mathbf{O}}=(A=0,X=1,T,\Delta,\mathbf{Z})$. We obtain

$$\begin{aligned}
&W(\mathbf{O};\boldsymbol{\phi}_0,H_{00},H_{01})[\mathbf{H}]-W(\tilde{\mathbf{O}};\boldsymbol{\phi}_0,H_{00},H_{01})[\mathbf{H}]\\
&=\left\{\frac{1}{\lambda_0}-\frac{\theta Te^{\beta^T\mathbf{Z}}}{1+\theta\lambda_0Te^{\beta^T\mathbf{Z}}+\theta H_0(T)e^{\zeta^T\mathbf{Z}}}\right\}h_1\\
&\quad+\left\{\frac{\Delta}{1+\theta}-\frac{\lambda_0Te^{\beta^T\mathbf{Z}}+H_0(T)e^{\zeta^T\mathbf{Z}}}{1+\theta\lambda_0Te^{\beta^T\mathbf{Z}}+\theta H_0(T)e^{\zeta^T\mathbf{Z}}}\right\}h_3\\
&\quad+\left\{1-\frac{\theta\lambda_0Te^{\beta^T\mathbf{Z}}}{1+\theta\lambda_0Te^{\beta^T\mathbf{Z}}+\theta H_0(T)e^{\zeta^T\mathbf{Z}}}\right\}\mathbf{Z}^T\mathbf{h}_4\\
&\quad-\frac{\theta H_0(T)e^{\zeta^T\mathbf{Z}}}{1+\theta\lambda_0Te^{\beta^T\mathbf{Z}}+\theta H_0(T)e^{\zeta^T\mathbf{Z}}}\mathbf{Z}^T\mathbf{h}_5\\
&\quad-\frac{\theta e^{\zeta^T\mathbf{Z}}}{1+\theta\lambda_0Te^{\beta^T\mathbf{Z}}+\theta H_0(T)e^{\zeta^T\mathbf{Z}}}\int_0^T h_6dH_0\\
&=0.
\end{aligned}$$

We then let $T=0$ and $\Delta=0$ and obtain $\frac{h_1}{\lambda_0}+\mathbf{Z}^T\mathbf{h}_4=0$. It follows from condition (C1) that $h_1=0$ and $\mathbf{h}_4=\mathbf{0}$. Next we let $T=0$ and $\Delta=1$ and obtain $h_3=0$. Therefore, with simple algebra, we obtain

$$-H_0(T)\mathbf{Z}^T\mathbf{h}_5-\int_0^T h_6dH_0=0.$$

Again condition (C1) implies that $\mathbf{h}_5=\mathbf{0}$ and $\int_0^T h_6dH_0=0$. As the second equality holds for any $T\in[0,\tau]$, immediately, condition (C5) implies $h_6(t)=0$ for any $t\in[0,\tau]$. Similarly, we can prove that $h_2=0$ and $h_7(t)=0$ for any $t\in[0,\tau]$.

We now have verified properties (P1)–(P4); Theorem 3.3.1 of [35] concludes that $\sqrt{n}(\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_0, \hat{H}_{0n} - H_{00}, \hat{H}_{1n} - H_{01})$ converges weakly to a tight Gaussian random element $-U'^{-1}\boldsymbol{\xi}$ in $l^\infty(\mathcal{H})$. Moreover, it can be shown that $\hat{\boldsymbol{\phi}}_n$ is an asymptotic linear estimator for $\boldsymbol{\phi}_0$ and that the corresponding influence functions are on the space spanned by the score functions. This implies that $\hat{\boldsymbol{\phi}}_n$ is semiparametrically efficient by semiparametric efficiency theory.

Acknowledgements

The authors are grateful to the associate editor and referees for their helpful comments.

References

1. Finkelstein DM, Schoenfeld DA, Stamenovic E. Analysis of multiple failure time data from an aids clinical trial. *Statistics in medicine* 1997; **16**(8):951–961.
2. Rudick RA, Stuart WH, Calabresi PA, Confavreux C, Galetta SL, Radue EW, Lublin FD, Weinstock-Guttman B, Wynn DR, Lynn F, Panzara MA. Natalizumab plus interferon beta-1a for relapsing multiple sclerosis. *New England Journal of Medicine* 2006; **354**(9):911–923.
3. Barai U, Teoh N. Multiple statistics for multiple events, with application to repeated infections in the growth factor studies. *Statistics in medicine* 1997; **16**(8):941–949.
4. Foley RN, Curtis BM, Parfrey PS. Hemoglobin targets and blood transfusions in hemodialysis patients without symptomatic cardiac disease receiving erythropoietin therapy. *Clinical Journal of the American Society of Nephrology* 2008; **3**(6): 1669–1675.
5. Hujoel P, Isokangas P, Tiekso J, Davis S, Lamont R, DeRouen T, Makinen K. A re-analysis of caries rates in a preventive trial using Poisson regression models. *Journal of Dental Research* 1994; **73**(2):573–579.
6. Zou G. A modified Poisson regression approach to prospective studies with binary data. *American Journal of Epidemiology* 2004; **159**(7):702–706.
7. Hodgson DC, Gilbert ES, Dores GM, Schonfeld SJ, Lynch CF, Storm H, Hall P, Langmark F, Pukkala E, Andersson M, Kaijser M. Long-term solid cancer risk among 5-year survivors of Hodgkin's lymphoma. *Journal of Clinical Oncology* 2007; **25**(12):1489–1497.
8. Gu K, Ng HKT, Tang ML, Schucany WR. Testing the ratio of two Poisson rates. *Biometrical Journal* 2008; **50**(2):283–298.
9. Li HQ, Tang ML, Wong WK. Confidence intervals for ratio of two Poisson rates using the method of variance estimates recovery. *Computational Statistics* 2014; **29**(3-4):869–889.
10. Wang MC, Qin J, Chiang CT. Analyzing recurrent event data with informative censoring. *Journal of the American Statistical Association* 2001; **96**(455):1057–1065.
11. Huang CY, Wang MC. Joint modeling and estimation for recurrent event processes and failure time data. *Journal of the American Statistical Association* 2004; **99**:1153–1165.
12. Liu L, Wolfe RA, Huang X. Shared frailty models for recurrent events and a terminal event. *Biometrics* 2004; **60**(3): 747–756.
13. Ye Y, Kalbfleisch JD, Schaubel DE. Semiparametric analysis of correlated recurrent and terminal events. *Biometrics* 2007; **63**(1):78–87.
14. Liu L, Huang X. Joint analysis of correlated repeated measures and recurrent events processes in the presence of death, with application to a study on acquired immune deficiency syndrome. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 2009; **58**(1):65–81.
15. Zeng D, Lin D. Semiparametric transformation models with random effects for joint analysis of recurrent and terminal events. *Biometrics* 2009; **65**(3):746–752.
16. Zhao X, Liu L, Liu Y, Xu W. Analysis of multivariate recurrent event data with time-dependent covariates and informative censoring. *Biometrical Journal* 2012; **54**(5):585–599.
17. Liu L, Huang X, Yaroshinsky A, Cormier JN. Joint frailty models for zero-inflated recurrent events in the presence of a terminal event. *Biometrics* 2016; **72**:204–214.
18. Cook RJ, Lawless J. *The Statistical Analysis of Recurrent Events*. Springer Science & Business Media, 2007.
19. Zhang Y, Jamshidian M. The gamma-frailty Poisson model for the nonparametric estimation of panel count data. *Biometrics* 2003; **59**(4):1099–1106.
20. Yao B, Wang L, He X. Semiparametric regression analysis of panel count data allowing for within-subject correlation. *Computational Statistics & Data Analysis* 2016; **97**:47–59.
21. Huang CY, Wang MC, Zhang Y. Analysing panel count data with informative observation times. *Biometrika* 2006; **93**(4):763–775.
22. Sun J, Tong X, He X. Regression analysis of panel count data with dependent observation times. *Biometrics* 2007; **63**(4):1053–1059.
23. He X, Tong X, Sun J. Semiparametric analysis of panel count data with correlated observation and follow-up times. *Lifetime data analysis* 2009; **15**(2):177–196.
24. Li N, Sun L, Sun J. Semiparametric transformation models for panel count data with dependent observation process. *Statistics in Biosciences* 2010; **2**(2):191–210.
25. Zhao X, Tong X. Semiparametric regression analysis of panel count data with informative observation times. *Computational Statistics & Data Analysis* 2011; **55**(1):291–300.

26. Wang X, Ma S, Yan J. Augmented estimating equations for semiparametric panel count regression with informative observation times and censoring time. *Statistica Sinica* 2013:359–381.
27. Zhao X, Tong X, Sun J. Robust estimation for panel count data with informative observation times. *Computational Statistics & Data Analysis* 2013; **57**(1):33–40.
28. Sun J, Zhao X. *Statistical Analysis of Panel Count Data*. Springer, 2013.
29. Murphy SA. Consistency in a proportional hazards model incorporating a random effect. *The Annals of Statistics* 1994; **22**:712–731.
30. Murphy SA. Asymptotic theory for the frailty model. *The Annals of Statistics* 1995; **23**:182–198.
31. Press WH, Teukolsky SA, Vetterling WT, Flannery BP. *Numerical Recipes in C: The Art of Scientific Computing,* Second Edition. Cambridge University Press: Cambridge, 1992.
32. Diao G, Lin D. A powerful and robust method for mapping quantitative trait loci in general pedigrees. *The American Journal of Human Genetics* 2005; **77**(1):97–111.
33. Diao G, Zeng D, Yang S. Efficient semiparametric estimation of short-term and long-term hazard ratios with right-censored data. *Biometrics* 2013; **69**(4):840–849.
34. Bickel PJ, Klaassen CAJ, Ritov Y, Wellner JA. *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins University Press: Baltimore, 1993, Ch. 3.
35. van der Vaart A, Wellner J. *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer-Verlag: New York, 1996.
36. Parner E. Asymptotic theory for the correlated gamma-frailty model. *The Annals of Statistics* 1998; **26**:183–214.
37. Murphy SA, van der Vaart AW. Semiparametric likelihood ratio inference. *The Annals of Statistics* 1997; **25**:1471–1509.
38. Lu M, Zhang Y, Huang J. Estimation of the mean function with panel count data using monotone polynomial splines. *Biometrika* 2007; **94**(3):705–718.
39. Nielsen J, Dean C. Clustered mixed nonhomogeneous Poisson process spline models for the analysis of recurrent event panel data. *Biometrics* 2008; **64**(3):751–761.
40. Hua L, Zhang Y, Tu W. A spline-based semiparametric sieve likelihood method for over-dispersed panel count data. *Canadian Journal of Statistics* 2014; **42**(2):217–245.
41. He X, Feng X, Tong X, Zhao X. Semiparametric partially linear varying coefficient models with panel count data. *Lifetime data analysis* 2016:1–28.
42. Zeng D, Lin DY. Maximum likelihood estimation in semiparametric regression models with censored data (with discussion). *Journal of the Royal Statistical Society: Series B* 2007; **69**:507–564.

Supporting information

Additional supporting information may be found online in the supporting information tab for this article.