Published in final edited form as:

Stat Med. 2017 May 10; 36(10): 1669–1680. doi:10.1002/sim.7217.

Regression Analysis of Mixed Panel Count Data with Dependent Terminal Events

Guanglei Yu^a, Liang Zhu^b, Yang Li^c, Jianguo Sun^{a,*†}, and Leslie L. Robison^d

^aDepartment of Statistics, University of Missouri-Columbia, MO, U.S.A

^bBiostatistics and Epidemiology Research Division, University of Texas Health Science Center at Houston, Houston, TX, U.S.A

^cDepartment of Mathematics and Statistics, The University of North Carolina at Charlotte, NC, U.S.A

^dDepartment of Epidemiology and Cancer Control, St. Jude Children's Research Hospital, Memphis, TN, U.S.A

Abstract

Event history studies are commonly conducted in many fields and a great deal of literature has been established for the analysis of the two types of data commonly arising from these studies: recurrent event data and panel count data. The former arises if all study subjects are followed continuously, while the latter means that each study subject is observed only at discrete time points. In reality, a third type of data, a mixture of the two types of the data above, may occur and furthermore, as with the first two types of the data, there may exist a dependent terminal event, which may preclude the occurrences of recurrent events of interest. This paper discusses regression analysis of mixed recurrent event and panel count data in the presence of a terminal event and an estimating equation-based approach is proposed for estimation of regression parameters of interest. In addition, the asymptotic properties of the proposed estimator are established and a simulation study conducted to assess the finite-sample performance of the proposed method suggests that it works well in practical situations. Finally the methodology is applied to a childhood cancer study that motivated this study.

Keywords

Estimating equation; Marginal modeling; Mixed data; Terminal event

1. Introduction

Event history studies are commonly conducted in many fields and usually yield two types of data, recurrent event data and panel count data [1, 2]. The former arises if all study subjects are followed continuously and thus gives complete information about the occurrence rates of

*Correspondence to: Jianguo Sun, 134E Middlebush Hall, Department of Statistics, University of Missouri-Columbia, MO, U.S.A, 65211.

†E-mail: sunj@missouri.com. Phone: (573)882-6667. Fax: (573)884-5524.

the recurrent events of interest. In contrast, the latter means that each study subject is observed only at discrete time points and hence only incomplete information is available for the occurrence rates of recurrent events. In practice, however, it is possible that some study subjects give complete information with the others giving incomplete information, or study subjects may provide complete information over certain time periods but only give incomplete information over the other time periods. That is, we have a third type of data, mixed recurrent event and panel count data [3]. In the following, we will discuss regression analysis of such mixed data in the presence of a dependent terminal event.

An example of mixed recurrent event and panel count data, which motivated this study, arose from the Childhood Cancer Survivor Study (CCSS), a longitudinal cohort study of childhood cancer survivors and healthy controls consisting of the siblings of a random sample of these survivors, conducted by St. Jude Children's Research Hospital [3, 4]. One main goal of the study is to assess the possible long-term effects of childhood cancer and cancer treatments on the subsequent reproductive function of the cancer survivors. For this, the pregnancy information of the participants was collected and used to compare the pregnancy rates of the cancer survivors and healthy controls. However, although some subjects provided complete recurrent event data, there existed some subjects who only provided incomplete panel count data over the whole follow-up period or some follow-up periods. That is, only mixed recurrent event and panel count data are available for the underlying pregnancy process. For the study of observed data, another complicating factor is premature death. Due to exposure to radiation or cytotoxic drugs during their cancer treatment, childhood cancer survivors may suffer some serious late effects such as second cancers or heart and lung problems and face an excess risk of early mortality compared to their healthy peers. In this example, the premature death prevents any further occurrences of pregnancies and should be considered as a terminal event in the recurrent event analysis.

A large literature has been established on the analysis of either recurrent event data or panel count data alone [1, 2] and also some methods have been developed for the analysis of these data with terminal events. For example, Huang and Wang [5], Ye et al. [6], Zeng and Cai [7], Zhu et al. [8] considered the analysis of recurrent event data with terminal events and proposed some frailty model approaches, which modeled the dependence between the recurrent and terminal events by using latent variables. Ghosh and Lin [9, 10], Zhao et al. [11], Zhao et al. [12] discussed the same problem but developed some marginal modeling approaches. Note that the latter four methods directly acknowledge the fact that there is no further occurrence of the events of interest after the terminal event, and also they leave the correlation between the underlying recurrent events and terminal events unspecified and thus can be more flexible. However, there does not seem to exist a method that can be used for the analysis of mixed recurrent event and panel count data in the presence of dependent terminal events. Corresponding to this issue, in the following, we will develop a marginal modeling approach for the situation.

The remainder of the article will be organized as follows. We will first begin in Section 2 with introducing some notation and models that will be used throughout the paper. The proposed estimating equation approach for estimation of regression parameters will then be presented in Section 3 and also in the section, the asymptotic properties of the proposed

estimator are established. In the approach, following Ghosh and Lin [10] and Zhao *et al.* [12], we will adopt the inverse probability of survival weighting technique to address the issue related to the terminal event. Section 4 reports some results from a simulation study conducted to assess the finite sample performance of the proposed inference procedure and they indicate that it seems to work well for practical situations. The proposed methodology is applied to the CCSS data described above in Section 5 and Section 6 contains some discussion and concluding remarks.

2. Notation and Models

Consider an event history study concerning recurrent events observed in n independent subjects. Let τ be a known constant representing the longest follow-up time, and for subject i , let $N_i(t)$ denote the total number of the occurrences of the event of interest up to time t and $O_i(t)$ the underlying observation process representing the cumulative number of potential observation times up to time t . Also for subject i , assume that there exists a sequence of observation time points $T_{i,0}=0 < T_{i,1} < T_{i,2} < \dots < T_{i,K_i} < \tau$, where K_i represents the total number of potential observation times up to τ , and over each interval $(T_{i,j-1}, T_{i,j}]$ or $(T_{i,K_i}, \tau]$, the subject gives either recurrent event data or panel count data. To describe this, define an indicator function $r_i(t)$ such that for $t \in (T_{i,j-1}, T_{i,j}]$ or $(T_{i,K_i}, \tau]$, $r_i(t) = 1$ if subject i gives recurrent event data and 0 otherwise, $j = 1, \dots, K_i$, $i = 1, \dots, n$. Also it is assumed that there exists a time of terminal event denoted by D_i which precludes the further jumps of the processes $N_i(t)$ and $O_i(t)$. That is, instead of $N_i(t)$ and $O_i(t)$, we really face the actual recurrent event and observation processes $N_i^*(t) = N_i(t \wedge D_i)$ and $O_i^*(t) = O_i(t \wedge D_i)$, which are constant after the terminal event occurs. Note that in reality, there may also exist an unrelated or noninformative censoring or follow-up time C_i such as an administrative stopping time. In these situations, the observations or follow-up will stop at $T_i^* = C_i \wedge D_i$ and we only observe $\tilde{N}_i(t) = N_i^*(t \wedge C_i)$ and $\tilde{O}_i(t) = O_i^*(t \wedge C_i)$.

Suppose that there exists a vector of external covariates [13] denoted by Z_i , $i = 1, \dots, n$. To describe covariate effects, we will consider the following model

$$E\{N_i^*(t) | D_i \geq t, Z_i\} = \exp\{\beta^T Z_i\} \mu_0(t), \quad (1)$$

where $\mu_0(t)$ denotes an unknown baseline mean function and β is a vector of regression parameters. That is, $N_i^*(t)$ follows the proportional means model before the terminal event. With respect to $O_i^*(t)$ and D_i we will assume that

$$E\{dO_i^*(t)\} = d\Gamma_0(t) \quad (2)$$

and

$$\lambda_i^d(t | Z_i) = \lambda_0^d(t) \exp\{\delta^T Z_i\}, \quad (3)$$

where $\Gamma_0(t)$ and $\lambda_0^d(t)$ are unknown mean and baseline hazard functions, respectively, and δ is a vector of regression parameters as β . Note that here we assume that D_i follows the proportional hazards model and the observation processes $O_i^*(t)$'s are homogeneous for simplicity. In addition, we consider a marginal model for $O_i^*(t)$ rather than a conditional model for $O_i^*(t)$ given $D_i \geq t$. Some comments on this can be found in Zhao *et al.* [12] and Luo and Huang [14].

Note that model (1) or similar models have been used for regression analysis of either recurrent event data or panel count data alone in the presence of terminal events. However, it is not straightforward to combine the methods developed for these situations together for regression analysis of mixed recurrent event and panel count data, partly due to the facts that one faces different data structures in different situations and as will be described next, the estimation procedures developed for these structures are different. For example, with recurrent event data, one does not have to deal with either the observation process $O_i^*(t)$ or the indicator function $r_i(t)$. In the following, we will assume that given Z_i , C_i is independent of $N_i^*(t)$, $O_i^*(t)$ and D_i , and that given $D_i \geq t$, $N_i^*(t)$, and $O_i^*(t)$ are independent.

3. Inference Procedure

In this section, we discuss estimation of regression parameters with primary interest on β since the estimation of regression parameter δ is straightforward with right-censored data under model (3) [15, 16]. To estimate β , we will construct an estimating equation and for this, we will look at the cases of recurrent event data and panel count data separately.

Note that in the case of recurrent event data, we have $r_i(t) = 1$ and similar as in Zhao *et al.* [11], one can show that

$$E\left\{\frac{I(t \leq T_i^*)}{S(t|Z_i)}\{\tilde{N}_i(t) - \exp\{\beta_0^T Z_i\}\mu_0(t)\}\right\}=0,$$

where β_0 denotes the true value of β and $S(t|Z_i)$ denotes the survival function of D_i given Z_i . This suggests that we can define the following zero-mean stochastic process

$$M_i^r(t; \beta_0) = \omega_i(t) \{\tilde{N}_i(t) - \exp\{\beta_0^T Z_i\}\mu_0(t)\},$$

where $\omega_i(t) = I(t \leq T_i^*)/S(t|Z_i)$.

For the case of panel count data with $r_i(t) = 0$, we can consider $N_i^*(t)dO_i^*(t)$ and $\tilde{N}_i(t)d\tilde{O}_i(t)$, and following Zhao *et al.* [12], one can show that

$$E\{N_i^*(t)dO_i^*(t)|Z_i\} = E\{N_i^*(t)|Z_i, D_i \geq t\}E\{dO_i^*(t)|Z_i\}$$

since

$$\begin{aligned}
 & E\{N_i^*(t)dO_i^*(t)|Z_i\} \\
 &= E\{E\{N_i^*(t)dO_i^*(t)|Z_i, D_i\}\} = E\{N_i^*(t)dO_i^*(t)|Z_i, D_i \geq t\}P\{D_i \geq t|Z_i\} \\
 &+ 0 \times P\{D_i < t|Z_i\} \\
 &= E\{N_i^*(t)|Z_i, D_i \geq t\}E\{dO_i^*(t)|Z_i, D_i \geq t\}P\{D_i \geq t|Z_i\} \\
 &= E\{N_i^*(t)|Z_i, D_i \geq t\}E\{dO_i^*(t)|Z_i\}
 \end{aligned}$$

and

$$E\{\tilde{N}_i(t)d\tilde{O}_i(t)|Z_i\} = E\{I(C_i \geq t)\exp\{\beta_0^T Z_i\}\mu_0(t)d\Gamma_0(t)\}.$$

However, the equation above cannot be used for the estimating equation directly since C_i and thus the term $I(C_i \geq t)$ are not observable for subjects with $D_i < C_i$. For this, motivated by the inverse probability of survival weighting technique, we propose to replace the indicator function $I(C_i \geq t)$ by $\omega_i(t) = I(T_i^* \geq t)/S(t|Z_i)$ since $E\{\omega_i(t)|Z_i\} = E\{I(C_i \geq t)|Z_i\}$. Therefore, we have the following zero-mean stochastic process

$$dM_i^p(t; \beta_0) = \tilde{N}_i(t)d\tilde{O}_i(t) - \omega_i(t)\exp\{\beta_0^T Z_i\}\mu_0(t)d\Gamma_0(t).$$

In general, for estimation of β for mixed data, we have the following estimating equations

$$\sum_{i=1}^n r_i(t)M_i^r(t; \beta) = 0, \quad \sum_{i=1}^n (1 - r_i(t))dM_i^p(t; \beta) = 0 \quad (4)$$

and

$$\sum_{i=1}^n \int_0^\tau r_i(t)Z_i M_i^r(t; \beta) dH(t) + (1 - r_i(t))Z_i dM_i^p(t; \beta) = 0, \quad (5)$$

where $H(t)$ is an increasing known weight function on $[0, \tau]$. Of course, in practice $S(t|Z_i)$ is also unknown and needs to be replaced by an estimator, say,

$$\hat{S}(t|Z_i) = \exp\left\{-\int_0^t \hat{\lambda}_0^d(s)\exp\{\hat{\delta}^T Z_i\}ds\right\},$$

where $\hat{\delta}$ denotes the estimator of δ obtained based on the partial likelihood and $\hat{\lambda}_0^d(s)$ is the Breslow's estimate of the baseline hazard function under model (3). Then one can solve the equations (4) and (5) with replacing $\omega_i(t)$ by $\hat{\omega}_i(t) = I(T_i^* \geq t)/\hat{S}_i(t|Z_i)$.

In practice, we can first solve equations (4) by fixing β and obtain

$$\hat{\mu}_0(t; \beta, \hat{\omega}(t)) = \frac{\sum_{i=1}^n r_i(t) \hat{\omega}_i(t) \tilde{N}_i(t)}{n \tilde{S}_r^{(0)}(t; \beta, \hat{\omega}(t))}$$

and

$$\hat{\mu}_0(t; \beta, \hat{\omega}(t)) d\hat{\Gamma}_0(t; \beta, \hat{\omega}(t)) = \frac{\sum_{i=1}^n (1 - r_i(t)) \tilde{N}_i(t) d\tilde{O}_i(t)}{n \tilde{S}_p^{(0)}(t; \beta, \hat{\omega}(t))},$$

where

$$\tilde{S}_r^{(d)}(t; \beta, \hat{\omega}(t)) = \frac{1}{n} \sum_{i=1}^n r_i(t) \hat{\omega}_i(t) \exp\{\beta^T Z_i\} Z_i^{\otimes d},$$

$$\tilde{S}_r^{(d)}(t; \beta, \hat{\omega}(t)) = \frac{1}{n} \sum_{i=1}^n (1 - r_i(t)) \hat{\omega}_i(t) \exp\{\beta^T Z_i\} Z_i^{\otimes d}, \quad d=0, 1, 2,$$

$\mathbf{a} \otimes^0 = \mathbf{1}$, $\mathbf{a} \otimes^1 = \mathbf{a}$, $\mathbf{a} \otimes^2 = \mathbf{a} \mathbf{a}'$, and $\omega(t) = (\omega_1(t), \dots, \omega_n(t))'$. By plugging the two estimators above into equation (5), we obtain

$$U(\beta) = \sum_{i=1}^n \int_0^T r_i(t) \hat{\omega}_i(t) (Z_i - \bar{Z}_r(t; \beta, \hat{\omega}(t))) \tilde{N}_i(t) dH(t) + (1 - r_i(t)) (Z_i - \bar{Z}_p(t; \beta, \hat{\omega}(t))) \tilde{N}_i(t) d\tilde{O}_i(t) = 0,$$

(6)

where $\bar{Z}_r(t; \beta, \hat{\omega}(t)) = \tilde{S}_r^{(1)}(t; \beta, \hat{\omega}(t)) / \tilde{S}_r^{(0)}(t; \beta, \hat{\omega}(t))$ and $\bar{Z}_p(t; \beta, \hat{\omega}(t)) = \tilde{S}_p^{(1)}(t; \beta, \hat{\omega}(t)) / \tilde{S}_p^{(0)}(t; \beta, \hat{\omega}(t))$.

Let $\hat{\beta}$ denote the estimator of β_0 given by solving equation (6). To establish the asymptotic properties of $\hat{\beta}$, define $N_i^d(t) = I(D_i \leq t, D_i \leq C_i)$,

$$M_i^d(t; \delta) = N_i^d(t) - \int_0^t I(T_i^* \geq s) \exp\{\delta^T Z_i\} d\Delta_0(s),$$

and $\bar{Z}_d(t; \hat{\delta}) = \tilde{S}_d^{(1)}(t; \hat{\delta}) / \tilde{S}_d^{(0)}(t; \hat{\delta})$, where $\Delta_0(t) = \int_0^t \lambda_0^d(s) ds$ and

$$\tilde{S}_d^{(d)}(t; \delta) = \frac{1}{n} \sum_{i=1}^n I(T_i^* \geq t) \exp\{\delta^T Z_i\} Z_i^{\otimes d}, \quad d=0, 1, 2.$$

Also define

$$\begin{aligned} & \sum_n \\ &= \frac{1}{n} \\ & \sum_{i=1}^n \left[\int_0^\tau r_i(t) (Z_i - \bar{Z}_r(t; \hat{\beta}, \hat{\omega}(t))) \widehat{M}_i^r(t) dH(t) + (1 - r_i(t)) (Z_i \right. \\ & \quad \left. - \bar{Z}_p(t; \hat{\beta}, \hat{\omega}(t))) d\widehat{M}_i^p(t) + \{ (Q_{1r} + Q_{1p})(Z_i - \bar{Z}_d(t; \hat{\delta})) \right. \\ & \quad \left. + \frac{Q_{2r}(t) + Q_{2p}(t)}{\tilde{S}_d^{(0)}(t; \hat{\delta})} \} d\widehat{M}_i^d(t) \right]^{\otimes 2}. \end{aligned}$$

In the above, $\widehat{M}_i^r(t)$ and $\widehat{M}_i^p(t)$ denote the estimates of $M_i^r(t; \beta_0)$, $M_i^p(t; \beta_0)$ and $M_i^d(t; \delta_0)$, respectively, defined as

$$\widehat{M}_i^r(t) = \hat{\omega}_i(t) \{ \tilde{N}_i(t) - \exp\{\hat{\beta}^T Z_i\} \hat{\mu}_0(t; \hat{\beta}, \hat{\omega}(t)) \},$$

$$\widehat{M}_i^p(t) = \tilde{N}_i(t) d\tilde{O}_i(t) - \hat{\omega}_i(t) \exp\{\hat{\beta}^T Z_i\} \hat{\mu}_0(t; \hat{\beta}, \hat{\omega}(t)) d\hat{\Gamma}_0(t; \hat{\beta}, \hat{\omega}(t)),$$

and

$$\widehat{M}_i^d(t) = N_i^d(t) - \int_0^t I(T_i^* \geq s) \exp\{\hat{\delta}^T Z_i\} d\hat{\Delta}_0(s; \hat{\delta})$$

with

$$\hat{\Delta}_0(t; \hat{\delta}) = \sum_{i=1}^n \int_0^t \frac{dN_i^d(s)}{n \tilde{S}_d^{(0)}(s; \hat{\delta})},$$

and δ_0 represents the true value of δ . Furthermore let

$$Q_{1r} = \frac{1}{n} \sum_{i=1}^n \int_0^\tau r_i(t) (Z_i - \bar{Z}_r(t; \hat{\beta}, \hat{\omega}(t))) \hat{g}^D(t; Z_i)^T \hat{\Omega}_D^{-1} \widehat{M}_i^r(t) dH(t),$$

$$Q_{1p} = -\frac{1}{n} \sum_{i=1}^n \int_0^\tau (1 - r_i(t)) (Z_i - \bar{Z}_p(t; \hat{\beta}, \hat{\omega}(t))) \hat{g}^D(t; Z_i)^T \hat{\Omega}_D^{-1} \hat{\omega}_i(t) \exp\{\hat{\beta}^T Z_i\} \hat{\mu}_0(t; \hat{\beta}, \hat{\omega}(t)) d\hat{\Gamma}_0(t; \hat{\beta}, \hat{\omega}(t)),$$

$$Q_{2r}(t) = \frac{1}{n} \sum_{i=1}^n \int_0^\tau r_i(s) (Z_i - \bar{Z}_r(s; \hat{\beta}, \hat{\omega}(s))) \exp\{\hat{\delta}^T Z_i\} I(t \leq s) \hat{M}_i^r(s) dH(s),$$

and

$$Q_{2p}(t) = -\frac{1}{n} \sum_{i=1}^n \int_0^\tau (1 - r_i(s)) (Z_i - \bar{Z}_p(s; \hat{\beta}, \hat{\omega}(s))) \exp\{\hat{\delta}^T Z_i\} I(t \leq s) \hat{\omega}_i(s) \exp\{\hat{\beta}^T Z_i\} \hat{\mu}_0(s; \hat{\beta}, \hat{\omega}(s)) d\hat{\Gamma}_0(s; \hat{\beta}, \hat{\omega}(s))$$

with

$$\hat{g}^D(t; Z_i) = \int_0^t \exp\{\hat{\delta}^T Z_i\} \{Z_i - \bar{Z}_d(s; \hat{\delta})\} d\hat{\Delta}_0(s),$$

and

$$\hat{\Omega}_D = \frac{1}{n} \sum_{i=1}^n \int_0^\tau \left\{ \frac{\tilde{S}_d^{(2)}(t; \hat{\delta})}{\tilde{S}_d^{(0)}(t; \hat{\delta})} - \bar{Z}_d(t; \hat{\delta})^{\otimes 2} \right\} dN_i^d(t).$$

In the Appendix, we show that under the regularity conditions (C1) – (C7) given there, $\hat{\beta}$ is consistent and one can approximate the distribution of $\sqrt{n}(\hat{\beta} - \beta_0)$ by the normal distribution with mean 0 and the variance-covariance matrix that can be estimated by

$\hat{A}_n^{-1} \sum_n \hat{A}_n^{-1}$, where

$$\begin{aligned} \hat{A}_n = & \frac{1}{n} \sum_{i=1}^n \int_0^\tau \{Z_i - \bar{Z}_r(t; \hat{\beta}, \hat{\omega}(t))\}^{\otimes 2} r_i(t) \hat{\omega}_i(t) \exp\{\hat{\beta}^T Z_i\} \hat{\mu}_0(t; \hat{\beta}, \hat{\omega}(t)) dH(t) \\ & + \{Z_i - \bar{Z}_p(t; \hat{\beta}, \hat{\omega}(t))\}^{\otimes 2} (1 - r_i(t)) \hat{\omega}_i(t) \exp\{\hat{\beta}^T Z_i\} \hat{\mu}_0(t; \hat{\beta}, \hat{\omega}(t)) d\hat{\Gamma}_0(t; \hat{\beta}, \hat{\omega}(t)). \end{aligned}$$

Note that if $r(t) = 1$ for all time t , the mixed data degenerate to recurrent event data and the proposed method share the same idea as that in Zhao *et al.* [11] when only time-invariant coefficients are considered. If $r(t) = 0$ for all time t , the observed mixed data reduce to panel count data and in this case, the proposed method actually coincides with the methodology developed by Zhao *et al.* [12]. Also note that to apply the estimation procedure above, one needs to select the weight function $H(t)$. For this, it is apparent that one natural choice is $H(t) = t$ and another one is $H(t) = 1/n \sum_{i=1}^n \tilde{N}_i(t)$. More comments on this can be found in Zhao *et al.* [11].

4. A Simulation Study

A simulation study was conducted to assess the finite sample performance of the estimation procedure proposed in the previous sections. In the study, we assumed that there was a single covariate Z_i following either the Bernoulli distribution with the success probability 0.5 or the uniform distribution over (0, 2). To generate the observed data, we first generated the censoring times C_i 's over {4, 4.5, 5} with equal probabilities. The longest follow-up time τ was set at 5. Also we generated the latent variables ν_i 's from the positive stable distribution with the Laplace transformation $L(s) = \exp(-s^\rho)$, where ρ is a constant with value 1 or 0.85. For given ν_i and Z_i , the terminal event time D_i was generated from a hazard function

$\lambda_i^d(t|Z_i, \nu_i) = \nu_i \lambda_0^d(t) \exp\{\delta_0^T Z_i\}$ with $\delta_0 = 0.05$ and $\lambda_0^d(t) = \lambda_0^d = 0.025$. It can be verified that D_i satisfies model (3) [12]. Note that under these settings, the censoring percentage on the terminal event times was between 10% and 15%, similar to that in the real data to be analyzed in the next section. Given ν_i , Z_i and T_i^* the recurrent event data were generated from a nonhomogeneous Poisson process with the mean function $\phi(\nu_i, t) \exp(\beta_0^T Z_i) \mu_0(t)$, where $\beta_0 = -0.5, 0$ or 0.5 , $\mu_0(t) = 0.5t$, and

$$\phi(\nu_i, t) = e^{-\nu_i} \exp\{(1 + \lambda_0^d \exp\{\delta_0^T Z_i\})^\rho - (\lambda_0^d \exp\{\delta_0^T Z_i\})^\rho\},$$

such that model (1) is satisfied. On the average, each subject has about 1 to 3 events, again similar to the real data to be analyzed in the next section.

To obtain the mixed recurrent event and panel count data, we first generated the observation times by assuming that given Z_i and T_i^* , $\tilde{O}_i(t)$ is the nonhomogeneous Poisson process with $E\{d\tilde{O}_i(t)|Z_i, T_i^*\} = 0.25S^{-1}(t|Z_i)I(T_i^* \geq t)$. That is, $\Gamma_0(t) = 0.25t$. Then for $t \in (T_{i,j-1}, T_{i,j}]$, we generated $r_i(t)$ from the Bernoulli distribution with the success probability p_r . That is, if $r_i(t) = 1$, we assumed that one observed recurrent event data over $(T_{i,j-1}, T_{i,j}]$ and otherwise, one only had panel count data or knew the number of the occurrences of the event of interest over the interval. The results given below are based on 1000 replications with the sample size $n = 200$ or 400 .

Table 1 presents the results obtained on estimation of β with $\rho = 1$, $p_r = 0.2, 0.5$ or 0.8 , the Z_i 's following the Bernoulli distribution, and the use of $H(t) = t$. They include the average of the biases given by the estimates minus the true value (BIAS), the average of the estimated standard errors (ESE), the sample standard errors (SSE) and the 95% empirical coverage probabilities (CP). Note that here ρ describes the correlation between the underlying recurrent event process and the terminal event and p_r represents the percentage of recurrent event data within the observed mixed data. One can see from the table that the results suggest that the proposed estimator seems to be unbiased and the variance estimation also seems to be reasonable. In addition, they indicate that the normal approximation to the distribution of the proposed estimator appears to be appropriate. Table 2 gives the result on estimation of β for the case where $\rho = 0.85$ with the other setups being the same as with Table 1, and the results for the case where the covariate was generated from the uniform

distribution on (0, 2) are presented in Tables 3 and 4. It is apparent that all results gave similar conclusions as with Table 1.

Note that for the analysis of mixed data, a commonly used naive approach is to coarsen the recurrent event data and then analyze the resulting panel count data with some existing method. To compare the proposed method to this naive method, in the simulation studies above, we also analyzed the generated data by using the method given in Zhao *et al.* [12] and included the estimation results in all four tables. One can see from these tables that as expected, the proposed method is more efficient than the naive approach and the efficiency gain depends on the proportion of the recurrent event data in the observed data.

5. An Application

In this section, we apply the methodology proposed in the previous sections to the CCSS data described in Section 1. As mentioned before, the study consists of childhood cancer survivors and healthy controls with a major goal being to assess the possible long-term effects of childhood cancer and cancer treatments on the subsequent reproductive function of the cancer survivors. In the study, a baseline summary questionnaire was distributed to more than 13,000 childhood cancer survivors who were diagnosed between 1970 and 1986 and had survived more than 5 years since diagnosis, and also to the siblings of a random sample of these childhood cancer survivors serving as a control group. The follow-up summary questionnaires were sent periodically thereafter. The information obtained from these summary questionnaires includes reports of all pregnancies and related outcomes. If a pregnancy was reported in any summary questionnaire, a detailed pregnancy questionnaire would be sent to the person to ask the precise age at pregnancy and other information. While both summary and pregnancy questionnaires were distributed between 1996 and 2000, not all participants who experienced a pregnancy completed the pregnancy questionnaire, so those participants provided only panel count data during some periods. Moreover, in some years (i.e., 2003, 2007), only one type of questionnaire was distributed, collecting either recurrent event data or panel count data since the last follow-up. As the result, the pregnancy data are mixed recurrent event and panel count data.

As one can expect, some subjects died during the follow-up and for these, the dates of death were obtained from their death certificates. For the analysis here, by following Zhu *et al.* [3], we will focus on a subgroup of the female participants who were at least 25 years old in 1996. The subgroup includes 3966 subjects with 2765 being childhood cancer survivors and the others being the siblings. Note that among these subjects, the majority of the survivors and siblings are not paired with only 410 pairs and all others being independent cancer survivors. For the observed data, the average pregnancy counts are 1.902, 1.684, and 2.403 for all subjects together, cancer survivors, and siblings, respectively. Among all participants, 376 died during the study, with 372 being in the cancer survivor group and 4 being in the sibling group. The ages at death range from 25 to 53 years old, with a median of 36, an age at which pregnancy is still possible.

Define $Z_i = 1$ if subject i is a cancer survivor and 0 otherwise. The application of the inference procedure proposed in the previous sections yielded $\hat{\beta} = -0.69$ with the estimated

standard error being 0.04. This corresponds to a p -value close to 0 for testing no difference between the pregnancy rates of the cancer survivors and control groups, and indicates that the pregnancy rate seems to be significantly lower for the cancer survivors than that for the siblings. In other words, the occurrence of childhood cancer and the treatments seem to have significantly negative effects on the subsequent reproductive function. In the analysis, we also obtained $\hat{\delta}=3.79$ with the estimated standard error of 0.50, and this suggests that the occurrence of childhood cancer and the treatments seem to have some significant negative effects on the lifetimes of the childhood cancer survivors too. For comparison, we also applied the naive approach described in the previous section and obtained $\hat{\beta}=-1.56$ with the estimated standard error being 0.11. In addition, the result above is consistent with but also suggests a stronger effect of the occurrence of childhood cancer and the treatments on the pregnancy than that obtained by Zhu *et al.* [3], who did not take into account the existence of the terminal event.

6. Discussion and Concluding Remarks

This paper discussed regression analysis of mixed recurrent event and panel count data in the presence of terminal events. As described above, such data occur when the observation schemes differ among subjects or from time period to time period. One major challenge for the analysis of these data is that one has to deal with different data structures and of course, the existence of terminal events can make the analysis even more complicated. For the problem, an estimating equation-based inference procedure on regression parameters was proposed with the use of the inverse probability of survival weighting technique, and the asymptotic properties of the proposed estimators were established. A simulation study was conducted for the assessment of the finite sample performance of the procedure and suggested that it seems to work well for practical situations.

As mentioned above, the proposed method is a marginal modeling approach and as an alternative, one could apply the joint modeling approach by using latent variables [6, 8]. In this latter case, typically one may need to specify the distribution of the latent variables and can apply the maximum likelihood estimation procedure. It is apparent that such procedures could give more efficient estimators of regression parameters but it may not be possible and practical to verify the assumed distribution of the latent variables. Also the interpretation of the regression parameters may not be as easy, as has been in the marginal modeling approach. In particular, the former acknowledges the fact that there is no further occurrence of the events of interest after the terminal event and leaves the correlation between the underlying recurrent event process and the terminal event unspecified. In contrast, the joint modeling approach may only describe certain or limited types of the correlation.

Note that in the proposed approach, for simplicity, it has been assumed that the observation process is independent of covariates, but it would be straightforward to generalize the proposed method to the situation where the observation process may depend on covariates. For this, one way is to employ the proportional rates or means model as model (1) to describe the relationship between the observation process and the covariates. Note that as mentioned above, we observe complete recurrent event data on the observation processes

$O_i^*(t)$'s and thus the inference about the assumed model is relatively easy and has been discussed by others [1]. Also for simplicity, we used the proportional means model and the proportional hazards model for the underlying recurrent event process and the terminal event, respectively. As alternatives, one could employ some other models such as the transformation means model for the former and the accelerated failure time model or the additive hazards model for the latter and develop similar inference procedures [2, 13]. In addition, in the preceding sections, we only considered time-independent covariates and it is straightforward to generalize the proposed method to the case with time-dependent covariates.

There exist several directions for future research. One is that in the analysis of CCSS, as pointed out by a reviewer, we simply treated each pregnancy as an instant event and it is clear that this is not true as each pregnancy has a duration period. In the literature of recurrent event data, this is often referred to as a recurrent episode [17, 18] and it is apparent that in general, the proposed method cannot apply to such situations and some new methods need to be developed. For the CCSS data, however, we believe that this should not affect the analysis results much since the time unit is in years and the events are relatively sparse per subject. Another issue that we did not address in the CCSS data analysis is the possible correlation between cancer survivors and their siblings. As mentioned above, the number of such subjects is small. In order to take such familial correlation into account, one faces clustered or correlated mixed recurrent event and panel count data and some new methods need to be developed.

Acknowledgments

The authors wish to thank the Editor, Dr. Ralph D'Agostino, the Associate Editor and two referees for their many helpful comments and suggestions that greatly improved the paper. This work was supported in part by the National Institutes of Health (R03 CA169150 and R21 CA198641) to Zhu and the funding from ALSAC and Cancer Center Support. The R code is available upon request to the corresponding author.

Appendix

We need the following regularity conditions to derive the asymptotic properties of the estimator $\hat{\beta}$.

- c1** $\{\tilde{N}_i(\cdot), \tilde{O}_i(\cdot), T_i^*, I(D_i \leq C_i), Z_i; i=1, \dots, n\}$ are independent and identically distributed.
- c2** There exists $\tau > 0$ such that $P(T_i^* \geq \tau | Z_i) > 0$.
- c3** Both $\tilde{N}_i(\tau)$ and $\tilde{O}_i(\tau)$ ($i = 1, \dots, n$) are bounded.
- c4** Each component of Z_i is bounded.
- c5** Weight function $H(t)$ converges to a nonrandom and bounded function $\tilde{H}(t)$ almost surely.

c6 The data type random process $r_i(t) (t \in [0, \tau])$ is conditionally independent of $N_i^*(t), O_i^*(t), T_i^*$, given Z_i and $0 < \inf_{0 \leq t \leq \tau} P(r_i(t) = 1 | Z_i) < \sup_{0 \leq t \leq \tau} P(r_i(t) = 1 | Z_i) < 1$.

c7 $A_{\beta_0} = E \left[\int_0^\tau \{Z_i - \bar{z}_r(t; \beta_0)\}^{\otimes 2} \omega_i(t) r_i(t) \exp\{\beta_0^T Z_i\} \mu_0(t) d\tilde{H}(t) + \{Z_i - \bar{z}_p(t; \beta_0)\}^{\otimes 2} \omega_i(t) (1 - r_i(t)) \exp\{\beta_0^T Z_i\} d\tilde{O}_i(t) \right]$ and $\Omega_D = E \left[\int_0^\tau \{Z_i - \bar{z}_d(t; \delta_0)\}^{\otimes 2} I(t \leq T_i^*) \exp\{\delta_0^T Z_i\} d\Delta_0(t) \right]$ are all positive definite, where $\bar{z}_r(t; \beta_0) = \lim_{n \rightarrow \infty} \bar{Z}_r(t; \beta_0, \omega(t))$, $\bar{z}_p(t; \beta_0) = \lim_{n \rightarrow \infty} \bar{Z}_p(t; \beta_0, \omega(t))$ and $\bar{z}_d(t; \delta_0) = \lim_{n \rightarrow \infty} \bar{Z}_d(t; \delta_0)$. Also,

$$s_r^{(d)}(t; \beta) = E\{\omega_i(t) r_i(t) \exp\{\beta^T Z_i\} Z_i^{\otimes d}\},$$

$$s_p^{(d)}(t; \beta) = E\{\omega_i(t) (1 - r_i(t)) \exp\{\beta^T Z_i\} Z_i^{\otimes d}\},$$

$$s_d^{(d)}(t; \delta) = E\{I(t \leq T_i^*) \exp\{\delta^T Z_i\} Z_i^{\otimes d}\}, d=0, 1, 2.$$

Consistency of $\hat{\beta}$

Define

$$\begin{aligned} X(\beta, \hat{\omega}) = & \frac{1}{n} \sum_{i=1}^n \int_0^\tau [(\beta - \beta_0)^T Z_i - \log\{\frac{\tilde{S}_r^{(0)}(t; \beta, \hat{\omega}(t))}{\tilde{S}_r^{(0)}(t; \beta_0, \hat{\omega}(t))}\}] \hat{\omega}_i(t) r_i(t) \tilde{N}_i(t) dH(t) \\ & + [(\beta - \beta_0)^T Z_i - \log\{\frac{\tilde{S}_p^{(0)}(t; \beta, \hat{\omega}(t))}{\tilde{S}_p^{(0)}(t; \beta_0, \hat{\omega}(t))}\}] (1 - r_i(t)) \tilde{N}_i(t) d\tilde{O}_i(t). \end{aligned}$$

From Strong Law of Large Numbers and consistency of $\hat{\delta}$ and $\hat{\Delta}_0(t; \hat{\delta})$, it is easy to see

$X(\beta, \hat{\omega}) \rightarrow \mathcal{X}(\beta)$, where

$$\begin{aligned} \mathcal{X}(\beta) = & E \left[\int_0^\tau [(\beta - \beta_0)^T Z_i - \log\{\frac{s_r^{(0)}(t; \beta)}{s_r^{(0)}(t; \beta_0)}\}] \omega_i(t) r_i(t) d\tilde{H}(t) + [(\beta - \beta_0)^T Z_i - \log\{\frac{s_p^{(0)}(t; \beta)}{s_p^{(0)}(t; \beta_0)}\}] (1 \right. \\ & \left. - r_i(t)) \tilde{N}_i(t) d\tilde{O}_i(t) \right] \end{aligned}$$

Following similar argument from Lin *et al.* [19], it follows that $\hat{\beta}$ is consistent.

Asymptotic normality of $\hat{\beta}$

Define

$$U(\beta_0; \hat{\omega}) = \sum_{i=1}^n \int_0^\tau r_i(t) \hat{\omega}_i(t) (Z_i - \bar{Z}_r(t; \beta_0, \hat{\omega})) \tilde{N}_i(t) dH(t) + (1 - r_i(t)) (Z_i - \bar{Z}_p(t; \beta_0, \hat{\omega})) \tilde{N}_i(t) d\tilde{O}_i(t),$$

then,

$$\begin{aligned} & n^{-1/2} U(\beta_0; \hat{\omega}) \\ &= n^{-1/2} \sum_{i=1}^n \int_0^\tau r_i(t) \hat{\omega}_i(t) (Z_i - \bar{Z}_r(t; \beta_0, \hat{\omega})) \tilde{N}_i(t) dH(t) + (1 - r_i(t)) (Z_i - \bar{Z}_p(t; \beta_0, \hat{\omega})) \tilde{N}_i(t) d\tilde{O}_i(t) \\ &= n^{-1/2} \sum_{i=1}^n \int_0^\tau r_i(t) (Z_i - \bar{z}_r(t; \beta_0)) M_i^r(t; \beta_0) d\tilde{H}(t) + (1 - r_i(t)) (Z_i - \bar{z}_p(t; \beta_0)) dM_i^p(t; \beta_0) \\ &\quad + r_i(t) (Z_i - \bar{z}_r(t; \beta_0)) (\hat{\omega}_i(t) - \omega_i(t)) \{ \tilde{N}_i(t) - \exp\{\beta_0^T Z_i\} \mu_0(t) \} d\tilde{H}(t) \\ &\quad + (1 - r_i(t)) (Z_i - \bar{z}_p(t; \beta_0)) (\omega_i(t) - \hat{\omega}_i(t)) \exp\{\beta_0^T Z_i\} \mu_0(t) d\Gamma_0(t) + o_p(1). \end{aligned}$$

It can be shown that

$$\omega_i(t) - \hat{\omega}_i(t) = I(T_i^* \geq t) \left(\frac{1}{S(t|Z_i)} - \frac{1}{\hat{S}(t|Z_i)} \right)$$

and

$$\begin{aligned} & n^{1/2} \left(\frac{1}{S(t|Z_i)} - \frac{1}{\hat{S}(t|Z_i)} \right) = -n^{1/2} \frac{1}{S(t|Z_i)^2} (S(t|Z_i) \\ &\quad - \hat{S}(t|Z_i)) \\ &\quad + o_p(1) = n^{1/2} \frac{1}{S(t|Z_i)} \left(\exp\left\{ - \left(\int_0^t \exp\{\hat{\delta}^T Z_i\} d\hat{\Delta}_0(s; \hat{\delta}) - \int_0^t \exp\{\delta_0^T Z_i\} d\Delta_0(s) \right) \right\} - 1 \right) \\ &\quad + o_p(1) \\ &= n^{1/2} \frac{1}{S(t|Z_i)} \left(\int_0^t \exp\{\delta_0^T Z_i\} d\Delta_0(s) - \int_0^t \exp\{\hat{\delta}^T Z_i\} d\hat{\Delta}_0(s; \hat{\delta}) \right) \\ &\quad + o_p(1) = -n^{1/2} \frac{1}{S(t|Z_i)} \left(\int_0^t \exp\{\hat{\delta}^T Z_i\} (d\hat{\Delta}_0(s; \hat{\delta}) - d\Delta_0(s)) \right. \\ &\quad \left. + (\exp\{\hat{\delta}^T Z_i\} - \exp\{\delta_0^T Z_i\}) d\Delta_0(s) \right) \\ &\quad + o_p(1) = -n^{1/2} \frac{1}{S(t|Z_i)} \left\{ \int_0^t \sum_{j=1}^n \frac{\exp\{\delta_0^T Z_i\} dM_j^d(s; \delta_0)}{ns_d^{(0)}(s; \delta_0)} \right. \\ &\quad \left. + g^D(t; Z_i) \Omega_D^{-1} \frac{1}{n} \sum_{j=1}^n \int_0^\tau \{Z_j - \bar{z}_d(s; \delta_0)\} dM_j^d(s; \delta_0) \right\} + o_p(1), \end{aligned} \tag{A1}$$

and

$$g^D(t; Z_i) = \int_0^t \exp\{\delta_0^T Z_i\} \{Z_i - \bar{z}_d(s; \delta_0)\} d\Delta_0(s).$$

By plugging the expression of (A1) back in $U(\beta_0; \hat{\omega})$ shown above, we obtain the following:

$$\begin{aligned} & n^{-1/2} U(\beta_0; \hat{\omega}) \\ &= n^{-1/2} \sum_{i=1}^n \int_0^\tau r_i(t) (Z_i - \bar{z}_r(t; \beta_0)) M_i^r(t; \beta_0) d\tilde{H}(t) + (1 - r_i(t)) (Z_i - \bar{z}_p(t; \beta_0)) dM_i^p(t; \beta_0) \\ &+ \{(Q_{1r} + Q_{1p})(Z_i - \bar{z}_d(t; \delta_0)) + \frac{Q_{2r}(t) + Q_{2p}(t)}{s_d^{(0)}(t; \delta_0)}\} dM_i^d(t; \delta_0) + o_p(1), \end{aligned}$$

(A2)

where

$$Q_{1r} = \frac{1}{n} \sum_{i=1}^n \int_0^\tau r_i(t) (Z_i - \bar{z}_r(t; \beta_0)) g^D(t; Z_i)^T \Omega_D^{-1} M_i^r(t; \beta_0) d\tilde{H}(t),$$

$$Q_{2r}(t) = \frac{1}{n} \sum_{i=1}^n \int_0^\tau r_i(s) (Z_i - \bar{z}_r(s; \beta_0)) \exp\{\delta_0^T Z_i\} I(t \leq s) M_i^r(s; \beta_0) d\tilde{H}(s),$$

$$Q_{1p} = -\frac{1}{n} \sum_{i=1}^n \int_0^\tau (1 - r_i(t)) (Z_i - \bar{z}_p(t; \beta_0)) g^D(t; Z_i)^T \Omega_D^{-1} \omega_i(t) \exp\{\beta_0^T Z_i\} \mu_0(t) d\Gamma_0(t),$$

$$Q_{2p}(t) = -\frac{1}{n} \sum_{i=1}^n \int_0^\tau (1 - r_i(s)) \exp\{\delta_0^T Z_i\} (Z_i - \bar{z}_p(s; \beta_0)) I(t \leq s) \omega_i(s) \exp\{\beta_0^T Z_i\} \mu_0(s) d\Gamma_0(s).$$

Then use of Taylor expansions of $U(\beta_0, \hat{\omega})$ and (A2) yields

$$n^{1/2}(\hat{\beta} - \beta_0) = A_{\beta_0}^{-1} n^{-1/2} U(\beta_0, \hat{\omega}) + o_p(1),$$

then the asymptotic normality holds from the multivariate central limit theorem.

References

1. Cook, RJ., Lawless, JF. The Analysis of Recurrent Event Data. Springer-Verlag; New York: 2007.
2. Sun, J., Zhao, X. Statistical Analysis of Panel Count Data. Springer; 2013.

3. Zhu L, Tong X, Zhao H, Sun J, Srivastava DK, Leisenring W, Robison LL. Statistical analysis of mixed recurrent event data with application to cancer survivor study. *Statistics in Medicine*. 2013; 32(11):1954–1963. [PubMed: 23139023]
4. Robison LL, Mertens AC, Boice JD, Breslow NE, Donaldson SS, Green DM, Li FP, Meadows AT, Mulvihill JJ, Neglia JP, Nesbit ME, Packer RJ, Potter JD, Sklar CA, Smith MA, Stovall M, Strong LC, Yasui Y, Zeltzer LK. Study design and cohort characteristics of the childhood cancer survivor study: A multi-institutional collaborative project. *Medical and Pediatric Oncology*. 2002; 38(4): 229–239. [PubMed: 11920786]
5. Huang CY, Wang MC. Joint modeling and estimation for recurrent event processes and failure time data. *Journal of the American Statistical Association*. 2004; 99(468):1153–1165. [PubMed: 24068850]
6. Ye YN, Kalbfleisch JD, Schaubel DE. Semiparametric analysis of correlated recurrent and terminal events. *Biometrics*. 2007; 63(1):78–87. [PubMed: 17447932]
7. Zeng D, Cai JW. A semiparametric additive rate model for recurrent events with an informative terminal event. *Biometrika*. 2010; 97(3):699–712. [PubMed: 23956458]
8. Zhu L, Sun J, Srivastava DK, Tong X, Leisenring W, Zhang H, Robison LL. Semiparametric transformation models for joint analysis of multivariate recurrent and terminal events. *Statistics in Medicine*. 2011; 30(25):3010–3023. [PubMed: 21786280]
9. Ghosh D, Lin DY. Nonparametric analysis of recurrent events and death. *Biometrics*. 2000; 56(2): 554–562. [PubMed: 10877316]
10. Ghosh D, Lin DY. Marginal regression models for recurrent and terminal events. *Statistica Sinica*. 2002; 12(3):663–688.
11. Zhao X, Zhou J, Sun L. Semiparametric transformation models with time-varying coefficients for recurrent and terminal events. *Biometrics*. 2011; 67(2):404–414. [PubMed: 20618309]
12. Zhao H, Li Y, Sun J. Analyzing panel count data with a dependent observation process and a terminal event. *Canadian Journal of Statistics*. 2013; 41(1):174–191.
13. Kalbfleisch, JD., Prentice, RL. *The Statistical Analysis of Failure Time Data*. John Wiley & Sons; New York: 2011.
14. Luo XH, Huang CY. A comparison of various rate functions of a recurrent event process in the presence of a terminal event. *Statistical Methods in Medical Research*. 2010; 19(2):167–182. [PubMed: 18445699]
15. Cox DR. Partial likelihood. *Biometrika*. 1975; 62(2):269–276.
16. Andersen PK, Gill RD. Cox's regression model for counting processes: a large sample study. *The Annals of Statistics*. 1982:1100–1120.
17. Zhu L, Zhao H, Sun J, Pounds S, Zhang H. Joint analysis of longitudinal data and recurrent episodes data with application to medical cost analysis. *Biometrical Journal*. 2013; 55(1):5–16. [PubMed: 23225670]
18. Hu XJ, Lorenzi M, Spinelli JJ, Ying SC, McBride ML. Analysis of recurrent events with non-negligible event duration, with application to assessing hospital utilization. *Lifetime Data Analysis*. 2011; 17(2):215–233. [PubMed: 20730625]
19. Lin DY, Wei LJ, Yang I, Ying Z. Semiparametric regression for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2000; 62(4):711–730.

Table 1Simulation results with $\rho = 1$ and $Z \sim B(1, 0.5)$

n	p_r	β_0	Proposed method				Naive method			
			BIAS	ESE	SSE	CP	BIAS	ESE	SSE	CP
200	0.2	0.5	-0.013	0.143	0.149	0.942	-0.011	0.164	0.166	0.948
		0	-0.011	0.156	0.168	0.937	-0.010	0.175	0.179	0.951
		-0.5	-0.012	0.175	0.181	0.943	-0.014	0.192	0.197	0.939
		0.5	-0.018	0.121	0.132	0.933	-0.020	0.163	0.169	0.937
	0.5	0	-0.016	0.135	0.142	0.939	-0.021	0.175	0.180	0.933
		-0.5	-0.018	0.153	0.162	0.927	-0.017	0.191	0.201	0.934
		0.5	-0.007	0.108	0.120	0.924	-0.010	0.163	0.173	0.930
		0	-0.012	0.120	0.129	0.919	-0.008	0.175	0.174	0.944
	0.8	-0.5	-0.009	0.138	0.148	0.932	-0.013	0.192	0.195	0.947
		0.5	0.002	0.102	0.106	0.940	0	0.116	0.117	0.951
		0	-0.011	0.110	0.116	0.930	-0.012	0.124	0.127	0.940
		-0.5	-0.007	0.124	0.127	0.936	-0.008	0.137	0.137	0.952
400	0.2	0.5	-0.004	0.087	0.089	0.943	-0.001	0.116	0.119	0.945
		0	-0.006	0.095	0.100	0.938	-0.008	0.124	0.126	0.946
		-0.5	-0.009	0.109	0.114	0.942	-0.010	0.136	0.136	0.953
		0.5	-0.004	0.076	0.081	0.928	-0.006	0.116	0.113	0.954
	0.5	0	-0.014	0.085	0.087	0.942	-0.018	0.124	0.123	0.950
		-0.5	-0.005	0.098	0.104	0.931	-0.010	0.136	0.138	0.945

Table 2Simulation results with $\rho = 0.85$ and $Z \sim B(1, 0.5)$

n	p_r	β_0	Proposed method				Naive method			
			BIAS	ESE	SSE	CP	BIAS	ESE	SSE	CP
200	0.2	0.5	-0.012	0.164	0.173	0.933	-0.011	0.184	0.190	0.936
		0	-0.015	0.176	0.182	0.938	-0.014	0.196	0.200	0.946
		-0.5	-0.022	0.193	0.210	0.924	-0.022	0.210	0.231	0.934
		0.5	-0.005	0.143	0.150	0.938	-0.004	0.184	0.184	0.950
	0.5	0	-0.015	0.155	0.163	0.937	-0.003	0.195	0.197	0.946
		-0.5	-0.007	0.171	0.180	0.938	-0.009	0.210	0.218	0.939
		0.5	-0.008	0.127	0.139	0.923	-0.012	0.184	0.182	0.958
		0	-0.013	0.139	0.153	0.920	-0.016	0.195	0.198	0.943
	0.8	-0.5	-0.025	0.155	0.169	0.920	-0.025	0.211	0.218	0.937
		0.5	-0.007	0.117	0.129	0.930	-0.004	0.131	0.138	0.929
		0	-0.015	0.125	0.131	0.939	-0.011	0.138	0.144	0.937
		-0.5	-0.010	0.128	0.142	0.941	-0.012	0.150	0.152	0.950
400	0.2	0.5	-0.009	0.101	0.113	0.921	-0.008	0.131	0.135	0.940
		0	-0.012	0.110	0.115	0.943	-0.013	0.139	0.140	0.948
		-0.5	-0.007	0.122	0.127	0.946	-0.004	0.150	0.153	0.946
		0.5	-0.008	0.090	0.098	0.927	-0.008	0.131	0.134	0.939
	0.5	0	-0.006	0.098	0.107	0.927	0	0.138	0.139	0.951
		-0.5	-0.013	0.110	0.117	0.931	-0.012	0.150	0.153	0.941

Table 3Simulation results with $\rho = 1$ and $Z \sim \mathcal{U}(0, 2)$

n	p_r	β_0	Proposed method				Naive method			
			BIAS	ESE	SSE	CP	BIAS	ESE	SSE	CP
200	0.2	0.5	-0.019	0.113	0.122	0.934	-0.019	0.133	0.139	0.943
		0	-0.015	0.134	0.152	0.901	-0.016	0.152	0.163	0.925
		-0.5	-0.014	0.166	0.177	0.929	-0.016	0.180	0.182	0.940
		0.5	-0.015	0.094	0.107	0.917	-0.009	0.134	0.143	0.934
	0.5	0	-0.021	0.116	0.123	0.929	-0.024	0.150	0.152	0.954
		-0.5	-0.021	0.148	0.158	0.926	-0.024	0.182	0.187	0.948
		0.5	-0.012	0.081	0.091	0.915	-0.016	0.134	0.137	0.937
		0	-0.013	0.103	0.109	0.921	-0.014	0.152	0.151	0.948
	0.8	-0.5	-0.020	0.134	0.142	0.935	-0.018	0.181	0.193	0.932
		0.5	-0.009	0.080	0.086	0.931	-0.008	0.095	0.098	0.944
		0	-0.009	0.096	0.093	0.957	-0.008	0.108	0.107	0.956
		-0.5	-0.011	0.119	0.127	0.920	-0.014	0.129	0.136	0.932
400	0.2	0.5	-0.009	0.067	0.073	0.926	-0.007	0.095	0.101	0.933
		0	-0.009	0.083	0.087	0.938	-0.009	0.108	0.110	0.939
		-0.5	-0.010	0.105	0.112	0.935	-0.003	0.129	0.135	0.939
		0.5	-0.012	0.058	0.066	0.918	-0.014	0.095	0.101	0.926
	0.5	0	-0.007	0.074	0.079	0.932	-0.009	0.108	0.108	0.949
		-0.5	-0.010	0.094	0.098	0.935	-0.010	0.129	0.133	0.945

Table 4Simulation results with $\rho = 0.85$ and $Z \sim \mathcal{U}(0, 2)$

n	p_r	β_0	Proposed method				Naive method			
			BIAS	ESE	SSE	CP	BIAS	ESE	SSE	CP
200	0.2	0.5	-0.012	0.133	0.149	0.924	-0.014	0.153	0.162	0.935
		0	-0.014	0.151	0.162	0.933	-0.015	0.169	0.171	0.949
		-0.5	-0.015	0.182	0.187	0.941	-0.017	0.197	0.201	0.944
		0.5	-0.021	0.114	0.127	0.913	-0.013	0.118	0.129	0.923
	0.5	0	-0.025	0.133	0.148	0.916	-0.026	0.168	0.179	0.933
		-0.5	-0.012	0.164	0.180	0.922	-0.016	0.198	0.215	0.925
		0.5	-0.017	0.101	0.116	0.913	-0.022	0.154	0.158	0.937
		0	-0.023	0.120	0.136	0.906	-0.029	0.167	0.178	0.932
	0.8	-0.5	-0.018	0.148	0.157	0.930	-0.020	0.197	0.207	0.935
		0.5	-0.019	0.095	0.098	0.934	-0.019	0.109	0.109	0.947
		0	-0.008	0.108	0.114	0.933	-0.010	0.120	0.120	0.942
		-0.5	-0.013	0.131	0.132	0.938	-0.010	0.141	0.144	0.951
400	0.2	0.5	-0.009	0.081	0.088	0.928	-0.014	0.108	0.109	0.949
		0	-0.006	0.095	0.102	0.930	-0.003	0.120	0.121	0.949
		-0.5	-0.009	0.116	0.121	0.928	-0.004	0.141	0.143	0.951
		0.5	-0.006	0.072	0.079	0.924	-0.007	0.109	0.112	0.936
	0.5	0	-0.012	0.085	0.092	0.928	-0.011	0.119	0.122	0.932
		-0.5	-0.010	0.105	0.112	0.932	-0.004	0.141	0.143	0.943