ORIGINAL PAPER

Semiparametric regression analysis of doubly censored recurrent event data

P. G. Sankaran[1] · S. Hari[1] · E. P. Sreedevi[1] (iD)

Abstract

Recurrent event data are common in survival and reliability studies, where a subject experiences the same type of event repeatedly. There are situations, in which the event of interest can be observed only if they belong to a window of observational range, leading to double censoring of recurrent event times. In this paper, we study recurrent event data subject to double censoring. We propose a proportional mean model for the analysis of doubly censored recurrent event data based on the mean function of the underlying recurrent event process. The estimators of the regression parameters and the baseline mean function are derived and their asymptotic properties are studied. A Monte Carlo simulation study is conducted to assess the finite sample behavior of the proposed estimators. Finally, the procedures are illustrated using two real-life data sets, one from a bladder cancer study and the other from a study on chronic granulomatous disease.

Keywords Counting process · Double censoring · Recurrent events · Mean function

1 Introduction

Reliability and survival studies often involve events that can occur more than once per subject, known as recurrent events. These types of outcomes are prevalent in various fields, such as biology, medicine, reliability engineering, and social sciences. Medical examples of recurrent events include tumor recurrences, sequences of asthmatic attacks, and multiple infection episodes. Recurrent events can also occur in other contexts, such as the recurrence of economic recessions, traffic accidents, or repeated breakdowns of devices like mobile phones or computers (Du & Lv, 2022).

✉ E. P. Sreedevi
 sreedeviep@gmail.com

1 Department of Statistics, Cochin University of Science and Technology, Cochin, India

Regression analysis methods for recurrent events can be classified into two categories: conditional methods and marginal techniques. Conditional methods involve modeling the intensity or hazard rate function (Prentice et al., 1981; Andersen & Gill, 1982), while marginal techniques focus on mean functions or rate functions (Pepe & Cai, 1993; Lawless & Nadeau, 1995; Lin et al., 2000). Marginal methods have the advantage of providing an easier interpretation of the mean number of recurrent events compared to the intensity of the event. Frailty models have been proposed for recurrent event data by Zeng and Lin (2007) and Sankaran and Anisha (2011), while Kelly and Lim (2000) provide a comparison of various approaches for recurrent event analysis. A comprehensive review of recurrent event analysis is presented by Cook and Lawless (2007). Su and Lin (2021) discussed a semiparametric rate model for recurrent event data with multiple event types, that incorporates cyclic or periodic components.

In many survival studies, the lifetimes of patients can be accurately observed only if it occurs within a window of observational time. The remaining lifetimes are only known to be either left censored or right censored. The data so obtained are referred to as doubly censored data. The analysis of doubly censored data has been studied by several researchers. Pioneering works include Turnbull (1974), Chang (1990), and Gu and Zhang (1993). Regression analysis for doubly censored data has been discussed by Cai and Cheng (2004) and semiparametric transformation models for doubly censored data were studied by Shen (2011). Ji et al. (2012) proposed a quantile regression method for doubly censored data. Proportional hazard models for the analysis of doubly censored data with multiple modes of failure have been proposed by Sankaran and Sreedevi (2016), while Shen and Chen (2018) and Shen (2022) introduced Aalen's linear model and the Cox–Aalen model for doubly censored data. Recently, Li et al. (2020) presented a regression analysis method for multivariate doubly censored data using flexible semiparametric transformation frailty models, and Choi et al. (2021) proposed the Buckley–James method for an accelerated failure time model under double random censoring. Additionally, Choi and Huang (2021) discussed nonparametric maximum-likelihood estimation of semiparametric transformation models for double censored data. A joint model for the birth distribution and lifetime distribution in the context of heterogeneous survival data under double truncation is studied by Dörre (2021).

While recurrent event times are often subject to right censoring, there are situations where the data may also have other types of incomplete information. For instance, recurrent event data with general interval censoring has been studied by Chen et al. (2005) and Shen and Cook (2015). If the event of interest can occur more than once per subject, and these event recurrences can be subjected to both left censoring and right censoring, we obtain doubly censored recurrent event data. In such scenarios, the recurrence of an event can only be observed within an observational window defined by left and right censoring random variables, denoted as U and V, respectively.

Even though recurrent event data exposed to different types of censoring have been studied, recurrent event data subject to double censoring has not been explored yet. This motivates us to propose a semiparametric regression model for the analysis of doubly censored recurrent event data, based on marginal mean function of the cumulative number of recurrent events. The mean function characterizes the subject's recurrence experience, and our proposed semiparametric regression model incorporates multi-

plicative covariate effects on the marginal mean function under double censoring. The effect of covariates on censoring random variables is also modeled using multiplicative intensity models.

The paper is organized as follows. In Sect. 2, we introduce the model and develop inference procedures to estimate regression parameters. The estimating equation approach is employed in inference problems. Asymptotic properties of the proposed estimators are discussed in detail. Estimation of the baseline cumulative mean function is also studied. A Monte Carlo simulation study is carried out in Sect. 3 to assess the finite sample behavior of the proposed estimators. The model is then applied to real-life data sets in Sect. 4. Finally, major conclusions of the study are presented in Sect. 5.

2 Model and inference procedures

Consider an event history study of some recurrent events that involve n independent subjects. Let $N_i(t)$ be the cumulative number of events that have occurred up to time t, for $0 \le t \le t^*$, $i = 1, 2, \ldots, n$ where t^* denotes the length of the study. Let T_{ij} be the recurrence times of the ith subject for $j = 1, 2, \ldots, m_i$; $i = 1, 2, \ldots, n$. Assume that T_{ij}'s are doubly censored by the random variables U_i and V_i; specifically left censored by U_i and right censored by V_i. This means that we can observe the exact event recurrence time T_{ij} only when it falls within the observational window $[U_i, V_i]$, where $U_i < V_i$, for $i = 1, 2, \ldots, n$. If $T_{ij} < U_i$, the recurrence time will be left-censored, and if $T_{ij} > V_i$, the recurrence time will be right-censored, where U_i and V_i are always observed. Denote $X_{ij} = \min\{\max(T_{ij}, U_i), V_i)\}$ for $j = 1, 2, \ldots, m_i$; $i = 1, 2, \ldots, n$. Let $\mathbf{Z}_i$ be a $p \times 1$ vector of covariates observed for each subject i, and we assume independence between T_i and (U_i, V_i) given $\mathbf{Z}_i$. Let $E(N_i(t)|\mathbf{Z}_i) = \mu(t|\mathbf{Z}_i)$ be the mean function of $N_i(t)$ conditional on $\mathbf{Z}_i$. We consider the proportional mean model to specify the covariate effect on recurrence times, given by

$$\mu(t|\mathbf{Z}_i) = \mu_0(t) \exp(\boldsymbol{\beta}^{\mathrm{T}} \mathbf{Z}_i), \tag{2.1}$$

where $\mu_0(t)$ is the unspecified baseline mean function and $\boldsymbol{\beta}$ is the $p \times 1$ vector of regression parameters. Furthermore, the covariate $\mathbf{Z}_i$ may also have an effect on the censoring random variables U_i and V_i.

To model the covariate effect on censoring random variables, we assume that the hazard rate function $\lambda_1(t)$ of U_i satisfies the proportional hazards (PH) model

$$\lambda_1(t|\mathbf{Z}_i) = \lambda_{10}(t) \exp(\boldsymbol{\gamma}^{\mathrm{T}} \mathbf{Z}_i), \tag{2.2}$$

and due to the strict order restriction between the censoring variables U_i and V_i, the hazard rate function $\lambda_2(t)$ of V_i is assumed to satisfy the proportional hazards (PH) model (as discussed in Wang et al. (2010))

$$\lambda_2(t|U_i, \mathbf{Z}_i) = I(t > U_i) \lambda_{20}(t) \exp(\boldsymbol{\tau}^{\mathrm{T}} \mathbf{Z}_i), \tag{2.3}$$

where $\lambda_{10}(t)$ and $\lambda_{20}(t)$ are completely unspecified baseline hazard functions, and $\boldsymbol{\gamma}$ and $\boldsymbol{\tau}$ are p-dimensional vector of regression parameters.

Define $\delta_{1ij} = I(T_{ij} \leq U_i)$, $\delta_{2ij} = I(U_i < T_{ij} \leq V_i)$ and $\delta_{3ij} = I(T_{ij} > V_i)$, where $I(A)$ is the indicator function of set A.

Denote

$$N_i^{(1)}(t) = \sum_{j=1}^{m_i} I(X_{ij} \leq t, \delta_{1ij} = 1) \text{ and } N_i^{(2)}(t) = \sum_{j=1}^{m_i} I(X_{ij} \leq t, \delta_{3ij} = 1).$$

be the number of left-censored recurrence times and right-censored recurrence times, respectively, for i^{th} individual. Then

$$E(dN_i^{(1)}(t)|\mathbf{Z}_i) = \lambda_1(t|\mathbf{Z}_i) \text{ and } E(dN_i^{(2)}(t)|U_i, \mathbf{Z}_i) = \lambda_2(t|U_i, \mathbf{Z}_i).$$

Also denote $N_i^*(t) = \sum_{j=1}^{m_i} I(X_{ij} \leq t)$ and $N_i(t) = \sum_{j=1}^{m_i} I(T_{ij} \leq t)$. Note that

$$P(X_{ij} \leq t \leq V_i|\mathbf{Z}_i) = P(U_i < t \leq V_i|\mathbf{Z}_i) \cdot P(T_{ij} \leq t|\mathbf{Z}_i). \tag{2.4}$$

Let $W(t|\mathbf{Z}_i) = P(U < t \leq V|\mathbf{Z}_i) = \bar{H}(t|\mathbf{Z}_i) - \bar{G}(t|\mathbf{Z}_i)$, where $\bar{G}(t|\mathbf{Z}_i) = P(U > t|\mathbf{Z}_i)$ and $\bar{H}(t|\mathbf{Z}_i) = P(V \geq t|U_i, \mathbf{Z}_i)$. Now, using the probability statement in (2.4), the consistent estimators of $\mu_0(t)$ and $\boldsymbol{\beta}$ are obtained by solving the following two estimating equations:

$$\sum_{i=1}^{n} [N_i^*(t) - W(t|\mathbf{Z}_i)\mu_0(t) \exp(\boldsymbol{\beta}^T \mathbf{Z}_i)] = 0 \tag{2.5}$$

and

$$\sum_{i=1}^{n} \int_{y_a}^{y_b} \mathbf{Z}_i [N_i^*(t) - W(t|\mathbf{Z}_i)\mu_0(t) \exp(\boldsymbol{\beta}^T \mathbf{Z}_i)] dk(t) = 0, \tag{2.6}$$

where y_a and y_b are prespecified constants, such that $P(X_{ij} \leq y_a)$ and $P(X_{ij} \geq y_b)$ are positive. Here, $k(t)$ is a data-dependent weight function, and in practice, we use $\hat{k}(t) = \frac{1}{N} \sum_{i=1}^{n} \sum_{j=1}^{m_i} I(X_{ij} \leq t)$, where $N = \sum_{i=1}^{n} m_i$. Thus, $\hat{k}(t)$ is an increasing weight function that uniformly converges to the deterministic function $k(t)$ within the interval $t \in [y_a, y_b]$.

The quantity $W(t|\mathbf{Z}_i)$ in (2.5) and (2.6) is unknown. To estimate $W(t|\mathbf{Z})$, we define $\Lambda_{10}(t) = \int_0^t \lambda_{10}(s)ds$ and $\Lambda_{20}(t) = \int_0^t \lambda_{20}(s)ds$ as the cumulative hazard rate functions of left censoring random variables U and right censoring random variables V, respectively. Then, the survival functions $\bar{G}(t|\mathbf{Z})$ and $\bar{H}(t|\mathbf{Z})$ are given by

$$\bar{G}(t|\mathbf{Z}) = \exp(-\Lambda_{10}(t)\exp(\boldsymbol{\gamma}^T \mathbf{Z}))$$

and

$$\bar{H}(t|\mathbf{Z}) = \exp(-\Lambda_{20}(t)\exp(\boldsymbol{\tau}^{\mathrm{T}}\mathbf{Z})).$$

To estimate $\boldsymbol{\gamma}$, we consider the equation

$$0 = U_1(\boldsymbol{\gamma}) = \sum_{i=1}^{n} \int_{y_a}^{y_b} \left[\mathbf{Z}_i - \frac{\sum_{i=1}^{n} I(U_i \geq t) \exp(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{Z}_i)\mathbf{Z}_i}{\sum_{i=1}^{n} I(U_i \geq t) \exp(\boldsymbol{\gamma}^{\mathrm{T}}\mathbf{Z}_i)} \right] dN_i^{(1)}(t). \tag{2.7}$$

To estimate $\boldsymbol{\tau}$, we consider the equation

$$0 = U_2(\boldsymbol{\tau}) = \sum_{i=1}^{n} \int_{y_a}^{y_b} \left[\mathbf{Z}_i - \frac{\sum_{i=1}^{n} I(V_i \geq t > U_i) \exp(\boldsymbol{\tau}^{\mathrm{T}}\mathbf{Z}_i)\mathbf{Z}_i}{\sum_{i=1}^{n} I(V_i \geq t > U_i) \exp(\boldsymbol{\tau}^{\mathrm{T}}\mathbf{Z}_i)} \right] dN_i^{(2)}(t). \tag{2.8}$$

Let $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\tau}}$ be the estimators of $\boldsymbol{\gamma}$ and $\boldsymbol{\tau}$ given by $U_1(\boldsymbol{\gamma}) = 0$ and $U_2(\boldsymbol{\tau}) = 0$. The estimators of $\Lambda_{10}(t)$ and $\Lambda_{20}(t)$ are Breslow type estimators. That is, for $\Lambda_{10}(t)$, the estimator is

$$\hat{\Lambda}_{10}(t) = \sum_{i=1}^{n} \int_{0}^{t} \frac{I(U_i \geq s)dN_i^{(1)}(s)}{\sum_{i=1}^{n} I(U_i \geq s) \exp(\hat{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{Z}_i)}. \tag{2.9}$$

For $\Lambda_{20}(t)$, the estimator is given by

$$\hat{\Lambda}_{20}(t) = \sum_{i=1}^{n} \int_{0}^{t} \frac{I(V_i \geq s > U_i)dN_i^{(2)}(s)}{\sum_{i=1}^{n} I(V_i \geq s) \exp(\hat{\boldsymbol{\tau}}^{\mathrm{T}}\mathbf{Z}_i)}. \tag{2.10}$$

Then, the survival functions $\bar{G}(t|\mathbf{Z})$ and $\bar{H}(t|\mathbf{Z})$ can be estimated as $\bar{G}_n(t|\mathbf{Z})$ and $\bar{H}_n(t|\mathbf{Z})$, where

$$\bar{G}_n(t|\mathbf{Z}) = \exp(-\hat{\Lambda}_{10}(t)\exp(\hat{\boldsymbol{\gamma}}^{\mathrm{T}}\mathbf{Z}))$$

and

$$\bar{H}_n(t|\mathbf{Z}) = \exp(-\hat{\Lambda}_{20}(t)\exp(\hat{\boldsymbol{\tau}}^{\mathrm{T}}\mathbf{Z})).$$

Now, by substituting $W(t|\mathbf{Z})$ with $\bar{W}_n(t|\mathbf{Z}) = \bar{H}_n(t|\mathbf{Z}) - \bar{G}_n(t|\mathbf{Z})$ in estimating equations (2.5) and (2.6), one can estimate $\mu_0(t)$ and $\boldsymbol{\beta}$.

We can estimate the baseline mean function $\mu_0(t)$ as

$$\hat{\mu}_0(t) = \frac{\sum_{i=1}^{n} N_i^*(t)}{\sum_{i=1}^{n} \bar{W}_n(t|\mathbf{Z}_i)\exp(\hat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{Z}_i)}. \tag{2.11}$$

Remark 2.1 When covariates do not have effect on the censoring times U and V, a simple estimate for $W(t|\mathbf{Z})$ is proposed as $W_n^*(t) = \frac{1}{n}\sum_{i=1}^{n} I(U_i < t \leq V_i)$. This can be substituted in (2.5) and (2.6) and the estimators of $\mu_0(t)$ and $\boldsymbol{\beta}$ can be obtained by solving those equations. We can note that, when $m_i = 1$ for all i, this situation reduces to the model studied in Cai and Cheng (2004).

2.1 Asymptotic properties

We now discuss the asymptotic properties of the estimators of regression parameters. We prove the consistency and asymptotic normality of the estimators of $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\tau}$. To establish the asymptotic properties, we define the martingale

$$M_i(t) = Y_i(t)[N_i^*(t) - \mu_0(t)W(t|\mathbf{Z}_i)\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_i)], \quad i = 1, 2, \ldots, n, \quad (2.12)$$

where $Y_i(t)$ represents the at-risk process, and we assume the following regularity conditions, similar to those given in Liu et al. (2010) and Du and Lv (2022):

C1: $\{N_i^*(.), N_i^{(1)}(.), N_i^{(2)}(.), Y_i(.), \mathbf{Z}_i, i = 1, 2, \ldots, n\}$ are independent and identically distributed, where $Y_i(t)$ is the at-risk process.

C2: $N_i^*(.)$, $N_i^{(1)}(.)$ and $N_i^{(2)}(.)$ are bounded by constants.

C3: $\mathbf{Z}_i, i = 1, 2, \ldots, n$ are of bounded total variation on $[0, t^*]$.

C4: The matrices B, P and A defined in Appendix are non-singular.

The consistency and asymptotic normality of $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\tau}}$, the regression coefficients for the left and right censoring times, respectively, are detailed in the Appendix.

Theorem 2.1 *Under the regularity conditions C1–C4, the estimator $\hat{\boldsymbol{\beta}}$ is strongly consistent and $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ can be approximated by a $p-$variate normal distribution with mean vector 0 and variance–covariance matrix given by Σ, where*

$$\Sigma = \frac{1}{n}A^{-1}\left(\sum_{i=1}^{n}\Omega_i\Omega_i^T\right)A^{-1},$$

where A and Ω_i are specified in Appendix in (A.7) and (A.9), respectively.

Theorem 2.2 *Under the regularity conditions C1–C4, $\sqrt{n}(\hat{\mu}_0(t) - \mu_0(t))$ converges weakly to a zero-mean Gaussian process with covariance function at (s, t) given by*

$$\Gamma(s, t) = E(\Psi_1(s)\Psi_1(t)),$$

where

$$\Psi_1(t) = \frac{M_1(t)}{n^{-1}\sum_{i=1}^{n} Y_i(t)W(t|\mathbf{Z}_i)\exp(\boldsymbol{\beta}^T\mathbf{Z}_i)}$$
$$- D(t, \boldsymbol{\beta})A^{-1}\int_{y_a}^{y_b}(\mathbf{Z}_1 - \bar{\mathbf{Z}}(t))M_1(t)dk(t),$$

with

$$D(t, \boldsymbol{\beta}) = \frac{\sum_{i=1}^n Y_i(t) W(t|\mathbf{Z}_i) \mathbf{Z}_i \mu_0(t) \exp(\boldsymbol{\beta}^T \mathbf{Z}_i)}{\sum_{i=1}^n Y_i(t) W(t|\mathbf{Z}_i) \exp(\boldsymbol{\beta}^T \mathbf{Z}_i)}. \tag{2.13}$$

A consistent estimator of $\Gamma(s, t)$ *is given by*

$$\hat{\Gamma}(s, t) = \frac{1}{n} \sum_{i=1}^n \hat{\Psi}_i(s) \hat{\Psi}_i(t),$$

where

$$\hat{\Psi}_i(t) = \frac{\hat{M}_i(t)}{n^{-1} \sum\limits_{i=1}^n Y_i(t) \bar{W}_n(t|\mathbf{Z}_i) \exp(\hat{\boldsymbol{\beta}}^T \mathbf{Z}_i)} - \hat{D}(t, \hat{\boldsymbol{\beta}}) \hat{A}^{-1} \int_{y_a}^{y_b} (\mathbf{Z}_i - \bar{\mathbf{Z}}(t)) \hat{M}_i(t) d\hat{k}(t),$$

with $\hat{M}_i(t)$ *and* $\hat{D}(t, \hat{\boldsymbol{\beta}})$ *are obtained from* (2.12) *and* (2.13)*, respectively, by replacing all the unknown quantities with their corresponding sample estimators. The terms* $\bar{\mathbf{Z}}(t)$ *and* $\hat{A}^{-1}$ *are defined in* (A.4) *and* (A.10)*, respectively.*

The proofs of the above theorems are given in Appendix. Note that the variance–covariance matrix of the estimator of $\boldsymbol{\beta}$ does not yield to a closed form. In practice, one can use bootstrap re-sampling technique to find the standard error of $\hat{\boldsymbol{\beta}}$. The confidence intervals of $\hat{\boldsymbol{\beta}}$ can be constructed using the asymptotic normality of the estimator $\hat{\boldsymbol{\beta}}$.

3 Simulation study

A Monte Carlo simulation study is carried out to assess the finite sample behavior of the proposed estimators. We consider two covariates Z_1 and Z_2, where Z_1 is generated from a Bernoulli distribution with a probability of success 0.5 and Z_2 is generated from a standard normal distribution. Based on these two covariates, under the proportional hazard assumption, the left and right censoring times U and V are generated separately, and then, the recurrence times are simulated accordingly. The data generation for (U, V) is independent of the recurrence times, given (Z_1, Z_2).

Following (2.2) and (2.3), we consider three different combinations for modeling the left and right censoring time. The left and right censoring times are assumed to follow the following three models:

(i) A Gompertz proportional hazard model with baseline hazard function

$$\lambda_{10}(t) = \theta_1 \exp\{\alpha_1 t\} \text{ and } \lambda_{20}(t) = \theta_2 \exp\{\alpha_2 t\},$$

where $\theta_1 = 0.2$, $\alpha_1 = 0.1$, $\theta_2 = 0.1$, and $\alpha_2 = 0.1$. We assume the regression coefficients $\boldsymbol{\gamma} = \boldsymbol{\tau} = (1, 2)^T$.

(ii) A Weibull proportional hazard model with with baseline hazard function

$$\lambda_{10}(t) = \theta_1 \alpha_1 t^{\alpha_1 - 1} \text{ and } \lambda_{20}(t) = \theta_2 \alpha_2 t^{\alpha_2 - 1},$$

where $\theta_1 = 0.2$, $\alpha_1 = 2$, $\theta_2 = 0.3$, and $\alpha_2 = 2.5$. Here, $\boldsymbol{\gamma} = \boldsymbol{\tau} = (1.5, 1)^{\mathrm{T}}$.

(iii) A baseline power model with baseline hazard function

$$\lambda_{10}(t) = t^{\alpha_1} \text{ and } \lambda_{20}(t) = t^{\alpha_2},$$

where $\alpha_1 = 1$, and $\alpha_2 = 2$. In this case, $\boldsymbol{\gamma} = \boldsymbol{\tau} = (1, 2)^{\mathrm{T}}$.

In all the cases, the left censoring times, U, take values in $[0, 5]$. To ensure a sufficient follow-up time for every subject, right censoring times, V, are generated between $U + 5$ and $U + 20$. The recurrence times follow a Gompertz proportional hazards model:

$$\lambda(t|\mathbf{Z}) = \theta \exp\{\alpha t\} \cdot \exp\{\beta_1 Z_1 + \beta_2 Z_2\},$$

with $\theta = 0.1$ and $\alpha = 0.1$ and $\lambda(t|\mathbf{Z})$ as the corresponding hazard rate function in presence of covariates. We generate 1000 data samples of sizes $n = 50, 100$, and 200 each with different choices of β_1 and β_2. In our simulation, we do not fix the censoring rates but rather consider random censoring proportions for different parameter combinations. Various combinations of parameters result in different proportions for left and right censoring times and varying levels of the average number of recurrences per subject. For example, here is a summary of the censoring proportions observed for some of the combinations in our study:

For Model (i) with a Gompertz proportional hazards model for left and right censoring times, with $\beta_1 = 1.3$ and $\beta_2 = 2$, the average number of events per subject is 21.14. The left censoring rate is found to be 8.91% of the total events, while the right censoring rate is 35.75% of the total events. The double censored data have an average of 11.48 events per subject. Under the same setting but with $\beta_1 = 0.2$ and $\beta_2 = 0.4$, the average number of events per subject reduces to 4.18. In this case, the left censoring rate is 10.47% and the right censoring rate is 25.49%. The double censored data have an average of 2.17 events per subject.

Similarly, for Model (ii) with $\beta_1 = 1.5$ and $\beta_2 = 1.9$, the average number of events per subject is 28.03. The left censoring rate is 10.56% and the right censoring rate is 8.51%. For the same model but with $\beta_1 = 0.2$ and $\beta_2 = 0.4$, the average number of events per subject is 2.55. In this case, the left censoring rate is 9.94% and the right censoring rate is 27.69%.

For Model (iii) with $\beta_1 = -0.5$ and $\beta_2 = -1.0$, the average number of events per subject is 1.83. The left censoring rate is 11.19% and the right censoring rate is 31.60%. When considering $\beta_1 = 1.3$ and $\beta_2 = 2.0$, the average number of events per subject is 24.66. The left censoring rate is 12.43% and the right censoring rate is 6.03%.

The absolute bias, mean squared error (MSE), and 95% coverage probability (CP) of the regression parameters for all the above models are calculated. Since, for all

Table 1 Absolute bias, MSE, and CP of $\hat{\boldsymbol{\beta}}$ for case (i)

β_1	β_2	n	Absolute bias		MSE		95 % CP	
			$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$
1.3	2	50	0.1114	0.1401	0.0297	0.0282	0.9470	0.9560
		100	0.0977	0.1202	0.0189	0.0188	0.9580	0.9590
		200	0.0812	0.1039	0.0135	0.0131	0.9610	0.9660
1.5	1.9	50	0.1451	0.1224	0.0427	0.0228	0.9360	0.9460
		100	0.1390	0.1074	0.0295	0.0155	0.9380	0.9510
		200	0.1299	0.0928	0.0213	0.0106	0.9410	0.9540
0.2	0.4	50	0.0299	0.0608	0.0252	0.0103	0.9460	0.9410
		100	0.0287	0.0526	0.0094	0.0062	0.9480	0.9440
		200	0.0215	0.0420	0.0056	0.0031	0.9540	0.9510
−0.5	-1	50	0.1205	0.2297	0.0505	0.0621	0.9510	0.9560
		100	0.1110	0.2123	0.0274	0.0588	0.9540	0.9580
		200	0.1003	0.2052	0.0216	0.0472	0.9660	0.9590

Table 2 Absolute bias, MSE, and CP of $\hat{\boldsymbol{\beta}}$ for case (ii)

β_1	β_2	n	Absolute bias		MSE		95 % CP	
			$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$
1.3	2	50	0.0759	0.2803	0.0387	0.1027	0.9410	0.9440
		100	0.0579	0.2444	0.0174	0.0723	0.9430	0.9460
		200	0.0430	0.2005	0.0096	0.0480	0.9470	0.9470
0.6	0.3	50	0.0544	0.0174	0.0330	0.0064	0.9360	0.9490
		100	0.0505	0.0125	0.0179	0.0034	0.9470	0.9520
		200	0.0438	0.0101	0.0096	0.0022	0.9540	0.9530
1.5	1.9	50	0.1166	0.2466	0.0459	0.0804	0.9410	0.9310
		100	0.1050	0.2145	0.0240	0.0556	0.9460	0.9440
		200	0.0828	0.1808	0.0139	0.0394	0.9560	0.9610
1.5	-2	50	0.1133	0.2543	0.0417	0.0856	0.9460	0.9320
		100	0.0784	0.2001	0.0222	0.0485	0.9470	0.9390
		200	0.0669	0.1731	0.0121	0.0355	0.9510	0.9450

choices of β_1 and β_2, the results are similar, we present the results for randomly chosen four combinations of β_1 and β_2, for each model in Tables 1–3.

Tables 1–3 demonstrate that both absolute bias and MSE decrease as the sample size increases, indicating the consistency of the proposed estimation procedure. Furthermore, the coverage probability stabilizes near the chosen significance level. We also note that the estimated baseline mean functions for different combinations are of a similar nature. Therefore, we present only the estimated baseline mean functions for case (i) with $\beta_1 = 1.5$ and $\beta_2 = 1.9$, for both the choices of $\bar{W}_n(t \mid \mathbf{Z})$ and $W_n^*(t \mid \mathbf{Z})$ in Fig. 1. In Fig. 1, the blue line represents the estimated values of $\mu_0(t)$

Table 3 Absolute bias, MSE, and CP of $\hat{\boldsymbol{\beta}}$ for case (iii)

β_1	β_2	n	Absolute bias		MSE		95 % CP	
			$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_1$	$\hat{\beta}_2$
1.3	2	50	0.1899	0.2059	0.0675	0.0848	0.9410	0.9400
		100	0.1402	0.1890	0.0438	0.0596	0.9470	0.9480
		200	0.1092	0.1471	0.0272	0.0437	0.9540	0.9560
1.5	1.9	50	0.1855	0.2245	0.0679	0.0641	0.9380	0.9340
		100	0.1363	0.1555	0.0584	0.0486	0.9410	0.9390
		200	0.1014	0.1189	0.0426	0.0365	0.9470	0.9450
0.3	0.5	50	0.0518	0.0364	0.0278	0.0083	0.9450	0.9410
		100	0.0489	0.0237	0.0139	0.0038	0.9490	0.9440
		200	0.0343	0.0153	0.0075	0.0029	0.9520	0.9480
−0.5	-1	50	0.0886	0.1545	0.0518	0.0471	0.9320	0.9450
		100	0.0642	0.1287	0.0244	0.0414	0.9400	0.9510
		200	0.0515	0.0896	0.0207	0.0362	0.9430	0.9540

using the weight function $\bar{W}_n(t \mid \mathbf{Z})$, green line represents the estimated values of $\mu_0(t)$ using the weight function $W_n^*(t \mid \mathbf{Z})$, and the red line represents the true values. We can observe that toward the end of the follow-up times, the estimator slightly overestimates the baseline mean function, for both choices of weight functions.

4 Data analysis

The inference procedure discussed in Sect. 2 is illustrated using two real-life data sets.

4.1 Multiple tumor recurrence data of bladder cancer patients

We apply the proposed methods to the multiple tumor recurrence data for patients with bladder cancer given in Wei et al. (1989). These data were obtained in a randomized clinical trial conducted by the Veterans Administration Co-operative Urological Research Group (VACURG) in which the patients with superficial bladder tumors were randomly assigned to one of three treatments: thiotepa, pyridoxine, and placebo. This data set is recently analyzed by Du and Lv (2022). The data are available in the R package `survival` (Therneau, 2023). In this analysis, we consider the patients treated with either placebo or thiotepa and having at least 1 month of follow-up, which results in the analysis of 47 patients assigned to placebo and 38 patients assigned to thiotepa. The right censoring times (V) are taken as the maximum follow-up time for each patient. Since the data are not doubly censored as such, we have chosen the left censoring times (U) randomly by observing the data. For patients with follow-up time greater than 8 months, U is generated randomly as a time less than or equal to 8 months. For all other patients, a time less than or equal to its right censoring time is
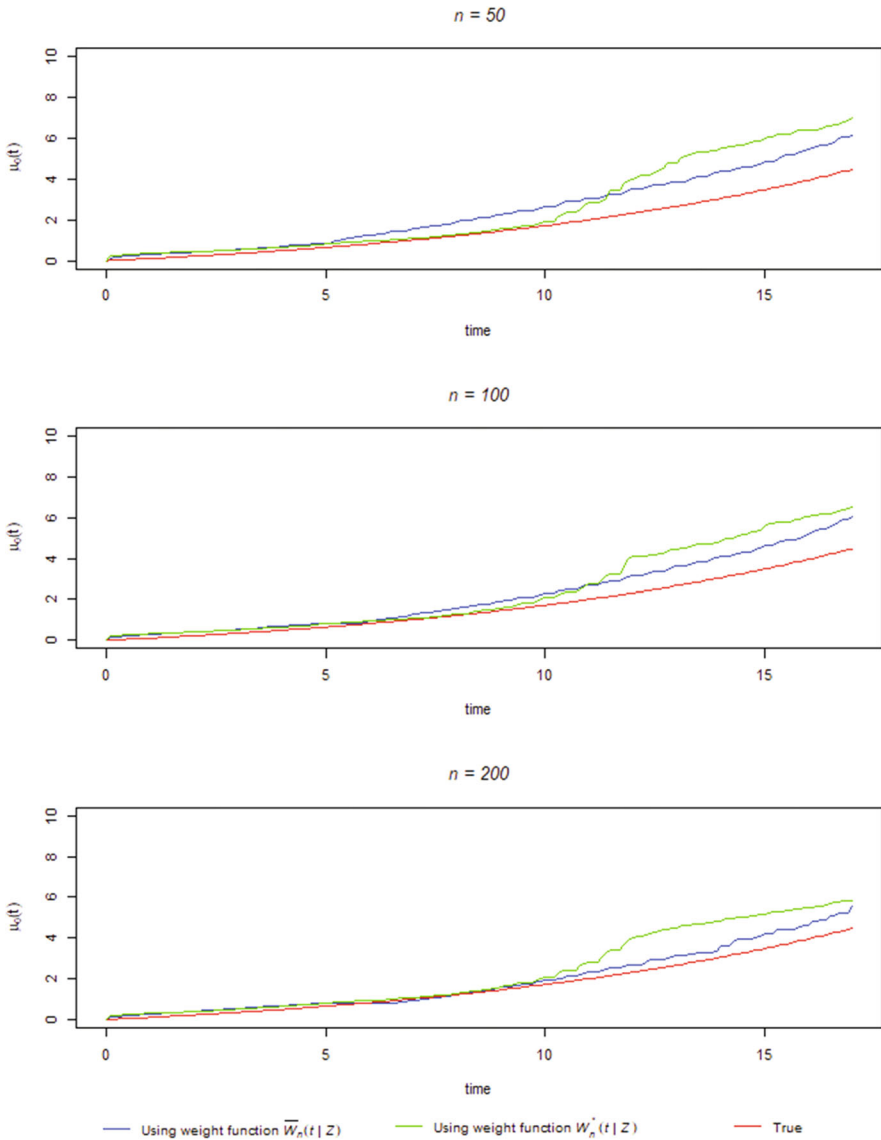
Fig. 1 Estimate of $\mu_0(t)$ for case (i) with $\beta_1 = 1.5$ and $\beta_2 = 1.9$ for sample sizes 50, 100, and 200

randomly assigned as its left censoring time U. In the new data, 14 out of 112 events were lost due to left censoring.

The main interest of this analysis is to study the influence of thiotepa on recurrence frequency and the dependence of such recurrence rates on the size of initial tumors. Therefore, the covariates considered are the treatment group (0—thiotepa; 1—placebo) (Z_1) and the number of initial tumors (Z_2) for each of the patients. In these data, we

Table 4 Estimates and SE of regression coefficients for the tumor recurrence data

Regression coefficient	Estimate	SE	PWP-TT Model	
			Estimate	SE
β_1	0.4032	0.1267	0.3805	0.2066
β_2	0.1422	0.0209	0.1682	0.0480
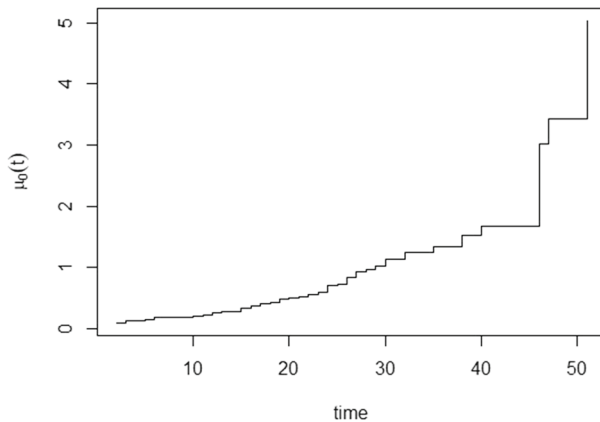
Fig. 2 Estimate of $\mu_0(t)$ for the tumor recurrence data

have the left censoring times U and the right censoring times V are independent of the covariate vector $\mathbf{Z} = (Z_1, Z_2)$.

The estimates of the model parameters, along with their corresponding standard errors (SE), are provided in Table 4. We also obtain estimates using the Prentice, Williams, and Peterson total time (PWP-TT) model (Prentice et al., 1981), which only considers right censoring and was fitted to the original dataset. From Table 4, it is evident that the estimates obtained from our proposed model align with those from the PWP-TT model. This agreement demonstrates the significant reduction in tumor recurrences for cancer patients with thiotepa treatment. The estimate of $\mu_0(t)$ is depicted in Fig. 2, showing an increase in the baseline mean function after 46 months.

4.2 Study of gamma interferon in chronic granulomatous disease

We consider the Chronic Granulomatous Disease (CGD) data described in Fleming and Harrington (1991) for illustration. The dataset is accessible within the R package `survival` (Therneau, 2023). CGD refers to a set of inherited rare immune system disorders characterized by recurrent pyogenic infections that typically appear early in life and may lead to death in childhood. The efficacy of gamma interferon in reducing the rate of serious infections in CGD patients was investigated in a double blinded clinical trial in which patients were randomly assigned to receive either gamma interferon or a placebo. Out of 128 patients enrolled, 63 were randomized into the gamma interferon group and the rest to the placebo group. Among the placebo group, 30

Table 5 Estimates and SE of regression coefficients for the CGD data

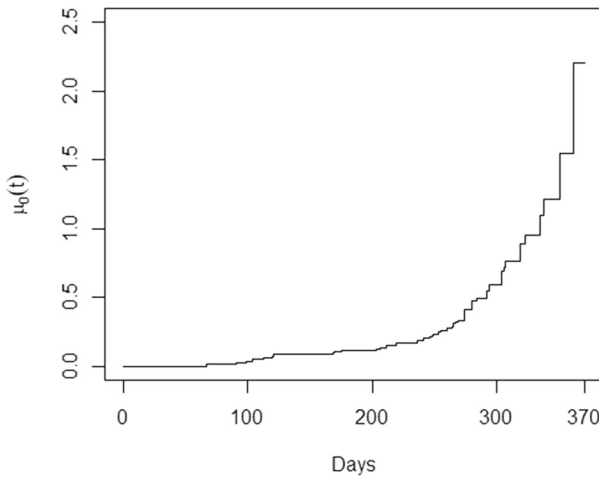| Regression coefficient	Estimate	SE	PWP-TT Model	
			Estimate	SE
β_1	1.0546	0.3604	1.1222	0.2614
β_2	−0.0369	0.0186	−0.0305	0.0131

Fig. 3 Estimate of $\mu_0(t)$ for the CGD data

patients experienced at least one infection, while 14 patients in the gamma interferon group encountered at least one infection.

We randomly chose left censoring times (U) by observing the data, because the data are not doubly censored as such. For patients having follow-up periods greater than 200 days, the left censoring time is randomly selected to be less than or equal to 200 days. All other patient's left censoring times are chosen randomly to be less than or equal to their right censoring time. In the modified data, 20 out of 76 events were lost due to left censoring. Several covariates were recorded, but we focus on two of them, treatment group Z_1, ($Z_1 = 0$ for gamma interferon and $Z_1 = 1$ for placebo) and the patient's age Z_2. We can see that the left and right censoring times are independent of the covariates for this data.

The estimates of the model parameters along with their standard errors (SE) are presented in Table 5. We also calculate the estimates of regression parameters using the PWP-TT model on the original data set, and the results are shown in Table 4. It is evident that the estimates obtained from our proposed model are consistent with those from the PWP-TT model. This shows that in people with CGD, gamma interferon decreases infection recurrence. Additionally, as patients get older, the rate of severe infections declines. The estimate of $\mu_0(t)$ is plotted in Fig. 3.

5 Conclusion

Recurrent event data subject to double censoring are often encountered in clinical and other lifetime studies. In this article, we proposed a semiparametric regression model for recurrent event data, which specifies the multiplicative covariate effect on the marginal mean function under double censoring. The estimating equation approach is employed for the estimation of regression parameters. Asymptotic properties of the proposed estimates are discussed in detail. The results of the Monte Carlo simulation study make sure that the procedures followed are efficient. The proposed method was demonstrated on two real data sets: the multiple tumor recurrence data for bladder cancer patients, and a study of gamma interferon in chronic granulomatous disease.

Several open problems remain to be addressed in the analysis of recurrent event data subject to double censoring. In this article, we considered time-independent covariates. However, recurrent event data often involve time-dependent covariates as well. Miloslavsky et al. (2004) proposed an inverse probability of censoring weighted estimator for the regression parameters in the Andersen–Gill model for recurrent event data in the presence of time-dependent covariates and dependent censoring. Analysis of multivariate recurrent data in the presence of time-dependent covariates was studied by Sun et al. (2009) and Zhao et al. (2012). Regression analysis of recurrent event data with time-dependent covariates and informative censoring was studied by Huang et al. (2010) which allows correlations between censoring times and recurrent event process via frailty. Therefore, it is of interest to develop methods for the regression analysis of doubly censored recurrent events in the presence of time-dependent covariates, and work in this direction will be studied separately.

In the analysis of recurrent event data, two common time scales are often employed: the total time and the time since the last event (gap time). Recurrent event data subject to double censoring can be modeled using the gap times between recurrent events. A semiparametric transformation model can be developed for such data type, which include proportional hazards and proportional odds models as special cases. Further, one can develop inference procedures for doubly censored recurrent event data with multiple modes of recurrence. The modeling and analysis of doubly censored recurrent event data using frailty models is not yet discussed in literature. These problems are under investigation and will be reported in separate studies.

Data availability The source of the data set used for illustration purpose is mentioned in the manuscript.

Declarations

Conflicts of interest The authors declare that there is no conflict of financial or non-financial interests that are directly or indirectly related to this research work.

Ethical statement The submitted work is original and has not been published elsewhere.

Appendix

A: Proofs of results in Sect. 2.1

We make use of the results in Cai and Cheng (2004), for doubly censored data in a non-recurrent setup to prove the theorems stated in Sect. 2.1. Sun and Wei (2000) proved the asymptotic results of the estimators of regression parameters of panel count data with covariate-dependent observation and censoring times, in a similar way. Furthermore, the consistency and asymptotic normality of $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\tau}}$, which are the regression coefficients in a proportional hazard model, have already been established in the literature and are stated below.

Theorem A.1 *Under the regularity conditions C1–C4, the estimator $\hat{\boldsymbol{\gamma}}$ is consistent. Further, $\sqrt{n}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})$ can be approximated by a $p-$variate normal distribution with mean vector 0 and variance–covariance matrix given by*

$$B^{-1}(\boldsymbol{\gamma}) \left(\sum_{i=1}^{n} d_i(\boldsymbol{\gamma}) d_i^T(\boldsymbol{\gamma}) \right) B^{-1}(\boldsymbol{\gamma})$$

where

$$B(\boldsymbol{\gamma}) = -\frac{\partial U_1(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}}$$

and

$$d_i(\boldsymbol{\gamma}) = \int_{y_a}^{y_b} \left[\mathbf{Z}_i - \frac{\sum_{i=1}^{n} I(U_i \geq t) \exp(\boldsymbol{\gamma}^T \mathbf{Z}_i) \mathbf{Z}_i}{\sum_{i=1}^{n} I(U_i \geq t) \exp(\boldsymbol{\gamma}^T \mathbf{Z}_i)} \right]$$
$$\left[dN_i^{(1)}(t) - \frac{I(U_i \geq t) \exp(\boldsymbol{\gamma}^T \mathbf{Z}_i)}{\sum_{i=1}^{n} I(U_i \geq t) \exp(\boldsymbol{\gamma}^T \mathbf{Z}_i)} d\bar{N}_1(t) \right] \quad (5.1)$$

with $\bar{N}_1(t) = \sum_{i=1}^{n} N_i^{(1)}(t)$.

Theorem A.2 *Under the regularity conditions C1–C4, the estimator $\hat{\boldsymbol{\tau}}$ is consistent. Also, $\sqrt{n}(\hat{\boldsymbol{\tau}} - \boldsymbol{\tau})$ can be approximated by a $p-$variate normal distribution with mean vector 0 and variance–covariance matrix given by*

$$P^{-1}(\boldsymbol{\tau}) \left(\sum_{i=1}^{n} e_i(\boldsymbol{\tau}) e_i^T(\boldsymbol{\tau}) \right) P^{-1}(\boldsymbol{\tau})$$

where

$$P(\boldsymbol{\tau}) = -\frac{\partial U_2(\boldsymbol{\tau})}{\partial \boldsymbol{\tau}}$$

and

$$e_i(\boldsymbol{\tau}) = \int_{y_a}^{y_b} \left[\mathbf{Z}_i - \frac{\sum_{i=1}^{n} I(V_i \geq t) \exp(\boldsymbol{\tau}^T \mathbf{Z}_i) \mathbf{Z}_i}{\sum_{i=1}^{n} I(V_i \geq t) \exp(\boldsymbol{\tau}^T \mathbf{Z}_i)} \right] \left[dN_i^{(2)}(t) - \frac{I(V_i \geq t) \exp(\boldsymbol{\tau}^T \mathbf{Z}_i)}{\sum_{i=1}^{n} I(V_i \geq t) \exp(\boldsymbol{\tau}^T \mathbf{Z}_i)} d\bar{N}_2(t) \right]$$

(A.1)

with $\bar{N}_2(t) = \sum_{i=1}^{n} N_i^{(2)}(t)$ (see Lawless and Nadeau (1995)).

Proof of Theorem 2.1

Proof Substituting (2.5) in (2.6), followed by simple algebraic manipulation, yields

$$U(\boldsymbol{\beta}) = \sum_{i=1}^{n} \int_{y_a}^{y_b} N_i^*(t)[\mathbf{Z}_i - \bar{\mathbf{Z}}(t)]dk(t) = 0,$$

(A.3)

where

$$\bar{\mathbf{Z}}(t) = \frac{\sum_{i=1}^{n} Y_i(t) \exp(\boldsymbol{\beta}^T \mathbf{Z}_i) \mathbf{Z}_i}{\sum_{i=1}^{n} Y_i(t) \exp(\boldsymbol{\beta}^T \mathbf{Z}_i)}.$$

(A.4)

Now, using the idea discussed in Du and Lv (2022), $U(\boldsymbol{\beta})$ can be written as

$$U(\boldsymbol{\beta}) = \sum_{i=1}^{n} \int_{y_a}^{y_b} [\mathbf{Z}_i - \bar{\mathbf{Z}}(t)]M_i(t)dk(t) = 0.$$

(A.5)

Using the Taylor expansion of $U(\boldsymbol{\beta})$ around $\boldsymbol{\beta}_0$, where $\boldsymbol{\beta}_0$ is the true value of $\boldsymbol{\beta}$, yields

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = \sqrt{n}A^{-1}U(\boldsymbol{\beta}_0) + o_p(1),$$

(A.6)

where

$$A = E\left(\int_{y_a}^{y_b} (\mathbf{Z}_i - \bar{z}(t))^T (\mathbf{Z}_i - z(t))\mu_0(t) \exp(\boldsymbol{\beta}^T \mathbf{Z}_i)W(t|\mathbf{Z}_i)dk(t) \right),$$

(A.7)

with $\bar{z}(t)$ as the limit of $\bar{\mathbf{Z}}(t)$ (Du & Lv, 2022). Further, note that, $\sup_t |\hat{\Lambda}_l(t|\mathbf{Z}) - \Lambda_l(t|\mathbf{Z})| \to 0$ for $l = 1, 2$, almost surely (see Lawless (2011)). Then, it is easy to follow that $\sup_t |\bar{G}_n(t|\mathbf{Z}) - \bar{G}(t|\mathbf{Z})| \to 0$ and $\sup_t |\bar{H}_n(t|\mathbf{Z}) - \bar{H}(t|\mathbf{Z})| \to 0$ almost

surely. Thus, $\sup_t |\bar{W}_n(t|\mathbf{Z}) - \bar{W}(t|\mathbf{Z})| \to 0$ almost surely. Along with this, using the arguments similar to that of Lin et al. (2000), it follows that $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ can be approximated by a $p-$variate normal distribution with mean vector 0 and variance–covariance matrix given by Σ, where

$$\Sigma = \frac{1}{n}A^{-1}\left(\sum_{i=1}^{n}\Omega_i\Omega_i^{\mathrm{T}}\right)A^{-1} \tag{A.8}$$

with

$$\Omega_i = \int_{y_a}^{y_b}(\mathbf{Z}_i - \bar{z}(t))\left[\sum_{j=1}^{m_i}I(X_{ij} \le t \le U_i) - W(t|\mathbf{Z}_i)\mu_0(t)\exp(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{Z}_i)\right]dk(t). \tag{A.9}$$

Now, a consistent estimator of Σ is given by $\frac{1}{n}\hat{A}^{-1}\left(\sum_{i=1}^{n}\hat{\Omega}_i\hat{\Omega}_i^{\mathrm{T}}\right)\hat{A}^{-1}$, where $\hat{\Omega}_i$ is obtained by replacing all the theoretical quantities in Ω_i with their estimates and

$$\hat{A}^{-1} = \frac{1}{n}\sum_{i=1}^{n}\int_{y_a}^{y_b}(\mathbf{Z}_i - \bar{z}(t))^{\mathrm{T}}(\mathbf{Z}_i - \bar{z}(t))\hat{\mu}_0(t)\exp(\hat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{Z}_i)\bar{W}(t|\mathbf{Z}_i)dk(t). \tag{A.10}$$

Similar results are derived in Cai and Cheng (2004) for double censoring data in a non-recurrent scenario. □

Proof of Theorem 2.2

Proof We first write

$$\hat{\mu}_0(t) - \mu_0(t) = (\hat{\mu}_0(t, \hat{\boldsymbol{\beta}}) - \hat{\mu}_0(t, \boldsymbol{\beta}_0)) + (\hat{\mu}_0(t, \boldsymbol{\beta}_0) - \mu_0(t)). \tag{A.11}$$

Using Taylor series expansion of $\hat{\mu}_0(t)$ in $\boldsymbol{\beta}_0$, the first term on the right side of A.11, equals to

$$\sqrt{n}(\hat{\mu}_0(t, \hat{\boldsymbol{\beta}}) - \hat{\mu}_0(t, \boldsymbol{\beta}_0)) = \frac{\partial\hat{\mu}_0(t, \boldsymbol{\beta}_0)}{\partial\boldsymbol{\beta}^{\mathrm{T}}}\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + o_p(1). \tag{A.12}$$

By the uniform strong law of large numbers and Lemma 1 of Lin et al. (2000), we obtain

$$\sup_{0\le t\le t^*}\left\|\frac{\partial\hat{\mu}_0(t, \boldsymbol{\beta}_0)}{\partial\boldsymbol{\beta}^{\mathrm{T}}} + D(t, \boldsymbol{\beta}_0)\right\| \to 0$$

almost surely, where $D(t, \boldsymbol{\beta}_0)$ is given in Theorem 2.2. Also, we have

$$\sup_t \left\| \bar{W}_n(t|\mathbf{Z}) - W(t|\mathbf{Z}) \right\| \to 0$$

almost surely. From (A.6)

$$\begin{aligned}
\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) &= \sqrt{n}A^{-1}U(\beta_0) + o_p(1) \\
&= A^{-1}\sqrt{n}\sum_{i=1}^{n}\int_{y_a}^{y_b} Y_i(t)\{\mathbf{Z}_i - \bar{\mathbf{Z}}(t)\}M_i(t)dk(t) + o_p(1).
\end{aligned}$$

(A.13)

Therefore, by (A.11)–(A.13), we get

$$\begin{aligned}
&\sqrt{n}\{\hat{\mu}_0(t, \hat{\boldsymbol{\beta}}) - \hat{\mu}_0(t, \boldsymbol{\beta}_0)\} \\
&= -D(t; \boldsymbol{\beta}_0)A^{-1}\sqrt{n}\sum_{i=1}^{n}\int_{y_a}^{y_b} Y_i(t)\{\mathbf{Z}_i - \bar{\mathbf{Z}}(t)\}M_i(t)dk(t) + o_p(1),
\end{aligned}$$
(A.14)

In addition, for $0 \le t \le \tau'$

$$\sqrt{n}\{\hat{\mu}_0(t; \boldsymbol{\beta}_0) - \mu_0(t)\} = n^{-1/2}\sum_{i=1}^{n}\frac{M_i(t; \boldsymbol{\beta}_0)}{n^{-1}\sum_{i=1}^{n}Y_i(t)\exp(\boldsymbol{\beta}^\mathrm{T}\mathbf{Z}_i)}.$$
(A.15)

Hence, Lemma 1 of Lin et al. (2000) implies that uniformly in t

$$\sqrt{n}\{\hat{\mu}_0(t; \boldsymbol{\beta}_0) - \mu_0(t)\} = n^{-1/2}\sum_{i=1}^{n}\frac{M_i(t; \boldsymbol{\beta}_0)}{\mathrm{E}\left[Y_i(t)\exp(\boldsymbol{\beta}^\mathrm{T}\mathbf{Z}_i)\right]} + o_p(1).$$
(A.16)

Finally, it follows from (A.11)–(A.16) that:

$$\sqrt{n}\{\hat{\mu}_0(t; \hat{\boldsymbol{\beta}}) - \mu_0(t)\} = \sqrt{n}\sum_{i=1}^{n}\psi_i(t) + o_p(1)$$
(A.17)

uniformly in t, where $\psi_i(t)$ is defined in Theorem 2.2. Hence, for each t, $\sqrt{n}\{\hat{\mu}_0(t; \hat{\boldsymbol{\beta}}) - \mu_0(t)\}$ follows an asymptotic behavior similar to a sum of independent and identically distributed mean-zero variables. By employing the multivariate central limit theorem, this expression converges, in finite-dimensional distribution, to a zero-mean Gaussian process. Utilizing modern empirical process theory, as described in Lin et al. (2000), we can ascertain that $\psi_i(t)$ is a tight function. Consequently, $\sqrt{n}\{\hat{\mu}_0(t; \hat{\boldsymbol{\beta}}) - \mu_0(t)\}$ is also a tight function and weakly converges to a mean-zero Gaussian process, with a covariance function $\Gamma(s, t) = E(\Psi_1(s)\Psi_1(t))$ at (s, t) . $\qquad\square$

References

Andersen, P. K., & Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study. *The Annals of Statistics, 10*(4), 1100–1120.

Cai, T., & Cheng, S. (2004). Semiparametric regression analysis for doubly censored data. *Biometrika, 91*(2), 277–290.

Chang, M. N. (1990). Weak convergence of a self-consistent estimator of the survival function with doubly censored data. *The Annals of Statistics, 18*(1), 391–404.

Chen, B. E., Cook, R. J., Lawless, J. F., & Zhan, M. (2005). Statistical methods for multivariate interval-censored recurrent events. *Statistics in Medicine, 24*(5), 671–691.

Choi, S., & Huang, X. (2021). Efficient inferences for linear transformation models with doubly censored data. *Communications in Statistics-Theory and Methods, 50*(9), 2188–2200.

Choi, T., Kim, A. K., & Choi, S. (2021). Semiparametric least-squares regression with doubly-censored data. *Computational Statistics & Data Analysis, 164*, 107306.

Cook, R. J., & Lawless, J. F. (2007). *The Statistical Analysis of Recurrent Events*. New York: Springer.

Dörre, A. (2021). Semiparametric likelihood inference for heterogeneous survival data under double truncation based on a Poisson birth process. *Japanese Journal of Statistics and Data Science, 4*(2), 1203–1226.

Du, Y., & Lv, Y. (2022). A flexible additive-multiplicative transformation mean model for recurrent event data. *Communications in Statistics-Theory and Methods, 51*(2), 328–339.

Fleming, T. R., & Harrington, D. P. (1991). *Counting Processes and Survival Analysis*. New Jersey: John Wiley & Sons.

Gu, M., & Zhang, C.-H. (1993). Asymptotic properties of self-consistent estimators based on doubly censored data. *The Annals of Statistics, 21*(2), 611–624.

Huang, C.-Y., Qin, J., & Wang, M.-C. (2010). Semiparametric analysis for recurrent event data with time-dependent covariates and informative censoring. *Biometrics, 66*(1), 39–49.

Ji, S., Peng, L., Cheng, Y., & Lai, H. (2012). Quantile regression for doubly censored data. *Biometrics, 68*(1), 101–112.

Kelly, P. J., & Lim, L. L. Y. (2000). Survival analysis for recurrent event data: an application to childhood infectious diseases. *Statistics in Medicine, 19*(1), 13–33.

Lawless, J. F. (2011). *Statistical Models and Methods for Lifetime Data*. New Jersey: John Wiley & Sons.

Lawless, J. F., & Nadeau, C. (1995). Some simple robust methods for the analysis of recurrent events. *Technometrics, 37*(2), 158–168.

Li, S., Hu, T., Tong, T., & Sun, J. (2020). Semiparametric regression analysis of multivariate doubly censored data. *Statistical Modelling, 20*(5), 502–526.

Lin, D. Y., Wei, L.-J., Yang, I., & Ying, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 62*(4), 711–730.

Liu, Y., Wu, Y., Cai, J., & Zhou, H. (2010). Additive-multiplicative rates model for recurrent events. *Lifetime Data Analysis, 16*(3), 353–373.

Miloslavsky, M., Keleş, S., Laan, M. J., & Butler, S. (2004). Recurrent events analysis in the presence of time-dependent covariates and dependent censoring. *Journal of the Royal Statistical Society Series B: Statistical Methodology, 66*(1), 239–257.

Pepe, M. S., & Cai, J. (1993). Some graphical displays and marginal regression analyses for recurrent failure times and time dependent covariates. *Journal of the American Statistical Association, 88*(423), 811–820.

Prentice, R. L., Williams, B. J., & Peterson, A. V. (1981). On the regression analysis of multivariate failure time data. *Biometrika, 68*(2), 373–379.

Sankaran, P. G., & Anisha, P. (2011). Shared frailty model for recurrent event data with multiple causes. *Journal of Applied Statistics, 38*(12), 2859–2868.

Sankaran, P. G., & Sreedevi, E. P. (2016). A proportional hazards model for the analysis of doubly censored competing risks data. *Communications in Statistics-Theory and Methods, 45*(10), 2975–2987.

Shen, P.-S. (2011). Semiparametric analysis of transformation models with doubly censored data. *Journal of Applied Statistics, 38*(4), 675–682.

Shen, P.-S. (2022). The Cox-Aalen model for doubly censored data. *Communications in Statistics-Theory and Methods, 51*(23), 8075–8092.

Shen, P.-S., & Chen, C.-M. (2018). Aalen's linear model for doubly censored data. *Statistics, 52*(6), 1328–1343.

Shen, H., & Cook, R. J. (2015). Analysis of interval-censored recurrent event processes subject to resolution. *Biometrical Journal, 57*(5), 725–742.

Su, C.-L., & Lin, F.-C. (2021). Analysis of cyclic recurrent event data with multiple event types. *Japanese Journal of Statistics and Data Science, 4*(2), 895–915.

Sun, J., & Wei, L. (2000). Regression analysis of panel count data with covariate-dependent observation and censoring times. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 62*(2), 293–302.

Sun, L., Zhu, L., & Sun, J. (2009). Regression analysis of multivariate recurrent event data with time-varying covariate effects. *Journal of Multivariate Aalysis, 100*(10), 2214–2223.

Therneau, T. M. (2023). *A Package for Survival Analysis in R*. R package version 3.5-7.

Turnbull, B. W. (1974). Nonparametric estimation of a survivorship function with doubly censored data. *Journal of the American Statistical Association, 69*(345), 169–173.

Wang, L., Sun, J., & Tong, X. (2010). Regression analysis of case II interval-censored failure time data with the additive hazards model. *Statistica Sinica, 20*(4), 1709.

Wei, L. J., Lin, D. Y., & Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American Statistical Association, 84*(408), 1065–1073.

Zeng, D., & Lin, D. (2007). Semiparametric transformation models with random effects for recurrent events. *Journal of the American Statistical Association, 102*(477), 167–180.

Zhao, X., Liu, L., Liu, Y., & Xu, W. (2012). Analysis of multivariate recurrent event data with time-dependent covariates and informative censoring. *Biometrical Journal, 54*(5), 585–599.